

## **TUGAS AKHIR – KS141501**

### **RANCANG BANGUN EKSTRAKSI TOPIK FITUR PRODUK DARI ULASAN PENGGUNA ONLINE DENGAN *LATENT DIRICHLET ALLOCATION***

**NATANAEL YABES WIRAWAN**  
**NRP. 5213 100 137**

**Dosen Pembimbing 1:**  
**Rully Agus Hendrawan, S.Kom, M.Eng**  
**NIP. 198112292005011002**

**JURUSAN SISTEM INFORMASI**  
**Fakultas Teknologi Informasi**  
**Insitut Teknologi Sepuluh Nopember**  
**Surabaya 2017**





**TUGAS AKHIR – KS141501**

**RANCANG BANGUN EKSTRAKSI TOPIK FITUR  
PRODUK DARI ULASAN PENGGUNA ONLINE DENGAN  
LATENT DIRICHLET ALLOCATION**

**NATANAEL YABES WIRAWAN**  
**NRP. 5213 100 137**

**Dosen Pembimbing 1:**  
**Rully Agus Hendrawan, S.Kom, M.Eng**  
**NIP. 198112292005011002**

**JURUSAN SISTEM INFORMASI**  
**Fakultas Teknologi Informasi**  
**Insitut Teknologi Sepuluh Nopember**  
**Surabaya 2017**





**FINAL PROJECT – KS141501**

***EXTRACTION OF TOPICAL PRODUCT FEATURE FROM  
ONLINE REVIEWS USING LATENT DIRICHLET  
ALLOCATION***

**NATANAEL YABES WIRAWAN  
NRP. 5213 100 137**

**Supervisor 1:  
Rully Agus Hendrawan, S.Kom, M.Eng  
NIP. 198112292005011002**

**INFORMATION SYSTEMS DEPARTMENT  
Faculty of Information Technology  
Insitut Teknologi Sepuluh Nopember  
Surabaya 2017**



## LEMBAR PENGESAHAN

### **RANCANG BANGUN EKSTRAKSI TOPIK FITUR PRODUK DARI ULASAN PENGGUNA ONLINE DENGAN LATENT DIRICHLET ALLOCATION**

#### **TUGAS AKHIR**

Disusun untuk memenuhi salah satu syarat  
memperoleh gelar Sarjana Komputer  
pada

**Jurusan Sistem Informasi**

**Fakultas Teknologi Informasi**

**Institut Teknologi Sepuluh Nopember**

Oleh:

**NATANAEL YABES WIRAWAN**

**5213 100 137**

**Surabaya, 16 Januari 2017**

**KETUA  
JURUSAN SISTEM INFORMASI**



**Dr. Ir. Aris Tjahyanto, M.Kom**

**NIP. 196503101991021001**



## LEMBAR PERSETUJUAN

### **RANCANG BANGUN EKSTRAKSI TOPIK FITUR PRODUK DARI ULASAN PENGGUNA ONLINE DENGAN *LATENT DIRICHLET ALLOCATION***

#### **TUGAS AKHIR**

Disusun untuk memenuhi salah satu syarat  
memperoleh gelar Sarjana Komputer  
pada

Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh:

**NATANAEL YABES WIRAWAN**  
5213 100 137

Disetujui Tim Penguji: Tanggal Ujian: 11 Januari 2017  
Periode Wisuda: Maret 2017

**Rully Agus Hendrawan, S.Kom, M.Eng**

(Pembimbing 1)

**Erma Suryani, ST, MT, Ph.D**

(Penguji 1)

**Arif Wibisono, S.Kom, M.Sc**

(Penguji 2)





# **RANCANG BANGUN EKSTRAKSI TOPIK FITUR PRODUK DARI ULASAN PENGGUNA ONLINE DENGAN LATENT DIRICHLET ALLOCATION**

**Nama Mahasiswa** : Natanael Yabes Wirawan  
**NRP** : 5213 100 137  
**Jurusan** : Sistem Informasi  
**Dosen Pembimbing** : Rully Agus Hendrawan, S.Kom,  
M.Eng

## **ABSTRAK**

*Berkembangnya teknologi Web 2.0 dalam dunia e-commerce menyebabkan timbulnya fenomena sosial commerce. Social commerce merupakan suatu kondisi dimana munculnya interaksi antara pengguna dengan situs, baik situs perusahaan atau organisasi maupun pribadi, ditandai dengan adanya sistem ulasan pada situs web. Ulasan produk merupakan salah satu fitur yang kerap dijumpai dalam situs-situs pasar online. Seringkali pula dalam membeli produk atau jasa di situs-situs marketplace, konsumen mempertimbangkan ulasan yang ditulis oleh pengguna lain yang membeli produk atau jasa pada situs-situs e-commerce. Dengan melihat ulasan produk, konsumen dapat melihat tanggapan pelanggan yang telah membeli produk tersebut kemudian menjustifikasi apakah fitur yang dijanjikan sudah sesuai dengan realitanya. Namun, untuk mengetahui fitur apa saja yang dibicarakan konsumen, pembeli harus melihat tiap ulasan yang ada pada produk yang ingin dibeli. Belum lagi membandingkan fitur pada produk serupa dengan merk yang berbeda. Untuk produk pada kategori populer saja, ulasan dapat mencapai ratusan bahkan ribuan ulasan. Tugas akhir ini dibuat dengan tujuan untuk menutupi kapasitas manusia untuk mencerna ulasan produk dalam jumlah besar. Luaran dari tugas akhir ini adalah berupa rancang bangun prototipe situs e-commerce yang dilengkapi*

*dengan kemampuan ekstraksi fitur produk dan orientasi opini. Data diperoleh dari salah satu produk dengan jumlah ulasan minimal 500 ulasan pada situs e-commerce terkemuka Amaazon.com. Luaran dari tugas akhir ini berupa rancang bangun prototipe sistem ekstraksi fitur dan orientasi opini yang mampu mendukung keputusan pembelian pelanggan. Hasil analisis terhadap produk Amazon Echo adalah bahwa produk tersebut memiliki speaker yang bagus, voice recognition dari amazon echo dapat memahami permintaan pengguna dengan tepat, produk amazon echo sangat personal dan layak untuk dibeli, aplikasi yang merupakan produk samping amazon echo mudah untuk digunakan bersamaan dengan amazon echo, product amazon echo sangat menyenangkan untuk digunakan, aplikasi amazon echo merupakan produk yang canggih dan keren dimata pengguna, produk amazon echo layak untuk dibeli dan sepadan dengan biaya yang dikeluarkan pengguna, dan produk amazon echo merupakan produk yang hebat dan menyenangkan untuk dipakai.*

***Kata kunci: social commerce, ekstraksi fitur, analisa sentiment, appraisal expression pattern, latent dirichlet allocation.***

**EXTRACTION OF TOPICAL PRODUCT FEATURE  
FROM ONLINE REVIEWS USING LATENT DIRICHLET  
ALLOCATION**

**Name** : Natanael Yabes Wirawan  
**Student Id.** : 5213 100 137  
**Department** : Information Systems  
**Supervisor** : Rully Agus Hendrawan, S.Kom,  
M.Eng

**ABSTRACT**

*The development of Web 2.0 technologies in the world of e-commerce led to the emergence of the phenomenon of social commerce. Social commerce is a condition in which the interaction between users and the site, both sites as well as private companies or organizations, characterized by a system of reviews on the website. Product reviews is one of the features that are often found in the websites of the online market. Often in e-commerce sites, consumers consider reviews written by other users who purchase products or services on the websites of e-commerce. By looking at product reviews, consumers can see the response customers who have purchased the product and then justify whether the promised features are in accordance with reality. However, to find out what features are talking about consumers, buyers should look every existing reviews on products to buy. Not to mention comparing the features of similar products with different brands. For products in popular categories alone, reviews can reach hundreds or even thousands of reviews. This final project was created with the purpose to cover the human capacity to digest large amounts of product reviews. The outcome of this thesis is to design a prototype in the form of e-commerce sites equipped with extraction capability product features and opinion orientation. Data obtained from leading e-commerce sites*

*Amaazon.com. The outcomes of this thesis is a prototype of a e-commerce systems which incorporates feature extraction and opinion mining. The e-commerce platform can support customers when making a purchase decisions. The result of LDA analysis throughout the Amazon Echo product is that these products have good speakers, that the voice recognition from amazon echo can interpret user requests very well, that amazon echo product is very personal and deserves to be bought, the side applications is easy to use in conjunction with amazon echo, amazon echo product is very pleasant to use, the amazon echo application is sophisticated and cool by the perspective of users, that amazon echo is eligible to be purchased within the budget of all users, and amazon echo is a great product and pleasant to be used.*

**Keywords:** *social commerce, ekstraksi fitur, analisa sentiment, appraisal expression pattern, latent dirichlet allocation.*

## KATA PENGANTAR

Puji dan syukur penulis limpahkan kepada Tuhan Yang Maha Esa. Atas rahmat dan karunia-Nya, penulis dapat menyelesaikan Buku Tugas Akhir dengan judul “Rancang Bangun Ekstraksi Topik Fitur Produk Dari Ulasan Pengguna Online Dengan Latent Dirichlet Allocation”. Dalam kesempatan ini, penulis menyampaikan terima kasih kepada:

1. Keluarga penulis, khususnya Bapak, Ibu, dan adik – adik tercinta yang senantiasa mendoakan kelancaran serta kesuksesan dalam pengerjaan Tugas Akhir.
2. Bapak Ir. Aris Tjahyanto, M.Kom. selaku Ketua Jurusan Sistem Informasi ITS Surabaya
3. Rully Agus Hendrawan S.Kom, M.Eng, selaku dosen wali sekaligus dosen pembimbing yang telah ikhlas meluangkan waktu beliau untuk membantu penulis dalam menyelesaikan tugas akhir ini.
4. Alitya Novianda Adityaputri yang senantiasa mengingatkan dan mendukung penulis menyelesaikan Tugas Akhir, bahkan disaat penulis merasa mustahil.
5. Rekan-rekan Lab Sistem Enterprise yang selalu ada untuk mendampingi dan mendukung penulis.
6. 5 Responden yang dengan sukarela membantu memberikan umpan balik dan masukkan sehingga penulis dapat menerima evaluasi aplikasi.
7. Seluruh teman – teman, Keluarga Beltranis yang selalu memberikan dukungan kepada penulis dan memberikan banyak cerita selama penulis melakukan studi.
8. Mas dan Mbak serta adik-adik jurusan Sistem Informasi yang telah memberikan semangat serta bantuan kepada penulis.
9. Seluruh civitas akademika Jurusan Sistem Informasi ITS atas dukungan yang diberikan selama pengerjaan Tugas Akhir

10. Dan seluruh pihak yang telah membantu penulis baik secara langsung maupun tidak langsung sehingga Tugas Akhir ini terselesaikan dengan baik.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna. Kritik dan saran sangat berarti bagi penulis untuk evaluasi penyusunan laporan di masa mendatang.

Surabaya, 06 Januari 2017,

Penulis.

## DAFTAR ISI

|                                                                 |                                     |
|-----------------------------------------------------------------|-------------------------------------|
| LEMBAR PENGESAHAN...                                            | <b>Error! Bookmark not defined.</b> |
| LEMBAR PERSETUJUAN..                                            | <b>Error! Bookmark not defined.</b> |
| KATA PENGANTAR .....                                            | vii                                 |
| DAFTAR GAMBAR .....                                             | xii                                 |
| DAFTAR PERSAMAAN .....                                          | xiv                                 |
| DAFTAR TABEL .....                                              | xv                                  |
| BAB I.....                                                      | 1                                   |
| PENDAHULUAN .....                                               | 1                                   |
| 1.1. Latar Belakang .....                                       | 1                                   |
| 1.2. Perumusan Masalah.....                                     | 6                                   |
| 1.3. Batasan Masalah.....                                       | 6                                   |
| 1.4. Tujuan Tugas Akhir.....                                    | 7                                   |
| 1.5. Manfaat Tugas Akhir.....                                   | 7                                   |
| 1.6. Relevansi Tugas Akhir .....                                | 8                                   |
| BAB II.....                                                     | 11                                  |
| TINJAUAN PUSTAKA.....                                           | 11                                  |
| 2.1. Penelitian Sebelumnya .....                                | 11                                  |
| 2.2. Dasar Teori.....                                           | 16                                  |
| 2.2.1. <i>E-Commerce</i> .....                                  | 17                                  |
| 2.2.2. <i>Social Commerce</i> .....                             | 23                                  |
| 2.2.3. <i>Web Scraping</i> .....                                | 25                                  |
| 2.2.4. <i>Standford Natural Language Processing (NLP)</i><br>36 |                                     |
| 2.2.5. <i>Part of Speech (POS) Tagging</i> .....                | 39                                  |
| 2.2.6. <i>Latent Dirichlet Allocation</i> .....                 | 44                                  |
| 2.2.7. <i>Appraisal Expression Pattern</i> .....                | 46                                  |
| 2.2.9. <i>Gibbs Sampling</i> .....                              | 53                                  |
| 2.2.10. <i>ICONIX</i> .....                                     | 55                                  |
| BAB III .....                                                   | 59                                  |
| METODOLOGI.....                                                 | 59                                  |
| 3.1. Tahapan Studi Literatur .....                              | 62                                  |
| 3.1.1. Mengkaji Metode Pengambilan Data.....                    | 62                                  |
| 3.1.2. Mengkaji Metode Pemrosesan Data .....                    | 62                                  |
| 3.1.3. Mengkaji Metode Ekstraksi Fitur dan Opini.               | 62                                  |
| 3.2. Tahapan Pengambilan Data .....                             | 63                                  |

|                                |                                                                                                                          |     |
|--------------------------------|--------------------------------------------------------------------------------------------------------------------------|-----|
| 3.3.                           | Tahapan Preproses Data .....                                                                                             | 63  |
| 3.4.                           | Tahapan Ekstraksi Fitur dan Opini .....                                                                                  | 64  |
| 3.5.                           | Tahapan Rancang Bangun Prototipe.....                                                                                    | 64  |
| BAB IV .....                   |                                                                                                                          | 67  |
| ANALISIS DAN PERANCANGAN ..... |                                                                                                                          | 67  |
| 4.1.                           | Pengambilan Data.....                                                                                                    | 67  |
| 4.1.1.                         | Identifikasi Karakteristik Situs.....                                                                                    | 67  |
| 4.1.2.                         | Menentukan Metadata Ulasan.....                                                                                          | 69  |
| 4.1.3.                         | Perancangan Basis Data Ulasan .....                                                                                      | 72  |
| 4.2.                           | Preproses Data .....                                                                                                     | 73  |
| 4.3.                           | Ekstraksi Fitur dan Opini Produk .....                                                                                   | 73  |
| 4.4.                           | Rancang Bangun Prototipe .....                                                                                           | 76  |
| 4.4.1.                         | Antar Muka Pengguna .....                                                                                                | 76  |
| 4.4.2.                         | Use Case .....                                                                                                           | 79  |
| 4.4.3.                         | Domain Model.....                                                                                                        | 82  |
| 4.4.4.                         | <i>Robustness Diagram</i> .....                                                                                          | 82  |
| 4.4.5.                         | Sequence Diagram .....                                                                                                   | 88  |
| 4.4.6.                         | Class Diagram.....                                                                                                       | 93  |
| BAB V .....                    |                                                                                                                          | 94  |
| IMPLEMENTASI .....             |                                                                                                                          | 95  |
| 5.1.                           | Lingkungan Implementasi .....                                                                                            | 95  |
| 5.2.                           | Struktur Direktori .....                                                                                                 | 95  |
| 5.3.                           | Pengambilan Data.....                                                                                                    | 96  |
| 5.3.1.                         | Mendapatkan jumlah paginasi pada halaman ulasan                                                                          | 98  |
| 5.3.2.                         | Melakukan Scraping dan Menyimpan hasil Scraping ke dalam basis data ulasan .....                                         | 99  |
| 5.4.                           | Preproses Data .....                                                                                                     | 102 |
| 5.4.1.                         | Menjalankan Stanford NLP sebagai service                                                                                 | 103 |
| 5.4.2.                         | Melakukan POS Tagging pada data ulasan.                                                                                  | 103 |
| 5.5.                           | Ekstraksi Fitur dan Opini.....                                                                                           | 104 |
| 5.5.1.                         | Melakukan Analisa LDA dan Mengurutkan Probabilitas Kata Kunci per Topik Berdasarkan Kata yang Paling Sering Muncul ..... | 107 |
| 5.5.2.                         | Melakukan Analisa LDA dengan <i>Gibbs Sampling</i>                                                                       | 109 |

|                                          |                                                                                 |     |
|------------------------------------------|---------------------------------------------------------------------------------|-----|
| 5.5.3.                                   | Melakukan Analisa <i>Gibbs Sampling</i> dengan <i>Sample Full Conditional</i> . | 111 |
| 5.6.                                     | Rancang Bangun Prototipe                                                        | 120 |
| 5.6.1.                                   | Implementasi Halaman Pencarian Produk ..                                        | 121 |
| 5.6.2.                                   | Implementasi Halaman Daftar Produk.....                                         | 122 |
| 5.6.3.                                   | Implementasi Halaman Detail Produk .....                                        | 123 |
| BAB VI                                   |                                                                                 | 126 |
| HASIL DAN PEMBAHASAN                     |                                                                                 | 127 |
| 6.1.                                     | Hasil Analisis LDA                                                              | 127 |
| 6.1.1.                                   | Hasil Kata Kunci Topik 1                                                        | 130 |
| 6.1.2.                                   | Hasil Kata Kunci Topik 2                                                        | 133 |
| 6.1.3.                                   | Hasil Kata Kunci Topik 3                                                        | 134 |
| 6.1.4.                                   | Hasil Kata Kunci Topik 4                                                        | 135 |
| 6.1.5.                                   | Hasil Kata Kunci Topik 5                                                        | 136 |
| 6.1.6.                                   | Hasil Kata Kunci Topik 6                                                        | 137 |
| 6.1.7.                                   | Hasil Kata Kunci Topik 7                                                        | 138 |
| 6.1.8.                                   | Hasil Kata Kunci Topik 8                                                        | 139 |
| 6.1.9.                                   | Hasil Kata Kunci Topik 9                                                        | 140 |
| 6.1.10.                                  | Hasil Kata Kunci Topik 10                                                       | 141 |
| 6.1.11.                                  | Hasil Kata Kunci Topik 11                                                       | 142 |
| 6.1.12.                                  | Hasil Kata Kunci Topik 12                                                       | 143 |
| 6.1.13.                                  | Hasil Kata Kunci Topik 13                                                       | 144 |
| 6.1.14.                                  | Hasil Kata Kunci Topik 14                                                       | 145 |
| 6.1.15.                                  | Hasil Kata Kunci Topik 15                                                       | 146 |
| 6.1.16.                                  | Hasil Kata Kunci Topik 16                                                       | 147 |
| 6.1.17.                                  | Hasil Kata Kunci Topik 17                                                       | 148 |
| 6.1.18.                                  | Hasil Kata Kunci Topik 18                                                       | 149 |
| 6.1.19.                                  | Hasil Kata Kunci Topik 19                                                       | 150 |
| 6.1.20.                                  | Hasil Kata Kunci Topik 20                                                       | 151 |
| 6.2.                                     | Hasil Uji Usabilitas                                                            | 153 |
| BAB VII                                  |                                                                                 | 155 |
| KESIMPULAN DAN SARAN                     |                                                                                 | 155 |
| 7.1.                                     | Kesimpulan                                                                      | 155 |
| 7.2.                                     | Saran                                                                           | 156 |
| DAFTAR PUSTAKA                           |                                                                                 | 157 |
| BIODATA PENULIS                          |                                                                                 | 163 |
| LAMPIRAN 1 USABILITY TEST PLAN DASHBOARD |                                                                                 | 164 |

|                                            |     |
|--------------------------------------------|-----|
| LAMPIRAN 3 CONSENT FORM.....               | 168 |
| LAMPIRAN 4 SYSTEM USABILITY SCALE FORM ... | 170 |

## **DAFTAR GAMBAR**

|                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------|----|
| Bagan 1 Notasi Blok Latent Dirichlet Allocation.....                                                               | 13 |
| Bagan 2 Notasi Pelat AEP-LDA .....                                                                                 | 15 |
| Bagan 3 Arsitektur Web 2.0. ....                                                                                   | 24 |
| Bagan 4 Request data via XMLHttpRequest pada salah satu website e-commerce di Indonesia.....                       | 30 |
| Bagan 5 Teknik Scraping dengan Pemrograman HTTP [22] .....                                                         | 31 |
| Bagan 6 Contoh HTML Parsing dengan Jaunt [23].....                                                                 | 32 |
| Bagan 7 Sebuah dokumen HTML .....                                                                                  | 33 |
| Bagan 8 Representasi Grafik DOM untuk dokumen HTML                                                                 | 33 |
| Bagan 9 Implementasi DOM parsing [25].....                                                                         | 34 |
| Bagan 10 Contoh implementasi Semantic Annotation pada dokumen HTML dengan format Microdata.....                    | 35 |
| Bagan 11 Contoh Implementasi Part of Speech pada ulasan pelanggan terhadap barang di marketplace online [30] ..... | 44 |
| Bagan 12 Notasi Pelat LDA [14].....                                                                                | 46 |
| Bagan 13 Grafik Dependensi antar kata [30].....                                                                    | 47 |
| Bagan 14 Notasi pelat AEP-LDA (atas) dibandingkan dengan LDA (bawah) [14].....                                     | 51 |
| Bagan 15 Proses dalam ICONIX [33] .....                                                                            | 56 |
| Bagan 16 Metodologi Tugas Akhir .....                                                                              | 61 |
| Bagan 17 Metodologi Pengambilan Data Ulasan .....                                                                  | 67 |
| Bagan 18 Tampilan halaman ulasan produk pada situs Amazon .....                                                    | 71 |
| Bagan 19 Arsitektur Sistem Ekstraksi Fitur dan Orientasi Opini .....                                               | 74 |
| Bagan 20 Diagram Alir Algoritma LDA .....                                                                          | 75 |
| Bagan 21 Antar Muka Halaman Pencarian.....                                                                         | 76 |
| Bagan 22 Halaman Daftar Produk.....                                                                                | 77 |
| Bagan 23 Halaman Detail Produk .....                                                                               | 78 |

|                                                                                       |     |
|---------------------------------------------------------------------------------------|-----|
| Bagan 24 Use Case Diagram untuk Sistem Ekstraksi Fitur dan Orientasi Opini.....       | 79  |
| Bagan 25 Domain model untuk aplikasi AmazoLDA .....                                   | 82  |
| Bagan 26 Robustness Analysis untuk UCC-1 Mencari Produk .....                         | 83  |
| Bagan 27 Robustness Analysis untuk UCC-2 Melihat Detail Produk.....                   | 84  |
| Bagan 28 Robustness Analysis untuk UCC-3 Melihat Fitur dan Opini.....                 | 85  |
| Bagan 29 Robustness Analysis untuk UCC-4 Melihat Ulasan .....                         | 86  |
| Bagan 30 Robustness Analysis untuk UCS-1 Melakukan POS Tagging .....                  | 87  |
| Bagan 31 Sequence Diagram untuk Robustness RBUCC-1 Mencari Produk.....                | 88  |
| Bagan 32 Sequence Diagram untuk Robustness RBUCC-2 Melihat Detail Produk .....        | 89  |
| Bagan 33 Sequence Diagram untuk Robustness RBUCC-3 Mencari Fitur dan Opini.....       | 90  |
| Bagan 34 Sequence Diagram untuk Robustness RBUCC-4 Melihat Ulasan .....               | 91  |
| Bagan 35 Sequence Diagram untuk Robustness RBUCS-1 Melakukan POS Tagging .....        | 92  |
| Bagan 36 Class Diagram aplikasi AmazoLDA .....                                        | 93  |
| Bagan 37 Struktur Direktori Aplikasi.....                                             | 96  |
| Bagan 38 Diagram Alir Proses Web Scraping .....                                       | 97  |
| Bagan 39 Konfigurasi MySQL.....                                                       | 98  |
| Bagan 40 Kode untuk mendapatkan jumlah paginasi pada halaman ulasan.....              | 99  |
| Bagan 41 Hasil web scraping ditampung dalam basis data ulasan .....                   | 100 |
| Bagan 42 Kode untuk melakukan scraping dan menyimpan ke dalam basis data ulasan ..... | 102 |
| Bagan 43 Kode untuk menjalankan server Standford NLP via konsole .....                | 103 |

|                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------|-----|
| Bagan 44 Menjalankan Server Stanford NLP via Konsole                                                  | 103 |
| Bagan 45 Kode untuk menjalankan POS Tagging pada Stanford NLP .....                                   | 104 |
| Bagan 46 Hasil POS Tagging.....                                                                       | 104 |
| Bagan 47 Kode Utama untuk melakukan analisa LDA dengan korpus berupa ulasan produk di Amazon.com..... | 106 |
| Bagan 48 Kode Analisa LDA dengan Gibbs Sampling .....                                                 | 109 |
| Bagan 49 Kode Analisa Gibbs Sampling dengan Sample Full Conditional.....                              | 116 |
| Bagan 50 Luaran dari analisis LDA .....                                                               | 120 |
| Bagan 51 Halaman Pencarian Produk .....                                                               | 127 |
| Bagan 52 Implementasi kode pada halaman pencarian .....                                               | 121 |
| Bagan 53 Halaman Daftar Produk.....                                                                   | 128 |
| Bagan 54 Implementasi kode pada halaman daftar produk                                                 | 123 |
| Bagan 55 Halaman Detail Produk .....                                                                  | 129 |
| Bagan 56 Implementasi kode pada halaman detail produk.                                                | 125 |
| Bagan 57 Grafik hasil analisis LDA.....                                                               | 129 |
| Bagan 58 Hasil Rekapitulasi SUS .....                                                                 | 154 |

## DAFTAR PERSAMAAN

|                                                                                               |    |
|-----------------------------------------------------------------------------------------------|----|
| Persamaan 1 Notasi Matematika dari LDA .....                                                  | 45 |
| Persamaan 2 Formula untuk menghitung Jumlah Shortest Dependency Path pada suatu dokumen ..... | 48 |
| Persamaan 3 Formula untuk menghitung Confidence Score                                         | 49 |
| Persamaan 4 Persamaan Multinomial untuk distribusi kata dalam kalimat s dalam dokumen d ..... | 52 |
| Persamaan 5 Probabilitas kata aspek .....                                                     | 53 |
| Persamaan 6 Probabilitas kata opini .....                                                     | 53 |
| Persamaan 7 Probabilitas background words .....                                               | 53 |

## DAFTAR TABEL

|                                                                                      |     |
|--------------------------------------------------------------------------------------|-----|
| Tabel 1 Relevansi Tugas Akhir Terhadap Kompetensi Jurusan Sistem Informasi ITS ..... | 8   |
| Tabel 2 Hasil Kajian Studi Literatur.....                                            | 11  |
| Tabel 3 Kriteria Kesuksesan E-Commerce [5] .....                                     | 18  |
| Tabel 4 Modifier pada Regex .....                                                    | 27  |
| Tabel 5 Brackets pada Regex .....                                                    | 27  |
| Tabel 6 Metacharacters pada Regex .....                                              | 28  |
| Tabel 7 Anotator yang disediakan oleh Stanford NLP [30]                              | 37  |
| Tabel 8 Macam-macam Part of Speech dalam Bahasa Inggris [32].....                    | 40  |
| Tabel 9 Deskripsi Produk .....                                                       | 68  |
| Tabel 10 Metadata ulasan.....                                                        | 69  |
| Tabel 11 Pemetaan Metadata Ulasan dengan Tag HTML DOM.....                           | 70  |
| Tabel 12 Deskripsi Tabel Ulasan .....                                                | 72  |
| Tabel 13Daftar Use Case Sistem Ekstraksi Fitur dan Orientasi Opini .....             | 80  |
| Tabel 14 Pemetaan Robustness Analysis dengan Use Case .                              | 82  |
| Tabel 14 Pemetaan Robustness Analysis dengan Sequence Diagram .....                  | 88  |
| Tabel 16 Spesifikasi Perangkat Keras .....                                           | 95  |
| Tabel 17 Spesifikasi Perangkat Lunak .....                                           | 95  |
| Tabel 18 Parameter Fungsi LDA.....                                                   | 106 |
| Tabel 19 Parameter LDA yang digunakan untuk analisis...                              | 130 |
| Tabel 20 Hasil Kata Kunci untuk Topik 1 .....                                        | 132 |
| Tabel 21 Hasil Kata Kunci untuk Topik 2 .....                                        | 133 |
| Tabel 22 Hasil Kata Kunci untuk Topik 3.....                                         | 134 |
| Tabel 23 Hasil Kata Kunci untuk Topik 4.....                                         | 135 |
| Tabel 24 Hasil Kata Kunci untuk Topik 5 .....                                        | 136 |
| Tabel 25 Hasil Kata Kunci untuk Topik 6.....                                         | 137 |
| Tabel 26 Hasil Kata Kunci untuk Topik 7 .....                                        | 138 |
| Tabel 27 Hasil Kata Kunci untuk Topik 8 .....                                        | 139 |
| Tabel 28 Hasil Kata Kunci untuk Topik 9.....                                         | 140 |
| Tabel 29 Hasil Kata Kunci untuk Topik 10.....                                        | 141 |

Tabel 30 Hasil Kata Kunci untuk Topik 11 ..... 142

Tabel 31 Hasil Kata Kunci untuk Topik 12 ..... 143

Tabel 32 Hasil Kata Kunci untuk Topik 13 ..... 144

Tabel 33 Hasil Kata Kunci untuk Topik 14 ..... 145

Tabel 34 Hasil Kata Kunci untuk Topik 15 ..... 146

Tabel 35 Hasil Kata Kunci untuk Topik 16 ..... 147

Tabel 36 Hasil Kata Kunci untuk Topik 17 ..... 148

Tabel 37 Hasil Kata Kunci untuk Topik 18 ..... 149

Tabel 38 Hasil Kata Kunci untuk Topik 19 ..... 150

Tabel 39 Hasil Kata Kunci untuk Topik 20 ..... 151

Tabel 40 Backlog Aplikasi ..... 153

*Halaman ini sengaja dikosongkan*



# **BAB I**

## **PENDAHULUAN**

Bab ini akan menjelaskan proses identifikasi masalah antara lain latar belakang permasalahan, perumusan masalah, batasan masalah, tujuan dari pembuatan tugas akhir, manfaat yang didapat dari tugas akhir, serta relevansi tugas akhir terhadap permasalahan yang ingin diselesaikan.

### **1.1. Latar Belakang**

*E-commerce* merupakan istilah yang kerap digunakan untuk merujuk pada proses perdagangan atau proses untuk memfasilitas perdagangan dengan menggunakan layanan jaringan komputer. Proses perdagangan *e-commerce* dapat terjadi melalui media internet, media sosial, maupun melalui aplikasi perangkat bergerak [1].

Saat ini, *e-commerce* sedang mengalami perkembangan yang pesat. Berdasarkan riset yang dilansir oleh ATKearney, transaksi yang terjadi dalam *e-commerce* pada tahun 2014 mengalami pertumbuhan sebesar 20% secara global atau hampir menyentuh nilai transaksi sebesar USD 840 Miliar dan diperkirakan akan terus tumbuh pada tahun-tahun mendatang [2]. Hal ini dapat terjadi dikarenakan oleh adanya ekspansi yang dilakukan ritel online ke negara lain. dan ritel yang selama ini berada pada pasar *commerce*, oleh karena perkembangan teknologi, akhirnya memutuskan akan memasuki pasar *e-commerce*.

Perkembangan pesat dalam dunia *e-commerce* tidak lepas dari faktor jumlah pengguna internet yang meningkat dari tahun ke tahun. Survei yang dilakukan oleh International Telecommunications Union menyatakan bahwa jumlah

pengguna internet global mencapai 3,5 miliar pengguna dimana 71% pengguna internet global tersebut berasal dari negara berkembang [3]. Peningkatan penggunaan internet juga diiringi dengan peningkatan jumlah pengguna perangkat bergerak. Bahkan pada tahun 2016, jumlah pengguna jaringan *mobile broadband* (3G) mencapai 84% dari total populasi dunia, diikuti pula dengan pertumbuhan pengguna jaringan *Long Term Evolution* (LTE/4G) yang mencapai 53% dari total populasi dunia. Akamai Technologies juga melansir bahwa angka konsumen yang melakukan transaksi dengan menggunakan perangkat bergerak melonjak sebanyak 250% selama 5 tahun terakhir [4].

Berkembangnya teknologi internet, situs web, dan perangkat bergerak telah merevolusi *e-commerce*. Pasar *e-commerce* telah mengalami perubahan yang signifikan melalui pengadopsian teknologi Web 2.0. Fenomena ini disebut sebagai *social commerce*, dimana terjadi utilisasi teknologi Web 2.0 di dalam *platform e-commerce*. Utilisasi teknologi Web 2.0 memungkinkan pengguna untuk saling berbagi informasi, seperti memberikan opini atau review atas suatu produk, rating, maupun rekomendasi kepada pengguna lain dalam suatu *platform e-commerce* [5].

Dengan berkembangnya fenomena *social commerce*, pembeli dapat memberikan ulasan terhadap suatu produk yang diperjual-belikan dalam situs *e-commerce*. Ulasan negatif yang berkualitas yang dilakukan pelanggan terhadap suatu produk mampu mempengaruhi perilaku pelanggan secara signifikan dalam membeli suatu produk di *marketplace e-commerce* [6]. Bahkan lebih jauh lagi, studi terkait evaluasi loyalitas pelanggan dengan menggunakan model DeLone dan McLean menyatakan bahwa kualitas informasi dari website *e-commerce*, berupa menawarkan fitur review pengguna dan evaluasi produk, berpengaruh positif terhadap perilaku belanja

hedonik [7]. Sehingga dapat diketahui bahwa ulasan produk berpengaruh terhadap keputusan pelanggan dalam membeli suatu produk dalam *marketplace e-commerce*.

Agar dapat mengetahui kualitas dari suatu produk maupun kualitas pelayanan toko penyedia produk dalam *marketplace e-commerce*, maka pelanggan perlu mempertimbangkan ulasan konsumen lain yang menggunakan produk serupa secara teliti. Namun tentu saja hal tersebut tidak efisien karena keterbatasan manusia dalam mengolah informasi [8]. Untuk kategori produk populer saja, ulasan dari penulis atau konsumen lain dapat mencapai ratusan bahkan ribuan. Maka dari itu, konsumen memerlukan sebuah sistem yang mampu mengekstraksi fitur beserta sentimen yang melekat pada produk atau jasa yang ditawarkan dalam situs-situs *e-commerce* dengan harapan bahwa proses mencerna informasi ulasan dapat dipersingkat sehingga mampu mendukung dalam pengambilan keputusan.

Beberapa penelitian sebelumnya terkait dengan analisa sentimen dan ekstraksi fitur yang telah mendahului tugas akhir ini telah dilakukan. Studi yang dilakukan oleh Minqing Hu dan Bing Liu mendemonstrasikan teknik data mining dengan menggunakan *association rule mining* untuk mengekstrak fitur yang melekat pada produk [9]. Fitur-fitur tersebut diekstrak dari ulasan produk yang ditulis oleh konsumen pada suatu situs *e-commerce* dengan menggunakan teknologi *text mining* dan *natural language processing* (NLP). Penelitian ini juga memaparkan arsitektur sistem untuk mengagregasi ulasan produk. Dengan rata-rata *recall* sebesar 80% dan rata-rata presisi sebesar 72%, hasil penelitian Minqing Hu dan Bing Liu dapat dikatakan akurat dan dapat dipraktikan dalam kasus nyata.

Penelitian terkait dengan *Latent Dirichlet Allocation* sangat populer terutama dalam penelitian-penelitian yang melibatkan dataset teks yang berukuran besar [10]. Blei, Andrew Ng., dan Michael I. Jordan merupakan peneliti yang pertama kali mengembangkan metode *Latent Dirichlet Allocation* (LDA) pada tahun 2003 silam. LDA adalah model statistik generatif untuk kumpulan data diskrit seperti korpus teks [11]. Hasil eksperimen terhadap penelitian Blei et.al. terhadap 5.225 abstrak karya tulis ilmiah C. Elegans juga menunjukkan bahwa metode LDA dapat mengungguli metode lainnya antara lain *Unigram* dan *Probabilistic Latent Semantic Indexing* (pLSI).

