

TUGAS AKHIR - KS141501

**PENILAIAN KUALITAS ULASAN PELANGGAN
BERDASARKAN KARAKTERISTIK STRUKTURAL,
METADATA, DAN KETERBACAAN**

***EVALUATION OF CUSTOMER REVIEWS
QUALITY BASED ON STRUCTURAL,
METADATA, AND READABILITY
CHARACTERISTICS***

RANI OKTAVIA
NRP 5213 100 122

Dosen Pembimbing
Rully Agus Hendrawan, S.Kom., M.Eng.

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017

TUGAS AKHIR - KS141501

**PENILAIAN KUALITAS ULASAN PELANGGAN
BERDASARKAN KARAKTERISTIK STRUKTURAL,
METADATA, DAN KETERBACAAN**

RANI OKTAVIA
NRP 5213 100 122

Dosen Pembimbing
Rully Agus Hendrawan, S.Kom., M.Eng.

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017

UNDERGRADUATE THESES - KS141501

***EVALUATION OF CUSTOMER REVIEWS
QUALITY BASED ON STRUCTURAL,
METADATA, AND READABILITY
CHARACTERISTICS***

**RANI OKTAVIA
NRP 5213 100 122**

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

**DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017**

Lembar Pengesahan
PENILAIAN KUALITAS ULASAN PELANGGAN
BERDASARKAN KARAKTERISTIK STRUKTURAL,
METADATA, DAN KETERBACAAN

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

RANI OKTAVIA
NRP. 5213100122

Surabaya, 17 Juli 2017

KEPALA
DEPARTEMEN SISTEM INFORMASI

Dr. Ir. Aris Tjahyanto, M.Kom
NIP. 196503101991021001



Lembar Persetujuan
PENILAIAN KUALITAS ULASAN PELANGGAN
BERDASARKAN KARAKTERISTIK STRUKTURAL,
METADATA, DAN KETERBACAAN

TUGAS AKHIR

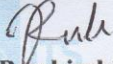
Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember


Oleh:


RANI OKTAVIA
NRP. 5213100122

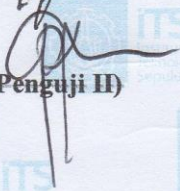
Disetujui Tim Penguji : Tanggal Ujian : 6 Juli 2017
Periode Wisuda: September 2017

Rully Agus Hendrawan, S.Kom., M.Eng.  (Pembimbing I)

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


(Penguji I)

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


(Penguji II)

**PENILAIAN KUALITAS ULASAN PELANGGAN
BERDASARKAN KARAKTERISTIK STRUKTURAL,
METADATA, DAN KETERBACAAN**

Nama Mahasiswa : Rani Oktavia
NRP : 5213100122
Jurusan : Sistem Informasi FTIF – ITS
**Pembimbing I : Rully Agus Hendrawan, S.Kom.,
M.Eng. Sc**

ABSTRAK

Ulasan online pelanggan memiliki peran penting dalam proses keputusan pembelian produk. Sekarang ini, semakin banyak ulasan online yang tersedia di pasar online sehingga pelanggan dapat membacanya untuk lebih memahami dengan baik mengenai produk atau jasa yang akan dibeli. Ulasan pelanggan menjadi informasi tambahan yang penting disamping informasi yang disediakan pasar online seperti deskripsi produk, ulasan dari ahli, dan rekomendasi dari sistem. Namun, karena semakin banyak ulasan pelanggan yang tersedia sekarang ini, muncullah pertanyaan apakah setiap ulasan tersebut berkualitas dan berguna bagi pelanggan lain. Berdasarkan hasil survei, sebanyak 87% pembeli akan membaca paling banyak 10 ulasan untuk menentukan keputusan pembelian. Hal ini dapat menimbulkan permasalahan karena adanya kemungkinan ulasan pelanggan yang baru ditulis dan berkualitas, namun belum populer sehingga tidak terbaca oleh pelanggan lain.

Oleh karena itu, diperlukan penilaian kualitas konten ulasan berdasarkan tiga karakteristik yaitu struktural, meta-data, dan keterbacaan menggunakan weighted sum yang dapat mengevaluasi beberapa alternatif berdasarkan kriteria tertentu. Kualitas konten ulasan juga akan dinilai menggunakan metode Support Vector Machine. Metode ini digunakan karena data memiliki fitur yang banyak yang

mencakup kategori struktural, meta-data, dan keterbacaan sehingga metode ini dapat digunakan untuk mengklasifikasi kualitas ulasan pelanggan. Hasil yang didapatkan menunjukkan bahwa perhitungan dengan weighted sum, diperoleh nilai kualitas ulasan tertinggi sebesar 0,736 dari skala 1 dan nilai kualitas ulasan terendah sebesar -0.104. Kategori yang paling mempengaruhi penilaian kualitas ulasan adalah nilai keterbacaan automated readability index, sedangkan kategori kegunaan tidak begitu mempengaruhi penilaian. Sedangkan berdasarkan hasil klasifikasi menggunakan support vector machine untuk memprediksi kegunaan suatu ulasan, diperoleh nilai keakuratan paling tinggi menggunakan kernel polynomial dengan nilai sebesar 94.4773%.

Kata kunci : Ulasan Online Pelanggan, Kualitas Ulasan, Weighted Sum, Support Vector Machine

***EVALUATION OF CUSTOMER REVIEWS
QUALITY BASED ON STRUCTURAL, METADATA,
AND READABILITY CHARACTERISTICS***

Nama Mahasiswa : Rani Oktavia
NRP : 5213100122
Jurusan : Sistem Informasi FTif – ITS
**Pembimbing I : Rully Agus Hendrawan, S.Kom.,
M.Eng. Sc**

ABSTRACT

Customer online review has the important role in product decisions process. Nowadays, there are a lot of customer online reviews available in online marketplace so that customer can read those reviews to better understand about product or service that will they purchase. Customer online reviews become important additional information besides product description, expert reviews, and recommendation from systems. However, there are a lot of reviews that become questions are these reviews have good quality and useful for other customers. Based on a survey, 87% customers will read at least 10 reviews before deciding to buy a product. This thing can become problematic because there's a possibility for new customer online reviews that have good quality but hasn't popular yet so this reviews might be miss read by other customers.

Therefore, it is necessary to evaluate the quality of content review based on three characteristics which are structural, metadata, and readability using weighted sum that can evaluate some alternatives based on certain criteria. The quality of content review will also be evaluated using Support Vector Machine. This method is used because the data has a lot of features including structural category, metadata, and readability so this method can be used for classifying the quality of online reviews. The results show that the calculation

of the weighted sum, obtained the highest quality review value of 0.736 from scale 1 and the lowest quality review value of -0.104. The category that most affects the quality rating of the review is the value of automated readability index readability, whereas the usability category does not significantly affect the assessment. While based on the classification results using the support vector machine to predict the usefulness of a review, obtained the highest accuracy value using polynomial kernel with a value of 94.4773%.

Keywords : *Customer Online Review, Quality of Reviews, Weighted Sum, Support Vector Machine*

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa, karena dengan limpahan rahmat-Nya penulis dapat menyelesaikan laporan tugas akhir dengan judul : **“PENILAIAN KUALITAS ULASAN PELANGGAN BERDASARKAN KARAKTERISTIK STRUKTURAL, METADATA, DAN KETERBACAAN”** yang merupakan salah satu karya penulis untuk Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Tugas akhir ini tidak akan pernah terwujud tanpa bantuan dan dukungan dari berbagai pihak yang sudah membantu penulis dalam menyelesaikan tugas akhir ini. Pada kesempatan ini penulis ingin mengucapkan terimakasih sebesar-besarnya kepada pihak di bawah ini, yaitu:

1. Orang tua dan adik penulis, Gusli Sarman, S.H., M.Kn., Z. Dearty Z., S.H., M.Kn, dan Cynthia Bella Permatasari yang telah memberikan dukungan, motivasi, kasih sayang, dan doa sehingga penulis dapat menyelesaikan pendidikan S1 dengan baik.
2. Pak Rully Agus Hendrawan, S.Kom., M.Eng, selaku dosen pembimbing dan kepala Lab Sistem Enterprise yang telah membimbing dengan baik sehingga penulis dapat menyelesaikan tugas akhir ini.
3. Bu Erma Suryani, ST., MT., Ph.D., Pak Arif Wibisono, S.Kom., M.Sc., dan Bu Amna Shifia Nisafani, S.Kom., M.Sc., sebagai dosen penguji yang telah memberikan evaluasi ke penulis demi terselesaikannya tugas akhir ini dengan baik.
4. Pak Edwin Riksakomara, S.Kom, M.T., selaku dosen wali yang telah membimbing dan mengarahkan penulis selama masa perkuliahan.
5. Muchammad Fahmi Zamroni, selaku teman dekat penulis yang telah memberikan motivasi, semangat, bantuan, dan

telah menemani penulis selama masa perkuliahan dan pengerjaan tugas akhir.

6. Pri Rezki, Hanum Fitriani, dan Arbintoro Mas sebagai sahabat penulis yang telah membantu penulis dalam proses penyelesaian tugas akhir.
7. Ashma Hanifah, Ari Agustina, Kevin Setiawan, Robbigh Faubendri, Ervi Ritya, dan Safrina Kharisma sebagai sahabat penulis yang saling mendukung dan memberikan semangat selama menjalani masa perkuliahan.
8. Teman-teman dan sahabat penulis selama masa perkuliahan, Marina Safitri, Shania Olivia, Siti Oryza, Almira Fiana Dhara, Bagas Ananta, Chandra Surya serta rekan-rekan mahasiswa Jurusan Sistem Informasi BELTRANIS lainnya dan anggota Lab Sistem Enterprise atas semua bantuan ketika masa perkuliahan.
9. Serta semua pihak yang telah membantu dalam pengerjaan Tugas Akhir ini yang belum mampu penulis sebutkan diatas.

Penulis menyadari bahwa Tugas Akhir ini masih belum sempurna dan memiliki banyak kekurangan di dalamnya. Dan oleh karena itu penulis meminta maaf atas segala kesalahan yang dibuat penulis dalam buku tugas akhir ini. Penulis membuka pintu selebar lebarnya bagi pihak yang ingin memberikan kritik dan saran, dan penelitian selanjutnya yang ingin menyempurnakan karya dari tugas akhir ini. Semoga buku tugas akhir ini bermanfaat bagi seluruh pembaca.

Surabaya, Juli 2017

Penulis

DAFTAR ISI

Lembar Pengesahan	ii
Lembar Persetujuan.....	iv
ABSTRAK.....	v
ABSTRACT.....	vii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xv
DAFTAR KODE.....	xvi
DAFTAR TABEL.....	xvii
BAB I PENDAHULUAN.....	1
1. 1. Latar Belakang.....	1
1. 2. Rumusan Masalah	5
1. 3. Batasan Masalah.....	5
1. 4. Tujuan.....	5
1. 5. Manfaat.....	5
1. 6. Relevansi	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Penelitian Sebelumnya	7
2.2. Dasar Teori.....	10
2. 2. 1. E-commerce	10
2. 2. 2. Pasar Online	11
2. 2. 3. Amazon	11
2. 2. 4. Web Scraping.....	12

2. 2. 5. Star Rating dan Review Extremity	13
2. 2. 6. Weighted Sum	14
2. 2. 7. Support Vector Machine.....	15
BAB III METODOLOGI	19
3.1. Studi Literatur	19
3.2. Pengumpulan Data.....	19
3.3. Preproses Data	21
3.4. Ekstraksi Fitur.....	21
3.4. 1.Flesch-Kincaid Reading Ease.....	21
3.4. 2.Automated Readability Index (ARI)	22
3.5. Penilaian dan Pengurutan Kualitas Ulasan Menggunakan Weighted Sum	22
3.6. Training.....	23
3.7. Testing	24
3.8. Analisis Penilaian Kualitas Ulasan	26
3.9. Penyusunan Buku Tugas Akhir	26
BAB IV PERANCANGAN	27
4.1. Pengambilan Data.....	27
4.2. Pemilihan Atribut.....	32
4.3. Perancangan Model	34
4. 3. 1. Perancangan Preproses Data.....	34
4. 3. 2. Perancangan Ekstraksi Fitur	35
4. 3. 3. Perancangan Pemrosesan Data	37
BAB V IMPLEMENTASI	41
5.1. Perangkat Penelitian	41
5.2. Praproses Data	41

5.3. Ekstraksi Fitur	42
5.3.1. Ekstraksi Fitur Struktural	42
5.3.2. Ekstraksi Fitur Metadata	46
5.3.3. Ekstraksi Fitur Keterbacaan	47
5.4. Pemrosesan Data	48
5.4.1. Penilaian dengan Weighted Sum	48
5.4.2. Klasifikasi dengan Support Vector Machineb	52
BAB VI HASIL DAN PEMBAHASAN	59
6.1. Data Ulasan	59
6.2. Ekstraksi Data.....	60
6.2.1. Data Ekstraksi Fitur Struktural	61
6.2.2. Data Ekstraksi Fitur Metadata.....	61
6.2.3. Data Ekstraksi Fitur Keterbacaan	62
6.3. Hasil Weighted Sum.....	63
6.3.1. Penyamaan Skala	63
6.3.2. Penentuan Bobot	63
6.3.3. Penilaian dan Pengurutan.....	64
6.4. Hasil Support Vector Machine	69
6.4.1. Data Normalisasi.....	69
6.4.2. Data Konversi	70
6.4.3. Training.....	71
6.4.4. Testing	78
6.5. Analisis Hasil	79
6.5.1. Hasil Penilaian dan Pengurutan Weighted Sum.....	79
6.5.2. Hasil Klasifikasi Support Vector Machine	87

BAB VII KESIMPULAN DAN SARAN	91
7.1. Kesimpulan.....	91
7.2. Saran	92
DAFTAR PUSTAKA.....	95
BIODATA PENULIS.....	99
LAMPIRAN A	100
LAMPIRAN B.....	104
LAMPIRAN C.....	117

DAFTAR GAMBAR

Gambar 2. 1. Hyperplane pada SVM sumber: stanford.edu...	16
Gambar 3. 1 Metodologi Penelitian	21
Gambar 3. 2 Syntax Permodelan pada LibSVM	24
Gambar 4. 1 Alur Pengambilan Data	27
Gambar 4. 2 Alur Ekstraksi Fitur	35
Gambar 6. 1 Percobaan Training Pertama	71
Gambar 6. 2 Model Pertama	72
Gambar 6. 3 Percobaan Training Kedua	74
Gambar 6. 4 Model Kedua	74
Gambar 6. 5 Percobaan Training Ketiga.....	75
Gambar 6. 6 Model Ketiga.....	75
Gambar 6. 7 Percobaan Training Keempat	76
Gambar 6. 8 Model Keempat	77
Gambar 6. 9 Percobaan Testing Pertama	78
Gambar 6. 10 Percobaan Testing Kedua.....	78
Gambar 6. 11 Percobaan Testing Ketiga.....	79
Gambar 6. 12 Percobaan Testing Keempat.....	79

DAFTAR KODE

Kode 5. 1 Method countWords.....	42
Kode 5. 2 Method countSentences	43
Kode 5. 3 Method countSyllables.....	43
Kode 5. 4 Method countit dan silente.....	45
Kode 5. 5 Ekstraksi Fitur Struktural	46
Kode 5. 6 Ekstraksi Fitur Metadata	47
Kode 5. 7 Ekstraksi Fitur Keterbacaan.....	48
Kode 5. 8 Konversi Data	54
Kode 5. 9 Pembagian Dataset.....	56
Kode 5. 10 Training SVM.....	56
Kode 5. 11 Testing SVM.....	58

DAFTAR TABEL

Tabel 3. 1 Confusion Matrix	25
Tabel 4. 1 Daftar Atribut Data Ulasan	28
Tabel 4. 2 Sampel Data Ulasan	29
Tabel 4. 3 Sampel Atribut Terpilih	32
Tabel 4. 4 Kriteria dan Bobot Penilaian	38
Tabel 5. 1 Bobot Weighted Sum	51
Tabel 5. 2 Pilihan SVM dan Kernel	57
Tabel 6. 1 Data Ulasan	59
Tabel 6. 2 Hasil Ekstraksi Fitur Struktural.....	61
Tabel 6. 3 Hasil Ekstraksi Fitur Metadata.....	62
Tabel 6. 4 Hasil Ekstraksi Fitur Keterbacaan.....	62
Tabel 6. 5 Hasil Penentuan Bobot.....	64
Tabel 6. 6 Data Sebelum Penyamaan Skala.....	65
Tabel 6. 7 Data Sesudah Penyamaan Skala.....	66
Tabel 6. 8 Hasil Penilaian Menggunakan Weighted Sum.....	67
Tabel 6. 9 Hasil Pengurutan Data.....	68
Tabel 6. 10 Data Normalisasi.....	69
Tabel 6. 11 Data Konversi.....	70
Tabel 6. 12 Tujuh Ulasan dengan Penilaian Tertinggi	79
Tabel 6. 13 Tujuh Ulasan dengan Penilaian Terendah.....	82
Tabel 6. 14 Hasil Klasifikasi SVM	87
Tabel A. 1 Hasil Penilaian dan Perhitungan Weighted Sum	102
Tabel B. 1 Data Review Setelah Dinilai dan Diurutkan.....	104
Tabel C. 1 Sparse Format Data	117

BAB I

PENDAHULUAN

Pada bab pendahuluan diuraikan proses identifikasi masalah pada penelitian ini yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan tugas akhir, dan manfaat tugas akhir.

1. 1. Latar Belakang

Sekarang ini, penggunaan situs online sebagai sumber pencarian informasi sudah semakin meluas ke berbagai aspek kehidupan manusia. Tidak hanya pencarian informasi seputar berita, trend global, atau keilmuan tetapi juga digunakan untuk pencarian informasi yang bisa mendukung transaksi di pasar online. Hal ini dikarenakan salah satu faktor yang mempengaruhi keputusan pembelian sebuah produk di pasar online adalah ketersediaan informasi yang cukup untuk menjadi bahan pertimbangan pembelian bagi pelanggan. Oleh karena itu, semakin banyak ulasan online yang dapat dibaca pelanggan sehingga mereka dapat lebih memahami dengan baik mengenai produk atau jasa yang akan dibeli [1].

Berdasarkan hasil survei yang dilakukan oleh BrightLocal mengenai “Local Consumer Review Survey” di tahun 2016, sebanyak 84% responden mempercayai ulasan online sama halnya dengan rekomendasi personal [2]. Penelitian sebelumnya juga telah meneliti peran dari ulasan online pelanggan terhadap suatu produk dilihat dari karakteristik pelanggan yang memberikan ulasan [3]. Selain itu, berdasarkan penelitian yang telah dilakukan terbukti bahwa ulasan pelanggan dapat memberikan dampak positif untuk penjualan [4]. Bahkan, penilaian ulasan yang sangat positif dapat mempengaruhi pertumbuhan penjualan produk dan

kualitas dari ulasan yang diukur berdasarkan penilaian kegunaan juga secara positif mempengaruhi penjualan [5].

Ulasan pelanggan secara umum dapat didefinisikan sebagai evaluasi pihak ketiga oleh pelanggan mengenai sebuah produk atau jasa yang ditampilkan di situs web dan bersebelahan dengan deskripsi produk guna untuk meningkatkan persepsi pelanggan [1]. Ulasan pelanggan merupakan opini dari pelanggan mengenai pengalaman mereka terhadap produk atau jasa yang dibeli. Ulasan pelanggan merupakan sumber yang dipercaya tinggi bagi pelanggan sebelum membeli sebuah produk [6]. Pelanggan biasanya mencari ulasan produk secara online mengenai informasi sebuah produk dan mengevaluasi alternatif produk lainnya. Sumber yang biasanya digunakan untuk mencari ulasan pelanggan adalah situs pasar online yang menyediakan fitur bagi pelanggan untuk dapat menuliskan ulasan produk dan memberikan nilai dengan skala 1-5 terhadap produk tersebut atau biasa disebut dengan istilah star rating. Dengan hal ini, mereka mendapatkan akses yang cukup banyak mengenai ulasan sebuah produk dari pelanggan lainnya melalui situs pasar online.

Kehadiran ulasan pelanggan di situs pasar online telah menunjukkan peningkatan persepsi pelanggan terhadap kegunaan dan kehadiran sosial dari situs tersebut [7]. Ulasan pelanggan mempunyai potensi tersendiri untuk menarik pelanggan ke situs tersebut, meningkatkan waktu yang dihabiskan di situs tersebut, dan secara tidak langsung membentuk komunitas antara sesama pelanggan yang sering membeli [8]. Namun, karena semakin banyaknya ulasan pelanggan yang tersedia di situs pasar online, munculah pertanyaan apakah setiap ulasan tersebut memang berarti terhadap sebuah produk dan berguna bagi pelanggan lain.

Salah satu pasar online raksasa, Amazon.com, memiliki fitur untuk mengevaluasi setiap ulasan pelanggan yang ada dengan pertanyaan “Was this review helpful?” di akhir ulasan. Setiap pengguna yang membaca ulasan tersebut dapat memberikan penilaian apakah ulasan tersebut berguna atau tidak bagi mereka. Amazon akan menempatkan ulasan dengan jumlah vote paling banyak di paling atas sehingga pengguna lain tahu ulasan mana yang paling berguna dari semuanya. Namun, penilaian seperti ini masih belum cukup untuk mengetahui apakah ulasan tersebut benar-benar berguna.

Hal ini dikarenakan ulasan yang sudah memiliki vote cukup banyak cenderung akan mendapatkan semakin banyak vote karena letak ulasan tersebut berada di barisan teratas sehingga dapat dengan mudah diakses pelanggan. Berbeda dengan ulasan baru yang masih memiliki vote sedikit sehingga ulasan tersebut akan terletak di barisan terbawah dan pelanggan lain tidak sampai membaca ulasan tersebut. Padahal belum tentu ulasan baru tersebut tidak lebih baik dari ulasan teratas yang memiliki banyak vote. Berdasarkan hasil survei yang dilakukan BrightLocal, sebanyak 87% pembeli akan membaca paling banyak 10 ulasan untuk menentukan keputusan pembelian [2]. Jadi, meskipun ulasan yang ditulis pelanggan itu sebenarnya berguna, tetapi karena ulasan tersebut tidak termasuk dalam barisan 10 ulasan teratas jadi ulasan tersebut tidak terbaca oleh pelanggan.

Selain itu, tidak jarang ditemukan pelanggan yang memberikan star rating paling rendah yaitu satu bintang namun tidak menuliskan ulasan dengan rinci mengenai penilaian itu. Sebaliknya, ada juga pelanggan yang memberikan star rating menengah yaitu tiga bintang namun ulasan yang ditulis sangat rinci dan dalam dibandingkan

pelanggan yang memberikan star rating tertinggi yaitu lima bintang namun tidak menulis ulasan dengan rinci. Ulasan dengan star rating paling rendah atau paling tinggi dikenal dengan istilah *review extremity* yang artinya pelanggan tersebut memberikan nilai yang sangat ekstrem terhadap suatu produk. Hal ini menjadi pertanyaan mengenai kualitas dari ulasan itu sendiri berdasarkan star rating dan ulasan yang ditulis pelanggan.

Ulasan yang ditulis pelanggan seharusnya bertindak sebagai pembenaran dari penilaian yang diberikan. Hal ini memungkinkan pengguna lainnya bisa menilai apakah ulasan tersebut adil atau tidak dan bisa menentukan apakah ulasan itu bisa dijadikan bahan pertimbangan pembelian produk tersebut. Maka dari itu, dibutuhkan analisis lebih dalam mengenai ulasan secara kontekstual dan bagaimana hubungannya dengan nilai kegunaan yang telah diberikan pengguna lain.

Oleh karena itu, peneliti akan meneliti kualitas konten ulasan yang ditulis pelanggan berdasarkan tiga karakteristik yaitu struktural, meta-data, dan keterbacaan. Kategori struktural meliputi panjang ulasan, jumlah kalimat, dan jumlah kata. Kategori meta-data meliputi star rating dan *review extremity*. Kategori keterbacaan meliputi keterbacaan ulasan menggunakan uji keterbacaan *Flesch-Kincaid Reading Ease* dan *Automated Readability Index*. Selanjutnya, ulasan akan diurutkan berdasarkan penilaian ulasan dari yang paling berkualitas sampai yang tidak berkualitas. Penilaian dilakukan menggunakan *weighted sum* yang dapat mengevaluasi beberapa alternatif berdasarkan kriteria tertentu dan metode *Support Vector Machine* yang digunakan untuk mengklasifikasi kualitas ulasan pelanggan.

1. 2. Rumusan Masalah

Pemilik/pengembang e-commerce dan konsumen membutuhkan cara yang mudah untuk membedakan kualitas ulasan pelanggan. Saat ini, ulasan pelanggan hanya dinilai berdasarkan jumlah vote kegunaan yang diberikan pelanggan lain. Sehingga vote yang banyak belum tentu menunjukkan kualitas ulasan. Akibatnya mungkin ada ulasan yang berkualitas namun tidak populer karena masih baru ditulis dan luput dari perhatian calon konsumen.

1. 3. Batasan Masalah

Batasan dalam pengerjaan tugas akhir ini adalah:

1. Sumber data yang digunakan adalah hasil dari *web scraping* situs belanja *online* internasional yaitu Amazon.
2. Masing-masing data ulasan setidaknya telah dinilai kegunaannya oleh pengguna lain.

1. 4. Tujuan

Tujuan pengerjaan tugas akhir ini adalah untuk menganalisis penilaian kualitas konten ulasan yang ditulis pelanggan berdasarkan tiga karakteristik yaitu struktural, meta-data, dan keterbacaan. Ulasan tersebut dinilai kualitasnya dan diurutkan dari ulasan yang paling berkualitas sampai yang tidak berkualitas.

1. 5. Manfaat

Manfaat dari pengerjaan tugas akhir ini adalah untuk dapat membedakan ulasan pelanggan yang berkualitas dan tidak berkualitas sehingga dapat memperkuat keputusan pembelian bagi pelanggan lain berdasarkan ulasan yang dibaca. Selain itu, ulasan juga diurutkan dari yang paling berkualitas sampai yang tidak berkualitas, sehingga ulasan yang berkualitas akan terbaca oleh pelanggan lain. Bagi pemilik/pengembang e-

commerce dapat menjadikan acuan untuk membuat pedoman bagi pelanggan untuk membuat ulasan yang berkualitas.

1. 6. Relevansi

Tugas akhir ini merupakan implementasi dari disiplin ilmu yang diajarkan selama perkuliahan di jurusan sistem informasi ITS. Tugas akhir ini mendukung penelitian dosen pembimbing dengan judul “*Crowd and Cloud Based Light-Weight Business Performance Management Tools*” dengan modul yang terdiri dari:

1. Katalog produk di *e-commerce* yang berpusat pada pengguna
2. *Lightweight Business Performance Management implementation*
3. Meningkatkan kegunaan pada *tools Business Performance Management*

Topik tugas akhir ini berkaitan dengan katalog produk di *e-commerce* yang berpusat pada pengguna. Mata kuliah yang berkaitan dengan topik tugas akhir ini adalah *e-business* yang membahas penggunaan data ulasan pelanggan di *e-commerce* dan sistem cerdas yang membahas penggunaan metode *support vector machine* untuk klasifikasi.

BAB II

TINJAUAN PUSTAKA

Pada bab tinjauan pustaka diuraikan studi sebelumnya dan dasar teori yang berkaitan dengan permasalahan tugas akhir.

2.1. Penelitian Sebelumnya

Telah banyak penelitian yang dilakukan dengan topik sejenis penelitian ini yaitu mengenai kegunaan ulasan pelanggan. Salah satunya adalah penelitian yang berjudul “*Predicting the helpfulness of online consumer reviews*”. Penelitian ini membuat model berdasarkan *machine learning* yang dapat memprediksikan kegunaan dari ulasan pelanggan menggunakan beberapa fitur tekstual seperti polaritas, subjektivitas, entropi, dan kemudahan membaca. Model yang dibuat akan secara otomatis menetapkan nilai kegunaan suatu ulasan yang baru saja diberikan sehingga ulasan tersebut mendapatkan kesempatan untuk dapat dibaca oleh pelanggan lain. Hasil dari penelitian ini dapat membantu pembeli untuk menulis ulasan yang lebih baik dan membantu pembeli lain untuk membuat keputusan pembelian [1]. Keterkaitan penelitian ini dengan topik tugas akhir adalah pembuatan model berdasarkan *machine learning* yang digunakan untuk memprediksi kegunaan dari ulasan pelanggan.

Selain itu terdapat pula penelitian serupa yang berjudul “*Predicting Helpfulness Ratings of Amazon Product Review*”. Penelitian ini mengembangkan *classifier* dari kumpulan data sebanyak 5000 ulasan produk secara acak dari 100.000 ulasan lebih untuk memprediksikan apakah ulasan yang ditulis berguna atau tidak berguna. Penelitian ini menggunakan metode Naive Bayes dan *Support Vector (SVM)* dengan berbagai macam kernel. Hasilnya dengan menggunakan SVM *polynomial* uji keakuratannya mencapai 76.6%. Sedangkan

metode Naive Bayes hanya berguna apabila ukuran sampel yang digunakan sangat kecil sehingga uji keakuratannya tidak lebih baik dibandingkan metode SVM *polynomial* [2]. Keterkaitan penelitian ini dengan topik tugas akhir adalah penggunaan metode SVM untuk memprediksikan kegunaan dari ulasan pelanggan.

Penelitian lain yang berjudul “*Ranking Online Customer Reviews with the SVR Model*” juga menggunakan metode turunan dari SVM yaitu *Support Vector Regression* untuk membuat sistem peringkat ulasan pelanggan. Data yang digunakan diambil dari Amazon.cn sebanyak 3730 ulasan sebanyak 8 kategori produk yaitu buku, kamera digital, PC tablet, tas, film, sepatu pria, mainan, dan telepon genggam. Ulasan yang diambil adalah ulasan yang telah memiliki penilaian berguna atau tidak berguna karena hasil penilaian bisa menjadi bias terhadap waktu penilaian dan jumlah pemilih. Penelitian ini menggunakan empat tipe indeks evaluasi dan membandingkan hasil regresi terhadap masing-masing indeks [3]. Keterkaitan penelitian ini dengan topik tugas akhir adalah penggunaan metode SVR untuk mengurutkan ulasan pelanggan berdasarkan penilaiannya.

Selain itu terdapat penelitian lain yang berjudul, “*Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content*”. Penelitian ini meneliti hubungan antara kegunaan ulasan *online*, penilaian, dan karakteristik kualitatif dari ulasan yang diukur menggunakan uji keterbacaan. Model teoritis yang digunakan berdasarkan tiga elemen yaitu kesesuaian, dapat dipahami, dan keekspresifan. Model ini divalidasi menggunakan empat pengukuran keterbacaan dasar pada *dataset* yang berisi 37.221 ulasan dari Amazon UK. Penelitian ini dapat membuktikan

saat suatu ulasan dianggap berguna bagi pelanggan yang telah membeli produk atau akan membelinya maka ulasan tersebut mempunyai kualitas karakteristik yang bagus. Dengan menggunakan rumus keterbacaan, penelitian ini dapat menganalisis ulasan dan menghasilkan hasil yang berhubungan dengan kegunaan ulasan tertentu yang telah dijabarkan di penelitian ini [4]. Keterkaitan penelitian ini dengan topik tugas akhir adalah penelitian berdasarkan karakteristik konteks yang diukur menggunakan uji keterbacaan.