Yue Lu, Qiaozhu Mei, dan Cheng Xiang Zhai melakukan studi empiris untuk membandingkan performa dari *Probabilistic Latent Semantic Analysis* (pLSA) dan LDA. Studi perbandingan dilakukan dengan melakukan tugas representatif dari kegiatan penggalian teks (*text mining*) antara lain klusterisasi dokumen, kategorisasi dokumen dan *ad-hoc retrieval*. Hasil penelitian menunjukkan bahwa metode LDA secara umum mengungguli PLSA pada ketiga tugas representatif teks mining [12]. Studi ini mendukung penelitian Blei, et.al.

Yohan Jo dan Alice Oh mengusulkan metode *Sentence LDA* (SLDA) yang kemudian dikembangkan lagi menjadi *Aspect and Sentiment Unification Model* (ASUM) [13]. SLDA merupakan model probalistik yang dikembangkan dari LDA dengan batasan bahwa semua kata dalam sebuah kalimat berasal dari satu sumber topik yang sama. Sementara ASUM merupakan perluasan dari SLDA yang menggabungkan aspek beserta sentimen terhadap aspek tersebut. Data yang digunakan untuk penelitian Alice Oh bersumber dari ulasan produk di Amazon.com. Hasil menunjukkan bahwa metode SLDA mengungguli metode LDA pada eksperimen penelusuran aspek. LDA memberikan hasil temuan aspek yang sifatnya

lebih umum dan tidak terlalu koheren dibandingkan dengan metode SLDA. Sementara pada eksperimen terhadap aspek yang berpasangan dengan sentimen, SLDA cenderung menemukan aspek berdasarkan kategori produk, sementara ASUM berhasil menemukan aspek dengan melihat sentimen yang mirip dari berbagai kategori produk. Pada eksperimen kata sentimen yang spesifik terhadap aspek, ASUM tidak mengungguli semua *classifier* yang ada, antara lain *baseline*, *lingpipe-unigram*, *lingpipe-bi*, ASUM, ASUM+, *Joint Sentiment/Topic* (JST+), dan *Topic Sentiment Mixture* (TSM+). Namun ASUM+ dapat mengungguli semua *classifier*. Kemudian untuk eksperimen kalifikasi sentimen, metode ASUM dan ASUM+ mengungguli model lainnya (JST+ dan TSM+) dengan tingkat akurasi klasifikasi yang cukup tinggi. Secara keseluruhan, ASUM menggabungkan aspek, sentimen, dan aspek-sentimen. Metode ASUM mampu mengidentifikasi sentimen yang melekat pada aspek produk. Dari hasil eksperimen di atas, dapat disimpulkan bahwa ASUM mengungguli kesemua metode dalam semua eksperimen yang ada.

Kemudian ada pula penelitian Xiaolin Zheng et.al., yang mengulas tentang pendekatan dalam mengekstrak fitur atau aspek, sekaligus kata sentimen yang melekat pada fitur atau aspek dari produk atau layanan terkait. Penelitian Xiaolin Zheng et.al. mengusulkan dua pendekatan antara lain pendekatan *unsupervised dependency analysis-based* dan *Appraisal Expression Patterns* (AEP) berbasis *Latent Dirichlet Allocation* (LDA) [14]. Hasil penelitian ini menunjukkan efektivitas dari pendekatan AEP-LDA dalam mengekstrak fitur dan sentimen yang melekat pada fitur. Pendekatan AEP-LDA juga mengungguli metode *Nearest* dan *Mutual Information*.

Dari hasil penelitian-penelitian terdahulu, tugas akhir ini ingin memanfaatkan arsitektur sistem ekstraksi fitur dan analisis

sentiment ulasan dari penelitian Minqing Hu dan Bing Liu [9]. Sementara metodologi ekstraksi fitur dan opini akan merujuk pada penelitian Yohan Jo dan Alice Oh [13]. Kemudian metodologi AEP-LDA berdasarkan penelitian Xiaolin Zhen et.al. akan digunakan sebagai metode ekstraksi fitur dan orientasi opini dari ulasan produk yang ada dengan pertimbangan bahwa metode AEP-LDA menghasilkan hasil yang lebih presisi dibandingkan dengan metode LDA dan SLDA [14]. Metode LDA sendiri lebih akurat daripada metode unigram dan pLSI. Data ulasan akan diambil dari situs e-commerce terkemuka, Amazon.com.

## **1.2. Perumusan Masalah**

Berdasarkan latar belakang tersebut, tugas akhir yang diajukan menitikberatkan permasalahan pada beberapa hal sebagai berikut.

1. Bagaimana mengekstrak fitur dan opini ulasan fitur produk dengan metode *Appraisal Expression Pattern* berbasis *Latent Dirichlet Allocation*.
2. Fitur apa sajakah dari produk tersebut yang sering dibahas pada ulasan?
3. Bagaimana orientasi / sentimen dari ulasan tersebut?

## **1.3. Batasan Masalah**

Berdasarkan rumusan masalah di atas, batasan masalah dalam tugas akhir ini dapat didefinisikan sebagai berikut.

1. Produk yang dianalisis minimal satu produk dengan jumlah ulasan minimal 500 ulasan.
2. Ulasan produk yang dianalisis adalah ulasan yang hanya ditulis dalam bahasa Inggris.

3. Sumber data yang yang diperoleh dalam studi kasus ini hanya berasal dari hasil web scraping pada situs <http://www.amazon.com>.

#### **1.4. Tujuan Tugas Akhir**

Berdasarkan rumusan masalah di atas, tujuan daripada tugas akhir ini adalah sebagai berikut.

1. Mengembangkan arsitektur aplikasi *e-commerce* yang mampu mengekstrak fitur dan orientasi opini produk dan berdasarkan ulasan yang telah ditulis oleh konsumen produk terkait menggunakan metode AEP-LDA.
2. Mengembangkan aplikasi *e-commerce* yang mampu mengekstrak fitur-fitur produk berdasarkan ulasan konsumen terhadap suatu produk.
3. Mengembangkan aplikasi *e-commerce* yang mampu menampilkan orientasi opini fitur produk berdasarkan ulasan konsumen terhadap suatu produk.

Luaran dari tugas akhir ini berupa aplikasi prototipe *e-commerce* yang mampu melakukan ekstraksi fitur dan melakukan analisa sentimen terhadap opini yang melekat pada fitur. Fitur dan hasil analisa sentimen akan di tampilkan pada aplikasi prototipe *e-commerce*.

#### **1.5. Manfaat Tugas Akhir**

Manfaat yang dapat diperoleh bagi pelaku *e-commerce* khususnya pembeli pada marketplace *e-commerce* antara lain adalah sebagai berikut.

1. Mampu mengetahui fitur unggulan berdasarkan ulasan produk oleh konsumen produk terkait.

2. Mampu mengetahui seberapa baik atau buruk suatu sentimen terhadap produk yang diulas oleh konsumen produk terkait.
3. Mampu membantu konsumen dalam mengambil keputusan untuk membeli atau tidak membeli suatu produk berdasarkan ulasan terhadap produk yang telah diaggregasi.

## 1.6. Relevansi Tugas Akhir

Relevansi tugas akhir akan membahas mata kuliah yang terkait dengan topik penelitian tugas akhir serta keterkaitan dengan roadmap laboratorium Sistem Enterprise Jurusan Sistem Informasi ITS. Mata kuliah yang berkaitan dengan topik tugas akhir ini telah dipetakan berdasarkan rumusan masalah, seperti yang ditunjukkan pada Tabel 1.

Penelitian ini adalah bagian dari topik ERP 3.0, yang berfokus pada pelanggan (*customer focus*). ERP 3.0 akan menciptakan *borderless enterprise* dengan mengintegrasikan / membuat *marketplace intelligence* / analitik dari e-commerce yang ada saat ini.

*Tabel 1 Relevansi Tugas Akhir Terhadap Kompetensi  
Jurusan Sistem Informasi ITS*

| No. | Rumusan Masalah                                                                                                            | Mata Kuliah                                                           |
|-----|----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| 1.  | Bagaimana cara mengambil data ulasan produk dari situs e-commerce kemudian merapkannya dan menyimpannya agar mudah diolah? | Desain Basis Data, Manajemen Basis Data, Integrasi Aplikasi Korporasi |
| 2.  | Fitur apa sajakah dari produk tersebut yang sering dibahas pada ulasan?                                                    | E-commerce, Sistem Cerdas                                             |

| No. | Rumusan Masalah                                                                                                                     | Mata Kuliah                                                                 |
|-----|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| 3.  | Bagaimana orientasi / sentimen dari ulasan tersebut?                                                                                | E-commerce, Sistem Cerdas                                                   |
| 4.  | Bagaimana rancang bangun dari aplikasi <i>e-commerce</i> secara keseluruhan yang dapat menampilkan fitur produk berdasarkan ulasan? | Arsitektur SI/TI dalam Perusahaan, Konstruksi dan Pengujian Perangkat Lunak |

***Halaman ini sengaja dikosongkan***

## BAB II

### TINJAUAN PUSTAKA

Bab ini akan menjelaskan mengenai profil perusahaan, baik dari profil perusahaan tempat dilaksanakannya kerja praktik, maupun perusahaan yang menjadi bahan kajian studi kasus.

#### 2.1. Penelitian Sebelumnya

Dalam penelitian tugas akhir ini, ada beberapa penelitian terdahulu yang dapat dijadikan sebagai bahan kajian studi literatur. Adapun penelitian sebelumnya mengenai gambaran umum penelitian, tujuan penelitian, hasil penelitian, dan keterkaitan penelitian akan diulas pada bab ini. Berikut adalah penelitian terdahulu terkait dengan tugas akhir ini.

*Tabel 2 Hasil Kajian Studi Literatur*

| Judul Penelitian 1                                                                                                                                |
|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>From e-commerce to social commerce; A close look at design features</i> [5]                                                                    |
| Tujuan                                                                                                                                            |
| Menunjukkan bahwa pasar <i>e-commerce</i> telah mengalami perubahan yang signifikan dengan berkembangnya teknologi Web 2.0.                       |
| Hasil                                                                                                                                             |
| Menghasilkan kerangka kerja dalam mendesain <i>platform e-commerce</i> yang bercirikan <i>social commerce</i>                                     |
| Keterkaitan Dengan Penelitian                                                                                                                     |
| Penelitian ini menunjukkan bahwa <i>social commerce</i> mempengaruhi cara pelanggan dalam berbelanja online. Sehingga tujuan dari tugas akhir ini |

adalah bagaimana cara memperbaiki sistem ulasan produk pada situs-situs e-commerce agar dibuat berdasarkan prinsip *social commerce*.

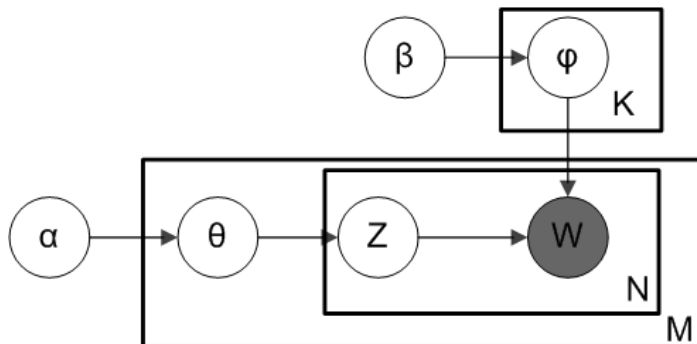
| Judul Penelitian 2                                                                                                                                                                                                                                                                                                    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Mining Opinion Features in Customer Reviews</i> [9]                                                                                                                                                                                                                                                                |
| Tujuan                                                                                                                                                                                                                                                                                                                |
| Mendemonstrasikan teknik <i>data mining</i> dengan menggunakan <i>association rule mining</i> untuk mengekstrak fitur yang melekat pada produk                                                                                                                                                                        |
| Hasil                                                                                                                                                                                                                                                                                                                 |
| <ol style="list-style-type: none"> <li>1. Arsitektur sistem untuk mengaggregasi ulasan produk</li> <li>2. Rata-rata <i>recall</i> sebesar 80% dan rata-rata presisi sebesar 72%</li> <li>3. Karena hasil <i>recall</i> dan presisi cukup bagus, maka penelitian ini dapat diaplikasikan dalam kasus nyata.</li> </ol> |
| Keterkaitan Dengan Penelitian                                                                                                                                                                                                                                                                                         |
| Penelitian ini memaparkan arsitektur sistem untuk mengekstrak fitur dari ulasan produk. Arsitektur tersebut akan diadopsi pada penelitian sebagai acuan dalam memproses data ulasan produk.                                                                                                                           |

| Judul Penelitian 3                                                                                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Latent Dirichlet Allocation</i> [11]                                                                                                                                                     |
| Tujuan                                                                                                                                                                                      |
| Menemukan model baru untuk menemukan deskripsi singkat yang memungkinkan efisiensi pemrosesan dokumen dalam jumlah besar serta menyatakan hubungan statistik sehingga dapat digunakan untuk |

klasifikasi, summarisasi, serta penilaian terhadap kesamaan dan relevansi topik yang terkandung dalam kumpulan dokumen.

### Hasil

Sebuah model matematis bernama *Latent Dirichlet Allocation*



Bagan 1 Notasi Blok Latent Dirichlet Allocation

Atau jika ditulis dalam notasi matematika:

$$p(w, z, \theta, \phi | \alpha, \beta) = p(\theta | \alpha) p(z | \theta) p(\phi | \beta) p(w | z, \phi)$$

#### Dimana:

$N$  = Jumlah kata dalam dokumen ( $N_{(m)}$ )

$M$  = Jumlah dokumen teridentifikasi

$K$  = Jumlah topik teridentifikasi

$W_{(n,m)}$  = kata-kata yang diobservasi dalam suatu korpus teks

$Z_{(n,m)}$  = topik yang melekat pada kata

$\Theta_{(m)}$  = proporsi topik per dokumen yang teridentifikasi

$\Phi_{(k)}$  = distribusi topik pada vocabulary

$\alpha, \beta$  = parameter dirichlet

### Keterkaitan Dengan Penelitian

Penelitian di atas memaparkan serta mendemonstrasikan implementasi dari *Latent Dirichlet Allocation* dalam permasalahan pemodelan topik. Model *Latent Dirichlet Allocation* akan digunakan untuk mengekstrak fitur-fitur apa saja yang sering dibicarakan konsumen. Dengan melihat probabilitas kata yang sering muncul dalam ulasan produk, maka dapat

ditarik kesimpulan mengenai fitur-fitur produk apa saja yang sering dibicarakan konsumen dan bagaimana sentimen masing-masing fitur yang dibicarakan oleh konsumen.

| Judul Penelitian 4                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Aspect and Sentiment Unification Model</i> [13]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Tujuan                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Mengusulkan pendekatan baru dalam ekstraksi fitur dan analisa sentiment menggunakan Latent Dirichlet Allocation, dinamakan dengan <i>Aspect and Sentiment Unification Model</i> (ASUM). Peneliti kemudian membandingkan dengan model yang ia kembangkan sebelumnya bernama <i>Sentence LDA</i> (SLDA).                                                                                                                                                                                                                                                                                                                                                  |
| Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| ASUM merupakan perluasan dari SLDA yang menggabungkan aspek beserta sentimen terhadap aspek tersebut. Hasil menunjukan bahwa dengan menggabungkan aspek beserta sentimen terhadap aspek tersebut, SLDA cenderung menemukan aspek berdasarkan kategori produk, sementara ASUM berhasil menemukan aspek dengan melihat sentimen yang mirip dari berbagai kategori produk. Secara keseluruhan, ASUM menggabungkan aspek, sentimen, dan aspek-sentimen. Metode ASUM mampu mengidentifikasi sentimen yang melekat pada aspek produk. Dari hasil eksperimen di atas, dapat disimpulkan bahwa ASUM mengungguli kesemua metode dalam semua eksperimen yang ada. |
| Keterkaitan Dengan Penelitian                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Penelitian Alice Oh menunjukan bahwa LDA maupun pengembangan menunjukan performa yang jauh lebih baik daripada metode <i>classifier</i> lainnya dalam kasus pemodelan topik. Ke depn, penelitian tugas akhir ini pun dapat dikembangkan dengan menggunakan model LDA yang telah dikembangkan.                                                                                                                                                                                                                                                                                                                                                           |

### Judul Penelitian 5

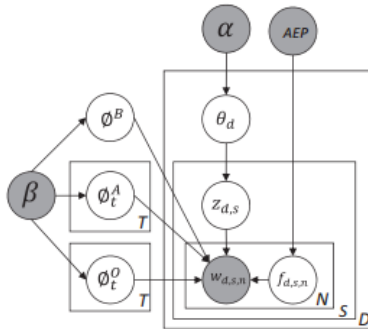
*Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification [14]*

### Tujuan

Mengusulkan dua pendekatan antara lain pendekatan *unsupervised dependency analysis-based* dan *Appraisal Expression Patterns (AEP)* berbasis *Latent Dirichlet Allocation (LDA)*

### Hasil

1. Model AEP-LDA sebagai pengembangan dari Smoothed LDA (SLDA)



Bagan 2 Notasi Pelat AEP-LDA

Berikut ini adalah notasi matematika untuk mendapatkan probabilitas kata aspek, opini, dan *background words* dalam suatu kalimat  $s$  dalam dokumen  $d$ .

$$P(f_{d,s,n} = 1 | AEP) \propto \max\{CS | SDP_h^n\}$$

$$P(f_{d,s,n} = 2 | AEP) \propto \max\{CS | SDP_n^h\}$$

$$P(f_{d,s,n} = 0 | AEP) \propto 1 - \max\{P(f_{d,s,n} = 1 | AEP), P(f_{d,s,n} = 2 | AEP)\}$$

**Dimana:**

$P(f_{d,s,n} = 0 | AEP) =$   
Probabilitas *background words*

$P(f_{d,s,n} = 1 | AEP) =$  Probabilitas kata aspek

$P(f_{d,s,n} = 2 | AEP) =$  Probabilitas kata opini

$SDP_h^n =$   
SDP antara kata ke- $n$  (kandidat kata sentimen) dan kata ke- $h$  (kandidat kata aspek) dalam kalimat  $s$  dalam dokumen  $d$ .

$SDP_n^h =$   
SDP antara kata ke- $h$  (kandidat kata aspek) dan kata ke- $n$  (kandidat kata sentimen) dalam kalimat  $s$  dalam dokumen  $d$ .

2. Hasil penelitian ini menunjukkan efektifitas dari pedekatan AEP-LDA dalam mengekstrak fitur dan sentimen yang melekat pada fitur. Pendekatan AEP-LDA juga mengungguli metode *Nearest* dan *Mutual Information*.

### **Keterkaitan Dengan Penelitian**

Metode AEP-LDA digunakan sebagai metode untuk mengekstrak fitur dari ulasan produk e-commerce, serta menentukan orientasi opini terhadap fitur berdasarkan ulasan produk.

## **2.2. Dasar Teori**

Dasar teori merupakan penjelasan mengenai teori-teori yang digunakan untuk mendukung pengerjaan tugas akhir. Terdapat beberapa teori yang digunakan dalam pengerjaan tugas akhir ini diantaranya *e-commerce*, *social commerce*, *web scraping*, *Stanford Natural Language Processing (NLP)*, *Part of Speech (POS) Tagging*, *Sentiment Analysis*, *Appraisal Expression Pattern*, *Latent Dirichlet Allocation*, dan *Gibbs Sampling*.

### 2.2.1. *E-Commerce*

*E-commerce* atau perdagangan elektronik merujuk pada model bisnis yang tidak lain adalah konsep umum yang melingkupi segala macam bentuk dari transaksi bisnis atau pelaksanaan pertukaran informasi dengan menggunakan teknologi informasi dan komunikasi [15].

Teknologi yang mengitari *e-commerce* sangat beragam antara lain *mobile commerce*, *electronic funds transfer*, *supply chain management*, *internet marketing*, *online transaction processing*, *electronic data interchange*, *inventory management systems*, dan *automated data collection systems*.

Bentuk-bentuk dari *e-commerce* pun bisa sangat beragam mulai dari situs belanja dari retailer menuju ke konsumen, sistem jual beli *business to business* (B2B), pengolahan data melalui kontak web dan sosial media, *electronic data interchange* pada lanskap *business to business* (B2B) hingga pertukaran mata uang yang dilakukan secara online.

Dengan menggunakan *e-commerce*, baik pembeli maupun penjual dapat berkomunikasi dan bertransaksi kapan pun dan dimanapun melalui jaringan internet. Sehingga *e-commerce* berdampak pada penurunan biaya infrastruktur, meleburnya batas-batas geografis, dan tumbuhnya peluang bisnis baru [16].

Kehadiran *e-commerce* dalam lanskap perdagangan telah mendorong timbulnya persaingan harga sebagaimana *platform e-commerce* mampu memberikan informasi yang lengkap mengenai spesifikasi produk beserta ulasan pengguna bagi pelanggan [17]. Informasi yang diterima oleh pelanggan

sangat banyak. Di situs-situs e-commerce terkemuka seperti Amazon.com, produk-produk yang dijual mencapai 391.337.440 per 1 Desember 2016 [18]. Belum lagi ulasan produk yang bisa mencapai puluhan ribu untuk tiap produknya.

E-commerce merupakan suatu ekosistem yang dikembangkan agar penjual dan pembeli dapat bertransaksi kapanpun dan dimanapun menggunakan segala macam perantara elektronik yang memungkinkan. Huang dan Benyoucef memaparkan kriteria-kriteria kesuksesan sebuah platform e-commerce berdasarkan Information Systems Success Model yang dikembangkan oleh Delone dan McLean [5]. Kriteria-kriteria tersebut dapat dilihat pada tabel 3.

*Tabel 3 Kriteria Kesuksesan E-Commerce [5]*

| Prinsip    | Karakteristik                 | Tujuan Bisnis                                                                    |
|------------|-------------------------------|----------------------------------------------------------------------------------|
| Usabilitas | Kemudahan Penggunaan          | Untuk membuat situs yang mudah digunakan.                                        |
|            | Interface yang mudah dipahami | Untuk membuat antar muka situs yang mudah dipahami.                              |
|            | Simplisitas                   | Untuk menyediakan struktur dan fungsi platform <i>e-commerce</i> yang sederhana. |
|            | Navigasi                      | Agar pengguna dapat melakukan navigasi di dalam situs.                           |
|            | Kontrol Pengguna              | Untuk mengizinkan pengguna masuk dan keluar dari situs atau                      |

| Prinsip | Karakteristik             | Tujuan Bisnis                                                                                                                                                                                                 |
|---------|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|         |                           | platform e-commerce.                                                                                                                                                                                          |
|         | Pencegahan Error          | Untuk mencegah agar pengguna tidak melakukan tindakan yang membuat error muncul.                                                                                                                              |
|         | Fungsi Help               | Untuk menawarkan pertolongan dan dokumentasi situs e-commerce bagi pengguna.                                                                                                                                  |
|         | Mudah dipahami            | Untuk membuat konten situs yang mudah dipahami.                                                                                                                                                               |
|         | Aksesibilitas             | Untuk memungkinkan website diakses oleh semua pengguna.                                                                                                                                                       |
|         | Kecepatan                 | Untuk membuat situs yang cepat untuk diunduh dan ditampilkan bagi pengguna.                                                                                                                                   |
|         | Visibilitas status sistem | Menginformasikan kepada user terkait dengan status situs atau platform saat ini. Contoh: mengirimkan notifikasi migrasi sistem kepada pengguna sehingga pengguna dihimbau untuk segera melakukan backup data. |

| Prinsip            | Karakteristik                                                                  | Tujuan Bisnis                                                                                 |
|--------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
|                    | Kemiripan dengan dunia nyata                                                   | Agar situs/platform yang dikembangkan mengikuti konvensi dunia nyata.                         |
|                    | Konsistensi                                                                    | Membuat komponen website dengan desain dan tata letak yang konsisten.                         |
|                    | Mengutamakan kemudahan untuk dikenali daripada kemudahan untuk diingat kembali | Untuk membuat informasi semakin mudah untuk diingat.                                          |
|                    | Desain yang estetik                                                            | Untuk membuat desain situs atau platform e-commerce yang memperhatikan estetika dan atraktif. |
|                    | Personalisasi                                                                  | Agar situs dapat dikustom sesuai dengan keperluan                                             |
| Kualitas Informasi | Relevansi                                                                      | Untuk menyediakan informasi yang relevan                                                      |
|                    | Akurasi                                                                        | Untuk menyediakan informasi yang akurat                                                       |
|                    | Kelengkapan                                                                    | Untuk menyediakan informasi yang lengkap                                                      |
|                    | Terperbaharui                                                                  | Untuk menyediakan informasi yang diperbaharui                                                 |

| Prinsip         | Karakteristik         | Tujuan Bisnis                                                                   |
|-----------------|-----------------------|---------------------------------------------------------------------------------|
|                 | Otoritas              | Untuk menjamin kerahasiaan pengguna                                             |
|                 | Kegunaan              | Untuk menyediakan informasi yang berguna                                        |
|                 | Kecukupan             | Untuk menyediakan informasi yang cukup                                          |
| Kualitas Sistem | Keamanan              | Menjamin bahwa sistem aman digunakan                                            |
|                 | Kemudahan Akses       | Memungkinkan akses yang cepat bagi semua pengguna                               |
|                 | Pemulihan dari error  | Untuk membantu sistem dalam pemulihan dari error                                |
|                 | Operasi dan komputasi | Untuk membuat sistem dan layanan yang tersedia mudah digunakan dan dioperasikan |
|                 | Tampilan              | Untuk mempersentasikan visual yang baik                                         |
|                 | Fungsionalitas        | Untuk menyediakan fungsi yang mencakupi semua tugas yang teridentifikasi        |
|                 | Sistem Pembayaran     | Untuk menyediakan sistem pembayaran                                             |

| Prinsip          | Karakteristik         | Tujuan Bisnis                                                                        |
|------------------|-----------------------|--------------------------------------------------------------------------------------|
|                  |                       | online yang aman dan mudah digunakan                                                 |
|                  | Sistem Pemesanan      | Untuk mengembangkan sistem pemesanan yang dapat melakukan pelacakan status pemesanan |
|                  | Konten                | Untuk menyediakan konten yang sesuai dengan harapan pelanggan                        |
| Kualitas Layanan | Responsif             | Untuk menyampaikan respon yang cepat dan tepat sesuai dengan kebutuhan user          |
|                  | Jaminan               | Untuk menyediakan dukungan dalam penyelesaian masalah pelanggan                      |
|                  | Empati                | Untuk memberikan perhatian dan empati bagi pengguna                                  |
|                  | Layanan yang proaktif | Untuk mendengarkan setiap umpan balik yang diberikan pengguna                        |
|                  | Reliabilitas          | Untuk menyampaikan layanan yang dapat diandalkan dan dapat dipercaya                 |
| Kesenangan       | Kenikmatan            | Untuk menyediakan pengalaman yang                                                    |

| Prinsip | Karakteristik          | Tujuan Bisnis                                           |
|---------|------------------------|---------------------------------------------------------|
|         |                        | menyeangkan bagi pengguna                               |
|         | Tampilan yang menarik  | Untuk menyediakan antar muka yang estetis bagi pengguna |
|         | Kontrol                | Memberikan kontrol bagi pengguna                        |
|         | Penasaran              | Untuk meningkatkan rasa penasaran bagi pengguna         |
|         | Ketertarikan intrinsik | Agar sesuai dengan ketertarikan user.                   |

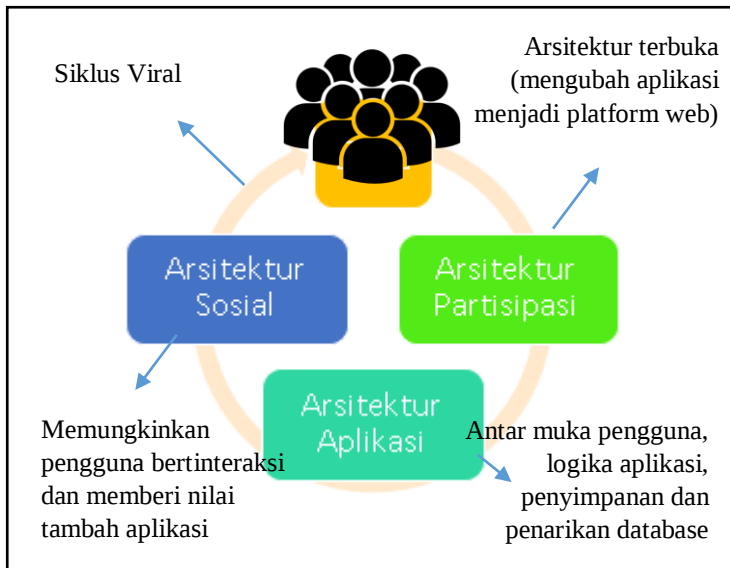
### 2.2.2. *Social Commerce*

*E-commerce* saat ini sedang mengalami evolusi melalui adaptasi teknologi dari kapabilitas Web 2.0. untuk meningkatkan partisipasi pelanggan dalam ekosistem *e-commerce* [5]. Fenomena inilah yang disebut sebagai *social commerce*. Ditambah lagi dengan adanya teknologi sosial media dan *mobile applications* yang semakin hari semakin berkembang, pasar e-commerce pun semakin hari dipenuhi oleh pelaku bisnis yang kreatif dan kolaboratif dalam membangun atau mengembangkan bisnis [19]. Sehingga pasar barang dan jasa telah bergeser menjadi pasar yang berfokus pada sosial dan digerakan oleh pengguna internet.

Web 2.0. merupakan suatu pendekatan pemberian layanan berbasis web menekankan pada kolaborasi online dan berbagi antar pengguna. Gambar di atas menjelaskan tentang arsitektur

dari Web 2.0.. Prinsip-prinsip Web 2.0 sebagaimana yang dikemukakan oleh O'Reilly adalah sebagai berikut [20]:

0. Web sebagai platform
1. Data sebagai pengendali utama
2. Efek jaringan diciptakan melalui arsitektur yang melibatkan partisipasi pengguna



*Bagan 3 Arsitektur Web 2.0.*

3. Inovasi dalam perakitan sistem serta situs disusun dengan menyatukan fitur dari pengembang yang terdistribusi dan independen (semacam model pengembangan "open source")
4. Model bisnis yang ringan, yang dikembangkan dengan gabungan isi dan layanan
5. Akhir dari siklus peluncuran (release cycle) perangkat lunak (perpetual beta)
6. Mudah untuk digunakan dan diadopsi oleh user

Adopsi teknologi Web 2.0 ke dalam *platform e-commerce* telah menimbulkan fenomena *social commerce*. *Social commerce* sendiri melibatkan disiplin ilmu yang beragam antara lain disiplin ilmu pemasaran, ilmu komputer, sosiologi, dan psikologi. Oleh karena karakteristik fenomena *social commerce* yang melibatkan multi disiplin ilmu, maka seringkali definisi dari fenomena *social commerce* menjadi bias.

Huang dan Benyoucef kemudian memberikan definisi bahwasannya *social commerce* merupakan aplikasi komersial berbasis internet yang mengutilisasi penggunaan teknologi sosial media dan Web 2.0 dalam upaya untuk menciptakan interaksi sosial sehingga user dapat memberikan konten berupa ulasan atau komentar agar memudahkan konsumen dalam mengambil keputusan pembelian produk baik berupa barang fisik atau jasa pada situs atau komunitas belanja online [5].

### **2.2.3. Web Scraping**

*Web scraping* merupakan teknik ekstraksi data yang bersumber dari halaman-halaman situs web. *Web scraping* dilakukan dengan cara menyalin data yang dikumpulkan dari halaman-halaman situs web yang tersebar di penjuru internet. Kemudian data-data salinan tersebut ditampung ke dalam media penyimpanan seperti basis data atau file *spreadsheet* untuk keperluan analisa lebih lanjut.

Halaman website pada umumnya dibuat dengan bahasa *markup* seperti HTML atau XHTML. Meskipun halaman web dengan format tersebut menyimpan banyak informasi, namun halaman web tersebut didesain untuk keperluan tampilan pada pengguna akhir dan susah untuk melakukan otomatisasi *scraping* pada web tersebut. Oleh karena itu, beberapa perusahaan telah

menyediakan *Application Programming Interface* (API) yang memudahkan proses *web scraping*.

Ragam bentuk *web scraping* bervariasi mulai dari cara manual yang melibatkan interaksi dan keputusan manusia, maupun yang terotomasi. Berikut adalah beberapa teknik *web scraping* yang cukup populer.

#### **2.2.3.1. Teknik Salin Manual**

Teknik ini merupakan teknik yang paling sederhana dalam *web scraping*. Pengguna hanya membuka atau mengunduh halaman website kemudian menyalin informasi-informasi yang dibutuhkan ke dalam media penyimpanan basis data, file *spreadsheet*, maupun media-media penyimpanan lainnya. Solusi ini dilakukan jika target situs web yang diinginkan memblokir atau tidak mengizinkan aktivitas *web scraping* terjadi pada situs web yang bersangkutan.

#### **2.2.3.2. Pencocokan Pola (*Regular Expression*)**

Merupakan salah satu teknik *web scraping* yang jamak dijumpai karena kemudahannya dalam melaksanakan tugas *web scraping*. Teknik ini dinamakan *regular expression* (regex), yakni melihat pola kata, digit, atau karakter yang muncul pada suatu dokumen di situs web. Bentuk dari regex sendiri adalah sebagai berikut [21].

/pola/modifier

Berikut ini adalah *modifer* yang digunakan dalam regex untuk melakukan pencarian secara global:

Tabel 4 Modifier pada Regex

| Modifier | Deskripsi                                                              |
|----------|------------------------------------------------------------------------|
| i        | Melakukan pencarian dengan mengabaikan besar kecilnya huruf.           |
| g        | Menemukan semua pola yang cocok. Mirip dengan fungsi <i>find all</i> . |
| m        | Melakukan pencocokan pola per baris.                                   |

Berikut ini adalah *brackets* dalam regex. *Brackets* digunakan untuk mencari jangkauan karakter.

Tabel 5 Brackets pada Regex

| Modifier | Deskripsi                                                                           |
|----------|-------------------------------------------------------------------------------------|
| [abc]    | Menemukan karakter yang ada sebagaimana yang didefinisikan dalam <i>bracket</i> .   |
| [^abc]   | Menemukan karakter yang ada selain yang didefinisikan dalam <i>bracket</i> .        |
| [0-9]    | Menemukan digit yang ada sebagaimana yang didefinisikan dalam <i>bracket</i> .      |
| [^0-9]   | Menemukan digit yang ada selain yang didefinisikan dalam <i>bracket</i> .           |
| (x y)    | Menemukan alternatif yang ada sebagaimana yang didefinisikan dalam <i>bracket</i> . |

Berikut ini adalah *metacharacters* dalam regex. *Metacharacters* merupakan karakter yang memiliki makna khusus.

Tabel 6 Metacharacters pada Regex

| <i>Metacharacter</i> | <i>Deskripsi</i>                                                                 |
|----------------------|----------------------------------------------------------------------------------|
| .                    | Menemukan karakter tunggal kecuali baris baru.                                   |
| \w                   | Menemukan karakter huruf                                                         |
| \W                   | Menemukan karakter non huruf                                                     |
| \d                   | Menemukan karakter digit                                                         |
| \D                   | Menemukan karakter non digit                                                     |
| \s                   | Menemukan karakter <i>whitespace</i>                                             |
| \S                   | Menemukan karakter non <i>whitespace</i>                                         |
| \b                   | Menemukan karakter pada awal / akhir kata                                        |
| \B                   | Menemukan karakter bukan pada awal / akhir kata                                  |
| \0                   | Menemukan karakter NULL                                                          |
| \n                   | Menemukan karakter baris baru                                                    |
| \f                   | Menemukan karakter <i>form feed</i>                                              |
| \r                   | Menemukan karakter <i>carriage return</i>                                        |
| \t                   | Menemukan karakter tab                                                           |
| \v                   | Menemukan karakter tab vertikal                                                  |
| \xxx                 | Menemukan karakter yang dispesifikasi dengan angka oktal                         |
| \xdd                 | Menemukan karakter yang dispesifikasi dengan angka heksadesimal dengan format dd |

| <i>Metacharacter</i> | <i>Deskripsi</i>                                                                   |
|----------------------|------------------------------------------------------------------------------------|
| \uxxxx               | Menemukan karakter yang dispesifikasi dengan angka heksadesimal dengan format xxxx |

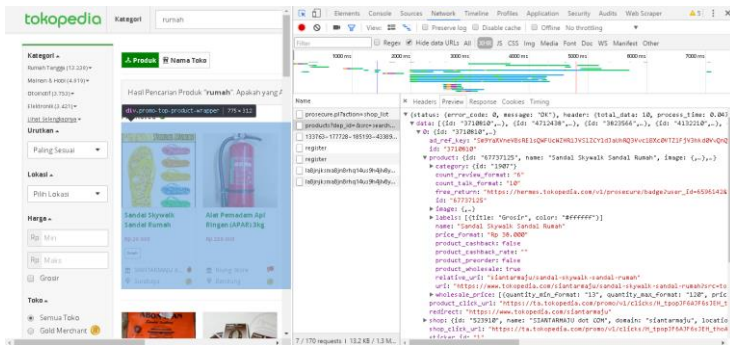
Berikut ini adalah *quantifiers* dalam regex.

| <i>Modifer</i> | <i>Deskripsi</i>                                                     |
|----------------|----------------------------------------------------------------------|
| n+             | Mencocokkan string yang mengandung paling tidak 1 karakter n         |
| n*             | Mencocokkan string yang mengandung 0 atau lebih dari 0 karakter n    |
| n?             | Mencocokkan string yang mengandung 0 atau 1 karakter n               |
| n{X}           | Mencocokkan string yang mengandung sekuensial X dari karakter n      |
| n{X,Y}         | Mencocokkan string yang mengandung sekuensial X ke Y dari karakter n |
| n\$            | Mencocokkan string yang berakhiran karakter n                        |
| ^n             | Mencocokkan string yang berawalan karakter n                         |
| ?=n            | Mencocokkan string yang diikuti oleh karakter n                      |
| ?!n            | Mencocokkan string yang tidak diikuti oleh karakter n                |

Teknik scraping dengan regex dapat dilakukan apabila teknik scraping dengan HTML parsing dan DOM parsing tidak mungkin dilakukan. Biasanya karena kode markup untuk menampilkan ke HTML dihasilkan dengan kode Javascript. Teknik ini juga efektif digunakan untuk memformat ulang bentuk data agar lebih mudah dibaca dan dapat disimpan sesuai dengan pola yang diinginkan melalui penerapan aturan regex.