Selain itu, terdapat juga penelitian dengan topik sejenis yang berjudul "*What makes a helpful online review? A study of customer reviews on Amazon.com*". Penelitian ini mengembangkan dan menguji model dari kegunaan ulasan pelanggan. Penelitian ini menggunakan 1.587 ulasan dari Amazon dengan enam jenis produk. Ulasan pelanggan tersebut diteliti berdasarkan penilaian ulasan, kedalaman ulasan, dan jenis produk yang mempengaruhi kegunaan dari ulasan. Hasil yang didapatkan adalah jenis produk (*search good* atau *experience good*) mempengaruhi pencarian informasi dan evaluasi yang dilakukan pelanggan. Penilaian dan panjang ulasan mempengaruhi ulasan informasi yang diberikan berdasarkan jenis produk. Ulasan dengan penilaian ekstrem tidak lebih berguna dibandingkan penilaian sedang untuk tipe produk *experience goods* dan tidak berlaku untuk produk *search goods*. Dan kedalaman ulasan mempengaruhi secara positif terhadap kegunaan ulasan untuk jenis *search goods* lebih besar dibandingkan *experience goods*. Penelitian ini dapat membantu pasar *online* untuk memahami peran dari ulasan pelanggan terhadap proses keputusan pembelian suatu produk. Maka dari itu, penelitian ini dapat dijadikan acuan bagi pasar *online* untuk mengembangkan pedoman bagaimana

membuat ulasan yang berharga [5]. Keterkaitan penelitian ini dengan topik tugas akhir adalah adanya hubungan antara penilaian ulasan, kedalaman ulasan, dan jenis produk yang mempengaruhi kegunaan ulasan.

Terdapat juga penelitian yang meneliti pentingnya volume ulasan dan peringkat terhadap peningkatan kinerja penjualan serta meneliti lebih lanjut peran dari kategori produk, pertanyaan yang terjawab, diskon, dan kegunaan ulasan. Penelitian yang berjudul “*Predicting online e-marketplace sales performances: A big data approach*” menggunakan data dari Amazon sebanyak 2939 ulasan menggunakan arsitektur *big data*. Hasil yang didapatkan adalah volume ulasan dan peringkat ulasan mempengaruhi secara signifikan penjualan untuk *search product*. Kegunaan ulasan juga secara signifikan mempengaruhi volume ulasan dan penjualan. Namun, pertanyaan yang terjawab dan diskon tidak mempengaruhi secara signifikan terhadap hubungan antara peringkat ulasan dan penjualan. Selain itu, penelitian ini juga menjabarkan arsitektur *big data* dan mengilustrasikan kegunaan teknologi *big data* untuk menguji kerangka kerja teoretis [6]. Keterkaitan penelitian ini dengan topik tugas akhir adalah peran dari kegunaan ulasan untuk meningkatkan penjualan.

2.2. Dasar Teori

Pada bagian ini dijelaskan mengenai teori-teori yang berkaitan dengan topik tugas akhir. Berikut adalah dasar teori yang berkaitan dengan tugas akhir ini:

2.2.1. E-commerce

Electronic commerce (E-commerce) adalah suatu istilah yang mendefinisikan aktivitas transaksi jual beli melalui media elektronik seperti internet. Istilah *e-commerce* pertama kali

dikenal pada tahun 1960 melalui pemanfaatan teknologi *Electronic Data Interchange* (EDI). Seiring dengan peningkatan penggunaan internet pada tahun 1990, istilah *e-commerce* semakin banyak digunakan. Amazon situs belanja buku *online* dan Ebay situs jual beli *online*, pertama kali beroperasi pada tahun 1995 [7].

Seiring dengan perkembangan penggunaan teknologi internet, sekarang ini semakin banyak situs *e-commerce* yang berada di pasaran baik itu yang diakses melalui *website* ataupun aplikasi *mobile*. Terdapat beberapa kategori *e-commerce* yaitu *business to business*, *business to consumer*, *consumer to business*, dan *consumer to consumer* [7].

2. 2. 2. Pasar Online

Pasar *online* atau *online marketplace* adalah tempat dimana pembeli dapat menemukan berbagai macam barang yang dijual oleh beberapa penjual dalam satu *platform*. Pemilik pasar *online* bertanggung jawab terhadap kelangsungan bisnis jual beli di *platform* miliknya, dalam hal ini adalah *website*. Pemilik juga memantau transaksi uang yang terjadi antara penjual dengan pembeli, sedangkan penjual akan memastikan langsung pengiriman barang yang dibeli ke pembeli. Hal ini membuat pembeli merasa lebih aman dengan transaksi yang dimediasi oleh *website* [8]. Pasar *online* merupakan tempat berlangsungnya *e-commerce* yaitu aktivitas transaksi jual beli melalui media elektronik. Terdapat setidaknya lima pasar *online* yang banyak dikenal masyarakat internasional yaitu Etsy, Amazon, Bonanza, Ruby Lane, dan eBay [9].

2. 2. 3. Amazon

Amazon adalah situs pasar *online* yang didirikan oleh Jeff Bezos pada tahun 1995 di Seattle, Washington. Pada awalnya,

Amazon hanya menjual buku saja namun Jeff Bezos menginginkan Amazon menjadi toko yang menjual apa saja. Dan akhirnya hingga saat ini, Amazon telah menjual berbagai macam produk mulai dari DVD, *audiobook*, *software*, *video games electronic*, pakaian, perabot, makanan, mainan, dan perhiasan. Amazon memiliki situs pasar *online* terpisah untuk beberapa negara yaitu Amerika Serikat, Inggris, Irlandia, Prancis, Kanada, Jerman, Italia, Spanyol, Belanda, Australia, Brasil, Jepang, India, Cina, dan Mexico. Amazon juga menyediakan layanan pengiriman internasional untuk negara tertentu dan produk tertentu [10].

Amazon merupakan pasar *online* terbesar di dunia berdasarkan total penjualan dan kapitalisasi pasar. Amazon menyediakan layanan jual beli produk melalui situsnya dimana pelanggan bisa membeli produk yang diinginkan. Selanjutnya produk tersebut akan dikirimkan oleh Amazon. Setelah sampai ke pelanggan, mereka bisa menuliskan ulasan mengenai produk yang telah dibeli. Pelanggan dapat memberikan penilaian terhadap produk dengan skala 1-5 bintang. Pelanggan lainnya dapat memberikan komentar atau memilih ulasan apabila ulasan tersebut dianggap berguna bagi pelanggan itu. Pada tahun 2010, Amazon disebut sebagai sumber tunggal terbesar ulasan pelanggan *online* [11].

2. 2. 4. Web Scraping

Web scraping adalah suatu teknik yang digunakan untuk mendapatkan informasi dari sebuah situs secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari *web scraping* adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya kedalam format berbeda. *Web scraping* berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi dari konten web yang tidak terstruktur biasanya

dalam format HTML, menjadi data terstruktur yang dapat disimpan dan dianalisis dalam *database* lokal pusat atau *spreadsheet* [12].

Manfaat dari *web scraping* yaitu membuat informasi yang dikeruk atau diambil lebih terfokus sehingga dapat memudahkan dalam melakukan pencarian sesuatu.

Web scraping memiliki sejumlah langkah meliputi [13] :

1. *Create Scraping Template*: Pembuat program mempelajari dokumen HTML dari situs yang akan diambil informasinya untuk *tag* HTML yang mengapit informasi yang akan diambil.
2. *Explore Site Navigation*: Pembuat program mempelajari teknik navigasi pada situs yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper* yang akan dibuat.
3. *Automate Navigation and Extraction*: Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, aplikasi *web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari situs yang ditentukan.

Extracted Data and Package History: Informasi yang didapat dari langkah tiga disimpan dalam tabel *database*.

2. 2. 5. Star Rating dan Review Extremity

Star rating adalah penilaian yang diberikan pelanggan terhadap suatu produk dari skala 1-5 bintang. Situs pasar *online* biasanya menggunakan *star rating* atau penilaian dengan bintang untuk mengevaluasi suatu produk. Sedangkan, *review extremity* adalah ulasan yang mempunyai *star rating* dengan nilai paling rendah (satu bintang) atau nilai paling tinggi (lima bintang). Ulasan dengan *star rating* paling rendah

mengindikasikan bahwa ulasan tersebut menggambarkan pandangan yang sangat negatif terhadap produk. Sebaliknya, ulasan dengan *star rating* paling tinggi mengindikasikan bahwa ulasan tersebut menggambarkan pandangan yang sangat positif terhadap produk. Nilai dari *review extremity* didapatkan dari perbedaan *star rating* yang diberikan pelanggan dengan rata-rata *star rating* suatu produk. *Star rating* dan *review extremity* merupakan variabel yang berpengaruh terhadap kegunaan ulasan pelanggan [5].

2. 2. 6. *Weighted Sum*

Weighted sum adalah metode untuk mengevaluasi alternatif berdasarkan kriteria dengan memberikan bobot pada kriteria tersebut. Metode ini dapat digunakan jika data yang digunakan memiliki satuan unit yang sama. Metode ini merupakan pendekatan yang paling mudah dan banyak digunakan di penelitian. Pembobotan dilakukan berdasarkan pentingnya kriteria tersebut untuk menentukan alternatif [14]. Model matematis dari metode ini adalah sebagai berikut:

$$A_i^{WSM-score} = \sum_{j=1}^n w_j a_{i,j}, \text{ untuk } i = 1, 2, 3, \dots, m \quad (2.1)$$

Misalnya terdapat masalah analisis keputusan multikriteria yang ditentukan oleh alternatif m dan kriteria keputusan n . Asumsikan semua kriteria merupakan kriteria yang memberikan keuntungan sehingga semakin tinggi nilai yang didapatkan maka hasilnya semakin baik. Selanjutnya w_j melambangkan bobot dari kriteria C_j dan a_{ij} adalah penilaian performa alternatif A_i saat dievaluasi dengan kriteria C_j [14]. Pada tugas akhir ini diasumsikan semakin besar nilai A_i maka semakin bagus kualitas dari ulasan pelanggan.

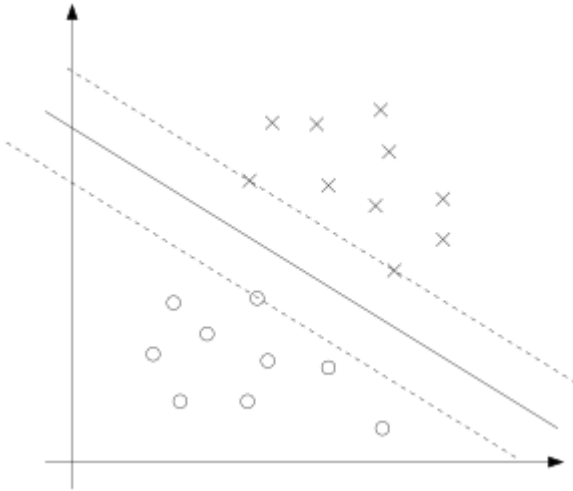
Terdapat beberapa metode untuk menentukan bobot kriteria, diantaranya adalah *equal weight method*, *rank sum weight method*, *rank exponent weight method*, dan *rank-order centroid method*. Pada penelitian sebelumnya yang membandingkan keakuratan masing-masing metode pembobotan dalam menghitung keputusan multikriteria, metode *rank-order centroid* (ROC) adalah metode yang paling baik dari segi keakuratan. Metode ini mempunyai pendekatan untuk menghasilkan estimasi bobot yang meminimalkan kesalahan dari masing-masing bobot dengan mengidentifikasi titik tengah dari semua kemungkinan bobot dengan memperhatikan urutan peringkat kriteria [15]. Rumus untuk menghitung bobot menggunakan metode ini adalah:

$$W_j (ROC) = \frac{1}{n} \sum_{k=j}^n \frac{1}{r_k} \quad (2.2)$$

dimana W adalah bobot, n adalah jumlah kriteria, dan r adalah urutan. Metode ini merefleksikan titik tengah dari simpleks yang didefinisikan oleh urutan kriteria. Semakin banyaknya kriteria, maka kesalahan dari urutan kriteria semakin berkurang.

2. 2. 7. Support Vector Machine

Support Vector Machine (SVM) adalah suatu algoritma yang mencoba menemukan pemisah linear (*hyperplane*) antara titik data dari dua kelas dalam ruang multidimensi. SVM merupakan algoritma klasifikasi *supervised machine learning* yang dapat digunakan untuk klasifikasi. Berbeda dengan algoritma klasifikasi lainnya, SVM biasanya menghasilkan hasil lebih baik pada dimensi yang tinggi dimana fitur angkanya cukup besar dan kerapatan datanya jauh [3]. Berikut adalah contoh penerapan SVM:



Gambar 2. 1. Hyperplane pada SVM
sumber: stanford.edu

Pada gambar diatas terdapat dua kelas data yang dilambangkan dengan simbol silang dan lingkaran. Kedua kelas tersebut dipisahkan oleh *hyperplane* yang dapat diletakkan dimana saja sebagai pemisah. *Hyperplane* yang dipilih adalah yang dapat memisahkan kelas dengan margin terbesar. Margin adalah jarak antara titik data yang terdekat dengan *hyperplane* untuk kedua kelas. Dengan memilih *hyperplane* dengan margin terbesar, maka dapat meningkatkan probabilitas pengelompokan secara benar untuk data *testing*.

SVM menggunakan formula $g(x) = W^T \phi(x) + b$ sebagai *hyperplane* pemisah linear, dimana w sebagai bobot vektor, b sebagai bias, dan ϕ sebagai kumpulan dimensi tinggi non-linear, w dan b didefinisikan sebagai data *training* yang dapat mengoptimalkan formula berikut [3]:

$$\min \frac{1}{2} W^t W + C \sum_{i=1}^N \varepsilon_i$$

$$s. t. \begin{cases} y_i g(x_i) \geq 1 - \varepsilon_i \\ \varepsilon_i \geq 0, i = 1 \dots N \end{cases} \quad (2.3)$$

dimana ε_i adalah variabel *slack*, C adalah koefisien, untuk semua sampel latihan (x_i, y_i) .

Dalam penerapannya, SVM menggunakan kernel untuk mendapatkan model yang dapat digunakan untuk memprediksi data. Beberapa kernel yang paling umum digunakan antaranya linear, polynomial, *radial basis function* (RBF), dan sigmoid. Perbedaan kernel terletak pada parameter yang digunakan untuk mendapatkan model. Adapun persamaan masing-masing kernel ditunjukkan pada formula berikut:

$$K(X_i, X_j) = \begin{cases} X_i \cdot X_j & (Linear) \\ (\gamma X_i \cdot X_j + C)^d & (Polynomial) \\ \exp(-\gamma |X_i \cdot X_j|^2) & (RBF) \\ \tanh(\gamma X_i \cdot X_j + C) & (Sigmoid) \end{cases}$$

Halaman ini sengaja dikosongkan

BAB III METODOLOGI

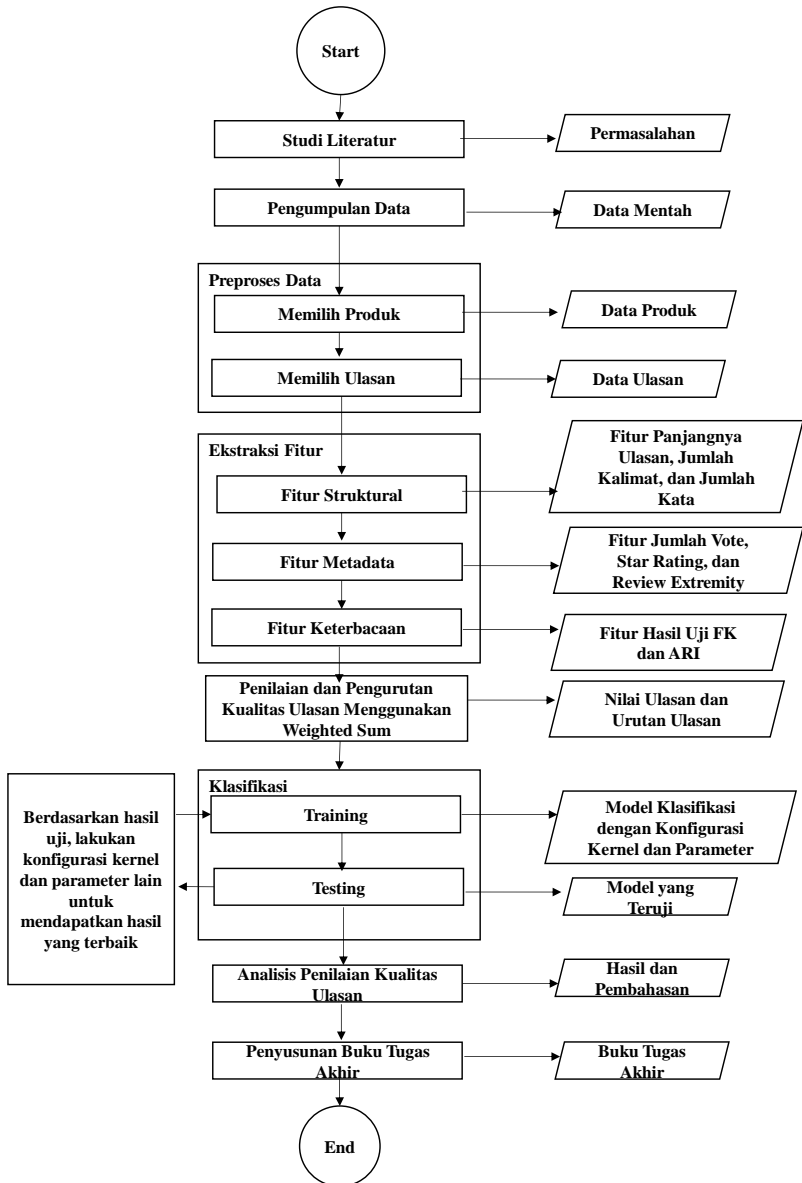
Pada bab ini dijelaskan mengenai gambaran metode dan alur pengerjaan tugas akhir. Penelitian ini menggunakan metode *Support Vector Machine* karena berdasarkan penelitian sebelumnya, metode ini dapat digunakan untuk memprediksi kegunaan ulasan pelanggan. Adapun tahapan dari penelitian ini digambarkan pada Gambar 3.1.

3.1. Studi Literatur

Studi literatur dilakukan untuk mendapatkan informasi, wawasan, dan pengetahuan yang berguna dalam pembuatan tugas akhir. Studi literatur yang dilakukan bertujuan untuk cara melakukan *web crawling* pada website Amazon dan penggunaan metode SVM untuk menganalisis data. Studi literatur bersumber dari berbagai macam media yaitu jurnal internasional, tugas akhir, *website*, serta dokumen lainnya terkait studi kasus. Luaran dari tahapan ini adalah identifikasi permasalahan yang akan diselesaikan.

3.2. Pengumpulan Data

Pengumpulan data merupakan proses untuk mengambil data ulasan yang ada terdapat pada situs Amazon dengan menggunakan alat *crawling*. Data yang diambil adalah nama produk, isi dari ulasan, *star rating*, dan total *votes*. Hasil dari *crawling* akan digunakan untuk analisis selanjutnya. Luaran dari tahapan ini adalah data mentah berupa data hasil *crawling* yang sudah dikelompokkan berdasarkan elemen *tag* pada halaman *website* dan disimpan dalam format csv.



Gambar 3. 1 Metodologi Penelitian

3.3. Preproses Data

Tahapan ini dimulai dari pemilihan produk yang memiliki setidaknya 10 ulasan. Hal ini dikarenakan untuk produk yang masih memiliki sedikit ulasan atau tidak ada ulasan sama sekali tidak perlu dimasukkan ke dalam model karena informasi yang tersedia masih sedikit. Setelah itu dilakukan pemilihan ulasan yang mempunyai setidaknya 1 jumlah *vote* kegunaan. Luaran dari tahapan ini adalah data yang telah disaring berdasarkan produk dan jumlah ulasan tertentu.

3.4. Ekstraksi Fitur

Berdasarkan data mentah yang telah didapatkan, selanjutnya dilakukan ekstraksi fitur agar data tersebut dapat dianalisis. Ekstraksi fitur yang dilakukan dimulai dari fitur struktural yaitu panjang ulasan, jumlah kalimat, dan jumlah kata. Hal ini dilakukan karena kegunaan dari sebuah ulasan dipengaruhi oleh panjangnya ulasan. Misalnya, ulasan pendek tidak akan memiliki informasi yang cukup banyak mengenai produk tersebut. Selanjutnya dilakukan ekstraksi fitur metadata yaitu *star rating* dan *review extremity*. Di setiap ulasan yang ditulis, pelanggan akan memberikan penilaian terhadap produk tersebut dengan penilaian bintang. Sedangkan *review extremity* didapatkan dari perbedaan *star rating* yang diberikan pelanggan dengan rata-rata *star rating* suatu produk. Terakhir, dilakukan ekstraksi fitur keterbacaan menggunakan uji keterbacaan *Flesch-Kincaid Reading Ease* dan *Automated Readability Indeks*. Adapun penjelasan dari uji ini adalah:

3.4. 1. Flesch-Kincaid Reading Ease

Flesch-Kincaid Reading Ease dilakukan untuk mengukur kerumitan dari sebuah teks. Pengujian ini menggunakan pengukuran linguistik inti berbasis suku kata per kata dan kata

per kalimat dari sebuah teks. Pengujian ini dilakukan untuk menentukan berapa lama waktu dibutuhkan untuk seseorang menilai teks. Adapun persamaan dari pengujian ini adalah:

$$FK = 0.39 \times \left(\frac{total_{kata}}{total_{kalimat}} \right) + 11.8 \times \left(\frac{total_{suku\ kata}}{total_{kata}} \right) - 15.59 \dots (3.1)$$

Untuk menghitung nilai *Flesch*, teks dari suatu ulasan akan dirinci menjadi kalimat, lalu menjadi kata, dan akhirnya menjadi suku kata yang dikombinasikan menggunakan konstanta dari rumus diatas. Semakin sedikit angka dari kata per kalimat, maka nilai dari uji keterbacaan *Flesch* akan semakin baik.

3.4. 2. Automated Readability Index (ARI)

Automated Readability Index dilakukan untuk mengukur kemudahan membaca. Pengukuran ini dilakukan untuk mengevaluasi keterbacaan dari teks berbahasa Inggris. Adapun persamaan dari pengujian ini adalah:

$$ARI = 4.71 \times \left(\frac{karakter}{kata} \right) + 0.5 \times \left(\frac{kata}{kalimat} \right) - 21.43 \dots (3.2)$$

Untuk menghitung ARI, pertama-tama hitung total karakter dan total kata suatu ulasan. Langkah ARI hampir sama dengan *Flesch*, namun ARI juga menghitung total karakter atau panjangnya ulasan. ARI dapat memberikan indikasi pengaruh dari panjangnya ulasan terhadap keterbacaan ulasan.

3.5. Penilaian dan Pengurutan Kualitas Ulasan Menggunakan *Weighted Sum*

Setelah didapatkan nilai untuk masing-masing fitur, langkah selanjutnya adalah melakukan penilaian dan pengurutan

kualitas ulasan menggunakan *weighted sum*. Metode ini dilakukan untuk dapat mengevaluasi beberapa alternatif berdasarkan kriteria yang ada dengan pembobotan kriteria. Alternatif yang dimaksud adalah berkualitas dan tidak berkualitas. Sedangkan kriteria yang digunakan mencakup tiga karakteristik yaitu struktural, metadata, dan keterbacaan. Semakin tinggi hasil penilaian menandakan bahwa semakin berkualitas ulasan tersebut. Luaran dari tahapan ini adalah hasil nilai ulasan dan urutan ulasan berdasarkan kualitas.

3.6. Training

Setelah itu data akan dibagi menjadi *data training* dan *data testing*. Lalu tahap selanjutnya adalah *training* dengan menggunakan *library SVM*. Cukup dengan memanggil *file properties*, maka *library* akan menjalankan iterasinya untuk menghasilkan model. Jika iterasi telah selesai maka akan didapatkan model optimal yang akan diuji di tahap *testing*.

SVM memiliki beberapa tipe *kernel* dengan parameter tertentu yang dapat digunakan untuk pembuatan model dari data *training*. Berikut ini adalah *syntax* yang dapat digunakan untuk pembuatan model data SVM:

```

Usage: svm-train [options] training_set_file [model_file]
options:
-s svm_type : set type of SVM (default 0)
  0 -- C-SVC          (multi-class classification)
  1 -- nu-SVC         (multi-class classification)
  2 -- one-class SVM
  3 -- epsilon-SVR    (regression)
  4 -- nu-SVR         (regression)
-t kernel_type : set type of kernel function (default 2)
  0 -- linear: u'*v
  1 -- polynomial: (gamma*u'*v + coef0)^degree
  2 -- radial basis function: exp(-gamma*|u-v|^2)
  3 -- sigmoid: tanh(gamma*u'*v + coef0)
  4 -- precomputed kernel (kernel values in training_set_file)
-d degree : set degree in kernel function (default 3)
-g gamma : set gamma in kernel function (default 1/num_features)
-r coef0 : set coef0 in kernel function (default 0)
-c cost : set the parameter C of C-SVC, epsilon-SVR, and nu-SVR (default 1)
-n nu : set the parameter nu of nu-SVC, one-class SVM, and nu-SVR (default 0.5)
-p epsilon : set the epsilon in loss function of epsilon-SVR (default 0.1)
-m cachesize : set cache memory size in MB (default 100)
-e epsilon : set tolerance of termination criterion (default 0.001)
-h shrinking : whether to use the shrinking heuristics, 0 or 1 (default 1)
-b probability_estimates : whether to train a SVC or SVR model for probability estimates, 0 or 1 (default 0)
-wi weight : set the parameter C of class i to weight*C, for C-SVC (default 1)
-v n: n-fold cross validation mode
-q : quiet mode (no outputs)

```

Gambar 3. 2 Syntax Permodelan pada LibSVM

Berdasarkan gambar diatas terdapat beberapa tipe SVM, tipe *kernel*, dan parameter yang dapat dipilih untuk menghasilkan model optimal. SVM dengan tipe *classification* dapat digunakan untuk memprediksi kegunaan dari sebuah ulasan, sedangkan SVM dengan tipe *regression* dapat digunakan untuk membuat peringkat beberapa ulasan berdasarkan nilai kegunaannya. Tipe *kernel* yang dapat digunakan adalah *linier*, *polynomial*, *radial basis*, dan *sigmoid*. Masing-masing *kernel* memiliki parameter yang dapat diubah untuk mendapatkan model optimal.

3.7. Testing

Setelah didapatkan model dari tahap *training*, tahap selanjutnya yaitu menguji model tersebut dengan *testing*. *Testing* dilakukan dengan memanggil *library* SVM untuk menguji model menggunakan data *testing*. Parameter yang digunakan sebagai evaluasi adalah akurasi. Akurasi dapat

didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual [20]. Untuk mengukur parameter ini digunakan *confusion matrix* sebagai alat ukur berbentuk matriks yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritme yang dipakai. Berikut adalah bentuk *confusion matrix*:

Tabel 3. 1 Confusion Matrix

	Prediksi (+)	Prediksi (-)	
Aktual (+)	<i>TP</i> (<i>True Positive</i>)	<i>FN</i> (<i>False Negative</i>)	Marginal
Aktual (-)	<i>FP</i> (<i>False Positive</i>)	<i>TN</i> (<i>True Negative</i>)	
	Marginal		Total

Aktual merupakan kelas sesungguhnya dari suatu titik data sedangkan prediksi merupakan kelas suatu titik data berdasarkan prediksi dari model. Nilai *true positive* merupakan titik data yang diklasifikasikan model sebagai positif dan kelas sesungguhnya memang positif sehingga prediksi dari model adalah benar. Nilai *true negative* merupakan titik data yang diklasifikasikan model sebagai negatif dan kelas sesungguhnya memang negatif sehingga prediksi dari model adalah benar. Nilai *false positive* merupakan titik data yang diklasifikasikan model sebagai positif dan kelas sesungguhnya adalah negatif sehingga prediksi dari model adalah salah. Nilai *false negative* merupakan titik data yang diklasifikasikan model sebagai negatif dan kelas sesungguhnya adalah positif sehingga prediksi dari model adalah salah. Tabel ini digunakan untuk menguji model yang didapatkan. Rumus yang digunakan untuk menghitung akurasi, presisi, dan *recall* adalah:

$$Akurasi = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah total data}} (3.3)$$

3.8. Analisis Penilaian Kualitas Ulasan

Setelah didapatkan model yang telah teruji, langkah selanjutnya adalah melakukan analisis penilaian kualitas ulasan. Luaran dari tahapan ini adalah analisis hasil yang didapatkan beserta pembahasan hasil tersebut apakah telah menjawab perumusan masalah.

3.9. Penyusunan Buku Tugas Akhir

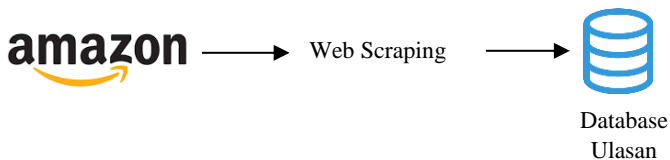
Penyusunan buku tugas akhir akan menghasilkan buku tugas akhir. Penyusunan buku ini dilakukan bersamaan dengan tahapan penelitian lainnya. Harapannya, buku tugas akhir ini dapat dijadikan referensi untuk tugas akhir lainnya.

BAB IV PERANCANGAN

Pada bab ini dijelaskan mengenai perancangan penelitian tugas akhir untuk dapat memberikan gambaran apa saja yang akan dikerjakan pada implementasi penelitian tugas akhir. Pada bab ini dijelaskan cara pemilihan subjek dan objek penelitian serta bagaimana penelitian akan dilakukan.

4.1. Pengambilan Data

Data merupakan salah satu komponen penting untuk dapat melakukan penelitian penilaian kualitas konten ulasan. Data ulasan merupakan dataset yang berupa tipe data teks dari ulasan suatu produk di website Amazon.



Gambar 4. 1 Alur Pengambilan Data

Sesuai dengan alur pengambilan data yang ditunjukkan pada Gambar 4.1, pengambilan data ulasan dilakukan oleh Natanael Yabes Wirawan (Sistem Informasi ITS 2013) dengan teknik *web scraping* menggunakan HTML dan DOM *parsing*. Data ulasan yang diambil adalah ulasan produk “Amazon Echo-Black” yang merupakan produk paling banyak terjual nomor satu di kategori *home automation hubs & controllers*. Selanjutnya, seluruh dataset akan disimpan ke dalam database MySQL. Data ulasan yang telah diambil memiliki beberapa atribut. Tabel 4.1 menunjukkan daftar nama atribut, tipe data, serta keterangan yang ada pada atribut.

Tabel 4. 1 Daftar Atribut Data Ulasan

Atribut	Tipe Data	Keterangan
review_title	Text	Judul ulasan
review_title-href	Text	Link ulasan
review_helpful	Text	Jumlah <i>vote</i> kegunaan yang diberikan oleh pembaca
reviewer	Text	Penulis ulasan
reviewer-href	Text	Profil penulis ulasan
review_date	Text	Tanggal dibuatnya ulasan
rating_value	Text	Penilaian yang diberikan penulis terhadap produk
review_verified	Text	Status pembelian penulis ulasan
review_detail	Text	Isi dari ulasan
title_text	Text	Nama produk yang diulas

Dari proses pengambilan data, terdapat sebanyak 2513 ulasan yang dikumpulkan. Tabel 4.3 menunjukkan sampel data ulasan yang diambil.