### 2.2.3.3. Pemrograman HTTP

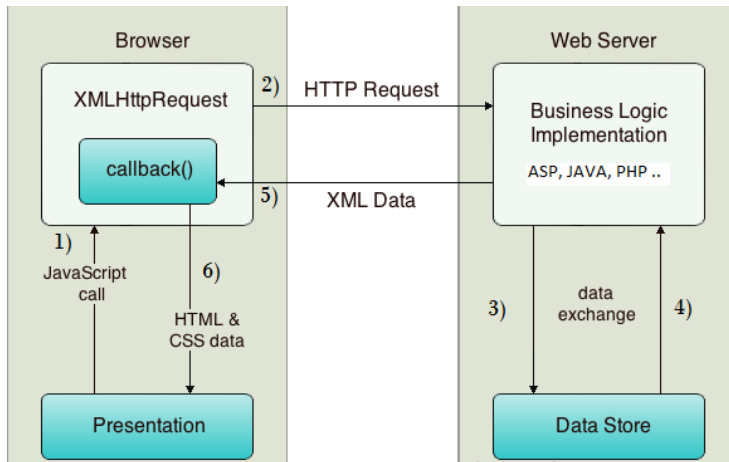
Teknik ini menggunakan teknik posting permintaan data menggunakan protokol HTTP (GET atau POST). Akan tetapi teknik ini tidak dapat diterapkan di semua website. Hanya web dengan keamanan rendah atau memang mengizinkan pengambilan data yang memungkinkan teknik ini digunakan. Teknik ini juga memungkinkan otomatisasi scraping. Gambar 3 menunjukkan ciri website yang dapat di-scraping dengan pemrograman HTTP.



Bagan 4 Request data via XMLHttpRequest pada salah satu website e-commerce di Indonesia

Sebuah request HTTP dilakukan oleh script program pada suatu URL. Request HTTP dilakukan dengan menggunakan XMLHttpRequest, sebuah API yang dirancang untuk berkomunikasi dengan server, mengambil data sesuai permintaan dan mengembalikannya ke browser klien [22]. Output dari XMLHttpRequest adalah berupa file yang pada umumnya berupa JSON atau XML. Nilai kembalian tersebut juga dapat disesuaikan melalui query URL. Selanjutnya data dari hasil query URL tersebut akan ditampilkan dalam array atau objek pemrograman untuk diolah lebih lanjut. Gambar 4

menunjukkan cara kerja teknik scraping dengan pemrograman HTTP.



*Bagan 5 Teknik Scraping dengan Pemrograman HTTP [22]*

#### 2.2.3.4. HTML Parsing

Pada website, data diambil dari basis data kemudian ditampilkan sebagai dokumen HTML. Seringkali data tersebut ditampilkan pada halaman serupa (paginasi) dan dengan menggunakan template tertentu (menggunakan HTML dan CSS Selector). Gambar 6 mencontohkan HTML parsing dengan menggunakan library Java Jaunt.

Teknik HTML parsing mengambil dokumen HTML dari suatu URL kemudian ditampung dalam objek dokumen. Selanjutnya, informasi dari HTML dokumen akan disaring dengan mencocokkan tag-tag HTML yang memiliki data sesuai keperluan. Pencocokan dapat dilakukan dengan Regex maupun pencocokan tag sederhana. Data-data dalam tag ini bisa diiterasi untuk ditampung ke dalam *array*, objek pemrograman,

maupun media penyimpanan seperti basis data, file *spreadsheet*, dan lain-lain.

```
//create new userAgent (headless browser)
UserAgent userAgent = new UserAgent();

//visit google
userAgent.visit("http://google.com");

//apply form input (starting at first
//editable field)
userAgent.doc.apply("butterflies");

//click submit button labelled "GoogleSearch"
userAgent.doc.submit("GoogleSearch");

//find search result links
Elements links = userAgent.doc.findEvery("<h3
class=r>").findEvery("<a>");

//print results
for(Element link : links)
System.out.println(link.getAt("href"));
```

*Bagan 6 Contoh HTML Parsing dengan Jaunt [23]*

### 2.2.3.5. DOM Parsing

World Wide Web Consortium mendefinisikan *Document Object Model* (DOM) sebagai *Application Programming Interface* untuk dokumen HTML dan XML yang valid. DOM mendefinisikan struktur logika dari dokumen dan bagaimana sebuah dokumen diakses dan dimanipulasi [24]. Gambar 7 adalah contoh dokumen HTML.

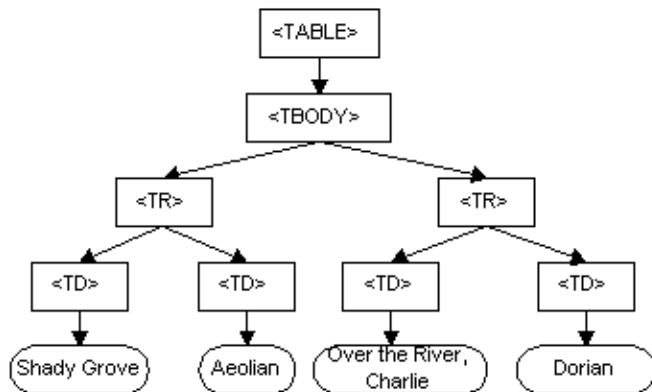
```

<TABLE>
<TBODY>
  <TR>
    <TD>Shady Grove</TD>
    <TD>Aeolian</TD>
  </TR>
  <TR>
    <TD>Over the River</TD>
    <TD>Dorian</TD>
  </TR>
</TBODY>
</TABLE>

```

*Bagan 7 Sebuah dokumen HTML*

Representasi grafik DOM dari tabel di atas adalah seperti yang terlihat pada gambar 7 berikut.



*Bagan 8 Representasi Grafik DOM untuk dokumen HTML*

Dengan menggunakan DOM, pengembang dapat membuat dokumen, menavigasi struktur dokumen, serta dapat

menambahkan, memodifikasi, atau menghapus elemen dan konten.

Beberapa *scraper* telah mendukung pencarian informasi dengan menggunakan DOM parsing. DOM parsing memiliki cara kerja yang serupa dengan HTML parsing, hanya saja, alih-alih menggunakan pencocokan pola dengan menggunakan regex atau pencocokan tag sederhana, sekarang pengguna dapat mencari data yang diperlukan dengan menembak tag-tag HTML menggunakan DOM *selector*. Gambar 8 menunjukkan implementasi dari teknik DOM *Parsing*.

```
File input =
new File("/tmp/input.html");

Document doc =
Jsoup.parse(input, "UTF-8",
"http://example.com/");

Element content =
doc.getElementById("content");

Elements links =
content.getElementsByTag("a");

for (Element link : links) {

    String linkHref = link.attr("href");

    String linkText = link.text();

}
```

*Bagan 9 Implementasi DOM parsing [25]*

#### **2.2.3.6. Semantic Annotation Recognizing**

Teknik ini melibatkan teknologi semantic web dalam melakukan web scraping. Sederhananya, sama halnya HTML parsing dan DOM parsing, pengguna mengambil informasi

berdasarkan tag-tag dalam dokumen HTML. Namun teknik *Semantic Annotation Recognizing* memiliki kelebihan lain karena teknik ini membaca tag-tag semantik yang didefinisikan oleh pengembang situs web dimana situs web yang bersangkutan menjadi target *scraping*. Tidak semua webstie dapat dijadikan target untuk melakukan *Semantic Annotation Recognizing*. Pengembang situs web harus mengimplementasikan *Semantic Annotaion Recognizing*.

Dalam kasus *web scraping*, pengembang website memberikan informasi tambahan berupa konsep pada teks atau konten lainnya. Dokumen yang di anotaskan dengan informasi tambahan ini digunakan agar mesin atau program dapat mengenali dan menginterpretasikan informasi. Gambar 9 merupakan contoh komponen semantik dalam suatu dokumen HTML.

```
The birds roosted at
<span class="geo">
  <span class="latitude">52.48</span>,
  <span class="longitude">-1.89</span>
</span>
```

*Bagan 10 Contoh implementasi Semantic Annotation pada dokumen HTML dengan format Microdata*

Pada contoh sebagaimana pada gambar 9, pengembang memberi informasi tambahan berupa lokasi geografis, yakni posisi garis lintang 52,48 ° sementara posisi garis bujur di -1,89 °. *Semantic Annotation* dapat mempermudah proses *scraping* sebagaimana informasi yang diperlukan telah didefinisikan dengan jelas.

Melakukan proses *anotattion* dapat dilakukan dengan berbagai macam algoritma antara lain *Pattern-based Annotation through*

*Knowledge on the Web* (PANKOW) Qtag, AeroDAML, dan lain-lain [26].

#### **2.2.3.7.     Analisa dengan *machine learning***

Ekstraksi konten atau informasi pada situs web merupakan salah satu teknologi kunci untuk mengenal lebih dalam mengenai dunia web yang luas. Sejauh ini, sistem ekstraksi konten web otomatis telah dipelajari dengan mendalam. Akan tetapi aplikasi ekstraksi konten web yang dikembangkan saat ini hanya berfokus pada pengambilan data terstruktur [27]. Contoh dari data terstruktur antara lain merupakan katalog produk, ulasan pelanggan, dan informasi sejenis.

*Web scraping* yang melibatkan pemanfaatan *machine learning* berfokus untuk menyelesaikan permasalahan dalam mengekstrak dan menganalisa data tidak terstruktur yang kerap dijumpai dalam surat kabar dan majalah digital, situs pembelajaran online, website pemerintah, dan situs-situs sejenis.

#### **2.2.4.     *Stanford Natural Language Processing* (NLP)**

*Natural Language Processing* (NLP) adalah bidang dari ilmu komputer, kecerdasan buatan, dan komputasi bahasa yang melibatkan interaksi antara komputer dengan bahasa manusia. Tantangan dari ilmu NLP adalah untuk memahami bahasa manusia, membuat komputer dapat menarik makna dari data masukkan yang berupa bahasa manusia. Untuk dapat melakukan penggalan opini dalam suatu dokumen, maka dibutuhkan beberapa proses awal untuk membuat struktur teks dan melakukan ekstraksi fitur dari dokumen-dokumen yang teridentifikasi. Bentuk tugas-tugas dari NLP antara lain tokenisasi, segmentasi kata-kata, Part of Speech tagging, dan parsing.

Tokenisasi merupakan teknik paling fundamental dalam NLP [28]. Tokenisasi memecah kalimat atau dokumen ke dalam wujud token (kata atau frasa). Dalam tokenisasi, *stop words* akan dihilangkan karena tidak memiliki makna berarti.

*Part of Speech* (POS) tagging merupakan proses untuk mengidentifikasi kata dari suatu kata tersebut dan menganalisa hubungan antara satu kata dengan kata lainnya [29]. Sederhananya proses ini mengkategorikan apakah suatu kata tergolong kata benda, kata kerja, kata sifat, dan lain-lain.

Stanford CoreNLP merupakan *library* Java yang dikembangkan oleh kelompok periset NLP di Universitas Stanford. *Library* ini menyediakan sekumpulan alat analisis yang dapat mengidentifikasi bentuk dasar dari sebuah kata, mengidentifikasi *Part of Speech*, mengidentifikasi struktur dari kalimat berupa frasa dan kata-kata, dependensi, analisa sentimen, dan seterusnya [30]. Stanford CoreNLP juga menyediakan sejumlah anotator yang dapat digunakan sesuai dengan keperluan analisa bahasa. Tabel berikut merupakan anotator yang disediakan oleh Stanford CoreNLP.

*Tabel 7 Anotator yang disediakan oleh Stanford NLP [30]*

| Nama Anotator | Dependensi | Deskripsi                                                         |
|---------------|------------|-------------------------------------------------------------------|
| Tokenize      | Tidak ada  | Melakukan tokenisasi terhadap teks yang dimasukkan sebagai input. |
| CleanXML      | Tokenize   | Menghilangkan token XML dari dokumen                              |
| Ssplit        | Tokenize   | Memecah belah token ke dalam kalimat                              |

| Nama Anotator | Dependensi                   | Deskripsi                                                                                                                                                                                               |
|---------------|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| POS           | Tokenize, Ssplit             | Memberi label pada token dengan POS tag.                                                                                                                                                                |
| Lemma         | Tokenize, Ssplit, POS        | Menghasilkan bentuk baku dari semua token pada korpus                                                                                                                                                   |
| NER           | Tokenize, Ssplit, POS, Lemma | Mengidentifikasi entitas bernama (orang, lokasi, organisasi, dan lain-lain), numerik (uang, angka, ordinal, persentase), dan temporal (waktu, tanggal, durasi, himpunan)                                |
| RegexNER      | ?                            | Implementasi NER menggunakan Java Regex                                                                                                                                                                 |
| Sentiment     | ?                            | Implementasi model sentimen Socher et.al. [31]                                                                                                                                                          |
| TrueCase      | ?                            | Mengidentifikasi bentuk kapital asli dari token-token yang terdapat dalam teks dimana informasi tentang bentuk kapitalnya tidak ada (contoh: pada teks yang hurufnya ditulis dalam huruf kapital saja). |
| Parse         | Tokenize, Ssplit             | Menyediakan analisis sintatik secara lengkap dengan memberikan representasi dependensi tiap kata kemudian menampilkannya menggunakan anotasi tree.                                                      |

| Nama Anotator | Dependensi                                          | Deskripsi                                                                                                                             |
|---------------|-----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| DepParse      | Tokenize, Ssplit, POS                               | Menyediakan parser dependensi sintatik yang cepat.                                                                                    |
| DCoref        | Tokenize, Ssplit, POS, Lemma, NER, Parse            | Mengimplementasikan resolusi <i>coreference</i> (hubungan antara beberapa istilah yang memiliki rujukan sama) pronominal dan nominal. |
| Relation      | Tokenize, Ssplit, POS, Lemma, NER, DepParse         | Menemukan relasi terhadap dua entitas berbeda.                                                                                        |
| NatLog        | Tokenize, Ssplit, POS, Lemma, NER, Parse / DepParse | Memberikan penilaian cakupan dan polaritas token menurut semantik logika natural.                                                     |
| Quote         | Tidak ada                                           | Mengambil karakter petik secara deterministik keluar dari teks.                                                                       |

Stanford NLP, selain tersedia dalam bentuk *source code*, juga tersedia *service* sebagaimana yang ada pada <http://nlp.stanford.edu:8080/corenlp/>.

### 2.2.5. *Part of Speech (POS) Tagging*

*Part of Speech (POS) Tagging* merupakan proses untuk membaca teks dalam suatu bahasa dan memberi label bagian mendasar dari kalimat (*part of speech*) untuk masing-masing kata. Kata-kata tersebut akan dikategorisasikan ke dalam bentuk kata-kata berikut.

Tabel 8 Macam-macam Part of Speech dalam Bahasa Inggris [32]

| Bentuk Kata                                                                                                                                                                                                                                                                                                                                 |  | Contoh                                                                                                                                                                                                                                                                                             |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Kata Benda</b>                                                                                                                                                                                                                                                                                                                           |  |                                                                                                                                                                                                                                                                                                    |
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- NN Noun, singular or mass</li> <li>- NNS Noun, plural</li> </ul>                                                                                                                                                                                              |  |                                                                                                                                                                                                                                                                                                    |
| Part of speech ini digunakan untuk menamai orang, benda, hewan, tempat, dan konsep abstrak. Macam-macam noun antara lain: <ul style="list-style-type: none"> <li>- <i>Countable dan Uncountable Noun</i></li> <li>- <i>Proper dan Common Noun</i></li> <li>- <i>Abstract dan Concrete Noun</i></li> <li>- <i>Collective Noun</i></li> </ul> |  | <ul style="list-style-type: none"> <li>- <i>book, house, car, love</i></li> <li>- <i>Your book is on the table.</i><br/>(Bukumu di atas meja.)</li> </ul>                                                                                                                                          |
| <b>Kata Ganti</b>                                                                                                                                                                                                                                                                                                                           |  |                                                                                                                                                                                                                                                                                                    |
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- NNP Proper noun, singular</li> <li>- NNPS Proper noun, plural</li> </ul>                                                                                                                                                                                      |  |                                                                                                                                                                                                                                                                                                    |
| Pronoun digunakan untuk menggantikan noun.<br><br>Part of speech ini bermanfaat untuk menghindari repetisi penggunaan noun. Macam-macam pronoun antara lain: <ul style="list-style-type: none"> <li>- <i>Personal Pronoun (I, you, they, she)</i></li> </ul>                                                                                |  | <ul style="list-style-type: none"> <li>- <i>It is on the table.</i><br/>(Itu di atas meja.)</li> <li>- <i>This is your cake.</i><br/>(Ini kuemu.)</li> <li>- <i>He didn't blame himself for the accident.</i><br/>(Dia tidak menyalahkan dirinya sendiri terhadap kecelakaan tersebut.)</li> </ul> |

| Bentuk Kata                                                                                                                                                                                                                                                                                                                                                                                                                              |  | Contoh                                                                                                                                                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>- <i>Demonstrative Pronoun (this, those)</i></li> <li>- <i>Interrogative Pronoun (who, what, which)</i></li> <li>- <i>Relative Pronoun (who, which, that)</i></li> <li>- <i>Indefinite Pronoun (anything, none)</i></li> <li>- <i>Reflexive Pronoun dan Intensive Pronoun (myself, yourself)</i></li> <li>- <i>Reciprocal Pronoun (each other, one another)</i></li> </ul>                        |  | <ul style="list-style-type: none"> <li>- <i>Friends help <b>each other</b>.</i><br/>(Teman membantu satu sama lain.)</li> </ul>                                                                |
| <b>Kata Kerja</b>                                                                                                                                                                                                                                                                                                                                                                                                                        |  |                                                                                                                                                                                                |
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- VB Verb, base form</li> <li>- VBD Verb, past tense</li> <li>- VBG Verb, gerund or present participle</li> <li>- VBN Verb, past participle</li> <li>- VBP Verb, non-3rd person singular present</li> <li>- VBZ Verb, 3rd person singular present</li> </ul>                                                                                                 |  |                                                                                                                                                                                                |
| Part of speech ini digunakan untuk menunjukkan tindakan dari subject, menunjukkan peristiwa, atau keadaan. Macam-macam verb antara lain: <ul style="list-style-type: none"> <li>- <i>Transitive dan Intransitive Verb</i></li> <li>- <i>Regular dan Irregular Verb</i></li> <li>- <i>Action dan Stative Verb</i></li> <li>- <i>Finite dan Non-Finite Verb</i></li> <li>- <i>Linking Verb</i></li> <li>- <i>Causative Verb</i></li> </ul> |  | <ul style="list-style-type: none"> <li>- <i>hit, touch, write</i></li> <li>- <i>You <b>hit</b> the nail on the head.</i><br/>(idiom: melakukan sesuatu dengan cara paling efektif.)</li> </ul> |
| <b>Kata Sifat</b>                                                                                                                                                                                                                                                                                                                                                                                                                        |  |                                                                                                                                                                                                |

| Bentuk Kata                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Contoh                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- JJ Adjective</li> <li>- JJR Adjective, comparative</li> <li>- JJS Adjective, superlative</li> <li>- LS List item marker</li> </ul>                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                   |
| Part of speech ini digunakan untuk menerangkan noun atau pronoun.                                                                                                                                                                                                                                                                                                                                                                                                                    | <ul style="list-style-type: none"> <li>- Vina is a <b>beautiful</b> girl.<br/>(Vina adalah gadis yang cantik.)</li> <li>- I am not <b>angry</b> with you.<br/>(Saya tidak marah padamu.)</li> </ul>                                                                               |
| <b>Kata Keterangan</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                   |
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- RB Adverb</li> <li>- RBR Adverb, comparative</li> <li>- RBS Adverb, superlative</li> </ul>                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                   |
| Part of speech ini digunakan untuk menerangkan verb, adjective, maupun adverb lain. Macam-macam adverb antara lain: <ul style="list-style-type: none"> <li>- <i>adverb of time</i> (yesterday, now)</li> <li>- <i>adverb of manner</i> (softly, quickly)</li> <li>- <i>adverb of degree</i> (very, so)</li> <li>- <i>adverb of modality</i> (likely, maybe)</li> <li>- <i>adverb of frequency</i> (always, sometimes)</li> <li>- <i>adverb of place</i> (here, somewhere)</li> </ul> | <ul style="list-style-type: none"> <li>- They met by chance <b>yesterday</b>.<br/>(Mereka bertemu tidak sengaja kemarin.)</li> <li>- He is not <b>here</b>.<br/>(Dia tidak disini)</li> <li>- I go fishing <b>sometimes</b>.<br/>(Saya pergi memancing kadang-kadang.)</li> </ul> |

| Bentuk Kata                                                                                                                                                                                                                                                                                                                                                                                                                               |  | Contoh                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| - <i>adverb of focus (also, only)</i>                                                                                                                                                                                                                                                                                                                                                                                                     |  |                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Preposisi (Kata Depan)</b>                                                                                                                                                                                                                                                                                                                                                                                                             |  |                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- IN Preposition or subordinating conjunction</li> </ul>                                                                                                                                                                                                                                                                                                      |  |                                                                                                                                                                                                                                                                                                                                                                    |
| Part of speech ini dikombinasikan dengan noun atau pronoun untuk membentuk phrase (prepositional phrase) yang menerangkan verb, noun, atau adjective                                                                                                                                                                                                                                                                                      |  | <ul style="list-style-type: none"> <li>- <i><b>in</b> the classroom, <b>on</b> the floor</i></li> <li>- <i>Is sleeping <b>on</b> the floor good for my back?</i><br/>(Apakah tidur di lantai baik untuk punggung saya?)</li> </ul>                                                                                                                                 |
| <b>Konjungsi (Kata Sambung)</b>                                                                                                                                                                                                                                                                                                                                                                                                           |  |                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Standford NLP Part of Speech:</b> <ul style="list-style-type: none"> <li>- CC Coordinating conjunction</li> </ul>                                                                                                                                                                                                                                                                                                                      |  |                                                                                                                                                                                                                                                                                                                                                                    |
| Part of speech ini digunakan untuk menghubungkan dua kata, frasa (phrase), klausa (clause), atau paragraf (paragraph). Kata ini terbagi menjadi: <ul style="list-style-type: none"> <li>- coordinate conjunction (and, but, yet)</li> <li>- correlative conjunction (either...or, rather...than, if...then)</li> <li>- subordinate conjunction (after, although, when)</li> <li>- adverbial conjunction (however, hence, then)</li> </ul> |  | <ul style="list-style-type: none"> <li>- <i>I like that car, <b>but</b> I don't have enough money to buy it.</i><br/>(Saya suka mobil itu, tapi saya tidak punya cukup uang untuk membelinya.)</li> <li>- <i>He would <b>rather</b> send a text message <b>than</b> make a phone call.</i><br/>(Dia lebih suka mengirim pesan text daripada menelepon.)</li> </ul> |

| Bentuk Kata                                                               |   | Contoh                                    |
|---------------------------------------------------------------------------|---|-------------------------------------------|
| Interjeksi (Kata Seru)                                                    |   |                                           |
| Stanford NLP <i>Part of Speech</i> :                                      |   |                                           |
| - UH Interjection                                                         |   |                                           |
| Interjection merupakan ucapan pendek digunakan untuk mengungkapkan emosi. | - | “Alas, he failed.”<br>- “Oy! Look at me!” |

Berikut ini adalah contoh implementasi *Part of Speech Tagging* pada salah satu ulasan pelanggan mengenai barang di *marketplace* online.

Part-of-Speech:

|   |                                                                                                      |
|---|------------------------------------------------------------------------------------------------------|
| 1 | So far I like it ... I bough a old beat up radio on CL for \$ 20 and restored it and added the Echo. |
| 2 | Attached you see the end result and people compliment me on it when they see it .                    |

Bagan 11 Contoh Implementasi *Part of Speech* pada ulasan pelanggan terhadap barang di *marketplace* online [30]

2.2.6. Latent Dirichlet Allocation

*Latent Dirichlet Allocation* (LDA) merupakan model generatif probabilistik untuk kumpulan data diskrit seperti korpus teks. LDA merupakan model Bayesian Hirarki tingkat tiga dimana setiap item pada himpunan data dimodelkan sebagai hasil dari penurunan dari beberapa topik [11].

Blei et.al. menuturkan bahwa LDA dikembangkan untuk menemukan deskripsi singkat yang memungkinkan efisiensi pemrosesan dokumen dalam jumlah besar serta menyatakan hubungan statistik sehingga dapat digunakan untuk klasifikasi,

sumarisasi, serta penilaian terhadap kesamaan dan relevansi topik yang terkandung dalam kumpulan dokumen.

Agar tidak bias, Blei et.al. juga mendefinisikan istilah penting terkait dengan model LDA. Berikut adalah definisi formal yang dipaparkan Blei et.al.

- **Kata** adalah satuan terkecil dalam data diskrit. Kata didefinisikan sebagai suatu item dari kumpulan kosakata  $\{1, \dots, V\}$ . Kata ke- $v$  dalam kosakata direpresentasikan dengan vektor  $w$  sebanyak  $V$  sedemikian sehingga  $w^v = 1$  dan  $w^u = 0$  dimana  $u \neq v$ .
- **Dokumen** adalah barisan  $N$  kata yang dinotasikan dengan  $w = (w_1, w_2, \dots, w_n)$ , dimana  $w_n$  merupakan kata ke  $n$  dalam suatu barisan kata.
- **Korpus** adalah kumpulan dari  $M$  dokumen yang dinotasikan dengan  $D = \{w_1, w_2, \dots, w_m\}$

LDA memiliki proses generatif sebagai berikut.

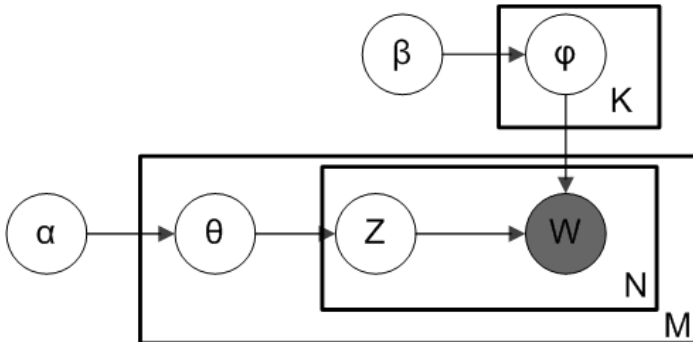
1. Untuk setiap dokumen  $w$  dalam korpus  $D$ 
  - a. Pilih  $N \sim \text{Poisson}(\xi)$
  - b. Pilih  $\theta \sim \text{Dir}(\alpha)$
2. Untuk setiap  $N$  kata dalam dokumen  $w_n$ :
  - a. Pilih topik  $z_n \sim \text{Multinomial}(\theta)$ .
  - b. Pilih kata  $w_n$  dari  $p(w_n | z_n, \beta)$

Jika ditulis dalam notasi matematika maka hasilnya dapat terlihat pada persamaan 1.

*Persamaan 1 Notasi Matematika dari LDA*

$$p(w, z, \theta, \phi | \alpha, \beta) = p(\theta | \alpha) p(z | \theta) p(\phi | \beta) p(w | z, \phi)$$

Gambar berikut adalah model LDA jika digambarkan dalam notasi pelat.



Bagan 12 Notasi Pelat LDA [14]

**Dimana:**

$N$  = Jumlah kata dalam dokumen ( $N_{(m)}$ )

$M$  = Jumlah dokumen teridentifikasi

$K$  = Jumlah topik teridentifikasi

$W_{(n,m)}$  = kata-kata yang diobservasi dalam suatu korpus teks

$Z_{(n,m)}$  = topik yang melekat pada kata

$\Theta_{(m)}$  = proporsi topik per dokumen yang teridentifikasi

$\Phi_{(k)}$  = distribusi topik pada vocabulary

$\alpha, \beta$  = parameter *dirichlet*

### 2.2.7. *Appraisal Expression Pattern*

*Appraisal Expression Pattern* (AEP) adalah sebuah istilah yang dikembangkan oleh Xiaolin Zheng et.al. sebagai satuan pengetahuan linguistik yang merangkum ekspresi penilaian atas suatu fitur dari objek yang dibicarakan [14].

Untuk memahami ekspresi penilaian atas suatu fitur dari objek yang dibicarakan, seseorang perlu tahu hubungan antara aspek



Untuk menghubungkan antara kata aspek dengan kata sentimen, Xiaolin Zheng et.al. mengembangkan *Shortest Dependency Path* (SDP), yakni jalur terpendek yang menghubungkan antara dua kata dalam grafik dependensi. Tujuan dibuatnya SDP adalah untuk menjalin hubungan antara aspek dan sentimen melalui jalur terpendek.

Nilai dari SDP didapatkan melalui kegiatan *unsupervised learning*. Langkah pertama dalam menentukan SDP adalah untuk menentukan kata-kata aspek dan sentimen kandidat. Umumnya kata aspek direpresentasikan dalam kata *nouns* (contoh: “*Amazon Echo*”, “*speaker*”) atau *pronouns* (contoh: “*it*”) dalam sebuah kalimat. Sehingga semua *nouns* dan *pronouns* diperlakukan sebagai kandidat kata aspek. Sementara kata-kata sentimen umumnya berupa *adjective* (contoh: “*good*”, “*noisy*”). Kata-kata sentimen ini kemudian dibandingkan dengan *lexicon sentiment* untuk menentukan bobot sentimen. Namun dalam studi kasus Tugas Akhir ini, sentimen didapatkan dengan melihat probabilitas seberapa banyak kata-kata sentimen diulang dalam dokumen (ulasan produk). Jumlah SDP yang mungkin terjadi dalam suatu dokumen dapat dihitung dengan mengalikan jumlah kata aspek dengan jumlah kata sentimen yang teridentifikasi pada dokumen sebagaimana yang terlihat pada persamaan 1.

*Persamaan 2 Formula untuk menghitung Jumlah Shortest Dependency Path pada suatu dokumen*

$$SDP = m \times n$$

### **Dimana:**

*SDP* = jumlah *Shortest Dependency Path* dalam dokumen

*m* = jumlah kata aspek teridentifikasi dalam dokumen

$n$  = jumlah kata sentimen teridentifikasi dalam dokumen

Untuk mengetahui seberapa banyak atau sering SDP muncul dalam suatu dokumen, Xiaolin Zheng et.al. juga mengembangkan parameter ukuran dinamakan *Confidence Score* (CS). Nilai CS merepresentasikan seberapa sering SDP muncul dalam suatu jalur dependensi. Nilai CS juga menentukan apakah suatu AEP bermakna atau tidak. Semakin besar nilai CS, maka AEP semakin bermakna. Persamaan 2 menggambarkan notasi untuk menghitung nilai CS.

*Persamaan 3 Formula untuk menghitung Confidence Score*

$$P(f_i) = \frac{\max\{\log(f_i) - \log(f_{thr}), 0\}}{\log(f_{max}) - \log(f_{thr})}$$

**Dimana:**

$P(f_i)$  = *Confidence Score*

$f_i$  = Frekuensi SDP ke  $i$

$f_{max}$  = Frekuensi SDP maksimal dalam himpunan jalur dependensi

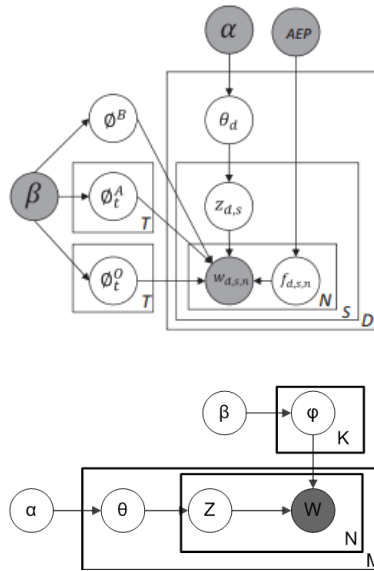
$f_{thr}$  = Nilai ambang yang menentukan frekuensi minimal jalur dependensi

Dalam persamaan 2, nilai denominator merupakan faktor normalisasi dari fungsi  $P(f_i)$ . Sementara nilai numeratornya merupakan nilai  $f_i$  dibandingkan dengan nilai  $f_{thr}$ . Jika frekuensi  $f_i$  lebih rendah dari  $f_{thr}$  maka SDP ke- $i$  jarang terlihat dalam himpunan jalur dependensi. Untuk menentukan nilai  $f_{thr}$ , maka asumsikan nilai  $P(f_i) = 0$ , kemudian lakukan iterasi. Maka akan di dapat nilai  $f_i$ . Nilai  $f_i$  terendah yang kemudian dijadikan nilai ambang.

AEP merupakan satuan pengetahuan linguistik yang merangkum ekspresi penilaian atas suatu fitur dari objek yang dibicarakan. AEP terdiri atas dua komponen yang saling berhubungan, yakni 1) SDP yang tidak lain adalah representasi umum tentang bagaimana seseorang mengekspresikan opini terhadap aspek suatu produk atau layanan. Kemudian 2) CS yang mengindikasikan nilai *confidence* terhadap SDP. Sehingga  $AEP = \langle SDP, CS \rangle$ .

AEP-LDA, sebagaimana yang dijelaskan oleh Xiaolin Zheng et.al. merupakan pengembangan dari LDA. Kesamaan AEP-LDA dengan LDA terletak pada bahwasannya keduanya melibatkan hubungan dependensi antar kata dan menggunakan informasi POS untuk merepresentasikan asosiasi antara kata aspek dengan kata sentimen. Sementara perbedaan kuncinya adalah sebagai berikut:

1. AEP diekstrak melalui *unsupervised learning*, sementara LDA biasa melalui *supervised learning*.
2. AEP digunakan untuk menganalisa perilaku pelanggan dalam mengekspresikan opini terhadap aspek suatu produk atau jasa. Sementara LDA biasa hanya dapat menentukan aturan gramatikal untuk mengidentifikasi kata fitur dan kata opini dalam kumpulan ulasan pelanggan.



Bagan 14 Notasi pelat AEP-LDA (atas) dibandingkan dengan LDA (bawah) [14]

Notasi pelat untuk AEP-LDA adalah sebagaimana yang terlihat pada gambar 12 (atas). Model AEP-LDA merupakan pengembangan dari model LDA. AEP-LDA mengasumsikan bahwa sejumlah  $T$  topik aspek dipertimbangkan dalam ulasan masing-masing produk atau layanan, dan pada tiap-tiap topik aspek terdapat beberapa kata-kata konsep aspek yang koheren antara yang satu dengan lainnya. Sebagai contoh kata “Amazon Echo” memiliki topik aspek berupa “voice” dan “price”. Selain itu, masing-masing topik aspek juga berkorespondensi dengan kata-kata sentimen yang mengikuti. Masing-masing topik kata sentimen mengandung kata sentimen yang merujuk pada aspek. Lebih jauh lagi, topik *background word* juga dipertimbangkan dalam model untuk mengumpulkan *background words* yang muncul dalam ulasan.