Tabel 4. 2 Sampel Data Ulasan

Sampel 1	
review_title	Scared to turn it on...
review_title-href	https://www.amazon.com/gp/customer-reviews/R2WP49R5JHS6O5/ref=cm_cr_arp_d_rvw_ttl?ie=UTF8&ASIN=B00X4WHP5E
review_helpful	0 of 12 people found the following review helpful
reviewer	Sam Casas
reviewer-href	https://www.amazon.com/gp/pdp/profile/A3LY1P8ITQFYSN/ref=cm_cr_arp_d_pdp?ie=UTF8
review_date	on September 30, 2015
rating_value	5.0 out of 5 stars
review_verified	Verified Purchase
review_detail	ready to put it through its paces....
title_text	Amazon Echo - Black
Sampel 2	
review_title	Awesome! ...and only slightly creepy.
review_title-href	https://www.amazon.com/gp/customer-reviews/RG89YY25W1UOL/ref=cm_cr_arp_d_rvw_ttl?ie=UTF8&ASIN=B00X4WHP5E
review_helpful	10 of 13 people found the following review helpful

reviewer	KungFuJosh
reviewer-href	https://www.amazon.com/gp/pdp/profile/A1GOI0YEMU8JCM/ref=cm_cr_arp_d_pdp?ie=UTF8
review_date	on November 24, 2016
rating_value	5.0 out of 5 stars
review_verified	Verified Purchase
review_detail	<p>We're in the kitchen finishing Thanksgiving dinner. My 2 year old daughter says "recksa." I said huh? This happens a couple times and then my wife says "are you saying ALEXA?" Which I followed up with "oh, you want to hear Sesame Street?"</p> <p>That was all it took. Alexa starts playing Sesame Street songs in the next room, and off goes my kid to dance and listen. It's past her bedtime, and she tricked us into getting Alexa to play songs for her.</p>
title_text	Amazon Echo - Black
Sampl 3	
review_title	Pretty great, and looking even better!
review_title-href	https://www.amazon.com/gp/customer-reviews/R3HNKEW04HBFQO/ref=cm_cr_arp_d_rvw_ttl?ie=UTF8&ASIN=B00X4WHP5E
review_helpful	7 of 8 people found the following review helpful

reviewer	Amazon Customer
reviewer-href	https://www.amazon.com/gp/pdp/profile/AVBII00AD5A22/ref=cm_cr_arp_d_pdp?ie=UTF8
review_date	on December 14, 2015
rating_value	4.4 out of 5 stars
review_verified	Verified Purchase
review_detail	<p>The speech recognition has been actually pretty good out of the box, and can train. The API is easy enough that this computer scientist was able to have a new skill uploaded in about 30 minutes, and Pandora, iHeartRadio, and other integrated devices have been great. This has turned into our kitchen timer, our shopping list, and our music device. We went from having the TV on every night, to just listening to music for "background noise". The number of things that have skills published for it are also pretty awesome, especially if you're trying to get into a smart(er) home. Most of the startups and Internet of Things vendors are also all publishing skills for the Echo, so I expect this to grow into a much better.....</p>
title_text	Amazon Echo – Black

4.2. Pemilihan Atribut

Tahap selanjutnya setelah pengambilan data adalah pemilihan atribut. Pemilihan atribut dilakukan untuk mempertimbangkan dan menyeleksi atribut apa saja yang akan digunakan untuk penelitian.

Berdasarkan data ulasan yang diperoleh dari Amazon, atribut yang dapat digunakan adalah atribut *review_helpful*, *rating_value*, dan *review_detail*. Atribut *review_helpful* dipilih karena atribut ini menunjukkan berapa banyak pembaca yang menganggap ulasan tersebut berguna. Atribut *rating_value* dipilih karena menunjukkan penilaian yang diberikan penulis terhadap produk. Atribut *review_detail* dipilih karena menunjukkan isi dari ulasan yang ditulis. Sehingga, semua atribut ini akan diolah dan dilakukan analisis. Tabel 4.3 menunjukkan sampel data sesuai dengan atribut yang dipilih untuk penelitian.

Tabel 4. 3 Sampel Atribut Terpilih

Sampel 1	
review_helpful	10 of 13 people found the following review helpful
rating_value	5.0 out of 5 stars
review_detail	ready to put it through its paces....
Sampel 2	
review_helpful	10 of 13 people found the following review helpful
rating_value	5.0 out of 5 stars
review_detail	We're in the kitchen finishing Thanksgiving dinner. My 2 year old daughter says "recksa." I said huh? This happens a couple times and then my wife says "are you saying ALEXA?" Which I followed up with "oh, you want to hear Sesame Street?" That was all it took. Alexa starts playing Sesame Street songs in the next room, and off

	goes my kid to dance and listen. It's past her bedtime, and she tricked us into getting Alexa to play songs for her.
Sample 3	
review_helpful	7 of 8 people found the following review helpful
rating_value	4.4 out of 5 stars
review_detail	<p>The speech recognition has been actually pretty good out of the box, and can train. The API is easy enough that this computer scientist was able to have a new skill uploaded in about 30 minutes, and Pandora, iHeartRadio, and other integrated devices have been great. This has turned into our kitchen timer, our shopping list, and our music device.</p> <p>We went from having the TV on every night, to just listening to music for "background noise". The number of things that have skills published for it are also pretty awesome, especially if you're trying to get into a smart(er) home. Most of the startups and Internet of Things vendors are also all publishing skills for the Echo, so I expect this to grow into a much better.....</p>

4.3. Perancangan Model

Tahap perancangan model mendefinisikan apa saja yang perlu dilakukan setelah pemilihan atribut data yang digunakan dalam penelitian. Terdapat tiga tahapan besar yaitu perancangan preproses data, perancangan ekstraksi fitur, dan perancangan pemrosesan data. Berikut ini merupakan tahapan yang ada pada setiap perancangan.

4. 3. 1. Perancangan Preproses Data

Data ulasan yang sudah terkumpul sebanyak 10363 akan dilakukan preproses data untuk disesuaikan dengan batasan penelitian ini. Data ulasan setidaknya harus memiliki satu *vote* yang diberikan pembaca lain. Hal ini mengingat bahwa atribut *vote* akan digunakan dalam pemrosesan data sehingga atribut *vote* harus memiliki nilai lebih besar dari 0. Setelah dilakukan seleksi tersisa 2513 data yang memiliki setidaknya satu *vote*. Sebagian besar ulasan yang tidak memiliki *vote* dikarenakan ulasan tersebut masih terbilang baru dengan rentang waktu mulai bulan Oktober sampai Desember 2016. Sedangkan ulasan yang sudah ada sebelum itu hampir sebagian besar sudah memiliki *vote*.

Selanjutnya, data yang sudah diseleksi akan dihitung rata-rata penilaian bintang yang diberikan oleh penulis. Untuk setiap data ulasan, akan dihitung rasio kegunaan yaitu dengan membandingkan jumlah *vote* yang didapatkan dengan total jumlah *vote* yang diberikan pembaca. Sehingga data ini dapat digunakan untuk tahap ekstraksi fitur.

4. 3. 2. Perancangan Ekstraksi Fitur

Ekstraksi fitur akan dilakukan menggunakan bahasa pemrograman Java. Gambar 4.2 menunjukkan alur ekstraksi fitur penelitian ini. Data yang telah tersimpan di database, akan diolah menggunakan bahasa pemrograman java. Setelah itu, data yang telah diolah akan disimpan ke tabel baru dalam database yang sama. Terdapat tiga fitur utama yang akan digunakan dalam penelitian ini.



Gambar 4. 5 Alur Ekstraksi Fitur

4. 3. 2. 1. Fitur Struktural

Ekstraksi fitur struktural akan dilakukan menggunakan bahasa pemrograman Java. Data ulasan yang telah terkumpul akan dihitung total karakter, total kata, total kalimat, dan total suku kata untuk masing-masing ulasan. Total karakter dihitung menggunakan method `length()`, method ini akan mengembalikan jumlah karakter yang terkandung dalam objek string. Total kata dihitung menggunakan `countWords()`, method ini akan menghitung jumlah kata yang terdapat dalam teks ulasan yang dipisahkan menggunakan spasi.

Total kalimat dihitung menggunakan `countSentences()`, method ini akan menghitung kalimat yang terdapat dalam teks dengan pemisah tanda baca di akhir kalimat yaitu tanda titik (`.`), tanda seru (`!`), tanda tanya (`?`), dan tanda titik dua (`:`). Total suku kata dihitung menggunakan

`countSyllables()`, method ini akan menghitung satu suara vokal dengan atau tanpa konsonan disekitarnya membentuk seluruh atau sebagian kata. Misalnya, dalam kata “*good*” terdapat satu suku kata dan dalam kata “*happy*” terdapat dua suku kata.

Sehingga, tahapan ini akan menghasilkan data kuantitatif yang merepresentasikan panjangnya ulasan, jumlah kalimat, jumlah kata, dan jumlah suku kata untuk masing-masing ulasan yang selanjutnya dapat digunakan pada tahap pemrosesan data.

4.3.2.2. Fitur Metadata

Ekstraksi fitur metadata akan dilakukan untuk menghasilkan rasio kegunaan, *star rating*, dan *review extremity*. Adapun rumus yang digunakan untuk menghitung rasio kegunaan ditunjukkan pada rumus dibawah ini.

$$\text{rasio kegunaan} = \frac{\text{jumlah vote yang didapat}}{\text{total jumlah vote yang diberikan}} \quad (4.1)$$

Jumlah *vote* yang didapat merepresentasikan berapa banyak pembaca yang menganggap ulasan tersebut berguna. Sedangkan total jumlah *vote* yang diberikan merepresentasikan berapa banyak pembaca yang memberikan *vote* berguna dan tidak berguna terhadap ulasan tersebut.

Star rating merupakan penilaian berbintang yang diberikan penulis terhadap produk tersebut mulai dari skala 1-5. Selain itu, penelitian ini juga memperhitungkan *review extremity* yang didapatkan dengan cara mengurangi *star rating* yang diberikan penulis dengan rata-rata *star rating* produk tersebut. Hal ini

diperhitungkan karena ulasan dengan penilaian ekstrem yang menyimpang dari rata-rata penilaian produk tersebut tidak akan berguna untuk pembaca lain.

4. 3. 2. 3. Fitur Ketebacaan

Ekstraksi fitur keterbacaan dilakukan untuk menilai kerumitan dan kekomprehensifan ulasan. Pada penelitian ini digunakan dua penilaian yaitu *Flesch Reading Ease Score* yang menilai kerumitan suatu teks dan *Automated Readability Index* yang menilai kekomprehensifan suatu teks. Pada tahapan ini, data yang digunakan adalah data dari hasil ekstraksi fitur struktural. Data tersebut telah merepresentasikan semua variabel yang dibutuhkan untuk menghitung penilaian ini, yaitu total karakter, total kata, total kalimat, dan total suku kata.

Setelah itu, dengan menggunakan bahasa pemrograman Java, masing-masing ulasan akan dinilai uji keterbacaannya. Sehingga, hasil dari tahapan ini adalah penilaian keterbacaan untuk ulasan yang akan digunakan untuk pemrosesan data.

4. 3. 3. Perancangan Pemrosesan Data

Data yang sudah dilakukan ekstraksi fitur selanjutnya akan dilakukan pemrosesan data. Pada tahapan ini, data disimpan dalam format CSV (*Comma Separated Value*) sehingga data dapat dilakukan pemrosesan data. Data memiliki beberapa atribut dengan nilai kuantitatif yang mencakup hasil dari ekstraksi fitur struktural, fitur metadata, dan fitur keterbacaan. Terdapat dua tahapan besar dalam pemrosesan data yaitu penilaian menggunakan *weighted sum* dan klasifikasi menggunakan *support vector machine*.

4.3.3.1. Perancangan Penilaian Weighted Sum

Data yang telah diekstrak selanjutnya akan dilakukan penilaian menggunakan *weighted sum*. Terdapat delapan kriteria penilaian dengan masing-masing bobot nya yang ditunjukkan pada tabel 4.4.

Tabel 4. 4 Kriteria dan Bobot Penilaian

No.	Kriteria	Bobot
1	Rasio kegunaan	0,34
2	Hasil uji FK	0,21
3	Hasil uji ARI	0,15
4	<i>Star rating</i>	0,11
5	<i>Review Extremity</i>	0,08
6	Panjang ulasan	0,06
7	Jumlah kalimat	0,03
8	Jumlah kata	0,02
Total		1

Bobot untuk masing-masing kriteria ditentukan metode *Rank-Order Centroid* (ROC). Metode ini memberikan bobot terhadap sejumlah kriteria yang diurutkan berdasarkan kepentingannya. Urutan masing-masing kriteria ditentukan berdasarkan hasil penelitian sebelumnya yang dilakukan oleh Nikolaos, dkk [4]. Rasio kegunaan adalah kriteria yang paling berpengaruh untuk menilai kualitas konten ulasan. Semakin besar nilai rasio kegunaan, maka ulasan tersebut semakin berkualitas. Selain itu, ulasan yang berkualitas juga memiliki nilai keterbacaan yang bagus, sehingga fitur keterbacaan memiliki bobot yang terbilang besar untuk penilaian kualitas ulasan. *Star rating* yang

diberikan penulis juga berpengaruh penting terhadap kualitas ulasan. *Review extremity* merupakan kriteria pendukung yang dapat memberikan gambaran mengenai penilaian *star rating* yang diberikan penulis dengan rata-rata penilaian keseluruhan ulasan. Sedangkan untuk fitur struktural memberikan hasil yang signifikan terhadap kualitas ulasan, namun tidak berpengaruh secara langsung. Sehingga, total bobot untuk fitur struktural adalah yang paling kecil diantara fitur lainnya.

Setelah masing-masing ulasan mendapatkan nilai, langkah selanjutnya adalah mengurutkan ulasan dari nilai yang terbesar hingga yang terkecil. Sehingga, hasil dari tahap ini adalah daftar urutan ulasan mulai dari yang paling berkualitas hingga yang paling tidak berkualitas.

Tabel 4. 5 Nilai Min dan Max Tiap Atribut

No.	Data	Min	Max
1	Jumlah karakter	20	9541
2	Jumlah kata	3	1635
3	Jumlah kalimat	1	141
4	Jumlah suku kata	5	2632
5	Helpful	0,12	1
6	Rating	4,4	5
7	Extremity	-0,46	0,14
8	FK	-2,81	54,47
9	ARI	-1,33	71,72

4. 3. 3. 2. Perancangan Klasifikasi SVM

Klasifikasi SVM dilakukan untuk dapat mengklasifikasikan ulasan yang baru ditulis penulis karena ulasan yang baru ditulis kemungkinan kecil mempunyai jumlah *vote* yang lebih banyak dibandingkan

ulasan lama. Padahal belum pasti jika ulasan yang baru tidak berkualitas. Klasifikasi dilakukan menggunakan *library* LIBSVM dengan bahasa pemrograman Java. Data akan dilakukan normalisasi agar memiliki *range* yang sama. Selanjutnya, data akan dikonversi formatnya menjadi format yang didukung oleh SVM yaitu *sparse format*. Lalu, data dibagi menjadi data *training* dan data *testing*. Selanjutnya, data *testing* akan dilatih menggunakan konfigurasi kernel dan parameter yang terdapat pada *library* sehingga mendapatkan model optimal. Selanjutnya, konfigurasi ini akan diuji menggunakan data *testing* untuk mendapatkan nilai keakuratan optimal untuk data ulasan ini.

BAB V IMPLEMENTASI

Pada bab ini dijelaskan mengenai proses implementasi penelitian. Bab ini membahas tentang perangkat penelitian yang digunakan, penerapan preproses data, penerapan ekstraksi fitur, serta penerapan pemrosesan data yang dilakukan menggunakan bahasa pemrograman Java.

5.1. Perangkat Penelitian

Dalam pelaksanaan penelitian ini, dibutuhkan perangkat-perangkat yang dapat mendukung proses yang ada pada setiap tahapan penelitian. Adapun perangkat-perangkat yang dibutuhkan meliputi perangkat keras dan perangkat lunak beserta spesifikasinya sebagai berikut:

5.1.1. Perangkat Keras

- CPU: Intel ® Core™ i3-2350M CPU @2.30 GHz,
RAM 4.00 GB, Harddisk 160 GB SDD
- Monitor: 14 inch 1440x900
- *Keyobard* dan *mouse*

5.1.2. Perangkat Lunak

- Sistem Operasi : Windows 10
- Bahasa Pemograman : Java
- *Tools* : Netbeans IDE 8.2
XAMPP
Ms. Excel 2016

5.2. Praproses Data

Data yang digunakan adalah data amazon yang dilakukan oleh Natanael Yabes Wirawan dengan teknik *scrapping*. Data ini kemudian diseleksi sesuai dengan kriteria yang dibutuhkan

yaitu masing-masing ulasan tidak boleh memiliki nilai kegunaan sebesar 0, sehingga terkumpulah 2513 data ulasan.

5.3. Ekstraksi Fitur

Ekstraksi fitur pada penelitian ini terdiri dari beberapa tahap antara lain ekstraksi fitur struktural, ekstraksi fitur metadata, dan ekstraksi fitur keterbacaan yang menggunakan bahasa pemrograman Java.