Proses generatif daripada model AEP-LDA dapat dijelaskan sebagai berikut:

1. Untuk setiap dokumen ulasan  $d$ :
  - Sebuah distribusi topik  $\theta_d \sim \text{Dir}(\alpha)$  dibuat.
2. Untuk setiap kalimat  $s$  dalam dokumen  $d$ :
  - Sebuah penugasan topik  $z_{d,s} \sim \text{Multi}(\theta_d)$  dibuat.
3. Untuk setiap kata  $n$  dalam kalimat  $s$  dalam dokumen  $d$ :
  - $w_{d,s,n}$  dibuat dalam basis  $f_{d,s,n}$ .

*Persamaan 4 Persamaan Multinomial untuk distribusi kata dalam kalimat  $s$  dalam dokumen  $d$*

$$w_{d,s,n} \sim \begin{cases} \text{Multi}(\emptyset^B), & \text{jika } f_{d,s,n} = 0 \\ \text{Multi}(\emptyset_t^A), & \text{jika } f_{d,s,n} = 1 \\ \text{Multi}(\emptyset_t^O), & \text{jika } f_{d,s,n} = 2 \end{cases}$$

Setiap kata dalam kalimat  $s$  dalam dokumen  $d$  bisa jadi berupa kata-kata aspek, kata-kata sentimen, atau *background words*. Sehingga  $f_{d,s,n}$  dibuat untuk setiap kata  $w_{d,s,n}$  untuk menentukan tipe  $w_{d,s,n}$ .  $f_{d,s,n}$  dibuat dari distribusi multinomial dari himpunan  $\{0,1,2\}$  dimana elemen  $0,1,2$  masing-masing merepresentasikan *background words*, kata aspek, dan kata sentimen. Karena setiap kata aspek berupa kata benda (*nouns*) sementara kata sentimen merupakan kata sifat (*adjective*), maka perlu diketahui hubungan dependensi untuk setiap kata aspek dan kata sentimen yang muncul.

Untuk semua kalimat dalam dokumen, semua pasangan kata *noun-adjective* diekstrak sebagai *tuple* (baris) aspek-opini yang dinotasikan dengan  $t = \langle n, h \rangle$ , dimana  $n$  atau  $h$  dapat berupa kandidat kata aspek atau kandidat kata sentimen. Sebuah kalimat dapat memiliki lebih dari satu *tuple* aspek-opini, dan sebuah kata dalam kalimat dapat muncul dalam lebih dari satu kandidat *tuple* aspek-opini. Notasi matematika untuk batasan di

atas adalah sebagaimana yang digambarkan dalam persamaan 4, 5, dan 6.

*Persamaan 5 Probabilitas kata aspek*

$$P(f_{d,s,n} = 1|AEP) \propto \max\{CS|SDP_h^n\}$$

*Persamaan 6 Probabilitas kata opini*

$$P(f_{d,s,n} = 2|AEP) \propto \max\{CS|SDP_n^h\}$$

*Persamaan 7 Probabilitas background words*

$$P(f_{d,s,n} = 0|AEP) \propto 1 - \max\{P(f_{d,s,n} = 1|AEP), P(f_{d,s,n} = 2|AEP)\}$$

Dimana:

$SDP_h^n$  = SDP antara kata ke- $n$  (kandidat kata sentimen) dan kata ke- $h$  (kandidat kata aspek) dalam kalimat  $s$  dalam dokumen  $d$ .

$SDP_n^h$  = SDP antara kata ke- $h$  (kandidat kata aspek) dan kata ke- $n$  (kandidat kata sentimen) dalam kalimat  $s$  dalam dokumen  $d$ .

### 2.2.9. *Gibbs Sampling*

*Gibbs sampling* adalah algoritma Markov chain Monte Carlo untuk memperoleh sampel dari distribusi probabilitas multivariat. *Gibbs sampling* mengambil sampel untuk membangkitkan nilai sampel berikutnya secara acak. Tujuan utama dari implementasi *Gibbs sampling* adalah untuk menemukan bentuk distribusi univariat. Dengan kata lain, setelah beberapa kali iterasi melalui rantai distribusi, *sampling* dari distribusi tersebut akan mendekati dengan *posterior* yang diinginkan.

Sebagai contoh, untuk melakukan *sampling* terhadap nilai  $x$  pada distribusi  $p(x) = p(x_1, \dots, x_m)$ , dimana tidak ada solusi utuh untuk  $p(x)$  namun terdapat hasil yang representatif, maka hasil

tersebut dapat ditemukan dengan menggunakan *Gibbs Sampling* dengan langkah-langkah sebagai berikut.

1. Buat variabel  $x_i$  secara acak.
2. Untuk  $t = 1, \dots, T$ :
  - 2.1.  $x_1^{t+1} \sim p(x_1 | x_2^{(t)}, x_3^{(t)}, \dots, x_m^{(t)})$
  - 2.2.  $x_2^{t+1} \sim p(x_2 | x_1^{(t+1)}, x_3^{(t)}, \dots, x_m^{(t)})$
  - 2.3.  $x_m^{t+1} \sim p(x_m | x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_{m-1}^{(t+1)})$

Prosedur ini diulang hingga nilai sample mendekati nilai distribusi sebenarnya. Dalam kasus LDA, nilai yang ingin diuji adalah porsi dokumen topik ( $\theta_d$ ), distribusi topik-kata  $\varphi^{(z)}$ , dan penugasan topik untuk tiap kata  $z_i$ . Rumus dari permasalahan LDA diturunkan menurut batasan *Gibbs Sampling* dan hasilnya adalah sebagai berikut:

Masukkan: kata  $w$  dalam dokumen  $d$   
 Luaran: penugasan topik  $z$  dan jumlah  $n_{d,k}$ ,  $n_{k,w}$ ,  $n_k$ .

Mulai:

Buat  $z$  secara acak dan naikan variabel increment

**Untuk** setiap iterasi, lakukan

*Word* <-  $w[i]$

*Topik* <-  $z[i]$

$n_{d, \text{topic}} = 1$ ;  $n_{\text{word}, \text{topic}} = 1$ ;  $n_{\text{topic}} = 1$

**Untuk**  $k = 0 \rightarrow K-1$ , lakukan

$$p(z=k) = (n_{d,k} + \alpha_k) \frac{n_{k,w} + \beta_w}{n_k + \beta \times W}$$

**Selesai**

*Topic* <- sample dari  $p(z)$

*Z[i]* <- *Topic*

$n_{d,topic} += 1; n_{word,topic} += 1; n_{topic} += 1$

**Selesai**

**Return**  $\mathbf{z}, n_{d,k}, n_{k,w}, n_k$ .

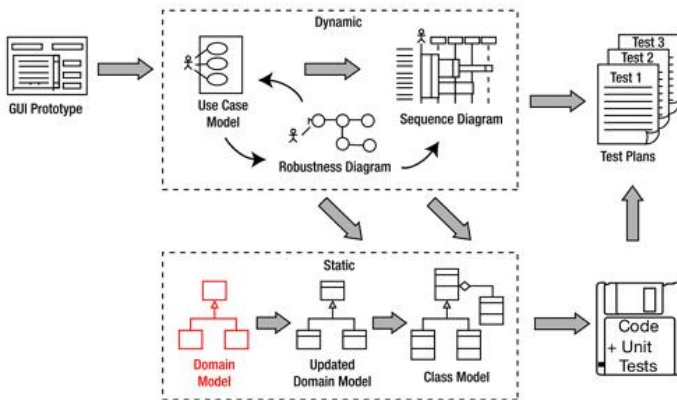
**Selesai**

*Bagan 15 Pseudo-code untuk Gibbs Sampling*

### 2.2.10. ICONIX

ICONIX merupakan metodologi pengembangan perangkat lunak yang mengakomodasi teknik *Rational Unified Process*, *Extreme Programming*, dan pengembangan perangkat lunak secara *Agile* [33]. ICONIX menekankan analisis *robustness*, sebuah metode untuk menjembatani kesenjangan antara proses analisis dan desain. Analisis *robustness* bertujuan untuk mengurangi keambiguitas dalam deskripsi *use case*.

Proses-proses yang terkandung dalam ICONIX terbagi ke dalam 4 tahapan yakni mengulas kebutuhan, mengulas desain awal, mengulas desain yang telah di detailkan, dan tahap pengembangan. Adapun penjelasan dari 4 tahapan tersebut adalah sebagai berikut.



Bagan 16 Proses dalam ICONIX [33]

### 2.2.10.1. Mengulas Kebutuhan

Tahapan ini merupakan langkah awal dalam pengembangan perangkat lunak dimana kebutuhan akan digali dan di analisa. Hasil dari penggalian kebutuhan berupa domain model dan prototipe antara muka sistem.

Prototipe antar muka dibuat untuk memudahkan proses pembuatan *domain model* dan *use case*. Domain model dapat digunakan untuk mengembangkan *use case*. *Use case* dikembangkan dengan mengidentifikasi aktor dan *use case* bisnis.

### 2.2.10.2. Mengulas Desain Awal

Setelah *use case* berhasil diidentifikasi, penjelasan mengenai *use case* bagaimana pengguna dengan sistem berinteraksi. Untuk menjembatani antara kebutuhan analisis dengan desain, maka digunakan analisa *robustness*. Analisa *robustness* digunakan untuk menemukan potensi kesalahan dalam deskripsi *use case*, kemudian domain model akan diperbaharui

untuk setiap kesalahan deskripsi *use case* yang teridentifikasi. Deskripsi *use case* digunakan untuk mengidentifikasi bagaimana pengguna berinteraksi dengan sistem. Deskripsi *use case* digunakan developer sebagai acuan yang ditujukan pada pelanggan dan memverifikasi analisis kebutuhannya sudah benar.

#### **2.2.10.3. Mengulas Desain yang Didetailkan**

Pada tahapan ini, domain model dan deskripsi *use case* dari langkah mengulas desain awal digunakan sebagai acuan dalam mendesain sistem yang akan dibangun. *Domain model* digunakan untuk membuat *class diagram*, sementara deskripsi *use case* digunakan untuk mengembangkan *sequence diagram*.

#### **2.2.10.4. Pengembangan Perangkat Lunak**

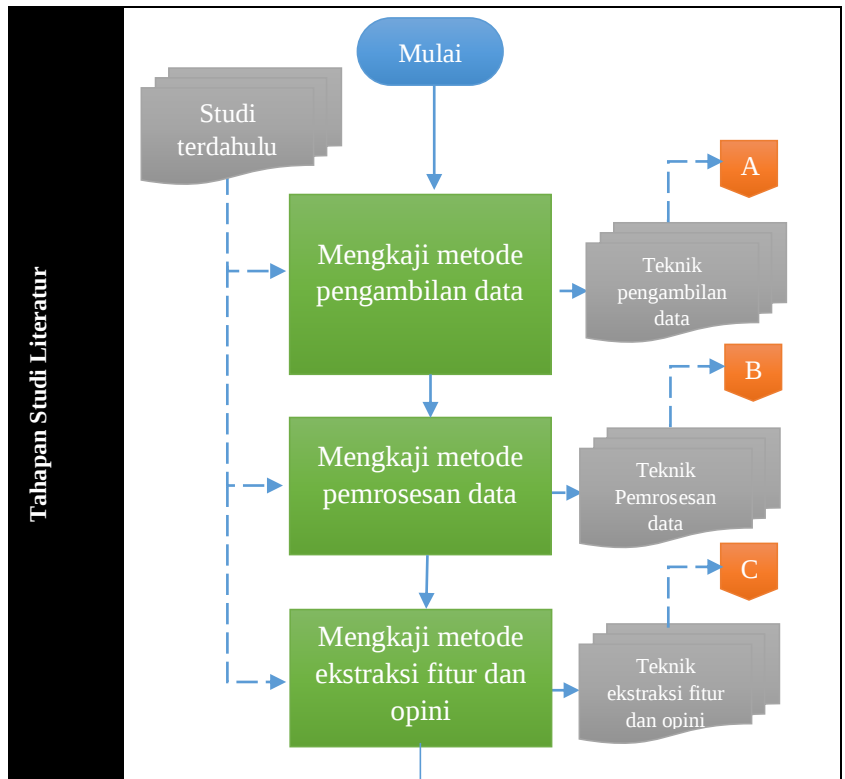
Unit test akan dibuat untuk menguji sistem agar sesuai dengan *use case* yang telah dibuat. Terakhir kode program dibuat dengan mengacu pada *class* dan *sequence diagram*.

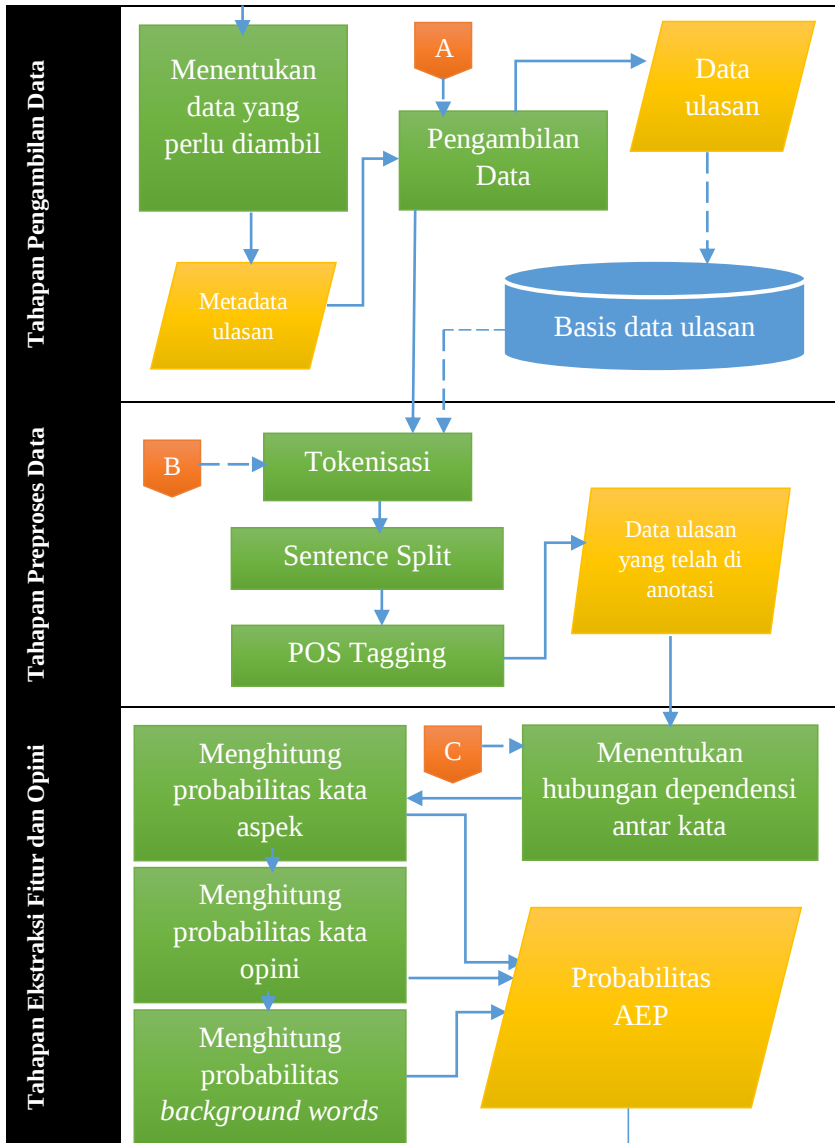
*Halaman ini sengaja dikosongkan*

## BAB III

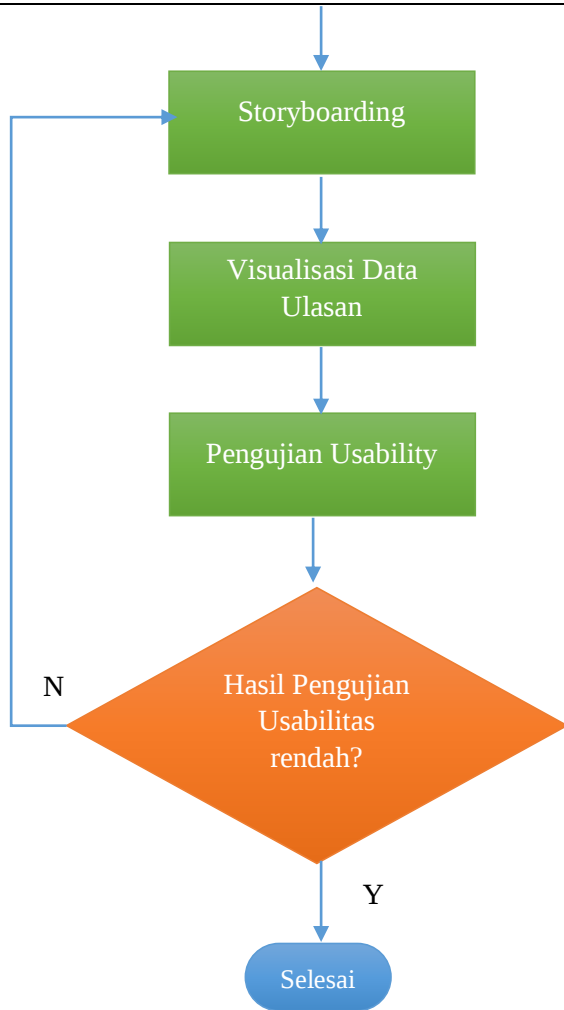
### METODOLOGI

Metodologi merupakan sebuah tahapan dalam penyelesaian permasalahan pada tugas akhir ini. Metodologi dapat digunakan sebagai panduan pengerjaan tugas akhir agar dapat diselesaikan secara terarah, teratur, dan sistematis. Gambar berikut merupakan bagan dari metodologi tugas akhir ini:





## Tahapan Rancang Bangun Prototipe



Bagan 17 Metodologi Tugas Akhir

### **3.1. Tahapan Studi Literatur**

Tahapan ini dilakukan untuk melakukan kajian atas tiga tugas utama dalam melakukan pemodelan topik pada studi kasus ulasan produk di situs Amazon.com. Kajian dilakukan dengan melihat dari berbagai penelitian sebelumnya terkait dengan masing-masing tugas. Tugas tersebut antara lain adalah sebagai berikut.

#### **3.1.1. Mengkaji Metode Pengambilan Data**

Metode pengambilan data pada situs web sangat beragam, sehingga segala teknik yang digunakan untuk melakukan pengambilan data baik secara manual (tanpa otomasi) maupun dengan teknik yang terotomasi. Selanjutnya semua teknik dirangkum sebagai kumpulan teknik pengambilan data untuk dikaji lebih lanjut terhadap studi kasus yang akan dilakukan, yakni pengambilan data pada situs belanja online Amazon.com.

#### **3.1.2. Mengkaji Metode Pemrosesan Data**

Metode pemrosesan data dilakukan untuk mengolah data agar dapat lebih mudah diolah dan juga dapat menghasilkan hasil analisa yang lebih akurat. Metode pemrosesan data sangat beragam, namun untuk studi kasus pada Tugas Akhir ini dititik beratkan pada teknik *Part of Speech Tagging*. Teknik tersebut kemudian dirangkum sebagai teknik pemrosesan data untuk dikaji lebih lanjut terhadap studi kasus yang akan dilakukan, yakni pemrosesan data pada situs belanja online Amazon.com.

#### **3.1.3. Mengkaji Metode Ekstraksi Fitur dan Opini**

Metode ekstraksi fitur dan opini dilakukan untuk memperoleh topik apa sajakah yang dibicarakan oleh konsumen seputar aspek produk beserta sentimen konsumen terhadap aspek yang

dibicarakan. Metode pemrosesan data sangat beragam, namun untuk studi kasus pada Tugas Akhir ini ditekankan pada algoritma *Latent Dirichlet Allocation* beserta cara implementasinya. Teknik tersebut kemudian dirangkum sebagai teknik ekstraksi fitur dan opini untuk dikaji lebih lanjut terhadap studi kasus yang akan dilakukan, yakni ekstraksi fitur dan opini ulasan produk oleh pelanggan pada situs belanja online Amazon.com.

### **3.2. Tahapan Pengambilan Data**

Tahapan ini dimulai dengan mempelajari kajian yang ada kemudian mulai menentukan data apa saja yang perlu diambil. Data yang perlu diambil kemudian didokumentasikan dalam bentuk metadata yang akan digunakan sebagai acuan dalam pengambilan data sekaligus membantu dalam perancangan basis data ulasan. Hasil pengambilan data berupa data ulasan akan ditampung dalam basis data ulasan yang telah dirancang sebelumnya.

### **3.3. Tahapan Preproses Data**

Tahapan ini dilakukan untuk mengetahui *Part of Speech* suatu kata. Tujuannya adalah untuk memahami hubungan antar kata dalam suatu kalimat sekaligus menghilangkan kata-kata yang tidak diperlukan. Kata-kata yang tidak diperlukan merupakan kata selain kata benda, kata sifat, dan kata kerja. Hal tersebut ditetapkan demikian mengingat kata aspek pada umumnya merupakan kata benda atau kata ganti benda, sementara kata opini biasanya berupa kata sifat. Tahapan ini dimulai dengan mempelajari kajian yang sudah dilakukan pada tahapan sebelumnya. Kemudian melakukan tokenisasi untuk setiap data yang muncul ke dalam bentuk kata per kata. Sentence split juga dilakukan karena masukkan dari fungsi LDA salah satunya berupa dokumen yang dipecah dalam bentuk kalimat per

kalimat. Selanjutnya POS tagging dilakukan per token yang teridentifikasi. Data ulasan yang dianotasi digunakan sebagai masukan pada tahapan berikutnya, yakni tahapan ekstraksi fitur dan opini.

### **3.4. Tahapan Ekstraksi Fitur dan Opini**

Tahapan ini dilakukan untuk mengetahui fitur apa sajakah yang dibicarakan pelanggan. Setelah mendapatkan data ulasan yang telah dianotasi dan mempelajari kajian mengenai teknik ekstraksi fitur dan opini, maka langkah selanjutnya adalah menghitung probabilitas kata aspek, kata opini, dan *background words*. Kemudian hasilnya adalah probabilitas *Appraisal Expression Pattern* (AEP), yakni probabilitas linguistik yang digunakan untuk mengetahui fitur apa saja yang dibicarakan oleh pelanggan terhadap aspek suatu produk berikut sentimen dari aspek yang dibicarakan pelanggan.

### **3.5. Tahapan Rancang Bangun Prototipe**

Setelah menyelesaikan hasil analisis, maka langkah selanjutnya adalah perlu membuat prototipe sistem *e-commerce* sederhana yang spesifik terhadap fungsi menampilkan ulasan. Tujuan utama dari Tugas Akhir ini adalah untuk memberikan pengalaman menganalisis produk yang baru dan lebih efisien. Sistem memaparkan kata kunci yang paling sering dibicarakan oleh konsumen terhadap suatu produk sehingga mereka dapat mendapatkan gambaran umum produk alih-alih membaca review secara satu per satu.

Prototipe dibuat dengan membuat draft antar muka pengguna kemudian membuat *use case* sesuai dengan antar muka yang teridentifikasi hingga membuat *class diagram*. Luaran dari tahapan rancang bangun prototipe adalah aplikasi sistem *e-commerce* yang mampu menampilkan aspek berikut opini pelanggan terhadap aspek yang dibicarakan terhadap suatu

produk pada ulasan di situs *e-commerce*. Selanjutnya, aplikasi akan diuji usabilitasnya. Apabila hasil ujian usabilitas lebih rendah dari yang diharapkan maka desain akan diubah sesuai dengan harapan pelanggan. Namun dalam konteks tugas akhir ini, hasil pengujian usabilitas akan digunakan sebagai bahan evaluasi perbaikan sistem di masa yang akan datang.

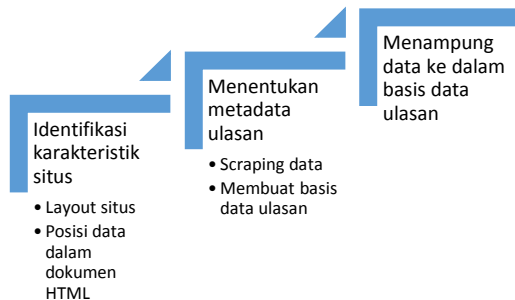
*Halaman ini sengaja dikosongkan*

## BAB IV

### ANALISIS DAN PERANCANGAN

#### 4.1. Pengambilan Data

Untuk melakukan proses pengambilan data, perlu diketahui karakteristik situs yang akan di-*scraping*. Selanjutnya menentukan metadata apa saja yang perlu dicantumkan untuk menyimpan data-data ulasan yang dibutuhkan. Metadata ulasan tersebut yang akan dijadikan acuan dalam melakukan *scraping* data berikut pengembangan basis data ulasan. Gambar 16 menerangkan metodologi pengambilan data ulasan.



Bagan 18 Metodologi Pengambilan Data Ulasan

##### 4.1.1. Identifikasi Karakteristik Situs

Pengambilan data dilakukan dengan teknik *web scraping* yang melibatkan HTML dan DOM *parsing*. Hal ini dikarenakan Amazon.com tidak memiliki API yang dapat diakses oleh siapapun. Untuk memiliki API Amazon, pengembang harus mendaftar sebagai partner Amazon. Pengambilan data dilakukan terhadap ulasan yang ditulis oleh pelanggan pada suatu produk di situs *marketplace* terkemuka Amazon.com.

Tabel 9 menjelaskan detail produk yang ulasannya akan diambil.

*Tabel 9 Deskripsi Produk*

| Aspek         | Deskripsi                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Nama Produk   | Amazon Echo - Black                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Produsen      | Amazon                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Kategori      | Home Audio Speakers                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| URL Produk    | <a href="https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/product-reviews/B00X4WHP5E/ref=cm_cr_arpd_view_opt_rvwr?ie=UTF8&amp;reviewerType=all_reviews&amp;showViewpoints=1&amp;sortBy=helpful&amp;pageNumber=1">https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/product-reviews/B00X4WHP5E/ref=cm_cr_arpd_view_opt_rvwr?ie=UTF8&amp;reviewerType=all_reviews&amp;showViewpoints=1&amp;sortBy=helpful&amp;pageNumber=1</a> |
| Jumlah review | 48.998                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Rating        | 4.4 dari 5                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

Produk Amazon Echo diambil sebagai contoh dari studi kasus ini karena selain ulasan yang dibuat oleh pengguna cukup banyak, masih banyak orang yang belum memahami produk amazon echo secara keseluruhan. Kelima responden yang menjadi subjek pengujian usabilitas kesemuanya juga tidak memiliki pengetahuan mengenai produk Amazon Echo.

Untuk melakukan proses pengambilan data dari ulasan menggunakan teknik HTML dan DOM parsing, perlu dipahami susunan atau tata letak dokumen HTML dari situs web yang dimaksud. Gambar 17 menerangkan tampilan dari halaman ulasan produk pada situs Amazon.com.

#### 4.1.2. Menentukan Metadata Ulasan

Langkah selanjutnya adalah menentukan DOM atau tag-tag HTML tempat melekatnya data yang diperlukan sesuai dengan metadata yang dibuat. Tabel 10 merupakan metadata ulasan yang dikembangkan berdasarkan halaman ulasan produk.

*Tabel 10 Metadata ulasan*

| Attribut                                 | Deskripsi                                                                                                                                                                        |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Judul Ulasan                             | Judul ulasan produk yang ditulis pelanggan. Berisikan keterangan ulasan singkat pelanggan                                                                                        |
| Pengulas                                 | Username pengulas yang memberikan ulasan pada produk                                                                                                                             |
| Tanggal Produk Diulas                    | Tanggal ulasan dibuat                                                                                                                                                            |
| Rating Ulasan                            | Nilai/skor rating produk per ulasan                                                                                                                                              |
| Pengulas merupakan pembeli terverifikasi | Pembeli yang terverifikasi memiliki label <i>verified user</i> . Diasumsikan bahwa <i>verified user</i> memiliki ulasan yang berdampak signifikan terhadap sentimen suatu produk |
| Detail Ulasan                            | Berisikan detail ulasan yang dituliskan pengulas. Berisi tentang keluhan, kekurangan, maupun kelebihan produk.                                                                   |
| Sentimen Ulasan                          | Apakah ulasan memiliki sentimen positif atau negatif                                                                                                                             |

Berdasarkan metadata yang ditetapkan, identifikasi tag-tag HTML atau DOM pada situs untuk mengambil data sesuai

keperluan. Tabel 11 memaparkan pemetaan antara tag HTML DOM dengan metadata yang telah dibuat.

*Tabel 11 Pemetaan Metadata Ulasan dengan Tag HTML DOM*

| Attribut                                 | Tag HTML                                                                                                                                                                             |
|------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Judul Ulasan                             | <pre>&lt;a data-hook="review-title"   class="a-size-base a-link-normal         review-title a-color-base a-         text-bold"&gt; ... &lt;/a&gt;</pre>                              |
| Pengulas                                 | <pre>&lt;a data-hook="review-author"   class="a-size-base a-link-normal         author"&gt; ... &lt;/a&gt;</pre>                                                                     |
| Tanggal Produk Diulas                    | <pre>&lt;span data-hook="review-date"   class="a-size-base a-color-         secondary review-         date"&gt; ... &lt;/span&gt;</pre>                                              |
| Rating Ulasan                            | <pre>&lt;i data-hook="review-star-rating"   class="a-icon a-icon-star a-star-5         review-rating"&gt;   &lt;span class="a-icon-alt"&gt;     ...   &lt;/span&gt; &lt;/i&gt;</pre> |
| Pengulas merupakan pembeli terverifikasi | <pre>&lt;span data-hook="avp-badge"   class="a-size-mini a-color-         state a-text-         bold"&gt; ... &lt;/span&gt;</pre>                                                    |

| Attribut        | Tag HTML                                                                                                                                                                                       |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Detail Ulasan   | <pre>&lt;span data-hook="review-body" class="a-size-base review- text"&gt; ... &lt;/span&gt;</pre>                                                                                             |
| Sentimen Ulasan | <p>Didapat dari query berdasarkan URL</p> <pre>https://www.amazon.com/{nama-produk}/product- reviews/{kode-ASIN} /?reviewerType=all_reviews &amp;pageNumber=1&amp; filterByStar=positive</pre> |

**amazon** all Q stylend WATCH NOW

Departments - Your Amazon.com Today's Deals Gift Cards & Registry Sell Help

Amazon Echo - Black - Customer Reviews

**Customer Reviews** 4.5 out of 5 stars 48,999

**Amazon Echo - Black** by Amazon

Color: Black Change

Price: \$179.99 - Free shipping with Amazon Prime

**Rate this item** Write a review

**Top positive review**  
See all 41,832 positive reviews +

**★★★★★ Uber Awesome. In fact you can even order an Uber ride from it!**  
By Michael S on July 24, 2015

LOVE OUR NEW ECHO!! I have been watching the reviews online and checking with friends that have purchased the Echo to see how much they liked or disliked its features. Last person I talked to went on and on about all the things there were using it for and that persuaded me it was time and Amazon Prime Day was the perfect opportunity to go for it. Amazon did a fantastic job of creating this tubular info-tuning command center. There are so many cool and awesome things it's able to do that it hit the highlights that work for our household. First, we love that it follows your voice in the room (the circle lighting will show which direction it is listening), the speaker is wonderfully balanced, so whether listening to music, the news or to Alexa speaking, I have nothing but high marks for its sound quality, given its size. Next, not on (hardly) need for Alexa to help set the mood in the house, room to room, all from a simple voice request. Even though we are fairly tech-savvy and have phone, tablets and watches that can control devices, (its HUGE) to just ask for something and a pleasant voice setting it in motion. I am a regular Amazon Prime shopper and reviewer, if my review helped you in making up your mind about whether to purchase the Echo, please mark the "helpful" button, it tells Amazon that folks are in fact reading reviews like this. Thanks!

**Top critical review**  
See all 7,187 critical reviews +

**★☆☆☆☆ Disappointing and frustrating - I'm returning Echo**  
By baskin on October 10, 2015

I ordered Echo specifically for music - I've had a huge music library on Amazon for years. But it didn't take long after installing Echo to find out that there is now an annual charge for having over 250 songs, so only a small portion of my library is available. But OK - since subscription is to be expected these days, I decided to pare down my library down to fit and get Echo to play only the songs that we like. But I found Alexa will not play them. When I ask to play something from my music library, it plays unfamiliar songs that I know are not in there. If I ask for one of my playlists instead, more often than not she'll just start playing crappy Prime music.

**Questions? Get fast answers from reviewers**  
What do you want to know about Amazon Echo - Black?  
See all 100+ answered questions Ask

**Customers also viewed these items**

- All-New Echo Dot (2nd Generation) - Black by Amazon \$49.99 ★★★★★ 10,284
- Amazon Tap - Alexa Enabled Portable Bluetooth Speaker by Amazon \$129.99 ★★★★★ 4,369
- Amazon Echo Dot Case (fits Echo Dot 2nd Generation only) - Indigo Fabric by Amazon \$14.99

**58 comments** 7,220 people found this helpful. Was this review helpful to you? Yes No Report abuse

**★★★★★ Already very practical for overcoming disability issues**

Bagan 19 Tampilan halaman ulasan produk pada situs Amazon

#### 4.1.3. Perancangan Basis Data Ulasan

Setelah mengetahui data ulasan apa saja yang ingin diolah, selanjutnya dilakukan proses perancangan basis data ulasan untuk menampung data-data ulasan yang diperlukan untuk kebutuhan analisis lebih lanjut. Berikut adalah skema basis data untuk menampung hasil *web scraping* terhadap data ulasan produk di situs Amazon.com.

*Tabel 12 Deskripsi Tabel Ulasan*

| Tabel: Review     |              |                                                                                                                              |
|-------------------|--------------|------------------------------------------------------------------------------------------------------------------------------|
| Attribut          | Tipe Data    | Deskripsi                                                                                                                    |
| Review_title      | VARCHAR(100) | Judul ulasan produk yang ditulis pelanggan. Berisikan keterangan ulasan singkat pelanggan                                    |
| Review_title-href | VARCHAR(100) | Link judul review                                                                                                            |
| Reviewer          | VARCHAR(100) | Username pengulas yang memberikan ulasan pada produk                                                                         |
| Reviewer-href     | VARCHAR(100) | Link profil pengulas                                                                                                         |
| Review_date       | VARCHAR(100) | Tanggal ulasan dibuat                                                                                                        |
| Rating_value      | VARCHAR(100) | Nilai/skor rating produk per ulasan                                                                                          |
| Review_verified   | VARCHAR(100) | Pembeli yang terverifikasi memiliki label <i>verified user</i> . Diasumsikan bahwa <i>verified user</i> memiliki ulasan yang |

| Tabel: Review |             |                                                                                                                |
|---------------|-------------|----------------------------------------------------------------------------------------------------------------|
| Attribut      | Tipe Data   | Deskripsi                                                                                                      |
|               |             | berdampak signifikan terhadap sentimen suatu produk                                                            |
| Review_detail | TEXT        | Berisikan detail ulasan yang dituliskan pengulas. Berisi tentang keluhan, kekurangan, maupun kelebihan produk. |
| Sentiment     | VARCHAR(45) | Apakah ulasan memiliki sentimen positif atau negatif                                                           |

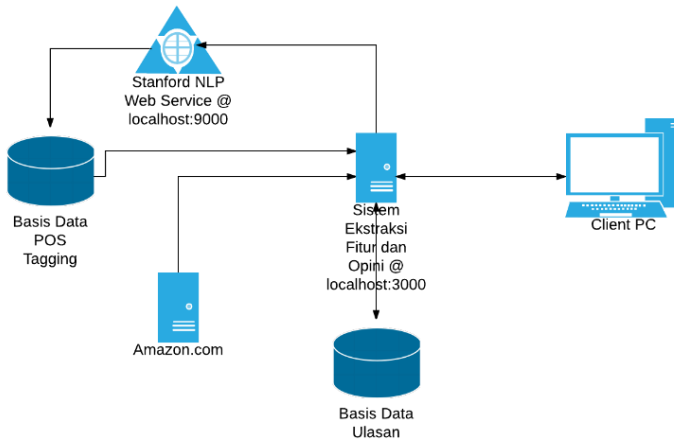
## 4.2. Preproses Data

Stanford NLP, selain tersedia dalam bentuk *source code*, juga tersedia *service* sebagaimana yang ada pada <http://nlp.stanford.edu:8080/corenlp/>. Dalam tugas akhir ini, tahap preproses data akan dijalankan bersamaan dengan tahapan ekstraksi fitur dan opini. Sehingga *service* Stanford NLP akan diintegrasikan ke dalam sistem visualisasi ekstraksi fitur dan orientasi opini. Gambar 20 adalah arsitektur integrasi antara sistem Stanford NLP dengan aplikasi sistem ekstraksi fitur dan orientasi opini.

## 4.3. Ekstraksi Fitur dan Opini Produk

Ekstraksi fitur dan opini dilakukan dengan mengimplementasikan LDA ke dalam kode program. LDA memiliki tiga parameter utama, yakni dokumen yang akan di proses, jumlah topik yang teridentifikasi dalam dokumen, dan jumlah kata kunci yang ingin ditampilkan. Salah satu

kelemahan dalam metode LDA adalah bahwa seseorang perlu menentukan jumlah topik yang ingin ditampilkan.

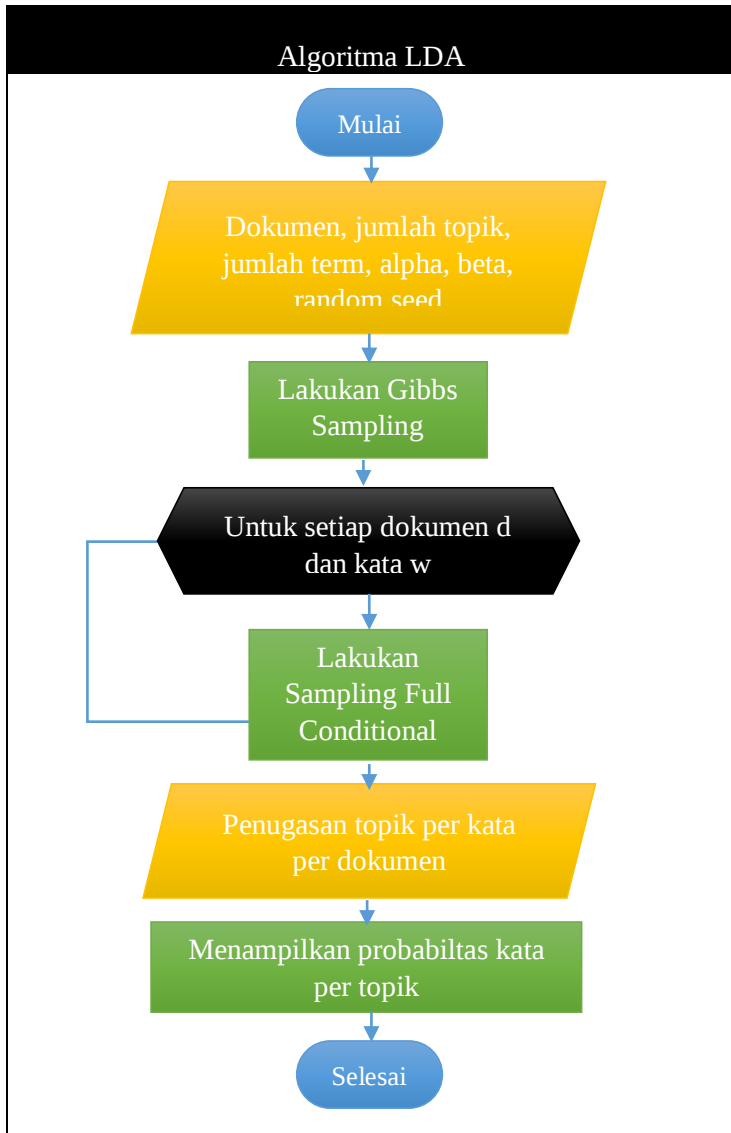


*Bagan 20 Arsitektur Sistem Ekstraksi Fitur dan Orientasi Opini*

LDA memiliki proses generatif sebagai berikut.

1. Untuk setiap dokumen  $w$  dalam korpus  $D$ 
  - a. Pilih  $N \sim \text{Poisson}(\xi)$
  - b. Pilih  $\theta \sim \text{Dir}(\alpha)$
2. Untuk setiap  $N$  kata dalam dokumen  $w_n$ :
  - a. Pilih topik  $z_n \sim \text{Multinomial}(\theta)$ .
  - b. Pilih kata  $w_n$  dari  $p(w_n | z_n, \beta)$

Bentuk implementasi dari algoritma LDA adalah terlihat sebagaimana pada gambar.



Bagan 21 Diagram Alir Algoritma LDA

#### 4.4. Rancang Bangun Prototipe

Sesuai dengan tahapan ICONIX, hal pertama yang akan dibuat adalah antar muka pengguna, kemudian diturunkan menjadi *use case diagram*. Selanjutnya *class diagram* dibuat dari hasil penurunan *use case diagram* dari antar muka pengguna yang telah dibuat. Karena sistem ini merupakan pengembangan yang sifatnya memberikan nilai tambah bagi pengguna, maka kebutuhan didasarkan pada studi terkait pemodelan topik dan manfaatnya apabila diimplementasikan pada kasus nyata.

##### 4.4.1. Antar Muka Pengguna

Berikut ini adalah desain antar muka pengguna. Antar muka pengguna yang dibuat terdiri dari tiga antar muka antara lain halaman pencarian, halaman daftar produk, dan halaman detail produk. Untuk setiap antar muka pengguna yang dibuat, akan dijadikan acuan dalam mengembangkan *use case*.

##### 4.4.1.1. Halaman Pencarian

|                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <p>AmazoLDA</p> <p>Sistem ekstraksi fitur dan opini untuk pengambilan keputusan pembelian pelanggan.</p> <div style="display: flex; justify-content: center; align-items: center; margin-top: 20px;"> <div style="border: 1px solid black; padding: 5px 20px; margin-right: 10px;">Masukkan Kata Kunci</div> <div style="background-color: black; color: white; padding: 5px 20px; border: 1px solid black;">Cari</div> </div> |
|                                                                                                                                                                                                                                                                                                                                                                                                                                |

*Bagan 22 Antar Muka Halaman Pencarian*

Halaman pada gambar 21 merupakan tampilan awal dari sistem AmazoLDA, nama aplikasi untuk sistem ekstraksi fitur dan opini untuk pengambilan keputusan pembelian pelanggan. Pengguna memasukkan kata kunci untuk menemukan barang dan klik cari untuk mencari barang yang ingin dicari dan dilihat lebih lanjut.

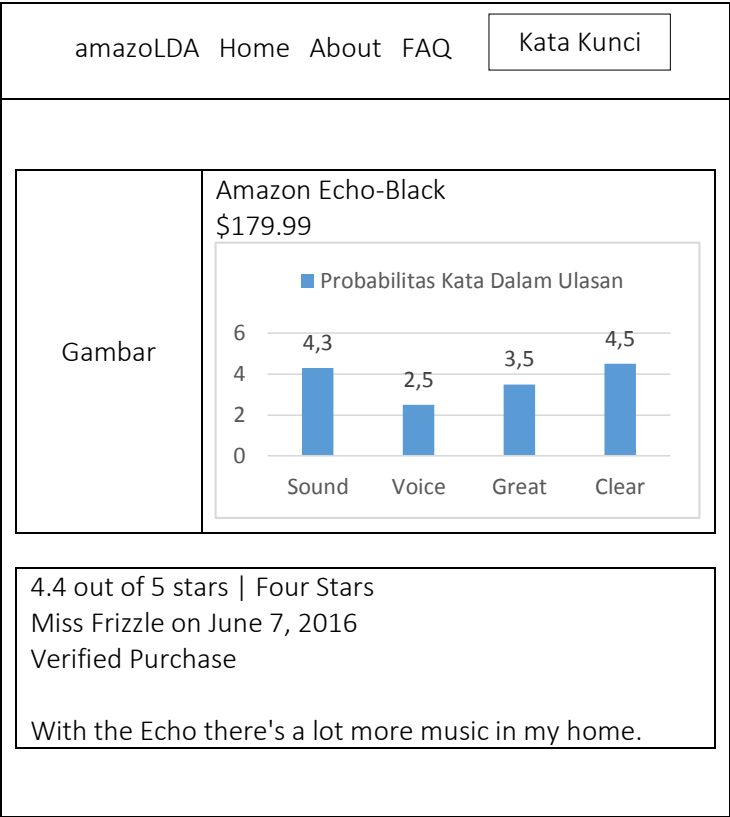
#### 4.4.1.2. Halaman Daftar Produk

|                         |                               |            |
|-------------------------|-------------------------------|------------|
| amazoLDA Home About FAQ |                               | Kata Kunci |
| Search "Kata Kunci"     |                               |            |
| Gambar                  | Amazon Echo-Black<br>\$179.99 |            |
| Gambar                  | Amazon Echo-White<br>\$179.99 |            |

*Bagan 23 Halaman Daftar Produk*

Halaman pada gambar 22 merupakan tampilan untuk menampilkan daftar barang sesuai dengan kata kunci pencarian. Data yang ditampilkan antara lain daftar produk berupa gambar, nama produk, dan harga produk. Pengguna dapat memilih produk yang dimaksud untuk dilihat detailnya lebih lanjut. Pengguna juga dapat mencari produk lain berdasarkan kata kunci apabila belum menemukan barang yang dimaksud. Kemudian menu navigasi dibuat untuk kembali ke halaman pencarian, menuju ke halaman tentang aplikasi, dan menuju ke halaman FAQ.

4.4.1.3.      **Halaman Detail Produk**

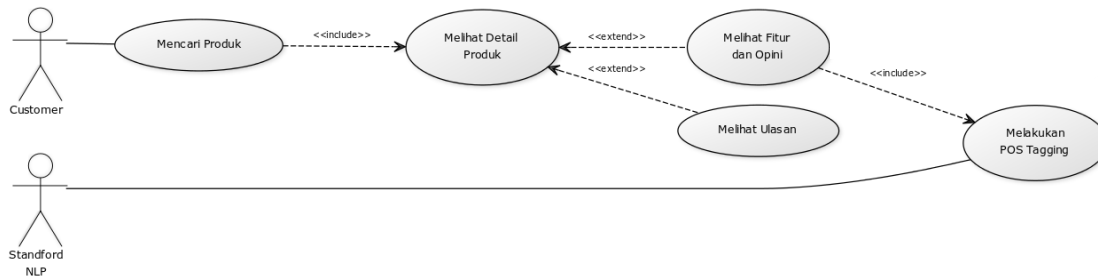


*Bagan 24 Halaman Detail Produk*

Halaman pada gambar 23 merupakan tampilan untuk menampilkan detail produk yang dipilih dari halaman daftar produk. Informasi produk seperti gambar, nama produk dan harga akan ditampilkan berikut hasil analisis dengan LDA. Wujudnya berupa grafik batang yang menampilkan kata kunci yang paling sering dibicarakan dalam ulasan produk.

#### 4.4.2. Use Case

*Use case* dibuat berdasarkan antar muka dan rancangan arsitektur integrasi sistem sebagaimana yang dibuat pada bab 4.2. Preproses data. Berikut ini adalah *use case* sistem secara keseluruhan.



*Bagan 25 Use Case Diagram untuk Sistem Ekstraksi Fitur dan Orientasi Opini*

#### 4.4.2.1. Daftar Use Case

*Tabel 13*Daftar Use Case Sistem Ekstraksi Fitur dan Orientasi Opini

| Kode  | Aktor         | Nama                    |
|-------|---------------|-------------------------|
| UCC-1 | Customer      | Mencari Produk          |
| UCC-2 | Customer      | Melihat Detail Produk   |
| UCC-3 | Customer      | Melihat Fitur dan Opini |
| UCC-4 | Customer      | Melihat Ulasan          |
| UCS-1 | Standford NLP | Melakukan POS Tagging   |

#### 4.4.2.2. Deskripsi Use Case

| A. UCC-1: Mencari Produk                                                                                                                                                                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Normal Flow:</b><br><br>User mencari produk berdasarkan kata kunci kemudian mengklik tombol cari. User kemudian diarahkan ke halaman daftar produk dan melihat daftar produk berdasarkan kata kunci yang ingin dicari. |
| <b>Alternate Flow:</b><br><br>Produk yang dicari tidak ditemukan dalam database: <ul style="list-style-type: none"><li>- Sistem menampilkan pesan error bahwa produk tidak dapat ditemukan</li></ul>                      |

| B. UCC-2: Melihat Detail Produk                                                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Normal Flow:</b><br><br>User mengklik produk untuk melihat detail produk pada halaman detail produk. Selanjutnya, user akan diarahkan pada halaman detail produk.                                                                                                               |
| <b>Alternate Flow:</b><br><br>Produk yang dilihat tidak memiliki ulasan yang cukup banyak sehingga sistem LDA tidak dapat berjalan sebagaimana mestinya. <ul style="list-style-type: none"><li>- Sistem menampilkan pesan error bahwa ulasan terlalu sedikit pada chart.</li></ul> |

### C. UCC-3: Melihat Fitur dan Opini

#### Normal Flow:

Pada halaman detail produk, user melihat grafik yang menampilkan fitur yang sering dibicarakan serta sentimen yang dibicarakan pada produk. Sistem mengkalkulasi probabilitas kata yang paling sering muncul dengan menggunakan algoritma *Latent Dirichlet Allocation*.

#### Alternate Flow:

Produk yang dilihat tidak memiliki ulasan yang cukup banyak sehingga sistem LDA tidak dapat berjalan sebagaimana mestinya.

- Sistem menampilkan pesan error bahwa ulasan terlalu sedikit pada chart.

### D. UCC-4: Melihat Ulasan

#### Normal Flow:

Pada halaman detail produk, user melihat daftar ulasan yang ditulis oleh pelanggan lain. Daftar ulasan akan menampilkan judul ulasan, nama pengulas, tanggal ulasan dibuat, user terverifikasi, dan detail ulasan

#### Alternate Flow:

Ulasan tidak dapat ditemukan

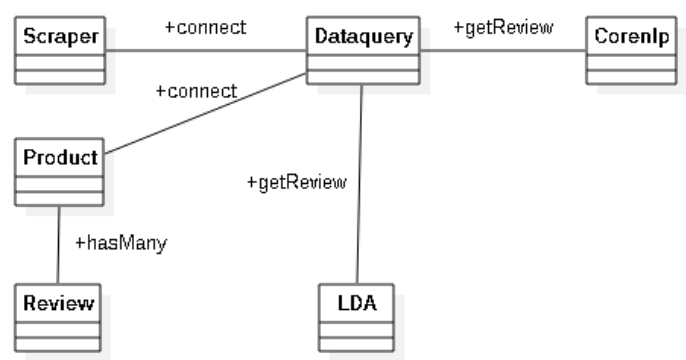
- Sistem menampilkan pesan error bahwa ulasan tidak ditemukan untuk produk tersebut.

### E. UCS-1: Melakukan POS Tagging

#### Normal Flow:

Pada halaman detail produk, sistem akan melakukan analisis LDA, namun sebelum itu, data ulasan juga akan dilakukan POS Tagging untuk membedakan mana yang kata sentimen dan mana yang kata opini.

4.4.3. Domain Model



Bagan 26 Domain model untuk aplikasi AmazoLDA

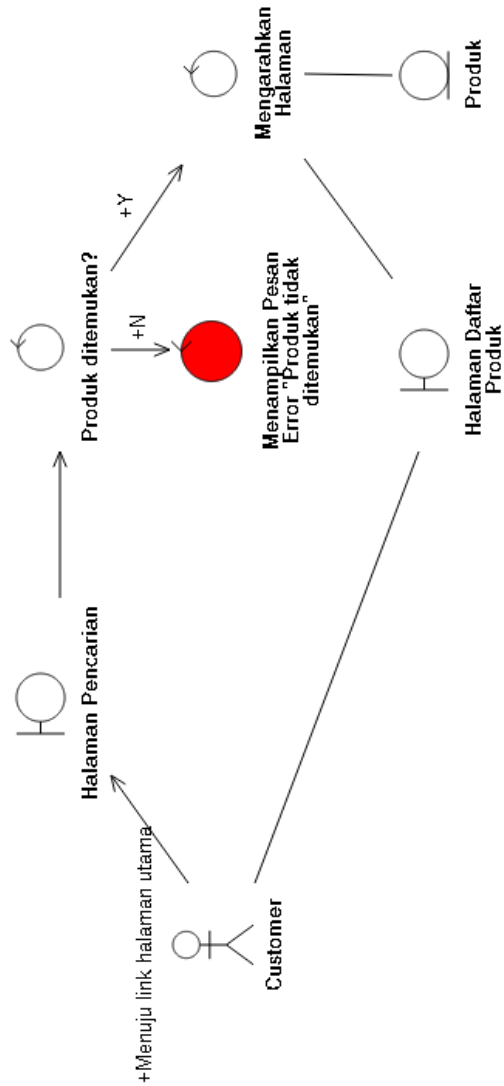
4.4.4. Robustness Diagram

Robustness diagram dibuat untuk mempermudah proses translasi dari use case menjadi sequence diagram. Tabel 14 menunjukkan pemetaan use case dengan robustness analysis. Berikut ini adalah diagram *robustness* untuk masing-masing *use case* yang teridentifikasi.

Tabel 14 Pemetaan Robustness Analysis dengan Use Case

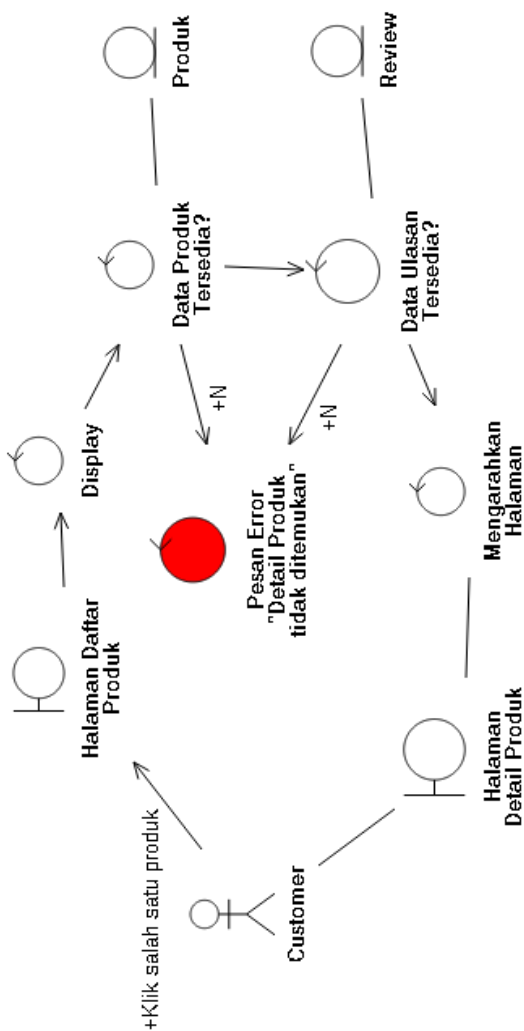
| Robustness Analysis | Use Case |
|---------------------|----------|
| RBUCC-1             | UCC-1    |
| RBUCC-2             | UCC-2    |
| RBUCC-3             | UCC-3    |
| RBUCC-4             | UCC-4    |
| RBUCS-1             | UCS-1    |

#### 4.4.4.1. Analisis Robustness UCC-1 Mencari Produk (RBUCC-1)



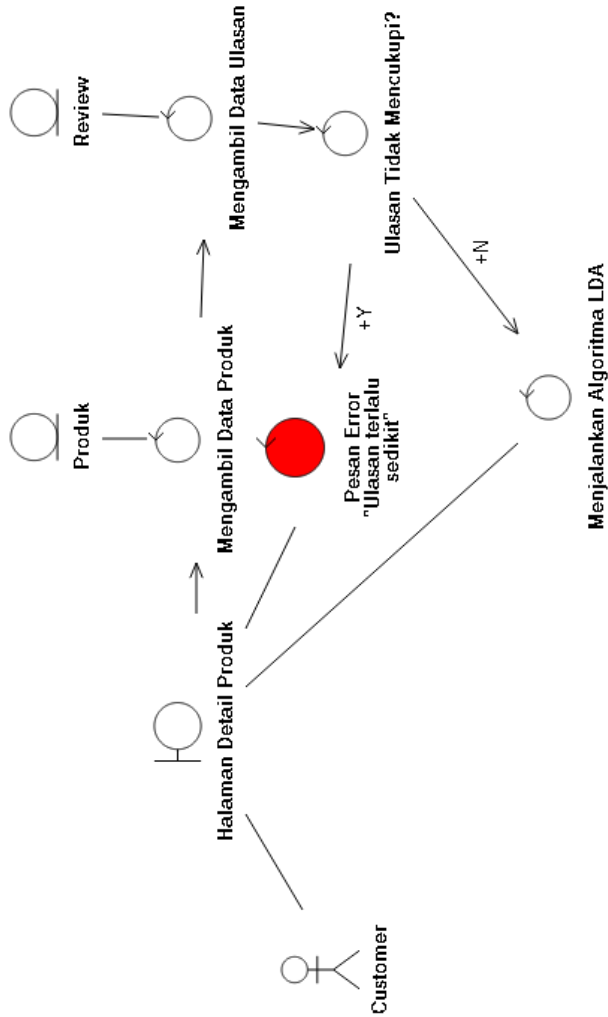
Bagan 27 Robustness Analysis untuk UCC-1 Mencari Produk

#### 4.4.4.2. Analisis Robustness UCC-2 Melihat Detail Produk (RBUCC-2)



Bagan 28 Robustness Analysis untuk UCC-2 Melihat Detail Produk

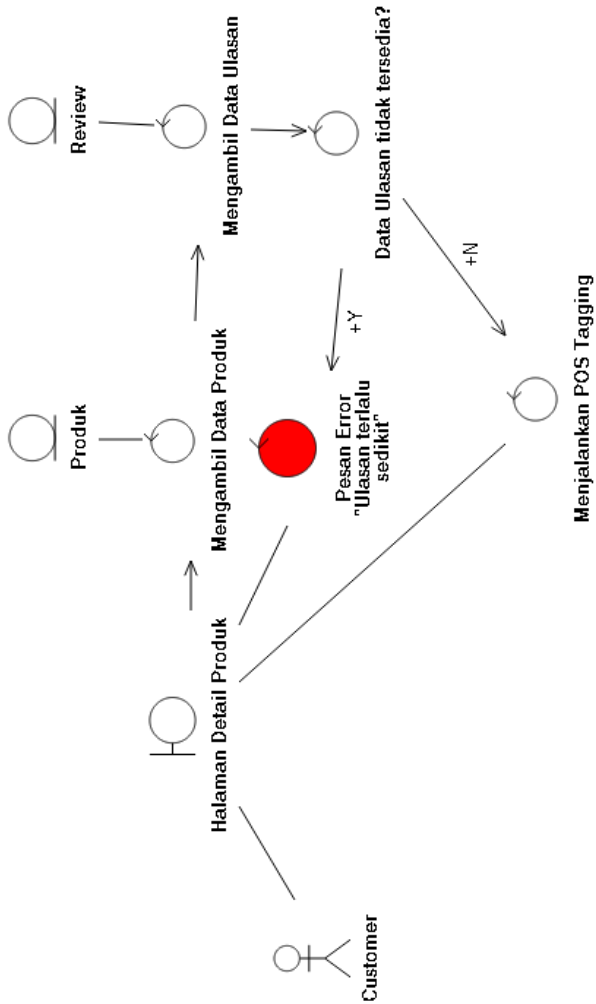
#### 4.4.4.3. Analisis *Robustness* UCC-3 Melihat Fitur dan Opini (RBUC-3)



Bagan 29 Robustness Analysis untuk UCC-3 Melihat Fitur dan Opini



#### 4.4.4.5. Analisis *Robustness* UCS-1 Melakukan POS Tagging (RBUCS-1)



Bagan 31 Robustness Analysis untuk UCS-1 Melakukan POS Tagging

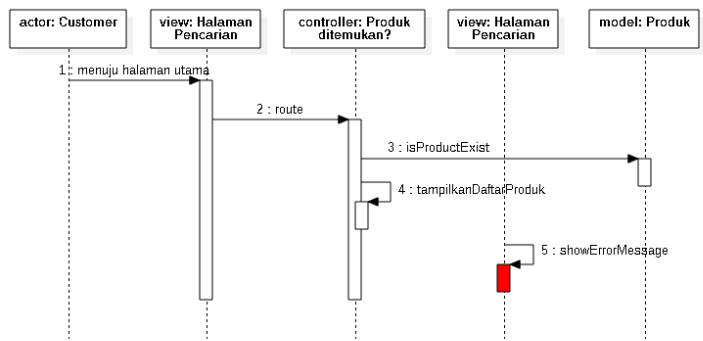
4.4.5. Sequence Diagram

Robustness diagram dibuat untuk mempermudah proses translasi dari use case menjadi sequence diagram. Tabel 14 menunjukkan pemetaan use case dengan robustness analysis. Berikut ini adalah diagram *robustness* untuk masing-masing *use case* yang teridentifikasi.

Tabel 15 Pemetaan Robustness Analysis dengan Sequence Diagram

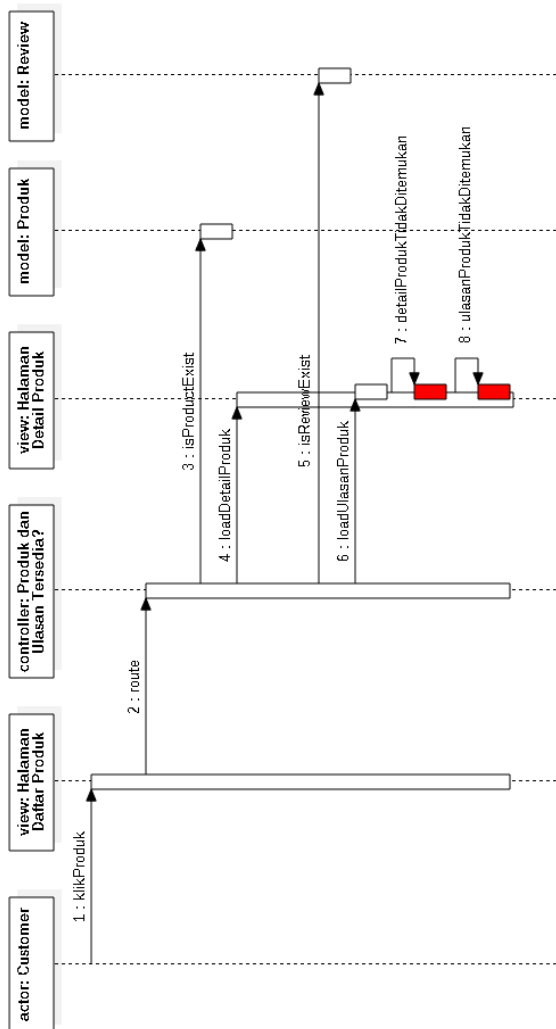
| Robustness Analysis | Sequence Diagram |
|---------------------|------------------|
| RBUCC-1             | SQRBUCC-1        |
| RBUCC-2             | SQRBUCC-2        |
| RBUCC-3             | SQRBUCC-3        |
| RBUCC-4             | SQRBUCC-4        |
| RBUCS-1             | SQRBUCS-1        |

4.4.5.1. Sequence Diagram RBUCC-1 Mencari Produk (SQRBUCC-1)



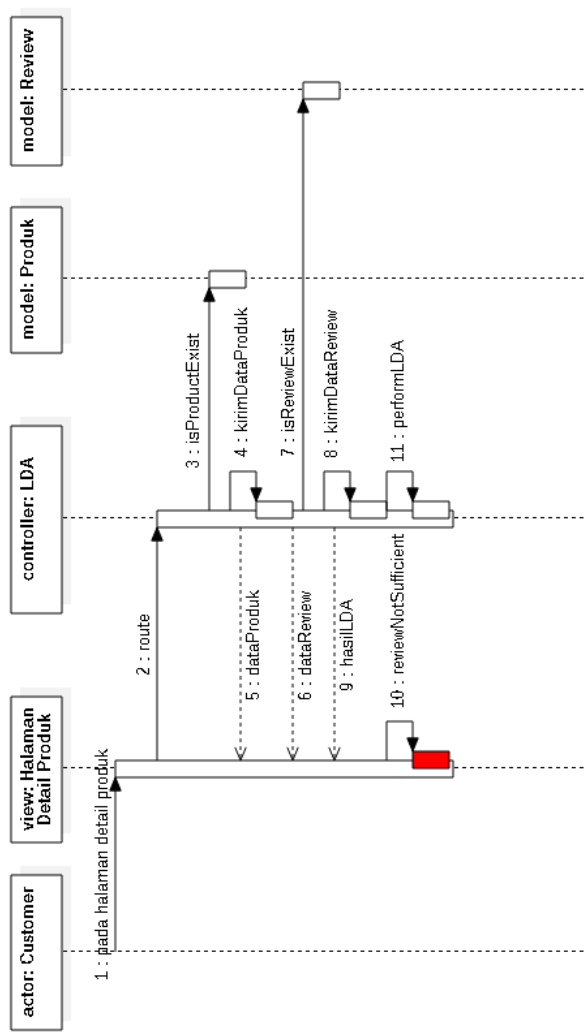
Bagan 32 Sequence Diagram untuk Robustness RBUCC-1 Mencari Produk

#### 4.4.5.2. Sequence Diagram RBUCC-2 Melihat Detail Produk (SQRBUCC-2)



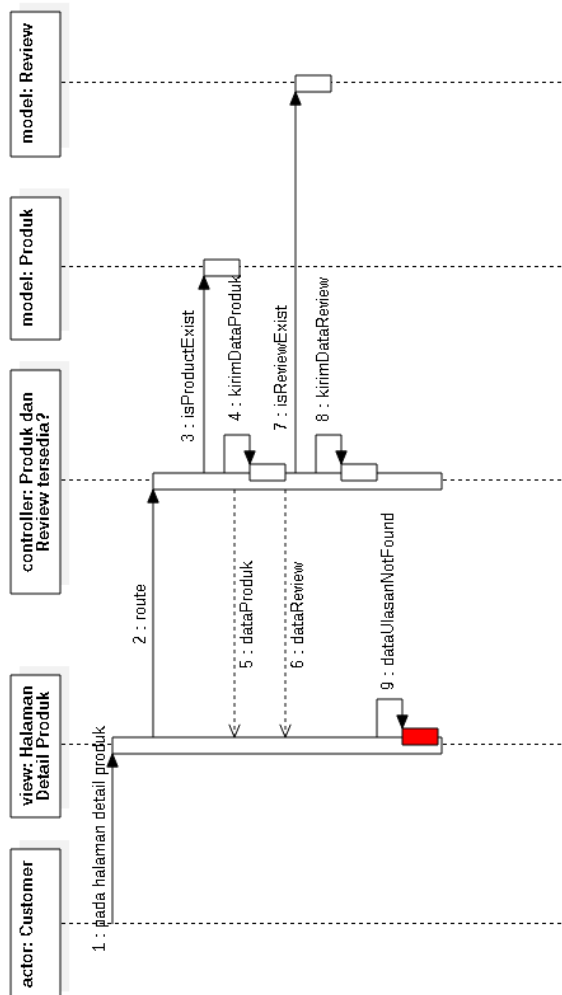
Bagan 33 Sequence Diagram untuk Robustness RBUCC-2 Melihat Detail Produk

4.4.5.3. **Sequence Diagram RBUCC-3 Mencari Fitur dan Opini (SQRBUCC-3)**



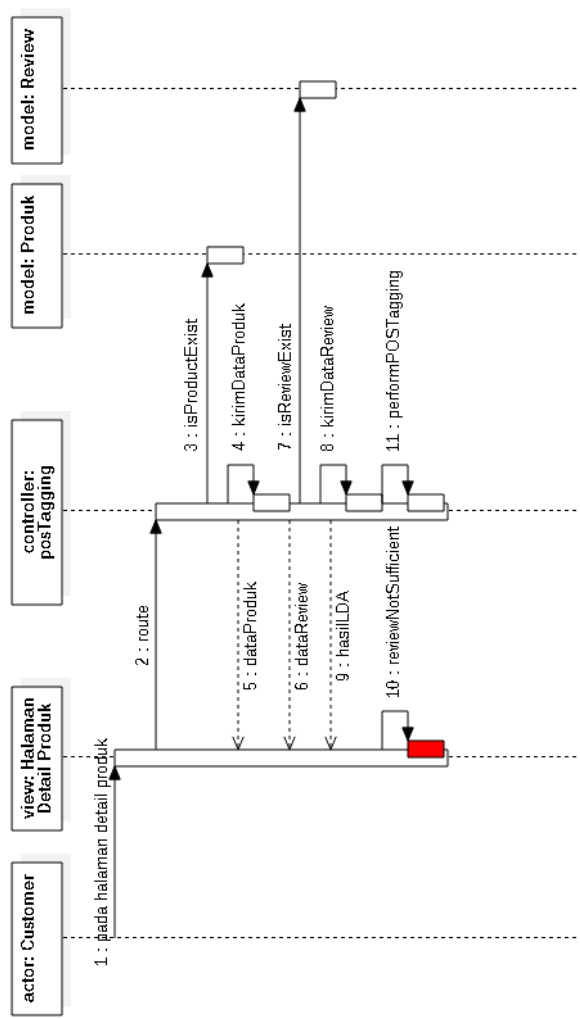
Bagan 34 Sequence Diagram untuk Robustness RBUCC-3 Mencari Fitur dan Opini

#### 4.4.5.4. Sequence Diagram RBUCC-4 Melihat Ulasan (SQRBUCC-4)



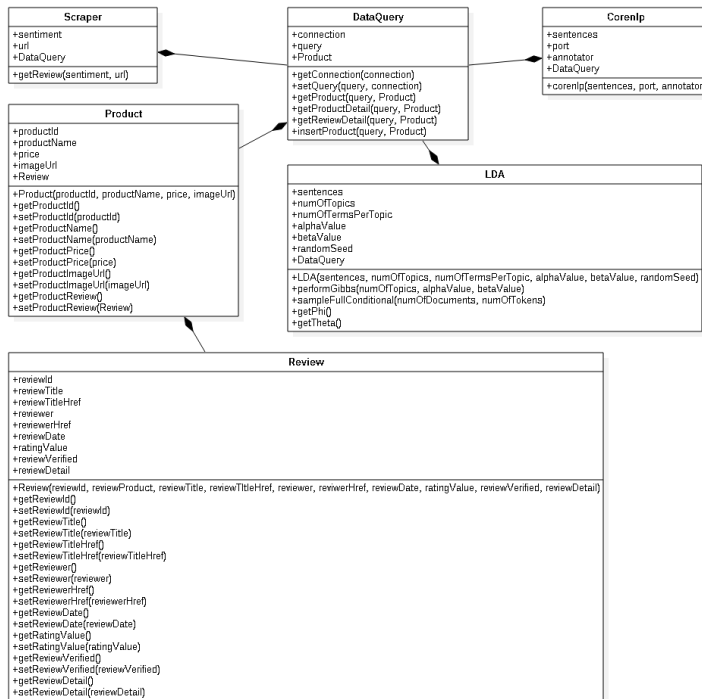
Bagan 35 Sequence Diagram untuk Robustness RBUCC-4 Melihat Ulasan

4.4.5.5. **Sequence Diagram RBUCS-1 Melakukan POS Tagging (SQRBUCS-1)**



Bagan 36 Sequence Diagram untuk Robustness RBUCS-1 Melakukan POS Tagging

#### 4.4.6. Class Diagram



Bagan 37 Class Diagram aplikasi AmazoLDA

*Halaman ini sengaja dikosongkan*

## **BAB V**

### **IMPLEMENTASI**

#### **5.1. Lingkungan Implementasi**

Pada pengembangan aplikasi, peneliti menggunakan perangkat keras dengan spesifikasi seperti pada Tabel 14. Sedangkan untuk perangkat lunak yang digunakan dalam pengembangan aplikasi adalah seperti pada Tabel 15.