5.3.1. Ekstraksi Fitur Struktural

Setiap ulasan dilakukan ekstraksi untuk menghitung total karakter, kata, kalimat, dan suku kata menggunakan bahasa pemrograman Java. Berikut ini adalah *source code* yang digunakan untuk melakukan ekstraksi tersebut.

```
public static int countWords(String str) {
    int count = 1;
    for (int i = 0; i <= str.length() - 1; i++) {
        if (str.charAt(i) == ' ' && str.charAt(i + 1) != ' ') {
            count++;
        }
    }
    return count;
}
```

Kode 5. 1 Method countWords

Method `countWords` digunakan untuk menghitung total kata yang ada pada suatu teks. Method ini akan membedakan satu kata dengan kata lainnya yang dipisahkan dengan spasi. Kode 5.1 menunjukkan *source code* yang digunakan.

```
public static int countSentences(String str) {
    String delimiters = ",,:?!";
    int sentenceCount = 0;
    int count = str.split("[!?:.]+").length;
    for (int i = 0; i < str.length() - 1; i++) {
```

```

        for (int j = 0; j < delimiters.length(); j++) {
            if (str.charAt(i) == delimiters.charAt(j)) {
                sentenceCount++;
            }
        }
    }
    return count;
}

```

Kode 5. 2 Method countSentences

Method `countSentences` yang ditunjukkan pada kode 5.2 digunakan untuk menghitung kalimat yang terdapat pada suatu teks. Method ini menghitung kalimat dengan cara memperhatikan tanda baca sebagai penentu akhir sebuah kalimat. Tanda baca yang dimaksud adalah tanda titik (.), tanda seru (!), tanda tanya (?), dan tanda titik dua (:).

```

public int countSyllables(String word) {
    int count = 0;
    word = word.toLowerCase();

    if (word.charAt(word.length() - 1) == 'e') {
        if (silente(word)) {
            String newword = word.substring(0,
word.length() - 1);
            count = count + countit(newword);
        } else {
            count++;
        }
    } else {
        count = count + countit(word);
    }
    return count;
}

```

Kode 5. 3 Method countSyllables

Method `countSyllables` yang ditunjukkan pada kode 5.3 digunakan untuk menghitung jumlah suku kata yang terdapat pada suatu teks. Data ulasan yang digunakan dalam penelitian ini ditulis dalam bahasa Inggris, sehingga peraturan penentuan suku kata yang digunakan sesuai dengan bahasa Inggris. Method ini akan menghitung satu suara vokal dengan atau tanpa konsonan disekitarnya membentuk seluruh atau sebagian kata. Misalnya, dalam kata “*good*” terdapat satu suku kata dan dalam kata “*happy*” terdapat dua suku kata.

```
private int countit(String word) {
    int count = 0;
    Pattern splitter =
Pattern.compile("[^aeiouy]*[aeiouy]+");
    Matcher m = splitter.matcher(word);

    while (m.find()) {
        count++;
    }
    return count;
}

private boolean silente(String word) {
    word = word.substring(0, word.length() - 1);

    Pattern yup = Pattern.compile("[aeiouy]");
    Matcher m = yup.matcher(word);

    if (m.find()) {
        return true;
    } else {
        return false;
    }
}
```


Kode 5. 4 Method countit dan silente

Selain itu, untuk menghitung suku kata diperlukan method lainnya untuk kasus tertentu. Kode 5.4 menunjukkan *method* yang dipakai untuk membedakan kasus dimana huruf “e” mati di akhir kalimat yang tidak dihitung sebagai suku kata dan huruf “e” hidup di akhir kalimat yang dihitung sebagai suku kata. Misalnya, kata “love” mengandung huruf “e” mati di akhir kalimat, sehingga kata tersebut hanya mempunyai satu suku kata. Sedangkan kata “employee” memiliki “e” hidup di akhir kalimat sehingga dihitung memiliki tiga suku kata.

```

while(rs.next()){
    String review = rs.getString("review_detail");
    Amazon amazon = new Amazon();
    System.out.println(review);
    System.out.println("Your sentence has " +
review.length() + " characters");
    System.out.println("Your sentence has " +
countWords(review) + " words.");
    System.out.println("Your sentence has " +
countSentences(review) + " sentences.");
    System.out.println("Your sentece has " +
amazon.countSyllables(review) + " sylablls.");

    String sqli = "INSERT INTO struktural (id,
karakter, kata, kalimat, suku_kata) "
+ "values (id,"+ review.length()+", "+
countWords(review)+", "+ countSentences(review)+", "+
amazon.countSyllables(review)+");"

    PreparedStatement statement =
conn.prepareStatement(sqli);
    statement.executeUpdate(sqli);
}

```

Kode 5. 5 Ekstraksi Fitur Struktural

Adapun *source code* yang digunakan untuk melakukan ekstraksi fitur struktural ditunjukkan pada gambar 4.7. Hasil dari tahapan ini adalah jumlah karakter, kata, kalimat, dan suku kata masing-masing ulasan yang disimpan dalam tabel di database.

5.3.2. Ekstraksi Fitur Metadata

Ekstraksi fitur metadata dilakukan untuk mendapatkan nilai rasio kegunaan, rating, dan extremity masing-masing ulasan. Pada data yang didapatkan, data total kegunaan berupa teks yang ditulis seperti ini, “10 out of 10 people found the following review helpful”. Informasi yang dibutuhkan hanya angka 10 dan 10 saja. Selain itu, data rating berupa teks yang ditulis seperti ini, “5.0 out of 5.0”. Informasi yang dibutuhkan hanya 5.0 saja. Sedangkan extremity merupakan hasil dari pengurangan rata-rata penilaian yang diberikan seluruh ulasan dengan penilaian yang diberikan penulis ulasan tersebut. Adapun *source code* yang digunakan untuk melakukan ekstraksi fitur metada ditunjukkan pada kode 5.6.

```
while (rs.next()) {  
    String helpful = rs.getString("review_helpful");  
    String value = rs.getString("review_value");  
    double a = helpful.charAt(0);  
    double b = helpful.charAt(5);  
    double c = a / b;  
    String g = value.substring(0, 3);  
    double a1 = Double.parseDouble(g);  
    double d = a1 - 4.863430163151559;
```

```

System.out.println("Helpful ratio: " + c);
System.out.println("Rating: " + a1);
System.out.println("Extremity " + d);

String sqli = "INSERT INTO metadata (id,
helpful, rating, extremity) "
+ "values (id," + c + ", " + g + ", " + d + ")";

PreparedStatement statement =
conn.prepareStatement(sqli);
statement.executeUpdate(sqli);
}

```

Kode 5. 6 Ekstraksi Fitur Metadata

5.3.3. Ekstraksi Fitur Keterbacaan

Ekstraksi fitur uji keterbacaan dilakukan dengan menggunakan data hasil ekstraksi fitur struktural. Lalu, masing-masing ulasan akan dilakukan uji keterbacaan sesuai dengan rumus Flesch Kincaid dan Automated Readability Index. Adapun *source code* yang digunakan ditunjukkan pada kode 5.7.

```

while(rs.next()){
    int chara = rs.getInt("karakter");
    int words = rs.getInt("kata");
    int sent = rs.getInt("kalimat");
    int syll = rs.getInt("suku_kata");
    double fk = 0.39 * words/sent + 11.8 * syll/words -
15.59 ;
    double ari = 4.71 * chara/words + 0.5 * words/sent
- 21.43;
    System.out.println("FK " + fk);
}

```

```

        System.out.println("ARI " + ari);

        String sqli = "INSERT INTO readability (id, fk,
        ari) " + "values (id,"+ fk+", "+ ari+)";
        PreparedStatement          statement          =
        conn.prepareStatement(sqli);
        statement.executeUpdate(sqli);}

```

Kode 5. 7 Ekstraksi Fitur Keterbacaan

5.4. Pemrosesan Data

Pemrosesan data pada penelitian ini terbagi menjadi dua tahap, yaitu penilaian dengan *weighted sum* dan klasifikasi dengan *support vector machine*.

5.4.1. Penilaian dengan Weighted Sum

Penilaian dengan *weighted sum* terdiri dari beberapa tahap yaitu penyamaan skala, penentuan bobot, dan penilaian. Pada subbab ini akan dijelaskan hasil implementasi penilaian dengan *weighted sum*.

5.4.1.1. Penyamaan Skala

Penilaian *weighted sum* dapat dilakukan jika semua atribut memiliki skala yang sama. Namun, data yang didapat tidak mempunyai skala yang sama, sehingga harus dilakukan penyamaan skala terlebih dahulu. Data memiliki 8 atribut yang mana 7 atribut memiliki skala yang tidak sama. Penyamaan skala dilakukan dengan cara membagi masing-masing nilai dengan nilai tertinggi yang ada di atribut tersebut. Sehingga data akan memiliki skala mulai dari 0 hingga 1.

Tabel 5.1 menunjukkan nilai masing-masing atribut saat belum dilakukan penyamaan skala. Atribut *fk* memiliki nilai tertinggi sebesar 54,47 dan nilai terendah sebesar -2,81 sehingga masing-masing nilai dibagi dengan nilai terbesar

untuk mendapatkan nilai dengan skala 0-1. Atribut ari memiliki nilai tertinggi sebesar 71,72 dan nilai terendah sebesar -1,33 sehingga masing-masing nilai dibagi dengan nilai terbesar untuk mendapatkan nilai dengan skala 0-1. Atribut rating memiliki nilai tertinggi sebesar 5 dan nilai terendah sebesar -4,4 sehingga masing-masing nilai dibagi dengan nilai terbesar untuk mendapatkan nilai dengan skala 0-1.

Atribut *extremity* memiliki nilai tertinggi sebesar 0,14 dan nilai terendah sebesar -0,46 sehingga masing-masing nilai dibagi dengan nilai terbesar untuk mendapatkan nilai dengan skala 0-1. Atribut karakter memiliki nilai tertinggi sebesar 9541 dan nilai terendah sebesar 20 sehingga masing-masing nilai dibagi dengan nilai terbesar untuk mendapatkan nilai dengan skala 0-1. Atribut kalimat memiliki nilai tertinggi sebesar 141 dan nilai terendah sebesar 1 sehingga masing-masing nilai dibagi dengan nilai terbesar untuk mendapatkan nilai dengan skala 0-1. Atribut kata memiliki nilai tertinggi sebesar 1635 dan nilai terendah sebesar 3 sehingga masing-masing nilai dibagi dengan nilai terbesar untuk mendapatkan nilai dengan skala 0-1. Data ini siap digunakan untuk dilakukan perhitungan *weighted sum*.

5.4.1.2. Penentuan Bobot

Penentuan bobot dilakukan menggunakan metode ROC. Metode ini memperhitungkan urutan kriteria mulai dari yang terpenting hingga yang kurang penting. Perhitungan masing-masing bobot dilakukan menggunakan rumus berikut:

$$W_j (ROC) = \frac{1}{n} \sum_{k=j}^n \frac{1}{r_k}$$

Sehingga, untuk rasio kegunaan yang berada di urutan pertama, perhitungan bobotnya adalah:

$$W_1 = \frac{1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8}{8} = 0,34$$

Untuk hasil uji FK, perhiungan bobotnya adalah:

$$W_2 = \frac{1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8}{8} = 0,21$$

Untuk hasil uji ARI, perhitungan bobotnya adalah:

$$W_3 = \frac{1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8}{8} = 0,15$$

Untuk *star rating*, perhitungan bobotnya adalah:

$$W_4 = \frac{1/4 + 1/5 + 1/6 + 1/7 + 1/8}{8} = 0,11$$

Untuk *review extremity* perhitungan bobotnya adalah:

$$W_5 = \frac{1/5 + 1/6 + 1/7 + 1/8}{8} = 0,08$$

Untuk panjang ulasan perhitungan bobotnya adalah:

$$W_6 = \frac{1/6 + 1/7 + 1/8}{8} = 0,06$$

Untuk jumlah kalimat perhitungan bobotnya adalah:

$$W_7 = \frac{1/7 + 1/8}{8} = 0,03$$

Untuk jumlah kata perhitungan bobotnya adalah:

$$W_8 = \frac{1/8}{8} = 0,02$$

Sehingga, dari perhitungan bobot masing-masing kriteria diperoleh bobot yang akan digunakan untuk perhitungan *weighted sum*. Tabel 5.3 menunjukkan hasil perhitungan bobot *weighted sum*.

Tabel 5. 1 Bobot Weighted Sum

No.	Kriteria	Bobot
1	Rasio kegunaan	0,34
2	Hasil uji FK	0,21
3	Hasil uji ARI	0,15
4	<i>Star rating</i>	0,11
5	<i>Review Extremity</i>	0,08
6	Panjang ulasan	0,06
7	Jumlah kalimat	0,03
8	Jumlah kata	0,02
Total		1

5.4.1.3. Penilaian dan Pengurutan Kualitas Ulasan Menggunakan *Weighted Sum*

Setelah data didapat dan bobot telah ditentukan, langkah selanjutnya adalah penilaian kualitas ulasan menggunakan *weighted sum*. Penilaian dilakukan menggunakan formula di Ms. Excel 2016 dengan mengacu pada rumus sebagai berikut:

$$A_i^{W_{SM-score}} = \sum_{j=1}^n w_j a_{i,j}, \text{ untuk } i = 1, 2, 3, \dots, m \text{ [14]}$$

dimana A adalah penilaian, w adalah bobot, dan a adalah alternatif. Pada penelitian ini terdapat delapan kriteria yaitu, rasio kegunaan, hasil uji FK, hasil uji ARI, *star rating*, *review*

extremity, panjang ulasan, jumlah kalimat, dan jumlah kata. Masing-masing kriteria memiliki bobot yang berbeda sesuai dengan urutan kepentingan dalam penilaian kualitas ulasan. Selanjutnya adalah melakukan penilaian dengan mengkalikan masing-masing nilai dengan bobotnya. Sebagai contoh untuk ulasan dengan id 1, perhitungannya adalah:

$$A_1 = 0,5 \times 0,34 + 0,1168 \times 0,21 + 0,1294 \times 0,15 + 1 \times 0,11 + 1 \times 0,08 \\ + 0,0875 \times 0,06 + 0,0851 \times 0,03 + 0,081 \times 0,02 \\ A_1 = 0,4135$$

Sehingga, ulasan dengan id 1 memiliki nilai sebesar 0,4135 dari nilai tertinggi 1. Setelah dilakukan penilaian, langkah selanjutnya adalah mengurutkan ulasan mulai dari yang memiliki nilai tertinggi hingga yang terkecil dengan pilihan *sort*. Sebagai contoh, id nomor 1 berada di urutan ke -1848 dari 2513.

5.4.2. Klasifikasi dengan Support Vector Machine

Langkah selanjutnya adalah klasifikasi dengan *support vector machine*. Data ulasan dibagi menjadi dua kelas berdasarkan rasio kegunaannya yaitu kelas *helpful* yang direpresentasikan dengan angka 1 dan kelas *unhelpful* yang direpresentasikan dengan angka 2. Batasan kelas adalah 0,5. Sebagai contoh, jika suatu ulasan memiliki rasio perbandingan 0,4 sehingga ulasan tersebut termasuk dalam kelas *unhelpful*. Jika suatu ulasan memiliki rasio perbandingan 0,6, maka ulasan tersebut termasuk dalam kelas *helpful*. Jika suatu ulasan memiliki rasio perbandingan 0,5 atau sama dengan batasan, maka ulasan tersebut termasuk dalam kelas *helpful*.

Dalam melakukan klasifikasi dengan *support vector machine*, dilakukan beberapa tahap seperti normalisasi data, konversi data, *training*, dan *testing*.

5.4.2.1. Normalisasi Data

Normalisasi data dilakukan agar setiap data memiliki *range* yang sama. Terdapat beberapa metode untuk melakukan normalisasi, namun metode umum yang sering dipakai adalah Z-score. Metode ini memperhitungkan nilai minimum, maksimum, dan standar deviasi yang ada pada data sehingga metode ini dianggap sebagai metode terbaik untuk melakukan normalisasi data.

Adapun rumus yang digunakan untuk melakukan normalisasi menggunakan Z-score adalah:

$$z_i = \frac{x_i - \bar{x}}{S}$$

dimana z_i adalah hasil normalisasi, x_i adalah data, \bar{x} adalah rata-rata, dan S adalah standar deviasi. Perhitungan normalisasi data dilakukan menggunakan Ms. Excel 2016.

5.4.2.2. Konversi Data

Setelah dilakukan normalisasi, langkah selanjutnya adalah konversi data. SVM membutuhkan data yang berformat *sparse data*, format ini tidak mengikutsertakan nilai 0. Konversi data dilakukan menggunakan bahasa pemrograman Java yang ditunjukkan oleh kode 5.8.

```
public static void write(String inputFile, String outputFile)
{
    try          (Stream<String>          stream          =
Files.lines(Paths.get(inputFile));
                BufferedWriter            bw=
Files.newBufferedWriter(Paths.get(outputFile),
StandardCharsets.UTF_8)){

        Object[] lines = stream.toArray();
        for (Object line : lines) {
```

```

        if (line instanceof String &&
(((String)line).trim().length(>0)){
            String[] data = ((String)line).split(",");
            if (data.length==8){
                String dataLine = "";
                if(data[7].contains("1")){
                    dataLine += "1";
                }
                else if (data[7].contains("2")){
                    dataLine += "2";
                }

                for (int i=0; i<7; i++){

                    double
attributeValue=Double.parseDouble(data[i]);
                    if(attributeValue!=0){
                        dataLine+="
"+(i+1)+":"+attributeValue;
                    }
                }
                System.out.println(dataLine);
                bw.write(dataLine);
                bw.newLine();
            }
        }
    } catch (IOException e){
        e.printStackTrace();
    }
}

```

Kode 5. 8 Konversi Data

5.4.2.3. Pembagian Dataset

Sebelum melakukan *training* dan *testing*, perlu dilakukan pembagian dataset yang membagi antara data untuk *training*

dan data untuk *testing*. Pembagian dilakukan menggunakan bahasa Java yang ditunjukkan oleh kode 5.9.

```

public static void main(String[] args) throws IOException {
    // TODO code application logic here
    randomize(80,

"C:\\Users\\Rani\\Documents\\NetBeansProjects\\svm\\src\\sv
m\\review.data",

"C:\\Users\\Rani\\Documents\\NetBeansProjects\\svm\\src\\sv
m\\review.train",

"C:\\Users\\Rani\\Documents\\NetBeansProjects\\svm\\src\\sv
m\\review.test");
    }
    public static void randomize(double percentageInTrain,
String inputFile, String outTrain, String outTest) throws
IOException
    {
        int ntotal = 0, ntrain = 0, ntest = 0;
        try (Stream<String> stream =
Files.lines(Paths.get(inputFile));
            BufferedWriter btrain =
Files.newBufferedWriter(Paths.get(outTrain),
StandardCharsets.UTF_8);
            BufferedWriter btest =
Files.newBufferedWriter(Paths.get(outTest),
StandardCharsets.UTF_8)) {
            Object[] lines = stream.toArray();
            for (Object line : lines) {
                if (line instanceof String && (((String)
line).trim().length() > 0)) {
                    String data = (String) line;
                    if (Math.random() < percentageInTrain/100) {

```

```

        btrain.write(data);
        btrain.newLine();
        ntrain++;
    } else {
        btest.write(data);
        btest.newLine();
        ntest++;
    }
    ntotal++;
}
}
}
System.out.println("The total number of instances in the
data: "+ntotal);
System.out.println("The number of data in the training
set: "+ntrain);
System.out.println("The number of data in the testing set:
"+ntest);
}
}
}

```

Kode 5. 9 Pembagian Dataset

Pertama-tama, data diacak dulu menggunakan metode *randomize*. Setelah itu data akan dibagi 80% menjadi data *training* dan 20% menjadi data *testing*. Selanjutnya data *training* dan *testing* akan disimpan untuk tahap selanjutnya.

5.4.2.4. Training

Data *training* selanjutnya akan dilatih menggunakan libSVM dengan metode *svm-train*. Metode ini dijalankan melalui *command prompt*. Adapun *syntax* untuk melakukan *training* ditunjukkan di kode 5.10

```
svm-train [options] training_set_file [model_file]
```

Kode 5. 10 Training SVM

Options yang dimaksud adalah pilihan svm dan kernel yang digunakan untuk membuat model SVM. Adapun pilihannya ditunjukkan oleh tabel 5.6.

Tabel 5. 2 Pilihan SVM dan Kernel

Typ	Pilihan	Keterangan
-s (svm)	0	C-SVC (multi-class classification)
	1	N-SVC (multi-class classification)
	2	One-class SVM
	3	Epsilon-SVR (regression)
	4	Nu-SVR (regression)
-t (kernel)	0	Linear: $u \cdot v$
	1	Polynomial: $(\gamma u \cdot v + \text{coef0})$
	2	Radial basis function: $\exp(-\gamma u-v ^2)$
	3	Sigmoid: $\tanh(\gamma u \cdot v + \text{coef0})$
	4	Precomputed kernel (kernel values in <code>training_set_file</code>)
-d	degree	Set degree in kernel function (default 3)
-g	gamma	Set gamma in kernel function (default $1/\text{num_features}$)
-r	coef0	Set coef0 in kernel function (default 0)
-c	cost	Set the parameter c of C-SVC, epsilon-SVR, and nu-SVR. (default 1)
-n	nu	Set the parameter nu of nu-SVC, one-class SVM, and nu-SVR (default 0.5)
-m	cacheSize	Set cache memory size in

		MB (default 100)
-e	epsilon	Set tolerance of termination criterion (default 0.001)
-h	shrinking	Whether to use the shrinking heuristics, 0 or 1 (default 1)
-b	probability_estimates	Whether to train SVC or SVR model for probability estimates, 0 or 1 (default 0)
-wi	weight	Set the parameter C of class i to $\text{weight}_i * C$, for C-SVC (default 1)
-v	n	n-fold cross validation mode
-q	Quiet mode	(no outputs)

5.4.2.5. *Testing*

Setelah didapatkan model, langkah selanjutnya adalah melakukan *testing* untuk melihat apakah model yang didapat bisa digunakan untuk data lain dengan nilai keakuratan tertentu. *Syntax* yang digunakan untuk melakukan *testing* adalah:

```
svm-predict [options] test_file_model output_file
```

Kode 5. 11 Testing SVM

Metode yang digunakan untuk *testing* adalah `svm-predict` dan selanjutnya memasukkan data *testing* dan data model yang telah didapatkan dari tahap *training*. Hasilnya akan dihitung berapa presentase keakuratan model untuk memprediksi data *testing*.

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan hasil serta analisis terhadap hasil yang diperoleh dari proses implementasi yang telah dibahas pada bab sebelumnya.

6.1. Data Ulasan

Setelah dilakukan pengambilan data ulasan, lalu data tersebut dipilih atributnya sehingga hanya atribut id, review_helpful, review_value, dan review_detail yang digunakan dalam penelitian ini. Adapun contoh data sampel dari data ulasan ditunjukkan pada tabel 6.1.

Tabel 6. 1 Data Ulasan

Sampel 1	
id	1
review_helpful	10 of 13 people found the following review helpful
rating_value	5.0 out of 5 stars
review_detail	ready to put it through its paces....
Sampel 2	
id	2
review_helpful	10 of 13 people found the following review helpful
rating_value	5.0 out of 5 stars
review_detail	<p>We're in the kitchen finishing Thanksgiving dinner. My 2 year old daughter says "recksa." I said huh? This happens a couple times and then my wife says "are you saying ALEXA?" Which I followed up with "oh, you want to hear Sesame Street?"</p> <p>That was all it took. Alexa starts playing Sesame Street songs in the next room, and off goes my kid to dance and listen. It's past her</p>

	bedtime, and she tricked us into getting Alexa to play songs for her.
Sampel 3	
id	3
review_helpful	7 of 8 people found the following review helpful
rating_value	4.4 out of 5 stars
review_detail	<p>The speech recognition has been actually pretty good out of the box, and can train. The API is easy enough that this computer scientist was able to have a new skill uploaded in about 30 minutes, and Pandora, iHeartRadio, and other integrated devices have been great. This has turned into our kitchen timer, our shopping list, and our music device.</p> <p>We went from having the TV on every night, to just listening to music for "background noise". The number of things that have skills published for it are also pretty awesome, especially if you're trying to get into a smart(er) home. Most of the startups and Internet of Things vendors are also all publishing skills for the Echo, so I expect this to grow into a much better.....</p>

6.2. Ekstraksi Data

Pada tahapan ini, dilakukan beberapa langkah untuk mengekstraksi data sehingga data bisa digunakan untuk pemrosesan data. Ekstraksi data terbagi menjadi tiga yaitu ekstraksi fitur struktural, ekstraksi fitur metada, dan ekstraksi fitur keterbacaan.

6.2.1. Data Ekstraksi Fitur Struktural

Ekstraksi fitur struktural menggunakan data dari atribut `review_detail` untuk dihitung total karakter, total kalimat, total paragraf, dan total silabus. Ekstraksi fitur dilakukan menggunakan pemrograman Java. Tabel 6.2 menunjukkan sampel data hasil ekstraksi data fitur struktural.

Tabel 6. 2 Hasil Ekstraksi Fitur Struktural

id	karakter	kata	kalimat
1	835	165	12
2	196	10	6
3	38	8	2
4	38	8	2
5	33	3	2
6	37	7	1
7	311	56	5
8	916	183	9
9	732	143	8
10	594	107	14
11	501	77	9
12	1031	195	9
13	430	77	6

6.2.2. Data Ekstraksi Fitur Metadata

Ekstraksi fitur metadata menggunakan data dari atribut `review_value` dan `review_helpful`. Data yang diambil adalah penilaian yang diberikan dan perbandingan kegunaan yang didapat. Ekstraksi ini dilakukan menggunakan pemrograman Java. Tabel 6.3 menunjukkan sampel data dari hasil ekstraksi fitur metadata.

Tabel 6. 3 Hasil Ekstraksi Fitur Metadata

id	helpful	Rating
1	0.5	5
2	0.1	5
3	0.083333333	5
4	0.083333333	5
5	0.076923077	5
6	0.076923077	5
7	0.076923077	5
8	0.076923077	4.4
9	0.076923077	5
10	0.076923077	5
11	0.076923077	4.4
12	0.066666667	5
13	0.066666667	5

6.2.3. Data Ekstraksi Fitur Keterbacaan

Ekstraksi fitur keterbacaan dilakukan menggunakan data hasil ekstraksi fitur struktural. Data tersebut akan dimasukkan kedalam rumus untuk dihitung uji keterbacaan FK dan ARI. Tabel 6.4 menunjukkan sampel data dari hasil ekstraksi fitur keterbacaan.

Tabel 6. 4 Hasil Ekstraksi Fitur Keterbacaan

id	fk	ari
1	6.364015152	9.280454545
2	39.34	71.71933333

3	0.72	2.9425
4	0.72	2.9425
5	28.26166667	31.13
6	2.311428571	6.965714286
7	6.267285714	10.32732143
8	9.556393443	12.31240437
9	8.792438811	11.61743007
10	5.476695594	8.538531375
11	10.5804329	13.49336219
12	9.924615385	14.30594872
13	7.957857143	11.28926407

6.3. Hasil *Weighted Sum*

Setelah melakukan ekstraksi, data tersebut digunakan untuk melakukan *weighted sum*. Berikut adalah hasil dari penilaian *weighted sum*

6.3.1. Penyesuaian Skala

Data yang sudah ada, pertama-tama dilakukan penyesuaian skala terlebih dahulu sehingga data tersebut dapat digunakan untuk penilaian. Tabel 6.6 menunjukkan sampel data yang belum dilakukan penyesuaian skala. Dapat dilihat bahwa nilai tiap atributnya berbeda. Lalu, di tabel 6.7 menunjukkan sampel data yang sudah dilakukan penyesuaian skala. Sehingga semua atribut rentangnya sama.

6.3.2. Penentuan Bobot

Setelah itu, bobot dihitung menggunakan rumus yang telah dijelaskan pada bab sebelumnya. Tabel 6.5 merupakan hasil dari perhitungan bobot masing-masing kriteria.

Tabel 6. 5 Hasil Penentuan Bobot

No.	Kriteria	Bobot
1	Rasio kegunaan	0,34
2	Hasil uji FK	0,21
3	Hasil uji ARI	0,15
4	<i>Star rating</i>	0,11
5	<i>Review Extremity</i>	0,08
6	Panjang ulasan	0,06
7	Jumlah kalimat	0,03
8	Jumlah kata	0,02
Total		1

6.3.3. Penilaian dan Pengurutan

Setelah dilakukan penyamaan skala, langkah selanjutnya adalah melakukan penilaian *weighted sum*. Penilaian dilakukan dengan memasukkan rumus di Ms. Excel. Hasil penilaian dapat dilihat di tabel 6.8. Setelah itu, baru dilakukan pengurutan dengan fungsi *sort* yang ada di Ms. Excel. Hasil pengurutan dapat dilihat di tabel 6.9.

Tabel 6. 6 Data Sebelum Penyamaan Skala

id	helpful	fk	ari	rating	extremity	karakter	kata	kalimat
1	0.5000	6.3640	9.2805	5	0.1366	835	165	12
2	0.1000	39.3400	71.7193	5	0.1366	196	10	6
3	0.0833	0.7200	2.9425	5	0.1366	38	8	2
4	0.0833	0.7200	2.9425	5	0.1366	38	8	2
5	0.0769	28.2617	31.1300	5	0.1366	33	3	2
6	0.0769	2.3114	6.9657	5	0.1366	37	7	1
7	0.0769	6.2673	10.3273	5	0.1366	311	56	5
8	0.0769	9.5564	12.3124	4.4	-0.4634	916	183	9
9	0.0769	8.7924	11.6174	5	0.1366	732	143	8
10	0.0769	5.4767	8.5385	5	0.1366	594	107	14
11	0.0769	10.5804	13.4934	4.4	-0.4634	501	77	9
12	0.0667	9.9246	14.3059	5	0.1366	1031	195	9
13	0.0667	7.9579	11.2893	5	0.1366	430	77	6

Tabel 6. 7 Data Sesudah Penyamaan Skala

id	Helpful	fk	ari	rating	extremity	karakter	kalimat	kata
1	0.5000	0.1168	0.1294	1	1	0.0875	0.0851	0.0881
2	0.1000	0.7223	1.0000	1	1	0.0205	0.0426	0.0175
3	0.0833	0.0132	0.0410	1	1	0.0040	0.0142	0.0038
4	0.0833	0.0132	0.0410	1	1	0.0040	0.0142	0.0038
5	0.0769	0.5189	0.4341	1	1	0.0035	0.0142	0.0042
6	0.0769	0.0424	0.0971	1	1	0.0039	0.0071	0.0034
7	0.0769	0.1151	0.1440	1	1	0.0326	0.0355	0.0315
8	0.0769	0.1754	0.1717	0.88	-3.3933	0.0960	0.0638	0.1014
9	0.0769	0.1614	0.1620	1	1	0.0767	0.0567	0.0802
10	0.0769	0.1005	0.1191	1	1	0.0623	0.0993	0.0623
11	0.0769	0.1943	0.1881	0.88	-3.3933	0.0525	0.0638	0.0566
12	0.0667	0.1822	0.1995	1	1	0.1081	0.0638	0.1071
13	0.0667	0.1461	0.1574	1	1	0.0451	0.0426	0.0460

Tabel 6. 8 Hasil Penilaian Menggunakan Weighted Sum

id	helpful	fk	ari	rating	extremity	karakter	kalimat	kata	penilaian
1	0.5000	0.1168	0.1294	1	1	0.0875	0.0851	0.0881	0.4135
2	0.1000	0.7223	1.0000	1	1	0.0205	0.0426	0.0175	0.5340
3	0.0833	0.0132	0.0410	1	1	0.0040	0.0142	0.0038	0.2280
4	0.0833	0.0132	0.0410	1	1	0.0040	0.0142	0.0038	0.2280
5	0.0769	0.5189	0.4341	1	1	0.0035	0.0142	0.0042	0.3942
6	0.0769	0.0424	0.0971	1	1	0.0039	0.0071	0.0034	0.2404
7	0.0769	0.1151	0.1440	1	1	0.0326	0.0355	0.0315	0.2661
8	0.0769	0.1754	0.1717	0.88	-3.393	0.0960	0.0638	0.1014	-0.0730
9	0.0769	0.1614	0.1620	1	1	0.0767	0.0567	0.0802	0.2827
10	0.0769	0.1005	0.1191	1	1	0.0623	0.0993	0.0623	0.2634
11	0.0769	0.1943	0.1881	0.88	-3.393	0.0525	0.0638	0.0566	-0.0695
12	0.0667	0.1822	0.1995	1	1	0.1081	0.0638	0.1071	0.2917
13	0.0667	0.1461	0.1574	1	1	0.0451	0.0426	0.0460	0.2725

Tabel 6. 9 Hasil Pengurutan Data

id	helpful	fk	ari	rating	extremity	karakter	kalimat	kata	penilaian	urutan
1	0.5000	0.1168	0.1294	1	1	0.0875	0.0851	0.0881	0.4135	1848
2	0.1000	0.7223	1.0000	1	1	0.0205	0.0426	0.0175	0.5340	1827
3	0.0833	0.0132	0.0410	1	1	0.0040	0.0142	0.0038	0.2280	2289
4	0.0833	0.0132	0.0410	1	1	0.0040	0.0142	0.0038	0.2280	2290
5	0.0769	0.5189	0.4341	1	1	0.0035	0.0142	0.0042	0.3942	1871
6	0.0769	0.0424	0.0971	1	1	0.0039	0.0071	0.0034	0.2404	2124
7	0.0769	0.1151	0.1440	1	1	0.0326	0.0355	0.0315	0.2661	1973
8	0.0769	0.1754	0.1717	0.88	-3.393	0.0960	0.0638	0.1014	-0.0730	2510
9	0.0769	0.1614	0.1620	1	1	0.0767	0.0567	0.0802	0.2827	1937
10	0.0769	0.1005	0.1191	1	1	0.0623	0.0993	0.0623	0.2634	1981
11	0.0769	0.1943	0.1881	0.88	-3.393	0.0525	0.0638	0.0566	-0.0695	2509
12	0.0667	0.1822	0.1995	1	1	0.1081	0.0638	0.1071	0.2917	1928
13	0.0667	0.1461	0.1574	1	1	0.0451	0.0426	0.0460	0.2725	1953

6.4. Hasil *Support Vector Machine*

Tahapan selanjutnya adalah melakukan klasifikasi menggunakan *support vector machine*. Tahapan ini terdiri dari beberapa langkah, berikut akan dibahas hasil yang didapat dari masing-masing langkah.