*Tabel 16 Spesifikasi Perangkat Keras*

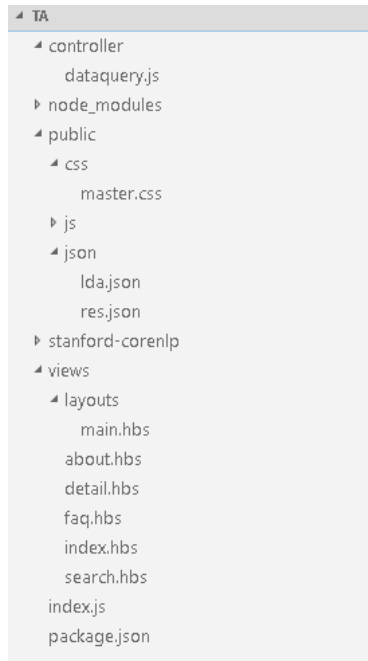
| Perangkat Keras | Spesifikasi                                     |
|-----------------|-------------------------------------------------|
| Laptop          | Prosesor: Intel® Core™ i7-2630QM CPU @ 2.00 GHz |
|                 | RAM: 4GB                                        |
|                 | Sistem Operasi Windows 8                        |

*Tabel 17 Spesifikasi Perangkat Lunak*

| Perangkat Lunak | Spesifikasi              |
|-----------------|--------------------------|
| Web Server      | NodeJS v6.9.2.           |
| Database Server | MySQL                    |
| Bahasa          | Javascript (Node JS)     |
| Framework       | Express JS, Stanford NLP |
| Editor          | Visual Studio Code       |

#### **5.2. Struktur Direktori**

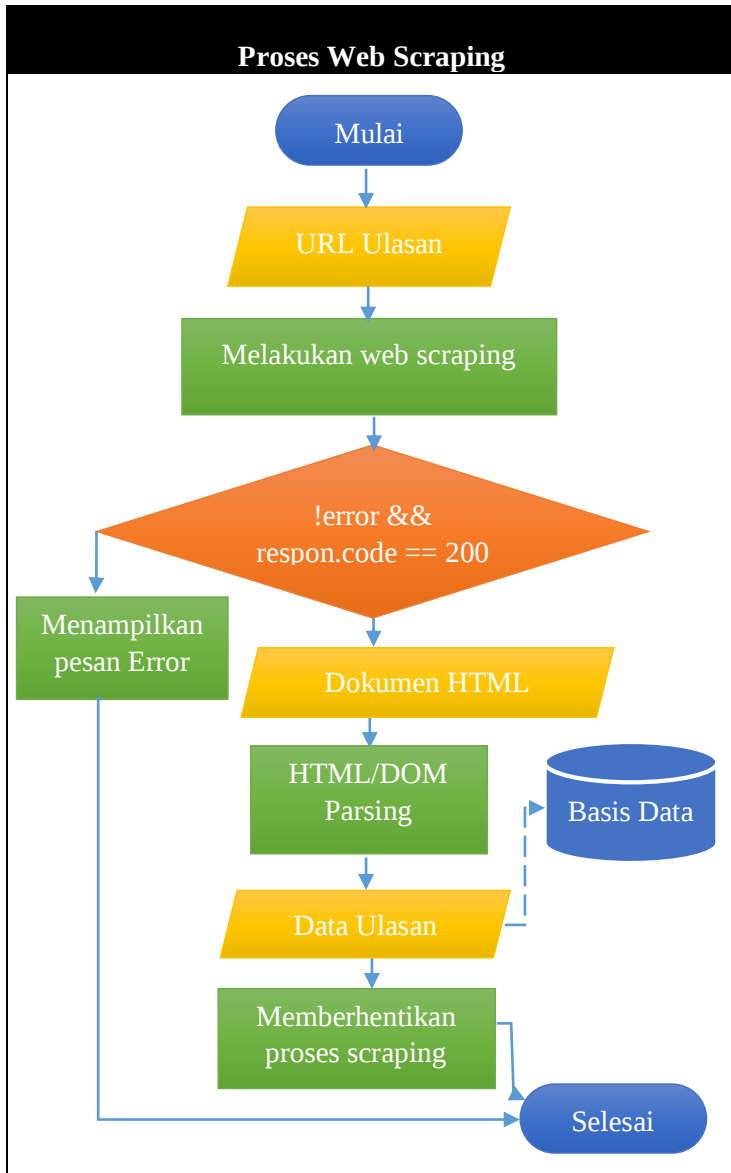
Berikut ini adalah struktur direktori untuk pengembangan aplikasi sistem ekstraksi fitur dan orientasi opini pada produk situs *e-commerce*.



*Bagan 38 Struktur Direktori Aplikasi*

### 5.3. Pengambilan Data

Setelah menentukan metadata ulasan berikut tag-tag HTML DOM yang perlu dirujuk, maka selanjutnya data diambil dengan melakukan *web scraping* pada URL ulasan produk. *Web scraping* dilakukan dengan menggunakan Request, sebuah *library* NodeJS dan cheerio, *library* NodeJS untuk HTML dan DOM *Parsing*. Proses dilakukan dengan melakukan request ke URL. Hasil nilai kembalian dari URL tersebut adalah berupa dokumen HTML. Selanjutnya data-data yang diperlukan diambil dengan melakukan filtrasi sesuai dengan tag-tag HTML atau DOM. Gambar 19 menunjukkan alur proses *web scraping* dengan NodeJS.



Bagan 39 Diagram Alir Proses Web Scraping

```

// Prerequisite modules
var request = require('request');
var cheerio = require('cheerio');
var mysql = require('mysql');

// Scrapper Global Variable
var scrapper = {};
var productName;
var customerName;
var reviewDate;
var reviewTitle;
var reviewDetails;
var verifiedPurchase;
var reviewSentiment;
var reviewRatings;
var sen = 'positive'; // change it with 'positive' or 'critical'

// Mysql Configuration
var connection = mysql.createConnection({
  host : 'localhost',
  user : 'root',
  password : '',
  database : 'amazolda'
});

```

*Bagan 40 Konfigurasi MySQL*

Kode pada gambar 22 merupakan konfigurasi untuk koneksi database dan menentukan variabel global dari Class *scrapper.js*. Class *scrapper.js* memiliki dependensi antara lain *request* untuk mendapat data dokumen HTML pada URL yang dituju, *cheerio* untuk melakukan *parsing* terhadap dokumen HTML yang telah diambil melalui modul *request*, dan terakhir modul MySQL untuk melakukan penyimpanan dalam database.

### 5.3.1. Mendapatkan jumlah paginasi pada halaman ulasan

```

var url = 'https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-
with-WiFi-Alexa/product-
reviews/B00X4WHP5E/ref=cm_cr_arp_d_paging_btm_1?ie=UTF8&rev
iewerType=all_reviews&showViewpoints=1&sortBy=recent&pageNum
ber=1&filterByStar='+sentiment+'

```

```

/**
 * Get Page Number and save the review into database
 *
 * @param {any} sentiment
 * @param {any} callback
 */
scraper.getReview = function (sentiment, callback) {
  request(url, function (error, response, html) {
    if (!error && response.statusCode == 200) {
      var $ = cheerio.load(html);
      var pageNum = $('li.page-button').children().last().text();
      return callback(pageNum);
    }
  });
}

```

Bagan 41 Kode untuk mendapatkan jumlah paginasi pada halaman ulasan

Kode pada gambar 21 digunakan untuk mendapatkan jumlah paginasi pada halaman ulasan. Semua ulasan pada produk akan diambil untuk kemudian ditampung ke dalam basis data ulasan. Namun karena sistem situs web Amazon.com menggunakan paginasi pada halaman ulasannya, maka perlu dilakukan *scraping* dengan teknik paginasi. Untuk itulah metode ini diperlukan untuk mendapatkan jumlah halaman paginasi pada halaman ulasan. Metode ini menerima input berupa sentimen dan fungsi *callback*. Sementara kembalian dari fungsi ini adalah berupa jumlah halaman paginasi pada halaman ulasan suatu produk.

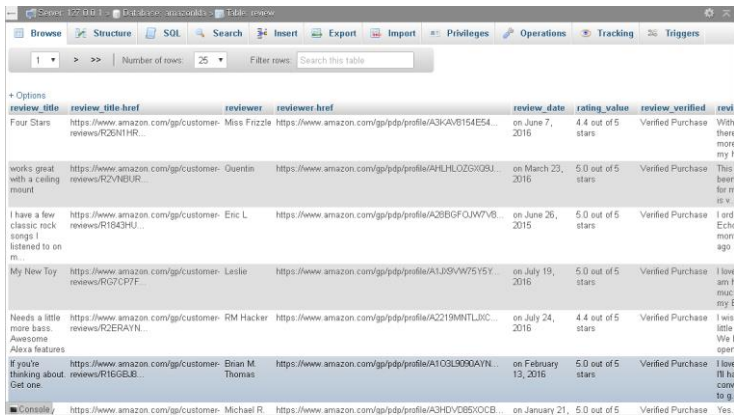
### 5.3.2. Melakukan Scraping dan Menyimpan hasil Scraping ke dalam basis data ulasan

Kode pada gambar 22 akan melakukan scraping sejumlah halaman paginasi yang telah diambil melalui fungsi *getReview*. Fungsi *getReview* memiliki dua parameter, yakni sentimen untuk menentukan mana review yang bersifat positif mana yang bersifat negatif (amazon.com telah melakukan pemilahan terhadap sentimen ulasan) dan *callback* untuk melakukan

*passing* nilai sehingga dapat digunakan dalam fungsi berikutnya.

Jumlah halaman paginasi ditampilkan dalam variabel *test*. Kemudian fungsi *request* akan mengambil halaman HTML sesuai dengan query URL yang ditetapkan. Setelah mendapatkan halaman sesuai dengan URL dan sentimen yang disampaikan, maka selanjutnya adalah melakukan ekstraksi data yang dibutuhkan untuk ditampilkan dalam basis data ulasan dengan menggunakan *library cheerio*.

Cheerio merupakan *library* NodeJS untuk melakukan HTML dan DOM *parsing*. Hasil dari ekstraksi data dari dokumen HTML akan ditampilkan dalam basis data dengan melakukan query *insert* pada basis data ulasan. Gambar 22 menjelaskan kode program untuk melakukan scraping dari halaman ulasan produk amazon.com kemudian menyimpan hasil scraping ke dalam basis data ulasan. Gambar 23 memaparkan hasil web scraping yang telah ditampilkan dalam basis data ulasan.



| review_title                                           | review_title_href                                                                                                         | reviewer        | reviewer_href                                                                                                     | review_date          | rating_value       | review_verified   | review_text             |
|--------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------|----------------------|--------------------|-------------------|-------------------------|
| Four Stars                                             | <a href="https://www.amazon.com/gp/customer-reviews/R26N1HR...">https://www.amazon.com/gp/customer-reviews/R26N1HR...</a> | Miss Frizzle    | <a href="https://www.amazon.com/gp/profile/A3KAV8154E54...">https://www.amazon.com/gp/profile/A3KAV8154E54...</a> | on June 7, 2016      | 4.4 out of 5 stars | Verified Purchase | With these more my h    |
| works great with a ceiling mount                       | <a href="https://www.amazon.com/gp/customer-reviews/R2YN8UR...">https://www.amazon.com/gp/customer-reviews/R2YN8UR...</a> | Quentin         | <a href="https://www.amazon.com/gp/profile/AHLH02G50BJ...">https://www.amazon.com/gp/profile/AHLH02G50BJ...</a>   | on March 23, 2016    | 5.0 out of 5 stars | Verified Purchase | This been for m (5 x)   |
| I have a few classic rock songs I listened to on it... | <a href="https://www.amazon.com/gp/customer-reviews/R1843HJ...">https://www.amazon.com/gp/customer-reviews/R1843HJ...</a> | Eric L.         | <a href="https://www.amazon.com/gp/profile/A28BGF0JW7V6...">https://www.amazon.com/gp/profile/A28BGF0JW7V6...</a> | on June 26, 2015     | 5.0 out of 5 stars | Verified Purchase | I ordi Echo mont ago i  |
| My New Toy                                             | <a href="https://www.amazon.com/gp/customer-reviews/R27CPTF...">https://www.amazon.com/gp/customer-reviews/R27CPTF...</a> | Leslie          | <a href="https://www.amazon.com/gp/profile/A1LD5VW75Y5Y...">https://www.amazon.com/gp/profile/A1LD5VW75Y5Y...</a> | on July 19, 2016     | 5.0 out of 5 stars | Verified Purchase | I love am h mual my E   |
| Needs a little more bass. Awesome Alexa features       | <a href="https://www.amazon.com/gp/customer-reviews/R2ERAYN...">https://www.amazon.com/gp/customer-reviews/R2ERAYN...</a> | RM Hacker       | <a href="https://www.amazon.com/gp/profile/A2219MNTL8IC...">https://www.amazon.com/gp/profile/A2219MNTL8IC...</a> | on July 24, 2016     | 4.4 out of 5 stars | Verified Purchase | I wist little Wp i open |
| If you're thinking about get one.                      | <a href="https://www.amazon.com/gp/customer-reviews/R16GBLB...">https://www.amazon.com/gp/customer-reviews/R16GBLB...</a> | Brian M. Thomas | <a href="https://www.amazon.com/gp/profile/A103L9000AYN...">https://www.amazon.com/gp/profile/A103L9000AYN...</a> | on February 13, 2016 | 5.0 out of 5 stars | Verified Purchase | I love it ha come to g  |
| Consolidate                                            | <a href="https://www.amazon.com/gp/customer-reviews/R26N1HR...">https://www.amazon.com/gp/customer-reviews/R26N1HR...</a> | Michael R.      | <a href="https://www.amazon.com/gp/profile/A3H0VD85XOCB...">https://www.amazon.com/gp/profile/A3H0VD85XOCB...</a> | on January 21, 2016  | 5.0 out of 5 stars | Verified Purchase | Yes.                    |

Bagan 42 Hasil web scraping ditampilkan dalam basis data ulasan

```

scraper.getReview(sen,function (val) {

    var test = parseInt(store(val).replace(/,/g, ""));

    var url = 'https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-
    with-WiFi-Alexa/product-
    reviews/B00X4WHP5E/ref=cm_cr_ar_p_d_paging_btm_1?ie=UTF8&rev
    ierType=all_reviews&showViewpoints=1&sortBy=recent&pageNum
    ber='+i+'&filterByStar='+sen+'

    for (var i = 0; i < test; i++) {
        request(url, function (error, response, html) {
            if (!error && response.statusCode == 200) {
                var $ = cheerio.load(html);

                // Get Review Title
                $('a.review-title').each(function(i, element){

                    productName = $('product-title').children().children().text();
                    customerName =
                $(this).parent().next().children().children().next().next().text();
                    reviewDate =
                $(this).parent().next().children().next().next().text().substr(3);
                    reviewTitle = $(this).text();
                    reviewDetails =
                $(this).parent().next().next().next().children().text();
                    verifiedPurchase =
                $(this).parent().next().next().children().next().next().children().children()
                .text();
                    reviewSentiment = sen;
                    reviewRatings =
                $(this).prev().prev().children().children().text().substr(0,1);
                    // console.log(productName + ' - ' + customerName + ' - ' +
                    reviewDate + ' - ' + reviewTitle + ' - ' + verifiedPurchase + ' - ' +
                    reviewSentiment + ' - ' + reviewRatings);

                    // Query
                    connection.query('INSERT INTO `review`(`review_id`,
                    `product_name`, `customer_name`, `review_date`, `review_title`,
                    `review_detail`, `verified_purchase`, `review_sentiment`,
                    `review_ratings`) +
                    'VALUES
                    (NULL,'+mysql.escape(productName)+' ,'+mysql.escape(customerName)
                    +' ,'+mysql.escape(reviewDate)+' ,'+mysql.escape(reviewTitle)+' ,'+mysql.
                    escape(reviewDetails)+' ,'+mysql.escape(verifiedPurchase)+' ,'+mysql.esc

```

```

ape(reviewSentiment)+' '+mysql.escape(reviewRatings)+'', function(err,
result) {
    if (err) throw err;
    console.log(reviewTitle + ' has been inserted into database
successfully.');
```

```

    });

});
}
});
}

});

```

*Bagan 43 Kode untuk melakukan scraping dan menyimpan ke dalam basis data ulasan*

#### 5.4. Preproses Data

Tahapan preproses data dilakukan untuk melakukan tokenisasi dan POS tagging pada data yang telah disimpan dalam basis data ulasan. Proses POS tagging dilakukan dengan menggunakan Stanford NLP, dan menurut dokumentasi, tokenisasi dan *sentence split* dilakukan bersamaan ketika anotator POS tagging dijalankan.

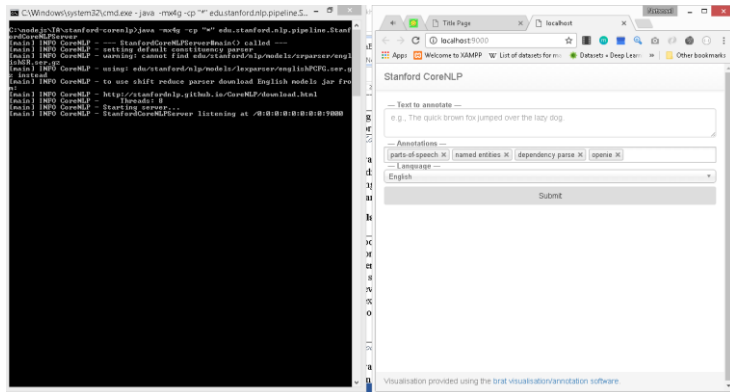
Proses POS tagging dalam Stanford NLP dilakukan dengan melakukan tokenisasi pada tiap-tiap kata yang terbentuk dalam ulasan. Masing-masing kata kemudian diberi label sesuai dengan referensi *Part of Speech* sebagaimana yang telah dibahas pada bab dasar teori. Setelah melakukan POS tagging, jenis-jenis kata untuk masing-masing token dapat diketahui. Karena pada dasarnya kata aspek berupa kata benda dan kata opini berupa kata sifat atau kata kerja, maka kata-kata selain kata benda, kata ganti benda, kata sifat, dan kata kerja akan dihilangkan dari detail ulasan.

### 5.4.1. Menjalankan Stanford NLP sebagai service

```
java -mx4g -cp "*" edu.stanford.nlp.pipeline.StanfordCoreNLPServer
```

*Bagan 44 Kode untuk menjalankan server Standford NLP via konsole*

Kode program pada gambar 25 digunakan untuk menjalankan server Stanford NLP. Gambar 26 menunjukkan bahwa server telah berhasil dijalankan dan dapat digunakan sesuai dengan kebutuhan pengembang. Server Stanford NLP dapat digunakan dengan mengakses <http://localhost:9000> atau juga dapat digunakan sebagai *webservice* sehingga API dapat digunakan dalam kode program. *Library* yang digunakan untuk dapat mengakses API Stanford NLP pada NodeJS adalah *corenlp-js-interface* yang dikembangkan oleh Noah Dessauer sebagaimana yang terlihat pada <https://www.npmjs.com/package/corenlp-js-interface>.



*Bagan 45 Menjalankan Server Standford NLP via Konsole*

### 5.4.2. Melakukan POS Tagging pada data ulasan.

```
app.get('/postagging', function(req, res){
  // Do Something...
  dataQuery.getReviewDetail(function (err,rows){
    // Do something ...
    var reviewString = joinResults(rows)
```

```

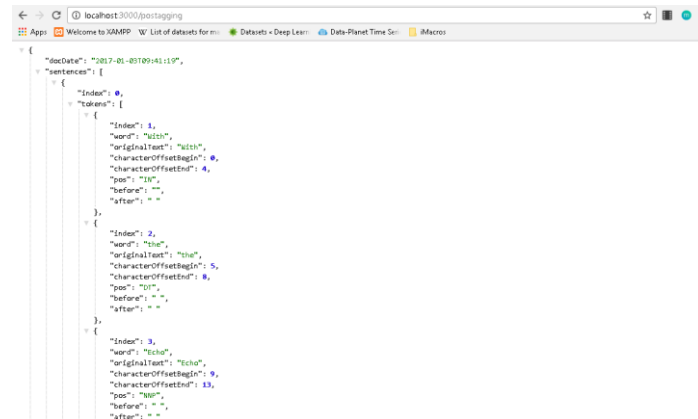
        var text = corenlp(reviewString,"pos");
        res.json(text)
    })
})

```

*Bagan 46 Kode untuk menjalankan POS Tagging pada Stanford NLP*

Kode program pada gambar 26 digunakan untuk menjalankan *library* Stanford NLP. *Library* ini akan menghasilkan file berupa JSON yang berisikan data ulasan yang telah ditokenisasi per kata.

Hasil dari preproses data akan ditampung ke dalam basis data POS tagging. Gambar 21 merupakan API yang dikembangkan untuk mengimplementasikan POS tagging ke data ulasan.



*Bagan 47 Hasil POS Tagging*

## 5.5. Ekstraksi Fitur dan Opini

Ekstraksi fitur dan opini dilakukan dengan mengimplementasikan LDA ke dalam kode program. Algoritma LDA diimplementasikan dengan menggunakan NodeJS.

LDA memiliki tiga parameter utama, yakni dokumen yang akan di proses, jumlah topik yang teridentifikasi dalam dokumen, dan jumlah kata kunci yang ingin ditampilkan. Salah satu kelemahan dalam metode LDA adalah bahwa seseorang perlu menentukan jumlah topik yang ingin ditampilkan.

LDA memiliki proses generatif sebagai berikut.

1. Untuk setiap dokumen  $w$  dalam korpus  $D$ 
  - a. Pilih  $N \sim \text{Poisson}(\xi)$
  - b. Pilih  $\theta \sim \text{Dir}(\alpha)$
2. Untuk setiap  $N$  kata dalam dokumen  $w_n$ :
  - a. Pilih topik  $z_n \sim \text{Multinomial}(\theta)$ .
  - b. Pilih kata  $w_n$  dari  $p(w_n | z_n, \beta)$

Bentuk implementasi dari algoritma LDA adalah sebagai berikut.

```
var process = function(sentences, numberOfTopics,
  numberOfTermsPerTopic, languages, alphaValue, betaValue,
  randomSeed) {

  var result = [];
  var documents = new Array();
  var f = {};
  var vocab=new Array();
  var vocabOrig = {};

  if (sentences && sentences.length > 0) {

    // lakukan tokenisasi dan stemming
    ...

    var V = vocab.length;
    var M = documents.length;
    var K = parseInt(numberOfTopics);
    var alpha = alphaValue || 0.1;
    var beta = betaValue || .01; words
```

```

// Melakukan gibbs sampling
lda.configure(documents,V,10000, 2000, 100, 10, randomSeed);
lda.gibbs(K, alpha, beta);

var theta = lda.getTheta();
var phi = lda.getPhi();

var text = "";

//tampung topik, kata kunci, beserta probabilitas dalam array
...

}
}

```

*Bagan 48 Kode Utama untuk melakukan analisa LDA dengan korpus berupa ulasan produk di Amazon.com*

Kode program pada gambar 49 merupakan implementasi LDA secara keseluruhan. Fungsi LDA yang dibuat mampu menampung dua parameter wajib dan lima parameter yang sifatnya opsional. Berikut ini adalah penjelasan dari masing-masing parameter.

*Tabel 18 Parameter Fungsi LDA*

| <b>Nama Parameter</b> | <b>Deskripsi</b>                                                                                                                                    |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| sentences             | Parameter ini menerima masukan berupa kumpulan kalimat yang akan dianalisa.                                                                         |
| numberOfTopics        | Parameter ini merupakan jumlah topik yang ingin ditampilkan sebagai hasil akhir dari analisa LDA.                                                   |
| numberOfTermsPerTopic | Parameter ini merupakan jumlah kata-kata dengan probabilitas yang paling sering muncul yang ingin ditampilkan sebagai hasil akhir dari analisa LDA. |
| languages             | Parameter untuk mengatur bahasa. Terdapat tiga bahasa yakni bahasa Inggris, bahasa Jerman, dan bahasa Spanyol.                                      |

| Nama Parameter | Deskripsi                                                                                                                                                                                                                                                              |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| alphaValue     | Parameter dirichlet LDA untuk distribusi probabilitas topik per dokumen                                                                                                                                                                                                |
| betaValue      | Parameter dirichlet LDA untuk distribusi probabilitas topik per kata                                                                                                                                                                                                   |
| randomSeed     | Karena iterasi LDA sifatnya unsupervised, maka setiap kali analisis LDA berjalan, nilai probabilitasnya selalu random. Parameter ini digunakan agar nilai acak dari iterasi LDA dapat diset permanen atau tidak memiliki variabilitas yang terlalu tinggi per iterasi. |

Setelah menentukan parameter yang akan dimasukkan ke dalam fungsi LDA, maka selanjutnya akan dilakukan proses tokenisasi pada masing-masing kata per kalimat yang teridentifikasi dalam dokumen. selanjutnya *Gibbs Sampling* dijalankan untuk mendapatkan nilai probabilitas kata kunci di masing-masing topik yang teridentifikasi. Nilai probabilitas tersebut kemudian diurutkan berdasarkan dari nilai probabilitas kata yang sering muncul dalam topik yang teridentifikasi. Hasil daripada *Gibbs Sampling* selanjutnya ditampung dalam array untuk dapat ditampilkan ke dalam antar muka.

### 5.5.1. Melakukan Tokenisasi Dokumen

Proses tokenisasi dokumen dilakukan untuk menghilangkan kata selain kata benda dan kata sifat serta menghilangkan tanda baca dan menemukan bentuk dasar kata. Tokenisasi dilakukan per kata per kalimat. Dokumen tersebut di tampung di dalam dua array. Pertama adalah array berisi kata-kata yang terkandung dalam dokumen. Array pertama digunakan sebagai indeks acuan pada array dokumen. Sementara array dokumen merupakan kumpulan indeks kata-kata yang telah ditokenisasi.

Indeks kata tersebut direpresentasikan dengan **dokumen** =  $\{0,1,2, \dots, vocab_{[n]}\}$ .

Persamaan berikut merepresentasikan bahwa masing-masing kata dalam suatu kalimat ditokenisasi menjadi satu dokumen.

$$Vocab = \begin{bmatrix} [0]Echo \\ [1]lot \\ [2]music \\ [3]home \\ [4]amazing \\ [5]responsive \\ [6]answer \\ [7]hear \\ [...] \dots \end{bmatrix}, Dokumen = \begin{bmatrix} 0 & 1 & \dots & \dots \\ 2 & 3 & \dots & \dots \\ 0 & 4 & 5 & \dots \\ 2 & 6 & 7 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

*Persamaan 8 Matriks vocabulary dan matriks dokumen*

Implementasi dalam bahasa pemrogramman adalah sebagai berikut.

```
var stopwords = new Array();

languages.forEach(function(value) {
  var stopwordsLang = require('./stopwords_' + value + ".js");
  stopwords = stopwords.concat(stopwordsLang.stop_words);
});

for(var i=0;i<sentences.length;i++) {
  if (sentences[i]=="") continue;
  documents[i] = new Array();

  var words = sentences[i].split(/[\s,\\'"]+/);

  if(!words) continue;
  for(var wc=0;wc<words.length;wc++) {
    var w=words[wc].toLowerCase().replace(/[^a-z'\u00C0-\u00ff ]+/g, "");
    var wStemmed = stem(w);
```

```

    if (w==" " || !wStemmed || w.length==1 ||
stopwords.indexOf(w.replace(" ", "")) > -1 ||
stopwords.indexOf(wStemmed) > -1 || w.indexOf("http")==0) continue;
    if (f[wStemmed]) {
        f[wStemmed]=f[wStemmed]+1;
    }
    else if(wStemmed) {
        f[wStemmed]=1;
        vocab.push(wStemmed);
        vocabOrig[wStemmed] = w;
    };
    documents[i].push(vocab.indexOf(wStemmed));
}

```

Bagan 49 Tokenisasi kata per kalimat ke dalam array vocab dan document

### 5.5.2. Melakukan Analisa LDA dengan *Gibbs Sampling*

Setelah melakukan tokenisasi terhadap dokumen, maka langkah selanjutnya adalah proses melakukan analisis LDA dengan metode *gibbs sampling*. Berikut adalah proses-proses yang terjadi selama melakukan *gibbs sampling*.

```

this.gibbs = function (K,alpha,beta) {
    this.createThetaSumAndPhiSum();
    this.initState(K);
    // Sample full conditional
    for (i = 0; i < this.ITERATIONS; i++) {
        for (var m = 0; m < this.z.length; m++) {
            for (var n = 0; n < this.z[m].length; n++) {
                var topic = this.sampleFullConditional(m, n);
                this.z[m][n] = topic;
            }
        }
        updateParams()
    }
}

```

Bagan 50 Kode Analisa LDA dengan Gibbs Sampling

Kode program pada gambar 30 menunjukkan implementasi fungsi *Gibbs Sampling* yang diturunkan dari fungsi LDA. Fungsi *Gibbs Sampling* digunakan untuk mengkalkulasi probabilitas topik  $z$  yang dimiliki oleh kata  $w_i$ . Masukkan untuk

fungsi *Gibbs Sampling* terdiri atas kata-kata  $w$  yang teridentifikasi dalam dokumen  $d$ . Luaran dari proses *gibbs sampling*. Proses diawali dengan menginisiasi topik  $z$ . Kemudian untuk tiap iterasi dan tiap jumlah kata yang teridentifikasi, kata-kata ditampung dalam barisan kata-kata, sementara distribusi topik ditampung dalam barisan distribusi topik. Kemudian untuk masing-masing topik yang teridentifikasi, hitung probabilitas topik dalam suatu dokumen dengan pendekatan *sample full conditional*. Luaran dari *gibbs sampling* adalah penugasan topik  $z$  dan menghitung jumlah topik dokumen, topik kalimat, dan topik dalam kata-kata.

Dengan penjelasan yang lebih detail, pertama, matriks *thetasum* dan *phisum* dibuat untuk menampung nilai probabilitas distribusi topik dalam dokumen dan distribusi kata dalam topik. Matriks *thetasum* dibuat dengan dimensi *jumlah dokumen*  $d \times$  *jumlah topik*  $t$  yang teridentifikasi dalam dokumen. Sementara matriks *phisum* dibuat dengan dimensi *jumlah topik*  $t$  yang teridentifikasi  $\times$  *jumlah vocabulary*  $w$  dalam matriks *vocab*. Bentuk matriks *thetasum* dan *phisum* dapat dijabarkan sebagai berikut.

$$thetasum = \begin{bmatrix} d_0t_0 & d_0t_1 & \dots & d_0t_y \\ d_1t_0 & d_1t_1 & \dots & d_1t_y \\ d_2t_0 & d_2t_2 & \dots & d_2t_y \\ \dots & \dots & \dots & \dots \\ d_xt_0 & d_xt_1 & \dots & d_xt_y \end{bmatrix}$$

Persamaan 9 Thetasum untuk menampung distribusi topik dalam dokumen

$$phisum = \begin{bmatrix} t_0w_1 & t_0w_2 & t_0w_3 & \dots & t_0w_y \\ t_1w_1 & t_0w_2 & t_0w_3 & \dots & t_1w_y \\ \dots & \dots & \dots & \dots & \dots \\ t_xw_1 & t_xw_2 & t_xw_3 & \dots & t_xw_y \end{bmatrix}$$

Persamaan 10 Phisum untuk menampung distribusi kata dalam topik

Implementasi dalam bahasa pemrograman adalah sebagai berikut.

```
var i;
this.K = K;
this.alpha = alpha;
this.beta = beta;
if (this.SAMPLE_LAG > 0) {
  this.thetasum = make2DArray(this.documents.length,this.K);
  this.phisum = make2DArray(this.K,this.V);
  this.numstats = 0;
}
```

*Bagan 51 Implementasi pembuatan matriks thetasum dan phisum*

### 5.5.2.1. Menginisasi Kondisi Awal

Kemudian langkah selanjutnya adalah membuat matriks **z**, **nw**, **nd**, **nwsun**, dan **ndsum**. Matriks **z** diperoleh dengan cara mengalikan jumlah topik  $K$  dengan angka random. Hasil pengalihan jumlah topik  $K$  akan menghasilkan angka 0 hingga sejumlah topik  $K$ . Sehingga untuk masing-masing vocabulary  $n$  dalam dokumen  $m$  topik  $z_{mn}$  akan dihasilkan. Bentuk dari matriks  $z$  merepresentasikan matriks dokumen. Hanya saja matriks  $z$  sudah menghubungkan masing-masing vocab dalam topik.

$$z = \begin{bmatrix} 0 & 1 & \dots & \dots \\ 0 & 0 & \dots & \dots \\ 0 & 1 & 0 & \dots \\ 1 & 1 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}, \text{dimana } K = 2$$

*Persamaan 11 Matriks  $z$  dimana nilai matriks diperoleh secara acak*

Matriks **nw** merupakan matriks yang digunakan untuk membuat hubungan topik untuk masing-masing kata. Isi matriks **nw** diperoleh dengan cara melihat *vocabulary* pada dokumen  $z$  kemudian memetakannya dengan matriks  $z$ . Berikut adalah contoh matriks **nw**.

$$nw = \begin{bmatrix} \textcolor{blue}{2} & 0 & \dots & w_0 t_y \\ 0 & 1 & \dots & w_1 t_y \\ 1 & 1 & \dots & w_2 t_y \\ 1 & 0 & \dots & w_3 t_y \\ 0 & \textcolor{red}{1} & \dots & w_4 t_y \\ 1 & 0 & \dots & w_5 t_y \\ 0 & 1 & \dots & w_6 t_y \\ 1 & 0 & \dots & w_7 t_y \\ \dots & \dots & \dots & \dots \\ w_x t_0 & w_x t_1 & \dots & w_x t_y \end{bmatrix}$$

$$D = \begin{bmatrix} \textcolor{blue}{0} & 1 & \dots & \dots \\ 2 & 3 & \dots & \dots \\ \textcolor{blue}{0} & \textcolor{red}{4} & 5 & \dots \\ 2 & 6 & 7 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad z = \begin{bmatrix} \textcolor{blue}{0} & 1 & \dots & \dots \\ 0 & 0 & \dots & \dots \\ \textcolor{blue}{0} & \textcolor{red}{1} & 0 & \dots \\ 1 & 1 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Persamaan 12 Matriks  $nw$  dengan matriks dokumen  $D$  dan matriks  $z$

Sebagai contoh, pada matriks  $nw$ , kosakata 4 memiliki topik 1 sebanyak 1. Nilai ini diperoleh dengan membandingkan dua matriks, yakni matriks dokumen  $d$  dan matriks topik  $z$ . Perhatikan pada matriks  $d$ , sejajarkan posisi vocab 4 dengan posisinya pada matriks  $z$ . Terlihat bahwa vocab 4 terikat dengan topik 1. Sehingga pada matriks  $nw$ , nilai topik 1 pada vocab 4 ditambahkan sebanyak 1. Proses ini dilakukan untuk semua vocab pada matriks *dokumen*.

Matriks  $nd$  merupakan matriks yang digunakan untuk membuat hubungan topik untuk masing-masing dokumen (kalimat). Isi matriks  $nd$  diperoleh dengan cara melihat distribusi topik pada dokumen  $z$ . Berikut adalah contoh matriks  $nd$ .

$$z = \begin{bmatrix} 0 & 1 & \dots & \dots \\ 0 & 0 & \dots & \dots \\ 0 & 1 & 0 & \dots \\ 1 & 1 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}, nd = \begin{bmatrix} 1 & 1 & \dots \\ 2 & 0 & \dots \\ 2 & 1 & \dots \\ 1 & 2 & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

*Persamaan 13 Matriks  $nd$  dengan matriks  $z$*

Nilai matriks  $nd$  diperoleh dengan cara menghitung jumlah masing-masing topik pada masing-masing dokumen. Contohnya pada dokumen 0, jumlah topik 0 dan topik 1 masing-masing adalah satu, maka pada matriks  $md$ , pada dokumen 0, nilai topik 0 adalah 1 dan nilai topik 1 adalah 1. Ulangi proses ini untuk setiap dokumen yang terbentuk pada matriks dokumen.

Matriks  $nwsun$  diperoleh dengan menjumlahkan *occurences* topik pada matriks  $z$ . Dapat terlihat jelas pada matriks  $z$ , topik 0 adalah sejumlah 6 dan topik 1 adalah sejumlah 4. Sehingga matriksnya adalah.

$$z = \begin{bmatrix} 0 & 1 & \dots & \dots \\ 0 & 0 & \dots & \dots \\ 0 & 1 & 0 & \dots \\ 1 & 1 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}, nwsun = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

*Persamaan 14 Matriks  $z$  dengan matriks  $nwsun$*

Sementara nilai  $ndsum$  diperoleh dengan menjumlahkan matriks  $nd$  per dokumen. Dapat terlihat jelas bahwa pada matriks  $nd$ , dokumen pertama berjumlah 2 topik, dokumen kedua berjumlah 2 topik, sementara dokumen ketiga dan keempat masing-masing berjumlah 3 topik. Matriks  $ndsum$  dapat dinotasikan sebagai berikut.

$$nd = \begin{bmatrix} 1 & 1 & \dots \\ 2 & 0 & \dots \\ 2 & 1 & \dots \\ 1 & 2 & \dots \\ \dots & \dots & \dots \end{bmatrix}, ndsum = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 3 \\ \dots \end{bmatrix}$$

Persamaan 15 Matriks *nd* dengan matriks *ndsum*

Implementasi dari inisiasi kondisi awal adalah sebagai berikut.

```

this.initialState = function (K) {
    var i;
    var M = this.documents.length;
    this.nw = make2DArray(this.V,K);
    this.nd = make2DArray(M,K);
    this.nwsum = makeArray(K);
    this.ndsum = makeArray(M);
    this.z = new Array(); for (i=0;i<M;i++) this.z[i]=new Array();
    for (var m = 0; m < M; m++) {
        var N = this.documents[m].length;
        this.z[m] = new Array();
        for (var n = 0; n < N; n++) {
            var topic = parseInt(""+(this.getRandom() * K));
            this.z[m][n] = topic;
            this.nw[this.documents[m][n]][topic]++;
            this.nd[m][topic]++;
            this.nwsum[topic]++;
        }
        this.ndsum[m] = N;
    }
}

```

Bagan 52 Implementasi inisiasi matriks *z*, *nw*, *nd*, *nwsum*, dan *ndsum*

#### 5.5.2.2. Melakukan Analisa Gibbs Sampling dengan Sample Full Conditional.

Setelah menginisiasi matriks topik *z*, matriks *nw*, *nd*, *nwsum*, dan *ndsum*, maka langkah selanjutnya adalah mengolah matriks tersebut dengan menggunakan fungsi *sample full conditional*. *Sample full conditional* adalah proses untuk memperbaiki distribusi topik dalam matriks *z* yang nilainya diinisiasi secara acak. Sehingga pertama-tama untuk masing-masing *vocab n*

dalam dokumen  $m$ , nilai matriks  $nw_{[dokumen(m,n)][z(m,n)]}$ ,  $nw_{[dokumen(m,n)][z(m,n)]}$ ,  $nd_{[dokumen(m,n)][z(m,n)]}$ ,  $nwsum_{[z(m,n)]}$ , dan  $nwsum_{[m]}$  dikurangi terlebih dahulu. Kemudian dibuatlah matriks  $p$  berdimensi sebesar jumlah topik yang teridentifikasi  $K$ . Untuk setiap topik  $K$  yang teridentifikasi, nilai dari  $p_{[k]}$  adalah sebagai berikut.

$$\sum_0^m \sum_0^n \sum_0^k p_k = \frac{(nw_{(doc(m,n),k)} + \beta)}{nwsum_k + V * \beta} \times \frac{(nd_{(m,k)} + \alpha)}{ndsum_m + K \times \alpha}$$

*Persamaan 16 Menghitung probabilitas untuk masing-masing vocab  $n$  dalam dokumen  $m$*

Kemudian nilai probabilitas  $p_k$  dibandingkan dengan nilai  $u$  yang merupakan nilai yang diperoleh dari  $p_{k-1}$  dikali dengan suatu nilai acak. Jika  $u < p_{topic}$ , maka iterasi selesai, jika tidak, maka jumlahkan nilai kembali pada matriks  $nw$ ,  $nd$ ,  $nwsum$ , dan  $ndsum$ . Hasil topik yang telah diperbaharui menjadi luaran dari fungsi *sample full conditional* ini.

Hasil distribusi topik setelah dilakukan *gibbs sampling* adalah sebagai berikut.

$$z = \begin{bmatrix} 0 & 0 & \dots & \dots \\ 1 & 1 & \dots & \dots \\ 0 & 0 & 0 & \dots \\ 1 & 1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

*Persamaan 17 Matriks distribusi topik  $z$  setelah diperbaharui dengan *sample full conditional**

Implementasi dari analisis *gibbs sampling* adalah sebagai berikut.

```
this.sampleFullConditional = function(m,n) {
  var topic = this.z[m][n];
  this.nw[this.documents[m][n]][topic]--;
```

```

this.nd[m][topic]--;
this.nwsum[topic]--;
this.ndsum[m]--;
var p = makeArray(this.K);
for (var k = 0; k < this.K; k++) {
    p[k] = (this.nw[this.documents[m][n]][k] + this.beta) /
    (this.nwsum[k] + this.V * this.beta)
    * (this.nd[m][k] + this.alpha) / (this.ndsum[m] + this.K *
    this.alpha);
}
for (var k = 1; k < p.length; k++) {
    p[k] += p[k - 1];
}
var u = this.getRandom() * p[this.K - 1];
for (topic = 0; topic < p.length; topic++) {
    if (u < p[topic])
        break;
}
this.nw[this.documents[m][n]][topic]++;
this.nd[m][topic]++;
this.nwsum[topic]++;
this.ndsum[m]++;
return topic;
}

```

*Bagan 53 Kode Analisa Gibbs Sampling dengan Sample Full Conditional*

Kode program pada gambar 31 menunjukkan implementasi dari *Gibbs Sampling* yang diturunkan dari fungsi LDA. Fungsi *sampleFullConditional* menerima masukan berupa jumlah dokumen dan jumlah kata yang teridentifikasi dalam dokumen. Masukan ini digunakan untuk menginisiasi matriks penugasan topik  $z$ . Kemudian untuk setiap topik yang ingin ditampilkan, lakukan perhitungan fungsi *Gibbs Sampling* yang diturunkan dari fungsi LDA. Luaran dari fungsi *sampleFullConditional* adalah matriks penugasan topik pada tiap dokumen dan tiap kata yang teridentifikasi.

### 5.5.3. Menghitung Probabilitas Kata Kunci per Topik Berdasarkan Matriks Distribusi Topik Z

Setelah mendapatkan probabilitas distribusi topik untuk ditampung dalam matriks  $z$ , maka langkah selanjutnya adalah untuk menghitung probabilitas untuk masing-masing kata kunci yang teridentifikasi dalam dokumen. Implementasi dalam bahasa pemrograman adalah sebagai berikut.

```

if ((i < this.BURN_IN) && (i % this.THIN_INTERVAL == 0))
{
    //document.write("B");
    this.dispcol++;
}
if ((i > this.BURN_IN) && (i % this.THIN_INTERVAL == 0))
{
    //document.write("S");
    this.dispcol++;
}
if ((i > this.BURN_IN) && (this.SAMPLE_LAG > 0) && (i %
this.SAMPLE_LAG == 0)) {
    this.updateParams();
    //document.write("|");
    if (i % this.THIN_INTERVAL != 0)
        this.dispcol++;
}
if (this.dispcol >= 100) {
    //document.write("*<br/>");
    this.dispcol = 0;
}

```

*Bagan 54 Implementasi perhitungan probabilitas kata per topik*

Fungsi *updateParams()* sebagaimana yang terlihat pada gambar 54 digunakan untuk menghitung probabilitas masing-masing kata yang berada pada topik tertentu dalam dokumen.

Untuk mendapatkan probabilitas tersebut, dapat diperoleh dengan membagi matriks  $nw$  dengan  $nws_{sum}$  dengan mempertimbangkan parameter dirichlet  $\beta$ . Berikut adalah formula untuk menghitung probabilitas masing-masing vocab

yang terhubung pada suatu topik untuk masing-masing topik yang teridentifikasi dalam dokumen.

$$\sum_0^k \sum_0^w phisum_{(k,w)} = \frac{(nw_{(w,k)} + \beta)}{nwsum_k + V + \beta}$$

*Persamaan 18 Menghitung nilai probabilitas kata per topik*

Bentuk implementasi dari fungsi phisum adalah sebagai berikut.

```

this.updateParams=function () {
    for (var m = 0; m < this.documents.length; m++) {
        for (var k = 0; k < this.K; k++) {
            this.thetasum[m][k] += (this.nd[m][k] +
this.alpha) / (this.ndsum[m] + this.K * this.alpha);
        }
    }
    for (var k = 0; k < this.K; k++) {
        for (var w = 0; w < this.V; w++) {
            this.phisum[k][w] += (this.nw[w][k] + this.beta) /
(this.nwsum[k] + this.V * this.beta);
        }
    }
    this.numstats++;
}

```

*Bagan 55 Implementasi fungsi update params untuk menghitung probabilitas kata per topik*

Selanjutnya, nilai dari matriks *phisum* akan dipetakan dengan *vocabulary* pada matriks *vocab*. Posisi probabilitas pada matriks phisum merepresentasikan probabilitas masing-masing vocab untuk tiap-tiap topik yang teridentifikasi. Sehingga, dengan menggabungkan matriks *vocab* dengan matriks *phisum*, maka akan di dapat *vocabulary* beserta probabilitas. Penggabungan matriks *vocab* dengan matriks *phisum* dilakukan sebanyak sejumlah topik yang teridentifikasi.

Bentuk implementasi dari pemetaan *vocab* terhadap probabilitas masing-masing vocab adalah sebagai berikut.

```
var topTerms=numberOfTermsPerTopic;
for (var k = 0; k < phi.length; k++) {
    var things = new Array();
    for (var w = 0; w < phi[k].length; w++) {

things.push(''+phi[k][w].toPrecision(2)+'_' +vocab[w]  +
''_ + vocabOrig[vocab[w]]);
    }
    things.sort().reverse();
    //console.log(things);
    if(topTerms>vocab.length) topTerms=vocab.length;

    //console.log("Topic ' + (k + 1));
    var row = [];

    for (var t = 0; t < topTerms; t++) {
        var topicTerm=things[t].split("_")[2];
        var prob=parseInt(things[t].split("_")[0]*100);
        if (prob<2) continue;

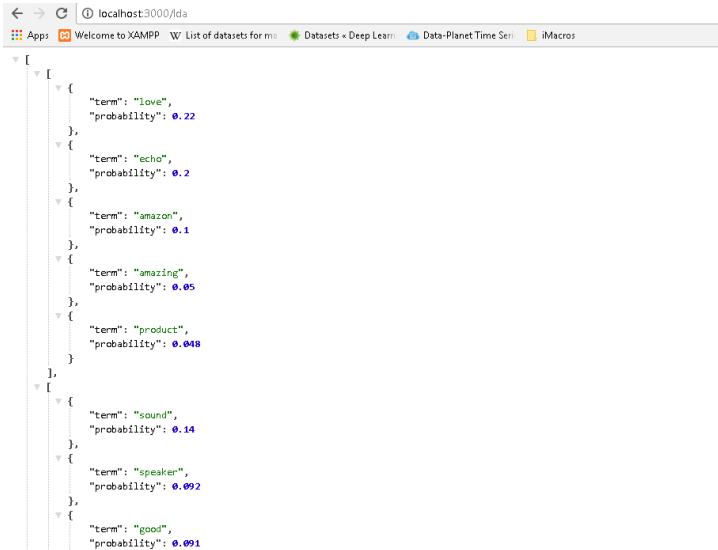
        var term = {};
        term.term = topicTerm;
        term.probability=parseFloat(things[t].split("_")[0]);
        row.push(term);
    }

    result.push(row);
}
```

*Bagan 56 Implementasi pemetaan kata per topik dalam matriks topik dengan matriks phisum*

Setelah melakukan pemetaan kata per topik dalam matriks topik terhadap matriks *phisum* ke dalam matriks *result*, maka nilai matriks *result* tersebut diurutkan mulai dari kata dengan

probabilitas tertinggi ke rendah. Hasil tersebut kemudian ditampung ke dalam array yang berisikan kata kunci yang telah dikategorikan per topik beserta nilai probabilitasnya. Berikut adalah contoh hasil LDA terhadap data ulasan produk Amazon Echo.



```

[
  {
    "term": "love",
    "probability": 0.22
  },
  {
    "term": "echo",
    "probability": 0.2
  },
  {
    "term": "amazon",
    "probability": 0.1
  },
  {
    "term": "amazing",
    "probability": 0.05
  },
  {
    "term": "product",
    "probability": 0.048
  }
],
[
  {
    "term": "sound",
    "probability": 0.14
  },
  {
    "term": "speaker",
    "probability": 0.092
  },
  {
    "term": "good",
    "probability": 0.091
  }
]

```

*Bagan 57 Luaran dari analisis LDA*

Gambar 54 menunjukkan luaran dari analisis LDA. Disini parameter yang diujikan merupakan jumlah topik dengan nilai 20 dan jumlah kata kunci 5 untuk masing-masing topik. Luaran berupa kata kunci dengan probabilitas yang paling sering muncul dalam dokumen serta nilai dari probabilitas per masing-masing kata itu sendiri.

## 5.6. Rancang Bangun Prototipe

Tahapan ini merupakan implementasi dari tahapan analisis dan desain ke dalam bentuk aplikasi sesuai dengan penjelasan lingkungan implementasi.

### 5.6.1. Implementasi Halaman Pencarian Produk

Halaman pencarian produk berperan sebagai halaman utama dari aplikasi amazolda. Pengguna mengentrikan kata kunci pada kotak pencarian untuk mendapatkan detail produk yang diinginkan. Selanjutnya sistem akan mencocokkan kata kunci pencarian dengan hasil pada basis data produk. Jika produk ditemukan, maka pengguna akan dialihkan pada halaman daftar produk sesuai dengan kata kunci yang dicari. Namun jika kata kunci produk tidak sesuai dengan nama produk yang terdapat dalam basis data ulasan, maka sistem akan menampilkan pesan *error* bahwa produk tidak ditemukan.

#### Index.js

```
app.get('/', function(req, res){

  // Render page
  res.render('index', {
    title: 'amazoLDA',
    description: "View and analyze product's features that customers talk about",
  })
})
```

#### Index.hbs

```
<div class="jumbotron">
  <div class="container text-center">
    <h1>{{title}}</h1>
    <p>{{description}}</p>
    <form action="/search/1" method="GET" class="form-inline"
role="form">
      <div class="input-group">
        <input type="text" class="form-control" id="query"
name="query" placeholder="Search Product">
        <span class="input-group-btn">
          <button type="submit" class="btn btn-
warning">Go!</button>
        </span>
      </div>
    </form>
  </div>
</div>
```

*Bagan 58 Implementasi kode pada halaman pencarian*

### 5.6.2. Implementasi Halaman Daftar Produk

Halaman daftar produk menampilkan daftar produk sesuai dengan kata kunci yang dientrikan oleh pengguna. Sistem akan menampilkan daftar produk dengan kata kunci yang sesuai dengan keinginan pengguna. Daftar produk diambil dengan melihat tabel produk pada basis data produk. Data yang akan ditampilkan dalam halaman daftar produk adalah nama produk, link detail produk, gambar produk, dan harga produk. Jika mengklik nama produk, maka pengguna akan dialihkan ke halaman detail produk yang dipilih.

#### Index.js

```
app.get('/search/:page', function(req, res){

  dataQuery.getProduct(function (err,rows){
    // Do something ...
    res.render('search',{
      title: 'amazoLDA',
      data: rows,
      query: req.param('query')
    })
  })

})
```

#### ./Controller/Dataquery.js

```
module.exports.getProduct = function (callback){
  // Do something ...
  setQuery('select * from product', function(err, rows){
    callback(err,rows)
  })
}
```

#### Search.hbs

```
<div id="content">

  <div class="container">
    <h1 class="text-center">Search "{{ query }}" in page {{ page
  }}</h1>
  </div>

  <div class="card">
    {{#each data}}
```

```

<div class="content">

  <div class="row">
    <div class="col-xs-3 col-sm-3 col-md-3 col-lg-3">
      
    </div>
    <div class="col-xs-9 col-sm-9 col-md-9 col-lg-9">
      <h2><a href="#" onclick='return performLDA("{{this.product_name}}");'>{{this.product_name}} </a></h2>
      <p>${{this.price}}</p>
    </div>
  </div>
  {{/each}}
</div>

</div>

```

*Bagan 59 Implementasi kode pada halaman daftar produk*

### 5.6.3. Implementasi Halaman Detail Produk

Halaman detail produk berisikan detail produk, ulasan produk, beserta grafik hasil analisis LDA. Detail produk yang ditampilkan berupa gambar produk, nama produk, dan harga produk. Sementara ulasan menampilkan semua ulasan yang telah ditulis oleh pemakai produk antara lain judul ulasan, pengulas, tanggal ulasan dibuat, detail ulasan, rating ulasan, dan keterangan apabila ulasan diulas oleh pembeli terverifikasi. Grafik analisis LDA menampilkan jumlah topik yang teridentifikasi dalam ulasan serta kata kunci terkait topik yang paling sering dibahas oleh konsumen.

**Index.js**

```

app.get('/lda/:productName', function(req, res){

  // Query data ulasan
  dataQuery.getReviewDetail(function (err,rows){

    // Join data ulasan menjadi satu dokumen

```

```

var reviewString = joinResults(rows)

// Menghilangkan tanda baca dan spasi
var documents = reviewString.match( /[\^\.\!\?]+\.[\!\?]+/g );

// Menjalankan fungsi Latent Dirichlet Allocation
var results = lda(documents,20,5)

var product = req.params.productName.replace('%20', ' ')
console.log(product)

dataQuery.getProductDetail(product.toString(),function
(error,details){
    // Render results to view
    res.render('detail', {
        title: 'amazoLDA',
        terms: results,
        productDetail: details
    })
})
})
})
})
})

```

**./Controller/Dataquery.js**

```

module.exports.getReviewDetail = function (callback){
    // Do something ...
    setQuery('select review_detail from review', function(err, rows){
        callback(err,rows)
    })
}

module.exports.getProductDetail = function (productName,callback){
    // Do something ...
    setQuery('select * from product join review on product.product_id =
review.review_product where product_name = '"+productName+"',
function(err, rows){
        callback(err,rows)
    })
}

```

**detail.hbs**

```

<div class="container detail">
    <div class="row">

        <div class="col-xs-6 col-sm-6 col-md-6 col-lg-6" id="image">

```

```


</div>

<div class="col-xs-6 col-sm-6 col-md-6 col-lg-6">
  <h2 class="product_name">{{productDetail.0.product_name}}
  @ ${{productDetail.0.price}}</h2>
  <div class="your-class">
    {{#each terms as |value key|}}

      <div>
        <h4>Topic: {{@index}}</h4>
        <canvas id="{{@index}}" class="barchart" width="100"
height="50"></canvas>
        {{#each this}}
          <div id="{{this.term}}" class="terms {{key}}"></div>
          <div id="{{this.probability}}" class="probability
{{key}}"></div>
          {{/each}}
        </div>
        {{/each}}
      </div>
    </div>
  </div>
</div>
<div class="card top-adjust">
  {{#each productDetail}}
    <div class="content review">
      <h3><span>{{this.rating_value}}</span>
      |
      {{this.review_title}}</h3>
      <p><a href="{{thisreviewer-href}}">{{thisreviewer}}</a>
      <span>{{thisreview_date}}</span></p>
      <p><span class="verified">{{thisreview_verified}}</span></p>
      <br>
      <p class="review-details">{{thisreview_detail}}</p>
    </div>
  {{/each}}
</div>
</div>

```

*Bagan 60 Implementasi kode pada halaman detail produk*

*Halaman ini sengaja dikosongkan*

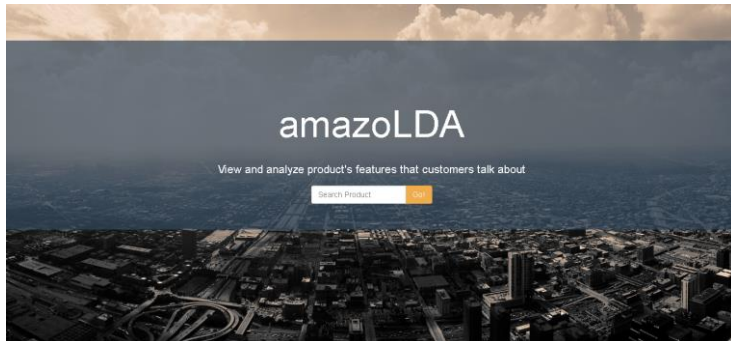
## BAB VI

### HASIL DAN PEMBAHASAN

#### 6.1. Hasil Prototipe Aplikasi

Hasil prototipe aplikasi berupa antar muka pengguna yang siap dipakai dengan kapabilitas terbatas hanya pada fitur pencarian produk dan melihat detail produk yang dipilih oleh pengguna. Terdapat tiga tampilan antar muka antara lain halaman pencarian produk, halaman daftar produk, dan halaman detail produk. Berikut adalah penjelasan mengenai masing-masing fitur halaman.

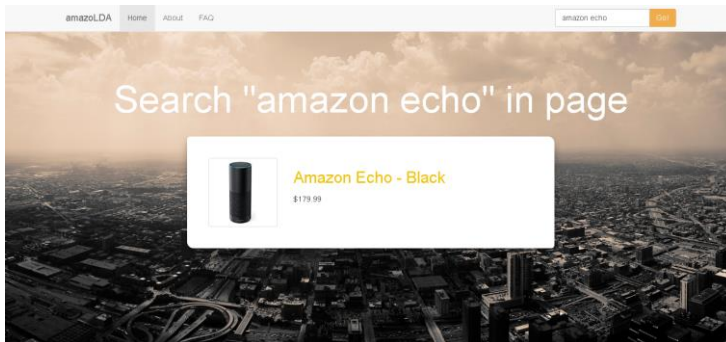
##### 6.1.1. Halaman Pencarian Produk



*Bagan 61 Halaman Pencarian Produk*

Halaman pada gambar 52 merupakan tampilan awal dari sistem AmazoLDA, nama aplikasi untuk sistem ekstraksi fitur dan opini untuk pengambilan keputusan pembelian pelanggan. Pengguna memasukkan kata kunci untuk menemukan barang dan klik cari untuk mencari barang yang ingin dicari dan dilihat lebih lanjut.

### 6.1.2. Halaman Daftar Produk

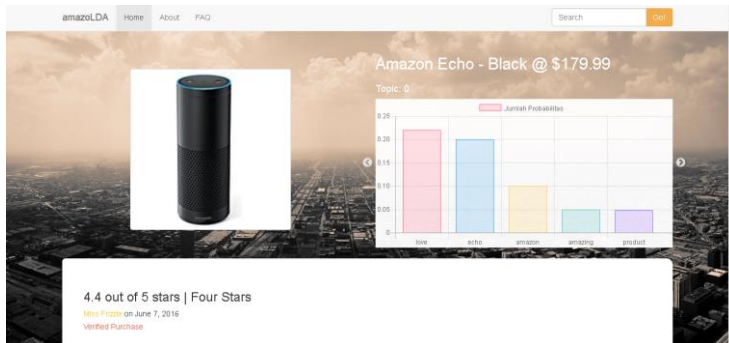


*Bagan 62 Halaman Daftar Produk*

Halaman pada gambar 41 merupakan tampilan untuk menampilkan daftar barang sesuai dengan kata kunci pencarian. Data yang ditampilkan antara lain daftar produk berupa gambar, nama produk, dan harga produk. Pengguna dapat memilih produk yang dimaksud untuk dilihat detailnya lebih lanjut. Pengguna juga dapat mencari produk lain berdasarkan kata kunci apabila belum menemukan barang yang dimaksud. Kemudian menu navigasi dibuat untuk kembali ke halaman pencarian, menuju ke halaman tentang aplikasi, dan menuju ke halaman FAQ.

### 6.1.3. Halaman Pencarian Produk

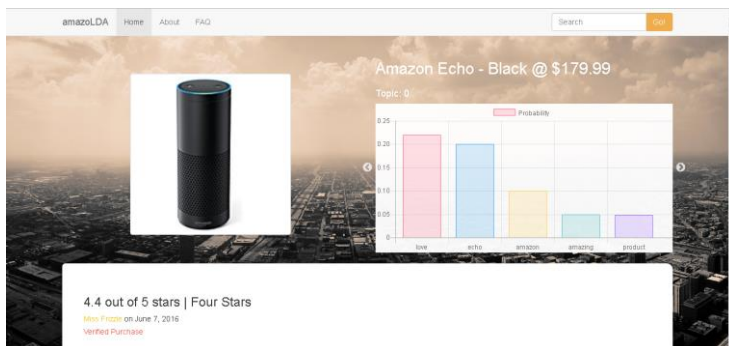
Halaman pada gambar 43 merupakan tampilan untuk menampilkan detail produk yang dipilih dari halaman daftar produk. Informasi produk seperti gambar, nama produk dan harga akan ditampilkan berikut hasil analisis dengan LDA. Wujudnya berupa grafik batang yang menampilkan kata kunci yang paling sering dibicarakan dalam ulasan produk.



Bagan 63 Halaman Detail Produk

## 6.2. Hasil Analisis LDA

Setelah memilih detail produk yang ingin dilihat, maka pengguna dapat melihat data ulasan beserta probabilitas kata yang paling sering muncul, Gambar 39 merupakan tampilan aplikasi dengan sistem ekstraksi fitur dan opini pada suatu produk.



Bagan 64 Grafik hasil analisis LDA

Tabel berikut menunjukkan parameter yang digunakan untuk melakukan analisis LDA.

Tabel 19 Parameter LDA yang digunakan untuk analisis

| Nama Parameter          | Nilai              |
|-------------------------|--------------------|
| Dokumen                 | 48.998 data ulasan |
| Jumlah Topik            | 20                 |
| Jumlah Kata dalam topik | 5                  |

Berikut adalah data berisi hasil analisis LDA yang menampilkan kata kunci dengan probabilitas yang paling sering muncul untuk masing-masing topik.

### 6.2.1. Parameter Analisis LDA

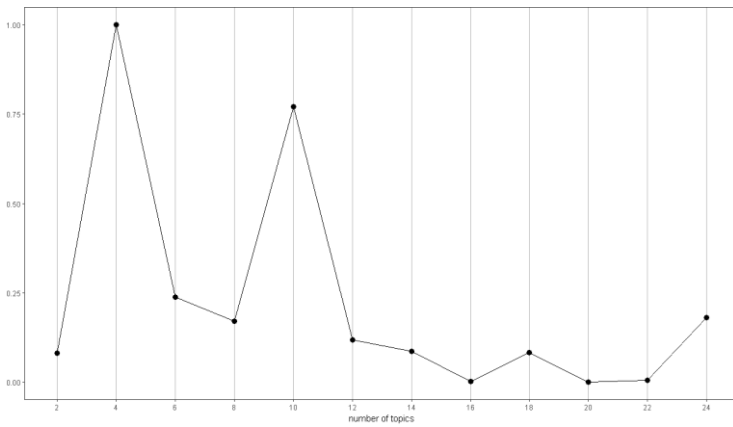
Parameter yang digunakan dalam analisis LDA antara lain adalah dokumen, jumlah topik dalam korpus dokumen yang teridentifikasi, dan jumlah kata kunci yang ingin ditampilkan per topik yang teridentifikasi. Berikut ini adalah penjelasan untuk masing-masing parameter.

Tabel 20 Parameter Analisis LDA

| Parameter                   | Deskripsi                                                                                 |
|-----------------------------|-------------------------------------------------------------------------------------------|
| Dokumen                     | Merupakan kumpulan teks yang akan dianalisis. Dokumen dipecah / ditokenisasi per kalimat. |
| Jumlah topik                | Jumlah topik yang teridentifikasi dalam dokumen.                                          |
| Jumlah kata kunci per topik | Jumlah kata kunci per topik yang ingin ditampilkan.                                       |

Untuk mengetahui jumlah topik yang optimum, maka perlu dilakukan pengujian sebelumnya. Pengujian dilakukan dengan melihat nilai minimum log likelihood sebagaimana yang diusulkan oleh Griffiths2004 [37].

Beberapa model LDA akan diuji dengan metrik-metrik sebagaimana yang diusulkan di atas. Model LDA diuji dengan menggunakan library R `ldatuning` [38]. Berikut adalah hasil pengujian parameter model LDA dengan menggunakan `ldatuning`.



*Bagan 65 Perhitungan log-likelihood untuk menentukan jumlah topik teridentifikasi dalam dokumen*

Dapat terlihat pada grafik pada fungsi maxima, nilai sempat mengalami jatuh kemudian naik lagi pada jumlah topik yang ke 20. Begitu pula dengan fungsi minima yang mencapai titik paling rendah pada topik ke 20. Sehingga dapat disimpulkan bahwa jumlah topik optimalnya adalah 20 – 22 topik.

### 6.2.2. Hasil Matriks LDA

Hasil dari analisis LDA berupa matriks yang berisi kumpulan kata yang dikategorikan per topik beserta probabilitas distribusi kata dalam suatu topik. Berikut adalah penjelasan kata-kata yang teridentifikasi untuk masing-masing topik.

### 6.2.2.1. Hasil Kata Kunci Topik 1

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 1

*Tabel 21 Hasil Kata Kunci untuk Topik 1*

|               |         |
|---------------|---------|
| 0/term        | love    |
| 0/probability | 0.22    |
| 1/term        | echo    |
| 1/probability | 0.2     |
| 2/term        | amazon  |
| 2/probability | 0.1     |
| 3/term        | amazing |
| 3/probability | 0.05    |
| 4/term        | product |
| 4/probability | 0.048   |

Pada topik pertama dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *love* disusul dengan *echo*, *amazon*, *amazing*, dan produk dengan probabilitas masing-masing adalah 0.22 untuk kata *love*, 0.2 untuk kata *echo*, 0.1 untuk kata *amazon*, 0.05 untuk kata *amazing* dan 0.048 untuk kata *product*. Hal ini mengindikasikan bahwa pengguna sangat menyukai produk amazon echo. Namun topik 1 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.2. Hasil Kata Kunci Topik 2

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 2

*Tabel 22 Hasil Kata Kunci untuk Topik 2*

|               |         |
|---------------|---------|
| 0/term        | sound   |
| 0/probability | 0.14    |
| 1/term        | speaker |
| 1/probability | 0.092   |
| 2/term        | good    |
| 2/probability | 0.091   |
| 3/term        | great   |
| 3/probability | 0.085   |
| 4/term        | quality |
| 4/probability | 0.061   |

Pada topik kedua dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *sound* disusul dengan *speaker*, *good*, *great*, dan *quality* dengan probabilitas masing-masing adalah 0.14 untuk kata *sound*, 0.092 untuk kata *speaker*, 0.091 untuk kata *good*, 0.085 untuk kata *great* dan 0.061 untuk kata *quality*. Hal ini mengindikasikan bahwa suara yang dihasilkan produk speaker amazon echo bagus. Topik 2 mulai memperlihatkan aspek dan sentimen ulasan terhadap fitur produk.

### 6.2.2.3. Hasil Kata Kunci Topik 3

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 3

*Tabel 23 Hasil Kata Kunci untuk Topik 3*

|               |             |
|---------------|-------------|
| 0/term        | voice       |
| 0/probability | 0.12        |
| 1/term        | alexa       |
| 1/probability | 0.092       |
| 2/term        | commands    |
| 2/probability | 0.066       |
| 3/term        | time        |
| 3/probability | 0.037       |
| 4/term        | recognition |
| 4/probability | 0.037       |

Pada topik ketiga dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *voice* disusul dengan *alexa*, *commands*, *time*, dan *recognition* dengan probabilitas masing-masing adalah 0.12 untuk kata *voice*, 0.092 untuk kata *alexa*, 0.066 untuk kata *commands*, 0.037 untuk kata *time* dan 0.037 untuk kata *recognition*. Namun topik 3 belum dapat menghasilkan aspek dari produk amazon.

#### 6.2.2.4. Hasil Kata Kunci Topik 4

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 4.

*Tabel 24 Hasil Kata Kunci untuk Topik 4*

|               |          |
|---------------|----------|
| 0/term        | echo     |
| 0/probability | 0.066    |
| 1/term        | amazon   |
| 1/probability | 0.056    |
| 2/term        | future   |
| 2/probability | 0.048    |
| 3/term        | learning |
| 3/probability | 0.045    |
| 4/term        | updates  |
| 4/probability | 0.043    |

Pada topik keempat dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *echo* disusul dengan *amazon*, *future*, *learning*, dan *updates* dengan probabilitas masing-masing adalah 0.066 untuk kata *echo*, 0.056 untuk kata *amazon*, 0.048 untuk kata *future*, 0.045 untuk kata *learning* dan 0.043 untuk kata *updates*. Namun topik 4 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.5. Hasil Kata Kunci Topik 5

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 5.

*Tabel 25 Hasil Kata Kunci untuk Topik 5*

|               |        |
|---------------|--------|
| 0/term        | echo   |
| 0/probability | 0.081  |
| 1/term        | alexa  |
| 1/probability | 0.076  |
| 2/term        | family |
| 2/probability | 0.066  |
| 3/term        | life   |
| 3/probability | 0.048  |
| 4/term        | day    |
| 4/probability | 0.042  |

Pada topik kelima dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *echo* disusul dengan *alexa*, *family*, *life*, dan *day* dengan probabilitas masing-masing adalah 0.081 untuk kata *echo*, 0.076 untuk kata *alexa*, 0.066 untuk kata *family*, 0.048 untuk kata *life* dan 0.042 untuk kata *day*. Namun topik 5 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.6. Hasil Kata Kunci Topik 6

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 6.

*Tabel 26 Hasil Kata Kunci untuk Topik 6*

|               |       |
|---------------|-------|
| 0/term        | love  |
| 0/probability | 0.19  |
| 1/term        | alexa |
| 1/probability | 0.11  |
| 2/term        | kids  |
| 2/probability | 0.059 |
| 3/term        | words |
| 3/probability | 0.031 |
| 4/term        | wife  |
| 4/probability | 0.029 |

Pada topik keenam dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *love* disusul dengan *alexa*, *kids*, *words*, dan *wife* dengan probabilitas masing-masing adalah 0.19 untuk kata *love*, 0.11 untuk kata *alexa*, 0.059 untuk kata *kids*, 0.031 untuk kata *words* dan 0.029 untuk kata *wife*. Namun topik 6 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.7. Hasil Kata Kunci Topik 7

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 7.

*Tabel 27 Hasil Kata Kunci untuk Topik 7*

|               |            |
|---------------|------------|
| 0/term        | questions  |
| 0/probability | 0.12       |
| 1/term        | alexa      |
| 1/probability | 0.09       |
| 2/term        | answer     |
| 2/probability | 0.087      |
| 3/term        | understand |
| 3/probability | 0.036      |
| 4/term        | talk       |
| 4/probability | 0.035      |

Pada topik ketujuh dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *questions* disusul dengan *alexa*, *answer*, *understand*, dan *talk* dengan probabilitas masing-masing adalah 0.12 untuk kata *questions*, 0.09 untuk kata *alexa*, 0.087 untuk kata *answer*, 0.036 untuk kata *understand* dan 0.035 untuk kata *talk*. Hal ini mengindikasikan bahwa product amazon echo dapat memahami permintaan pengguna dengan tepat.

### 6.2.2.8. Hasil Kata Kunci Topik 8

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 8.

*Tabel 28 Hasil Kata Kunci untuk Topik 8*

|               |            |
|---------------|------------|
| 0/term        | echo       |
| 0/probability | 0.069      |
| 1/term        | reviews    |
| 1/probability | 0.057      |
| 2/term        | recommend  |
| 2/probability | 0.049      |
| 3/term        | personable |
| 3/probability | 0.045      |
| 4/term        | purchase   |
| 4/probability | 0.032      |

Pada topik kedelapan dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *echo* disusul dengan *reviews*, *recommend*, *personable*, dan *purchase* dengan probabilitas masing-masing adalah 0.069 untuk kata *questions*, 0.057 untuk kata *reviews*, 0.049 untuk kata *recommend*, 0.045 untuk kata *personable* dan 0.032 untuk kata *purchase*. Hal ini mengindikasikan bahwa product amazon echo sangat personal dan layak untuk direkomendasikan dalam katalog pembelian.

### 6.2.2.9. Hasil Kata Kunci Topik 9

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 9.

*Tabel 29 Hasil Kata Kunci untuk Topik 9*

|               |        |
|---------------|--------|
| 0/term        | love   |
| 0/probability | 0.11   |
| 1/term        | echo   |
| 1/probability | 0.078  |
| 2/term        | bought |
| 2/probability | 0.078  |
| 3/term        | gift   |
| 3/probability | 0.069  |
| 4/term        | year   |
| 4/probability | 0.039  |

Pada topik kesembilan dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *love* disusul dengan *echo*, *bought*, *gift*, dan *year* dengan probabilitas masing-masing adalah 0.11 untuk kata *love*, 0.078 untuk kata *echo*, 0.078 untuk kata *gift*, 0.069 untuk kata *gift* dan 0.039 untuk kata *year*. Namun topik 9 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.10. Hasil Kata Kunci Topik 10

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 10.

*Tabel 30 Hasil Kata Kunci untuk Topik 10*

|               |          |
|---------------|----------|
| 0/term        | list     |
| 0/probability | 0.17     |
| 1/term        | shopping |
| 1/probability | 0.11     |
| 2/term        | add      |
| 2/probability | 0.066    |
| 3/term        | items    |
| 3/probability | 0.035    |
| 4/term        | app      |
| 4/probability | 0.031    |

Pada topik kesepuluh dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *list* disusul dengan *shopping*, *add*, *items*, dan *app* dengan probabilitas masing-masing adalah 0.17 untuk kata *list*, 0.11 untuk kata *shopping*, 0.066 untuk kata *add*, 0.035 untuk kata *items* dan 0.031 untuk kata *app*. Namun topik 10 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.11. Hasil Kata Kunci Topik 11

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 11.

*Tabel 31 Hasil Kata Kunci untuk Topik 11*

|               |        |
|---------------|--------|
| 0/term        | easy   |
| 0/probability | 0.065  |
| 1/term        | app    |
| 1/probability | 0.065  |
| 2/term        | works  |
| 2/probability | 0.054  |
| 3/term        | set    |
| 3/probability | 0.042  |
| 4/term        | skills |
| 4/probability | 0.027  |

Pada topik ketujuh dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *easy* disusul dengan *app*, *works*, *set*, dan *skills* dengan probabilitas masing-masing adalah 0.065 untuk kata *easy*, 0.065 untuk kata *app*, 0.054 untuk kata *works*, 0.042 untuk kata *set* dan 0.027 untuk kata *skills*. Hal ini mengindikasikan bahwa aplikasi yang merupakan product samping amazon echo mudah untuk digunakan bersamaan dengan amazon echo.

### 6.2.2.12. Hasil Kata Kunci Topik 12

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 12.

*Tabel 32 Hasil Kata Kunci untuk Topik 12*

|               |          |
|---------------|----------|
| 0/term        | weather  |
| 0/probability | 0.094    |
| 1/term        | music    |
| 1/probability | 0.085    |
| 2/term        | news     |
| 2/probability | 0.082    |
| 3/term        | play     |
| 3/probability | 0.066    |
| 4/term        | mornings |
| 4/probability | 0.029    |

Pada topik keduabelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *weather* disusul dengan *music*, *news*, *play*, dan *mornings* dengan probabilitas masing-masing adalah 0.094 untuk kata *weather*, 0.085 untuk kata *music*, 0.082 untuk kata *news*, 0.066 untuk kata *play*, dan 0.029 untuk kata *mornings*. Namun topik 12 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.13. Hasil Kata Kunci Topik 13

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 13.

*Tabel 33 Hasil Kata Kunci untuk Topik 13*

|               |         |
|---------------|---------|
| 0/term        | lights  |
| 0/probability | 0.11    |
| 1/term        | control |
| 1/probability | 0.089   |
| 2/term        | home    |
| 2/probability | 0.064   |
| 3/term        | turn    |
| 3/probability | 0.049   |
| 4/term        | hue     |
| 4/probability | 0.034   |

Pada topik ketigabelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *lights* disusul dengan *control*, *home*, *turn*, dan *hue* dengan probabilitas masing-masing adalah 0.11 untuk kata *lights*, 0.089 untuk kata *control*, 0.064 untuk kata *home*, 0.049 untuk kata *turn*, dan 0.034 untuk kata *hue*. Namun topik 13 belum dapat menghasilkan aspek dari produk amazon.

#### 6.2.2.14. Hasil Kata Kunci Topik 14

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 14.

*Tabel 34 Hasil Kata Kunci untuk Topik 14*

|               |         |
|---------------|---------|
| 0/term        | devices |
| 0/probability | 0.084   |
| 1/term        | home    |
| 1/probability | 0.054   |
| 2/term        | echo    |
| 2/probability | 0.044   |
| 3/term        | time    |
| 3/probability | 0.041   |
| 4/term        | fun     |
| 4/probability | 0.038   |

Pada topik keempatbelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *devices* disusul dengan *home*, *echo*, *time*, dan *fun* dengan probabilitas masing-masing adalah 0.084 untuk kata *devices*, 0.054 untuk kata *home*, 0.044 untuk kata *echo*, 0.041 untuk kata *time*, dan 0.038 untuk kata *fun*. Hal ini menunjukkan bahwa product Amazon Echo sangat menyenangkan untuk digunakan..

### 6.2.2.15. Hasil Kata Kunci Topik 15

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 15.

*Tabel 35 Hasil Kata Kunci untuk Topik 15*

|               |            |
|---------------|------------|
| 0/term        | devices    |
| 0/probability | 0.053      |
| 1/term        | amazing    |
| 1/probability | 0.049      |
| 2/term        | phone      |
| 2/probability | 0.039      |
| 3/term        | cool       |
| 3/probability | 0.035      |
| 4/term        | technology |
| 4/probability | 0.026      |

Pada topik kelimabelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *devices* disusul dengan *amazing*, *phone*, *cool*, dan *techonology* dengan probabilitas masing-masing adalah 0.053 untuk kata *devices*, 0.054 untuk kata *home*, 0.044 untuk kata *echo*, 0.041 untuk kata *time*, dan 0.038 untuk kata *fun*. Hal ini menunjukkan bahwa product Amazon Echo sangat menyenangkan untuk digunakan..

### 6.2.2.16. Hasil Kata Kunci Topik 16

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 16.

*Tabel 36 Hasil Kata Kunci untuk Topik 16*

|               |          |
|---------------|----------|
| 0/term        | music    |
| 0/probability | 0.17     |
| 1/term        | play     |
| 1/probability | 0.11     |
| 2/term        | prime    |
| 2/probability | 0.071    |
| 3/term        | amazon   |
| 3/probability | 0.046    |
| 4/term        | listened |
| 4/probability | 0.045    |

Pada topik keenambelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *music* disusul dengan *play*, *prime*, *amazon*, dan *listened* dengan probabilitas masing-masing adalah 0.17 untuk kata *music*, 0.11 untuk kata *play*, 0.071 untuk kata *prime*, 0.046 untuk kata *amazon*, dan 0.045 untuk kata *fun*. Namun topik 16 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.17. Hasil Kata Kunci Topik 17

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 17.

*Tabel 37 Hasil Kata Kunci untuk Topik 17*

|               |        |
|---------------|--------|
| 0/term        | music  |
| 0/probability | 0.11   |
| 1/term        | set    |
| 1/probability | 0.075  |
| 2/term        | timers |
| 2/probability | 0.069  |
| 3/term        | alarm  |
| 3/probability | 0.058  |
| 4/term        | play   |
| 4/probability | 0.051  |

Pada topik keenambelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *music* disusul dengan *set*, *timers*, *alarm*, dan *play* dengan probabilitas masing-masing adalah 0.11 untuk kata *music*, 0.075 untuk kata *set*, 0.069 untuk kata *timers*, 0.058 untuk kata *alarm*, dan 0.058 untuk kata *play*. Namun topik 17 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.18. Hasil Kata Kunci Topik 18

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 18.

*Tabel 38 Hasil Kata Kunci untuk Topik 18*

|               |          |
|---------------|----------|
| 0/term        | echo     |
| 0/probability | 0.079    |
| 1/term        | amazon   |
| 1/probability | 0.062    |
| 2/term        | purchase |
| 2/probability | 0.06     |
| 3/term        | worth    |
| 3/probability | 0.032    |
| 4/term        | price    |
| 4/probability | 0.032    |

Pada topik ketujuhbelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *echo* disusul dengan *amazon*, *purchase*, *worth*, dan *price* dengan probabilitas masing-masing adalah 0.079 untuk kata *echo*, 0.062 untuk kata *amazon*, 0.06 untuk kata *purchase*, 0.032 untuk kata *worth*, dan 0.032 untuk kata *price*. Hal ini menunjukkan bahwa product amazon echo layak dibeli dan sepadan dengan harga yang diperlukan untuk membeli amazon echo.

### 6.2.2.19. Hasil Kata Kunci Topik 19

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 19.

*Tabel 39 Hasil Kata Kunci untuk Topik 19*

|               |        |
|---------------|--------|
| 0/term        | alexa  |
| 0/probability | 0.087  |
| 1/term        | room   |
| 1/probability | 0.062  |
| 2/term        | house  |
| 2/probability | 0.055  |
| 3/term        | living |
| 3/probability | 0.05   |
| 4/term        | turn   |
| 4/probability | 0.034  |

Pada topik ketujuhbelas dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *alexa* disusul dengan *room*, *house*, *living*, dan *turn* dengan probabilitas masing-masing adalah 0.087 untuk kata *alexa*, 0.062 untuk kata *room*, 0.055 untuk kata *house*, 0.05 untuk kata *living*, dan 0.034 untuk kata *turn*. Namun topik 19 belum dapat menghasilkan aspek dari produk amazon.

### 6.2.2.20. Hasil Kata Kunci Topik 20

Berikut adalah hasil kata kunci dengan probabilitas kemunculan kata pada dokumen untuk topik 20.

*Tabel 40 Hasil Kata Kunci untuk Topik 20*

|               |         |
|---------------|---------|
| 0/term        | great   |
| 0/probability | 0.19    |
| 1/term        | product |
| 1/probability | 0.11    |
| 2/term        | works   |
| 2/probability | 0.098   |
| 3/term        | fun     |
| 3/probability | 0.062   |
| 4/term        | awesome |
| 4/probability | 0.062   |

Pada topik kedupuluh dapat dilihat bahwa kata kunci yang paling sering muncul dalam dokumen adalah kata *great* disusul dengan *product*, *works*, *fun*, dan *awesome* dengan probabilitas masing-masing adalah 0.19 untuk kata *great*, 0.11 untuk kata *product*, 0.098 untuk kata *works*, 0.062 untuk kata *fun*, dan 0.062 untuk kata *awesome*. Hal ini menunjukkan bahwa product amazon echo merupakan produk yang hebat dan menyenangkan untuk dipakai.

**6.2.3. Hasil Ekstraksi Fitur Produk**

**6.2.4. Hasil Orientasi Opini Fitur Produk**

### 6.3. Hasil Uji Usabilitas

Uji usabilitas digunakan untuk mengetahui apakah fitur yang diusulkan akan membantu pelanggan dalam mengambil keputusan pembelian pelanggan berdasarkan ulasan produk yang ada. Uji usabilitas dilakukan pada tanggal 5 Januari 2017 sebanyak 5 responden dengan kondisi sistem telah jadi dan matang (NodeJS + MySQL).

Beberapa evaluasi yang diperlukan dijabarkan pada tabel beriku yang akan dijadikan backlog untuk iterasi berikutnya.

*Tabel 41 Backlog Aplikasi*

| ID   | Backlog                                                                                                                                           |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| B001 | Memperbaiki sistem grafik agar dapat menampilkan legenda, judul grafik, dan penjelasan grafik                                                     |
| B001 | Memperbaiki sistem grafik agar menamai topik berdasarkan fitur. Contoh: kata kunci fitur yang ditampilkan diangkat menjadi topik yang dibicarakan |
| B001 | Memperbaiki sistem LDA agar menampilkan topik beserta kata kunci yang lebih sesuai                                                                |

Hasil dari *System Usability Scale* yang diberikan menghasilkan kemudahan dalam penggunaan sebesar 60% sebagaimana yang terlihat pada tabel berikut. Hasil tersebut dijabarkan dalam rata-rata pengguna setuju dengan pernyataan yang diajukan pada tabel berikutnya.

| Peserta                     | Jenis Kelamin | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10   |     |
|-----------------------------|---------------|----|----|----|----|----|----|----|----|----|-------|-----|
| Alitya Novianda Adityaputri | P             | 5  | 5  | 1  | 4  | 3  | 1  | 2  | 3  | 4  | 2     | 60% |
| Tesar Akram                 | L             | 4  | 4  | 2  | 4  | 4  | 1  | 3  | 4  | 4  | 2     | 64% |
| Faisal Setia Putra          | L             | 4  | 5  | 1  | 3  | 3  | 1  | 3  | 3  | 4  | 3     | 60% |
| Mochamad Fahmi Zamroni      | L             | 5  | 4  | 2  | 3  | 3  | 1  | 4  | 3  | 3  | 1     | 58% |
| Ashma Hanifah               | P             | 4  | 5  | 1  | 2  | 4  | 1  | 3  | 4  | 4  | 1     | 58% |
|                             |               |    |    |    |    |    |    |    |    |    | Total | 60% |

*Bagan 66 Hasil Rekapitulasi SUS*

| Pernyataan                                                                                   | Justifikasi  |
|----------------------------------------------------------------------------------------------|--------------|
| Pengguna berpikir bahwa saya ingin menggunakan sistem ini sesering mungkin                   | Setuju       |
| Pengguna berpikir sistem ini sangat rumit dan tidak perlu                                    | Tidak Setuju |
| Pengguna berpikir sistem ini mudah digunakan                                                 | Setuju       |
| Pengguna berpikir saya akan membutuhkan bantuan seseorang untuk dapat menggunakan sistem ini | Tidak Setuju |
| Pengguna menemukan seluruh fungsi yang telah terintegrasi dengan sangat baik                 | Setuju       |
| Pengguna berpikir terlalu banyak inkonsistensi dari sistem ini                               | Tidak Setuju |
| Pengguna berpikir banyak orang yang akan dapat belajar menggunakan sistem ini dengan cepat   | Setuju       |
| Pengguna berpikir bahwa sistem ini sangat sulit untuk digunakan                              | Tidak Setuju |
| Pengguna merasa sangat percaya diri dalam menggunakan sistem ini                             | Setuju       |
| Pengguna perlu belajar banyak hal sebelum dapat menggunakan sistem ini                       | Tidak Setuju |

## BAB VII

### KESIMPULAN DAN SARAN

Pada bab ini akan dijelaskan mengenai kesimpulan dari hasil penelitian dan saran untuk pengembangan penelitian di kemudian hari.

#### 7.1. Kesimpulan

Berdasarkan dengan pengerjaan tugas akhir dengan judul "Rancang Bangun Ekstraksi Topik Fitur Produk Dari Ulasan Pengguna Online Dengan Latent Dirichlet Allocation" yang telah dilakukan dapat disimpulkan beberapa hal sebagai berikut:

1. Dari tahapan analisis, didapatkan metadata ulasan yang kemudian diturunkan menjadi rancangan basis data ulasan dan digunakan sebagai acuan untuk melakukan proses pengambilan data.
2. Dari hasil pengujian LDA, didapatkan bahwa dari 20 topik yang dihasilkan 8 topik memberikan aspek dan sentimen terhadap produk amazon echo. Indikasi produk tersebut memiliki sentimen yang positif antara lain:
  - a. Product amazon echo memiliki *speaker* yang bagus.
  - b. *Voice recognition* dari amazon echo dapat memahami permintaan pengguna dengan tepat.
  - c. Produk amazon echo sangat personal dan layak untuk dibeli.

- d. Aplikasi yang merupakan produk samping amazon echo mudah untuk digunakan bersamaan dengan amazon echo.
- e. Product amazon echo sangat menyenangkan untuk digunakan.
- f. Aplikasi amazon echo merupakan produk yang canggih dan keren dimata pengguna.
- g. Produk amazon echo layak untuk dibeli dan sepadan dengan biaya yang dikeluarkan pengguna.
- h. Produk amazon echo merupakan produk yang hebat dan menyenangkan untuk dipakai.
- i. Hasil *System Usability Scale* yang dilakukan terhadap 5 orang, didapatkan bahwa sistem telah mencapai 60% telah memenuhi kemudahan dalam penggunaan aplikasi.

## 7.2. Saran

Saran penulis untuk penelitian dan pengembangan selanjutnya adalah sebagai berikut:

1. Pada milestone berikutnya, target yang perlu diselesaikan adlah product backlog hasil pengujian usabilitas
2. Menguji parameter LDA untuk beberapa skenario, dimana nilai yang diujikan adalah jumlah topik dan jumlah kata yang muncul dalam topik
3. Membuat semacam fitur *autocomplete* ketika membuat ulasan berdasarkan topik yang paling sering dibicarakan pelanggan.

## DAFTAR PUSTAKA

- [1] R. Buettner, "Predicting User Behavior in Electronic Markets Based on Personality-Mining in Large Online Social Network: A Personality Based Product Recommender Framework," Munich, 2016.
- [2] ATKearney, The 2015 Global Retail E-commerce Index, ATKearney, 2015.
- [3] International Telecommunications Union, "Global ICT Developments," 2016. [Online]. Available: <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>. [Diakses 6 October 2016].
- [4] Akamai Technologies, "Performance Matters: 9 Key Consumer Insights," [Online]. Available: <https://www.akamai.com/us/en/multimedia/documents/content/akamai-performance-matters-key-consumer-insights-ebook.pdf>. [Diakses 6 October 2016].
- [5] Z. Huang dan M. Benyoucef, "From e-commerce to social commerce; A close look at design features," *Electronic Commerce Research and Applications* 12, pp. 246-259, 2015.
- [6] J. Lee, D. Park dan I. Han, "The effect of negative online consumer reviews on product attitude: An information processing view," *Electronic Commerce Research and Application*, pp. 341-352, 2008.
- [7] C. Kim, R. D. Gliiers dan N. Shun, "Factors influencing Internet Shopping Value and Customer Repurchase

- Intention,” *Electronic Commerce Research and Applications*, pp. 374-387, 2012.
- [8] T. Klingberg, “Limitaions in information processing in the human brain: neuroimaging of dual task performance and working memory tasks,” *Brain Research*, pp. 95-102, 2000.
- [9] M. Hu dan B. Liu, “Mining Opinion Features in Customer Reviews,” *Artificial Intelligence*, pp. 755-760, 2004.
- [10] A. T. a. O. R. Zaian, “Mining Contentious Documents ] using an Unsupervised Topci Model Based Approach,” *Knowledge and Information Systems*, pp. 537-560, 2015.
- [11] D. M. Blei, A. Y. Ng dan M. .. Jordan, “Latent Dirichlet ] Allocation,” *Journal of Machine Learning Research* 3, pp. 993-1022, 2003.
- [12] Y. Lu dan M. Q. a. C.Zhai, “Investigating task ] performance of probabilistic topic models: an emperical study of PLSA and LDA,” *Information Retreival*, pp. 178-203, 2011.
- [13] Y. Jo dan A. Oh, “Aspect and Sentiment Unification ] Model,” dalam *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011.
- [14] X. Zheng et.al, “Incorporating appraisal expression ] patterns into topic modeling for aspect and sentiment word idenfication,” *Knowledge Based Systems*, pp. 28-47, 2014.

- [15 D. Whiteley, E-commerce: Strategy, Technologies and  
] Applications, McGraw-Hill, 2000.
  
- [16 M. Khosrow-Pour, Consumer Behavior, Organizational  
] Development, and Electronic Commerce: Emerging  
Issues for Advancing Modern Socioeconomies, Heshey:  
IGI Publishing, 2008.
  
- [17 M. a. C. Q. Schniederjans, E-Commerce operations  
] management, Singapore: World Scientific, 2002.
  
- [18 Data Bites, “How many products does Amazon sell,”  
] Data Bites Production, 1 December 2016. [Online].  
Available: <https://www.scrapehero.com/how-many-products-does-amazon-sell-august-2016/>. [Diakses 15  
December 2016].
  
- [19 R. T. B. R. I. a. B. J. Wigand, “Web 2.0 and beyond:  
] implications for electronic commerce,” dalam  
*Proceedings of the 10th International Conference on  
Electronic Commerce*, New York, 2008.
  
- [20 T. O'Reilly, “What Is Web 2.0: Design Patterns and  
] Business Models for the Next Generation of Software,”  
O'Reilly, 9 September 2005. [Online]. Available:  
<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>. [Diakses 31 December 2016].
  
- [21 W3School, “RegExp Reference,” W3School, [Online].  
] Available:  
[http://www.w3schools.com/jsref/jsref\\_obj\\_regexp.asp](http://www.w3schools.com/jsref/jsref_obj_regexp.asp).  
[Diakses 31 December 2016].
  
- [22 Mozilla Firefox, “XMLHttpRequest,” Mozilla Firefox  
] Production, [Online]. Available:

<https://developer.mozilla.org/en-US/docs/Web/API/XMLHttpRequest>. [Diakses 31 December 2016].

[23 Jaunt, “Jaunt Java Web Scraping & JSON Querying,”  
] Jaunt, [Online]. Available: <http://jaunt-api.com/>.  
[Diakses 1 January 2016].

[24 World Wide Web Consortium, “What is the Document  
] Object Model?,” World Wide Web Consortium  
Production, 13 November 2000. [Online]. Available:  
[https://www.w3.org/TR/DOM-Level-2-  
Core/introduction.html](https://www.w3.org/TR/DOM-Level-2-Core/introduction.html). [Diakses 1 January 2016].

[25 JSoup, “Use DOM methods to navigate a document,”  
] JSoup Production, [Online]. Available:  
[https://jsoup.org/cookbook/extracting-data/dom-  
navigation](https://jsoup.org/cookbook/extracting-data/dom-navigation). [Diakses 1 January 2016].

[26 T. Slimani, “Semantic Annotation: The Mainstay of  
] Semantic Web,” 2004. [Online]. Available:  
<https://arxiv.org/ftp/arxiv/papers/1312/1312.4794.pdf>.  
[Diakses 1 January 2016].

[27 Z. Zhou dan M. Mashuq, “Web Content Extraction  
] Through Machine Learning,” 2013. [Online]. Available:  
[http://cs229.stanford.edu/proj2013/ZhouMashuq-  
WebContentExtractionThroughMachineLearning.pdf](http://cs229.stanford.edu/proj2013/ZhouMashuq-WebContentExtractionThroughMachineLearning.pdf).  
[Diakses 1 January 2016].

[28 S. Shiliang, L. Chen dan C. Junyu, “A Review of Natural  
] Language Processing Techniques for Opinion Mining  
Systems,” *Information Fusion*, vol. 25, p. October, 2016.

- [29 R. Miktov, The Oxford Handbook of Computational  
] Linguistics, Oxford: OUP Oxford, 2003.
  
- [30 C. D. M. S. J. B. J. F. S. J. B. a. D. M. Manning, “The  
] Stanford CoreNLP Natural Language Processing  
Toolkit,” dalam *In Proceedings of the 52nd Annual  
Meeting of the Association for Computational  
Linguistics: System Demonstrations*, 2014.
  
- [31 A. P. J. Y. W. J. C. C. D. M. A. Y. N. a. C. P. Richard  
] Socher, “Recursive Deep Models for Semantic  
Compositionality Over a Sentiment Treebank,” dalam  
*Proceedings of the conference on empirical methods in  
natural language processing (EMNLP)*, 2013.
  
- [32 Oxford Dictionaries, “Word classes (or parts of speech),”  
] Oxford Dictionaries, [Online]. Available:  
[https://en.oxforddictionaries.com/grammar/word-classes-  
or-parts-of-speech](https://en.oxforddictionaries.com/grammar/word-classes-or-parts-of-speech). [Diakses 1 January 2016].
  
- [33 D. S. M. & C.-C. M. Rosenberg, Agile Development  
] with ICONIX Process, Apress, 2005.
  
- [34 K. Liu, L. Xu dan J. Zhao, “Co-Extracting Opinion  
] Targets and Opinion Words from Online Reviews Based  
on the Word Alignment Model,” *IEEE Transactions on  
Knowledge and Data Engineering*, pp. 636-650, 2015.
  
- [35 W. H. D. a. E. R. McLean, “Measuring e-Commerce  
] Success: Applying the,” *International Journal of  
Electronic Commerce*, vol. 9, no. 1, pp. 31-47, 2004.

*Halaman ini sengaja dikosongkan*

## BIODATA PENULIS



Penulis merupakan kelahiran Surabaya, 8 April 1995. Merupakan anak pertama dari tiga bersaudara. Pendidikan formal ditempuh dari SD Kristen Gloria II Surabaya lulus pada tahun 2007, SMP Kristen Gloria I Surabaya lulus pada tahun 2010, dan SMA Kristen Gloria I Surabaya lulus pada tahun 2013.

Ketertarikan penulis akan dunia komputer menghantarkannya ke pendidikan tinggi di Jurusan Sistem Informasi Institut Teknologi Sepuluh Nopember (ITS) Surabaya pada tahun 2013. Selama menjadi mahasiswa, penulis aktif dalam kegiatan kemahasiswaan seperti bergabung dalam Google Student Groups pada tahun 2014. Selain itu, penulis pernah terlibat dalam proyek Monvis Musrenbang pada tahun 2015 yang merupakan penelitian Jurusan Sistem Informasi untuk mengembangkan sistem informasi geografis yang dapat melacak progress usulan masyarakat yang dijalankan pemerintah. Penulis juga pernah terlibat dalam pengembangan start-up BantuTemu.in untuk mengembangkan sistem pelaporan barang hilang pada tahun 2015. Terakhir penulis menjalani kegiatan magang di PT. Abyor International sebagai programmer ABAP, mengembangkan fungsionalitas sistem SAP sesuai dengan spesifikasi yang diberikan.

Ketertarikan penulis akan dunia pemodelan topik muncul ketika mengambil mata kuliah integrasi aplikasi korporasi dan e-bisnis yang mempelajari karakteristik data pada *e-commerce* serta integrasi sistem. Pengalaman itu juga yang menjadi inspirasi penulis dalam penelitian pemodelan topik. Penulis dapat dihubungi melalui email [yabes.wirawan@gmail.com](mailto:yabes.wirawan@gmail.com).

## LAMPIRAN 1 USABILITY TEST PLAN DASHBOARD

|                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                        |                                                                                                                                                                                                                |                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| <b>Author:</b><br>Natanael Yabes Wirawan                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                      | <b>Contact:</b><br><a href="mailto:yabes.wirawan@gmail.com">yabes.wirawan@gmail.com</a>                                                                                |                                                                                                                                                                                                                | <b>Date:</b><br>06 Januari 2017                     |
| <b>Product Under Test:</b><br>Website AmazoLDA                                                                                                                                                                                                                                     | <b>TestObjectives</b> <ul style="list-style-type: none"><li>- Apakah user paham dengan tujuan dibuatnya aplikasi</li><li>- Apakah fitur ekstraksi fitur dan orientasi opini membantu user dalam mengambil keputusan pelanggan</li><li>- Apakah pengguna merasa nyaman menggunakan aplikasi</li></ul> | <b>Participant</b> <ul style="list-style-type: none"><li>- 5 orang</li><li>- Pernah membeli barang di situs online</li><li>- Terdiri dari laki dan perempuan</li></ul> | <b>Test Tasks</b> <ul style="list-style-type: none"><li>- Buka halaman utama</li><li>- Mencari barang dengan kata kunci “Amazon echo”</li><li>- Melihat detail produk</li><li>- Keluar dari aplikasi</li></ul> | <b>PIC</b><br>Natanael Yabes (PM, Moderator)        |
|                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                      | <b>Equipment</b> <ul style="list-style-type: none"><li>- Laptop dan aplikasi</li><li>- Form responden</li></ul>                                                        |                                                                                                                                                                                                                | <b>Location &amp; Dates</b><br>Jumat 6 Januari 2017 |
| <b>Business Case</b><br>Mendapatkan masukan untuk pengembangan berikutnya                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                        |                                                                                                                                                                                                                |                                                     |
| <b>Procedure</b> <div><div>Selamat datang /consent form 0-5 menit</div><div>Pre test interview 5-10 menit</div><div>Pengerjaan Test task 3 menit</div><div>Post test questionnaire 10 menit</div><div>Post test interview 5 menit</div><div>Pemberian insentif 3 menit</div></div> |                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                        |                                                                                                                                                                                                                |                                                     |

## LAMPIRAN 2 USABILITY TEST SCENARIO

|                                                              |
|--------------------------------------------------------------|
| <b>Web browser harus membuka halaman Google atau halaman</b> |
|--------------------------------------------------------------|

Hi, \_\_\_\_\_. nama saya \_\_\_\_\_, dan saya akan menemani Anda melalui sesi hari ini.

Sebelum kita mulai, saya punya beberapa informasi untuk Anda, dan saya akan membacanya untuk memastikan bahwa saya menyampaikan semuanya.

Anda mungkin sudah memiliki pemikiran yang baik tentang mengapa kami meminta Anda di sini, tapi biarkan saya menjelaskan kembali sebentar. Kami meminta orang untuk mencoba menggunakan sebuah situs Web yang kami kerjakan sehingga kita bisa melihat apakah web tersebut bekerja sesuai dengan tujuannya. sesi akan memakan waktu kurang lebih sekitar satu jam.

Hal pertama yang saya ingin perjas adalah bahwa kami sedang menguji situs, bukan Anda. Anda tidak akan dapat melakukan kesalahan di sini. Bahkan, ini mungkin adalah salah satu hari ini di mana Anda tidak perlu khawatir tentang membuat kesalahan.

Ketika Anda menggunakan situs ini, saya akan meminta Anda sebanyak mungkin untuk mencoba untuk berpikir keras: untuk mengatakan apa yang Anda cari di website, apa yang Anda coba lakukan, dan apa yang Anda pikirkan. Ini akan menjadi bantuan besar bagi kami.

Juga, jangan khawatir bahwa Anda akan menyakiti perasaan kami. Kami melakukan ini untuk meningkatkan Web kami, sehingga kami perlu mendengar reaksi jujur Anda.

Jika Anda memiliki pertanyaan selama menjalankan web, tanyakan saja. Saya mungkin tidak dapat menjawab seluruhnya, karena kami tertarik pada bagaimana orang-orang lakukan ketika mereka tidak memiliki seseorang yang duduk di samping mereka untuk membantu.

Tapi jika Anda masih memiliki pertanyaan ketika sudah selesai saya akan mencoba untuk menjawab pertanyaan yang ada. Dan jika Anda perlu mengambil istirahat pada setiap titik, Anda bisa memberitahukan saya.

Anda mungkin telah memperhatikan kamera. Dengan izin Anda, kami akan merekam apa yang terjadi di layar dan percakapan yang akan kita lakukan. Rekaman hanya akan digunakan untuk membantu kami mencari cara untuk meningkatkan website, dan itu tidak akan dilihat oleh siapapun kecuali orang yang bekerja pada proyek ini. Dan itu membantu saya, karena saya tidak perlu mengambil banyak catatan.

Dan Jika Anda mau, saya akan meminta Anda untuk menandatangani formulir izin sederhana bagi kita. Itu hanya mengatakan bahwa kami memiliki izin untuk merekam Anda, dan bahwa rekaman hanya akan dilihat oleh orang-orang yang bekerja pada proyek ini.

|                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p style="text-align: center;"><b>Beri mereka formulir izin perekaman dan pena</b><br/><b>Sementara mereka menandatangani, MULAI RECORDER</b><br/><b>SCREEN</b></p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Apakah Anda memiliki pertanyaan sejauh ini?

BAIK. Sebelum kita memulai untuk melihat website, saya ingin menanyakan Anda beberapa pertanyaan singkat.

Pertama, apa pekerjaan Anda? Apa yang Anda lakukan sepanjang hari?

Sekarang, kira-kira berapa jam seminggu sekali - hanya estimasi kasarnya - Anda menghabiskan waktu menggunakan Internet, termasuk Web browsing dan email, di tempat kerja dan di rumah?

Jenis situs apa yang Anda cari di saat Anda menelusuri Web?

Apakah Anda memiliki situs web favorit?

OK bagus. Kami sudah selesai dengan pertanyaan, dan kita bisa mulai melihat website.

**Buka Halaman Website <http://localhost:3000>**

Sekarang saya akan meminta Anda untuk mencoba melakukan beberapa tugas-tugas tertentu. Saya akan membaca setiap satu secara keras dan memberikan salinan cetak dari tugas tersebut.

Saya juga akan meminta Anda untuk melakukan tugas-tugas ini tanpa menggunakan fitur Search atau bantuan lainnya. Kami akan mempelajari lebih banyak tentang seberapa baik situs bekerja dengan cara seperti itu.

**Berikanlah peserta skenario pertama, dan baca dengan keras. Biarkan pengguna untuk melakukannya sampai Anda tidak merasa itu menghasilkan nilai atau pengguna menjadi sangat frustrasi. Ulangi untuk setiap tugas atau sampai waktu habis.**

Terima kasih, Anda telah sangat membantu.

Apakah Anda memiliki pertanyaan untuk saya, sekarang kita sudah selesai?

**Berikan insentif mereka, atau mengingatkan mereka akan dikirim kepada mereka. Hentikan perekam layar dan simpan file.  
Terima kasih mereka dan mengantar mereka keluar.**

## LAMPIRAN 3 CONSENT FORM

### Usability Testing

#### Pendahuluan

Testing ini dilakukan untuk melakukan penilaian terhadap kegunaan (Usabilitas) dari aplikasi yang akan dibuat oleh peneliti.

Aplikasi yang akan dibuat diberi nama AmazoLDA, aplikasi sistem ekstraksi fitur dan orientasi opini terhadap ulasan produk amazon.com menggunakan *Appraisal Expression Pattern* berbasis *Latent Dirichlet Allocation*.

Pengembangan aplikasi AmazoLDA memiliki tujuan sebagai berikut:

1. Untuk mengetahui fitur unggulan berdasarkan ulasan produk oleh konsumen produk terkait.
2. Untuk mengetahui seberapa baik atau buruk suatu sentimen terhadap produk yang diulas oleh konsumen produk terkait.
3. Untuk membantu konsumen dalam mengambil keputusan untuk membeli atau tidak membeli suatu produk berdasarkan ulasan terhadap produk yang telah diaggregasi.

Usability testing ini akan memakan waktu sekitar 30 menit dan media yang digunakan adalah paper prototype dari harapan hasil akhir dari aplikasi yang akan dibuat. seluruh fitur dan fungsi dari aplikasi dibuat berdasarkan literatur dan penelitian sebelumnya yang terkait dengan penelitian ini.

Terdapat 3 Bagian pertanyaan yang harus dijawab. Pada bagian pertama menggunakan tipe soal likert dengan skala 1 –

5 yang bertujuan mengukur usabilitas dari aplikasi. Pada bagian kedua menggunakan tipe soal bebas, untuk mengambil evaluasi dan saran untuk pengembangan aplikasi.

Terima kasih atas partisipasi anda.

---

### **FORM PERSETUJUAN**

Saya mengerti bahwa partisipasi dalam studi kegunaan ini bersifat sukarela.

Silakan isi form berikut untuk menunjukkan bahwa Anda telah membaca dan Anda memahami informasi pada formulir ini dan bahwa setiap pertanyaan yang mungkin Anda miliki tentang sesi telah dijawab.

Tanggal : \_\_\_\_\_

Nama Lengkap : \_\_\_\_\_

Jenis Kelamin : \_\_\_\_\_

Jurusan/Angkatan : \_\_\_\_\_

Tanda tangan : \_\_\_\_\_

## LAMPIRAN 4 SYSTEM USABILITY SCALE FORM

**Instruksi :** Isikan setiap jawaban pernyataan dengan memberikan tanda *checklist* (√) atau *cross* (x) pada salah satu kolom 1 – 5 yang menurut Anda paling sesuai dengan pendapat Anda.

1 = Sangat tidak setuju

2 = Tidak Setuju

3 = Netral

4 = Setuju

5 = Sangat setuju

| Pernyataan                                                                               | 1 | 2 | 3 | 4 | 5 |
|------------------------------------------------------------------------------------------|---|---|---|---|---|
| Saya berpikir bahwa saya ingin menggunakan sistem ini sesering mungkin                   |   |   |   |   |   |
| Saya berpikir sistem ini sangat rumit dan tidak perlu                                    |   |   |   |   |   |
| Saya berpikir sistem ini mudah digunakan                                                 |   |   |   |   |   |
| Saya berpikir saya akan membutuhkan bantuan seseorang untuk dapat menggunakan sistem ini |   |   |   |   |   |
| Saya menemukan seluruh fungsi yang telah terintegrasi dengan sangat baik                 |   |   |   |   |   |
| Saya berpikir terlalu banyak inkonsistensi dari sistem ini                               |   |   |   |   |   |
| Saya berpikir banyak orang yang akan dapat belajar menggunakan sistem ini dengan cepat   |   |   |   |   |   |
| Saya berpikir bahwa sistem ini sangat sulit untuk digunakan                              |   |   |   |   |   |
| Saya merasa sangat percaya diri dalam menggunakan sistem ini                             |   |   |   |   |   |
| Saya perlu belajar banyak hal sebelum dapat menggunakan sistem ini                       |   |   |   |   |   |

## **Bagian II Evaluasi**

1. Adakah fungsi dari aplikasi yang anda rasa tidak perlu ada? sebutkan

---

---

2. Adakah fungsi dari sistem yang menurut Anda penting namun belum terdapat pada aplikasi? Sebutkan

---

---

3. Adakah fungsi dari sistem yang menurut Anda penting namun belum terdapat pada aplikasi? Sebutkan

---

---