6.4.1. Data Normalisasi

Normalisasi dilakukan menggunakan metode z-score. Hasil normalisasi ditunjukkan di tabel 6.10. Hanya ada 7 atribut saja yang dinormalisasi yaitu atribut rating, extremity, fk, ari, karkater, kata, dan kalimat. Sedangkan atribut helpful tidak dilakukan normalisasi karena atribut tersebut akan dijadikan acuan untuk pembagian kelas klasifikasi.

Tabel 6. 10 Data Normalisasi

rating	extremity	fk	ari	karakter	kata	kalimat
0.5427	0.5427	-0.6068	-0.5284	-0.0336	0.0268	0.0546
0.5427	0.5427	10.6290	16.4744	-0.7522	-0.9373	-0.4630
0.5427	0.5427	-2.5299	-2.2543	-0.9298	-0.9497	-0.8081
0.5427	0.5427	-2.5299	-2.2543	-0.9298	-0.9497	-0.8081
0.5427	0.5427	6.8543	5.4215	-0.9355	-0.9808	-0.8081
0.5427	0.5427	-1.9876	-1.1587	-0.9310	-0.9560	-0.8944
0.5427	0.5427	-0.6398	-0.2433	-0.6228	-0.6512	-0.5493
-1.8417	-1.8417	0.4809	0.2973	0.0575	0.1388	-0.2042
0.5427	0.5427	0.2206	0.1080	-0.1494	-0.1100	-0.2905
0.5427	0.5427	-0.9092	-0.7304	-0.3046	-0.3340	0.2271
-1.8417	-1.8417	0.8298	0.6189	-0.4092	-0.5206	-0.2042
0.5427	0.5427	0.6064	0.8401	0.1868	0.2134	-0.2042

6.4.2. Data Konversi

Setelah dilakukan normalisasi, data selanjutnya dikonversi menjadi *sparse format* yang merupakan format data untuk SVM. Tabel 6.11 menunjukkan bentuk data yang telah diubah menjadi *sparse format*.

Tabel 6. 11 Data Konversi

2	1:0.542748735	2:0.542748735	3:-0.606825779
4:	-0.5283659	5:-0.0335889	6:0.026816224 7:0.054581977
2	1:0.542748735	2:0.542748735	3:10.6289626
4:	16.47440594	5:-0.752168329	6:-0.93730798
7:	-0.463019941		
2	1:0.542748735	2:0.542748735	3:-2.529890705
4:	-2.254258439	5:-0.929845245	6:-0.949748292
7:	-0.808087887		
2	1:0.542748735	2:0.542748735	3:-2.529890705
4:	-2.254258439	5:-0.929845245	6:-0.949748292
7:	-0.808087887		

Untuk atribut *helpful*, dikonversi menjadi nilai 1 untuk kelas yang *helpful* dan nilai 2 untuk kelas yang *unhelpful*. Nilai inilah yang menandakan data tersebut berada dalam kelas tertentu dan nilai ini terletak di urutan terdepan. Setelah itu, data selanjutnya merepresentasikan nilai atribut *rating*, atribut ini dianggap sebagai atribut pertama sehingga penulisannya menjadi 1:0.542748735, dimana 1 menandakan atribut pertama dan 0.542748735 adalah nilai atribut tersebut. Selanjutnya atribut *extremity* dianggap sebagai atribut kedua juga ditulis kedalam penulisan 2:0.542748735, dimana 2

menandakan atribut kedua dan 0.542748735 menandakan nilai atribut tersebut. Begitu seterusnya untuk atribut fk, ari, karakter, kata, dan kalimat.

6.4.3. *Training*

Langkah selanjutnya adalah melakukan *training* menggunakan *command prompt* dengan libSVM. *Training* dilakukan menggunakan metode svm-train dengan mencoba berbagai tipe svm dan tipe kernel. Berikut adalah *training* yang dilakukan untuk mendapatkan model svm:

6.4.3.1. Percobaan *Training* Pertama

Pada percobaan pertama, pilihan yang dipilih adalah tipe svm C-SVC dengan kernel linear. Gambar 6.1 menunjukkan proses dari percobaan *training* pertama.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-train -s 0
-t 0 review.train.data review.model1.data
*
optimization finished, #iter = 400
nu = 0.106680
obj = -214.000006, rho = 1.000067
nSV = 253, nBSV = 201
Total nSV = 253
```

Gambar 6. 1 Percobaan *Training* Pertama

Setelah melakukan *training*, model yang didapat akan disimpan ke dalam *file* review.model1.data. Adapun isi dari *file* tersebut ditunjukkan pada gambar 6.2.

```

svm_type c_svc
kernel_type linear
nr_class 2
total_sv 253
rho 1.00007
label 2 1
nr_sv 107 146
SV

```

Gambar 6. 2 Model Pertama

Berdasarkan gambar 6.1, proses *training* terjadi dengan iterasi sebanyak 400 kali. Nu merupakan parameter SVM, pada proses ini parameter nu sebesar 0,106680. Obj adalah nilai tujuan optimal dari permasalahan SVM, pada proses ini nilai obj sebesar -214,000006. Rho adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 1,000067. Lalu, nSV adalah banyaknya *support vector*, pada proses ini sebanyak 253 dan nBSV adalah banyaknya *bounded support vector*, pada proses ini sebanyak 201.

Sedangkan *output* model dari hasil *training* ini ditunjukkan pada gambar 6.2. Tipe kernel adalah linear, nr_class menggambarkan banyaknya kelas pada *training* ini yaitu 2, kelas *helpful* dan kelas *unhelpful*. Total_sv menyatakan banyaknya *support vector* yaitu sebanyak 253. Rho adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 1,00007. Lalu, proses ini mempunyai *support vector* untuk kelas 2 sebanyak 107 dan *support vector* untuk kelas 1 sebanyak 146.

6.4.3.2. Percobaan *Training* Kedua

Pada percobaan pertama, pilihan yang dipilih adalah tipe svm C-SVC dengan kernel polynomial. Gambar 6.3 menunjukkan proses dari percobaan *training* kedua.

Setelah melakukan *training*, model yang didapat akan disimpan ke dalam *file* `review.model2.data`. Adapun isi dari *file* tersebut ditunjukkan pada gambar 6.4.

Berdasarkan gambar 6.3, proses *training* terjadi dengan iterasi sebanyak 475304 kali. ν merupakan parameter SVM, pada proses ini parameter ν sebesar 0,105537. C adalah nilai tujuan optimal dari permasalahan SVM, pada proses ini nilai C sebesar -211,403914. ρ adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 1,002398. Lalu, nSV adalah banyaknya *support vector*, pada proses ini sebanyak 239 dan $nBSV$ adalah banyaknya *bounded support vector*, pada proses ini sebanyak 195.

Sedangkan *output* model dari hasil *training* ini ditunjukkan pada gambar 6.4. Tipe kernel adalah polynomial dengan parameter *degree* sebesar 3, *gamma* sebesar 0,142857, *coef0* sebesar 0, *nr_class* menggambarkan banyaknya kelas pada *training* ini yaitu 2, kelas *helpful* dan kelas *unhelpful*. *Total_sv* menyatakan banyaknya *support vector* yaitu sebanyak 239. ρ adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 1,0024. Lalu, proses ini mempunyai *support vector* untuk kelas 2 sebanyak 107 dan *support vector* untuk kelas 1 sebanyak 132.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-train -s 0
-t 1 review.train.data review.model2.data
.....
.....
.....
.....
.....
WARNING: using -h 0 may be faster
*.....
.....
WARNING: using -h 0 may be faster
*....
WARNING: using -h 0 may be faster
*
optimization finished, #iter = 475304
nu = 0.105537
obj = -211.403914, rho = 1.002398
nSV = 239, nBSV = 195
Total nSV = 239
```

Gambar 6.3 Percobaan *Training* Kedua

```
svm_type c_svc
kernel_type polynomial
degree 3
gamma 0.142857
coef0 0
nr_class 2
total_sv 239
rho 1.0024
label 2 1
nr_sv 107 132
SV
```

Gambar 6.4 Model Kedua

6.4.3.3. Percobaan *Training* Ketiga

Pada percobaan pertama, pilihan yang dipilih adalah tipe svm C-SVC dengan kernel *radial basis function*. Gambar 6.5 menunjukkan proses dari percobaan *training* ketiga.

```

C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-train -s 0
-t 2 review.train.data review.model3.data
*..
WARNING: using -h 0 may be faster
*
optimization finished, #iter = 2515
nu = 0.106680
obj = -213.241706, rho = 0.950187
nSV = 404, nBSV = 143
Total nSV = 404

```

Gambar 6. 5 Percobaan *Training* Ketiga

Setelah melakukan *training*, model yang didapat akan disimpan ke dalam *file* review.model3.data. Adapun isi dari *file* tersebut ditunjukkan pada gambar 6.6.

```

svm_type c_svc
kernel_type rbf
gamma 0.142857
nr_class 2
total_sv 404
rho 0.950187
label 2 1
nr_sv 107 297
SV

```

Gambar 6. 6 Model Ketiga

Berdasarkan gambar 6.5, proses *training* terjadi dengan iterasi sebanyak 2515 kali. Nu merupakan parameter SVM, pada proses ini parameter nu sebesar 0,106680. Obj adalah nilai tujuan optimal dari permasalahan SVM, pada proses ini nilai obj sebesar -213,241706. Rho adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 0,950187. Lalu, nSV adalah banyaknya *support vector*, pada proses ini sebanyak 404 dan nBSV adalah banyaknya *bounded support vector*, pada proses ini sebanyak 143.

Sedangkan *output* model dari hasil *training* ini ditunjukkan pada gambar 6.6. Tipe kernel adalah *radial basis function* (rbf) dengan parameter *gamma* sebesar 0,142857, *nr_class* menggambarkan banyaknya kelas pada *training* ini yaitu 2, kelas *helpful* dan kelas *unhelpful*. *Total_sv* menyatakan banyaknya *support vector* yaitu sebanyak 404. *Rho* adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 0,950187. Lalu, proses ini mempunyai *support vector* untuk kelas 2 sebanyak 107 dan *support vector* untuk kelas 1 sebanyak 297.

6.4.3.4. Percobaan *Training* Keempat

Pada percobaan pertama, pilihan yang dipilih adalah tipe svm C-SVC dengan kernel sigmoid. Gambar 6.7 menunjukkan proses dari percobaan *training* keempat.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-train -s 0
-t 3 review.train.data review.model4.data
*
optimization finished, #iter = 321
nu = 0.101695
obj = -1112.693830, rho = 16.878168
nSV = 205, nBSV = 202
Total nSV = 205
```

Gambar 6. 7 Percobaan *Training* Keempat

Setelah melakukan *training*, model yang didapat akan disimpan ke dalam *file* review.model4.data. Adapun isi dari *file* tersebut ditunjukkan pada gambar 6.8.


```

svm_type c_svc
kernel_type sigmoid
gamma 0.142857
coef0 0
nr_class 2
total_sv 205
rho 16.8782
label 2 1
nr_sv 102 103
SV

```

Gambar 6. 8 Model Keempat

Berdasarkan gambar 6.7, proses *training* terjadi dengan iterasi sebanyak 321 kali. Nu merupakan parameter SVM, pada proses ini parameter nu sebesar 0,101695. Obj adalah nilai tujuan optimal dari permasalahan SVM, pada proses ini nilai obj sebesar -1112,693830. Rho adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 16,878168. Lalu, nSV adalah banyaknya *support vector*, pada proses ini sebanyak 205 dan nBSV adalah banyaknya *bounded support vector*, pada proses ini sebanyak 202.

Sedangkan *output* model dari hasil *training* ini ditunjukkan pada gambar 6.8. Tipe kernel adalah sigmoid dengan parameter *gamma* sebesar 0,142857, *coef0* sebesar 0, *nr_class* menggambarkan banyaknya kelas pada *training* ini yaitu 2, kelas *helpful* dan kelas *unhelpful*. *Total_sv* menyatakan banyaknya *support vector* yaitu sebanyak 404. Rho adalah bias dari fungsi $\text{sgn}(w^T x - \rho)$ yaitu sebesar 16,8782. Lalu, proses ini mempunyai *support vector* untuk kelas 2 sebanyak 102 dan *support vector* untuk kelas 1 sebanyak 103.

6.4.4. *Testing*

Setelah dilakukan *training*, akan didapat data model yang digunakan untuk melakukan *testing*. *Testing* dilakukan dengan menggunakan metode svm-predict pada *command prompt*. Berikut adalah *testing* yang dilakukan berdasarkan hasil dari model *training* sebelumnya:

6.4.4.1. Percobaan *Testing* Pertama

Pada percobaan *testing* pertama, hasil keakuratan yang didapatkan sebesar 94.4773% atau model dapat memprediksi 479 data dengan benar dari total 507 data. Hasil ini ditunjukkan oleh gambar 6.5.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-predict re
view.test.data review.model1.data review.result1.data
Accuracy = 94.4773% (479/507) (classification)
```

Gambar 6. 9 Percobaan *Testing* Pertama

6.4.4.2. Percobaan *Testing* Kedua

Pada percobaan *testing* kedua, hasil keakuratan yang didapatkan sebesar 94.6746% atau model dapat memprediksi 480 data dengan benar dari total 507 data. Hasil ini ditunjukkan oleh gambar 6.6.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-predict re
view.test.data review.model2.data review.result2.data
Accuracy = 94.6746% (480/507) (classification)
```

Gambar 6. 10 Percobaan *Testing* Kedua

6.4.4.3. Percobaan *Testing* Ketiga

Pada percobaan *testing* pertama, hasil keakuratan yang didapatkan sebesar 94.4773% atau model dapat memprediksi 479 data dengan benar dari total 507 data. Hasil ini ditunjukkan oleh gambar 6.7.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-predict re
view.test.data review.model3.data review.result3.data
Accuracy = 94.4773% (479/507) (classification)
```

Gambar 6. 11 Percobaan *Testing* Ketiga

6.4.4.4. Percobaan *Testing* Keempat

Pada percobaan *testing* pertama, hasil keakuratan yang didapatkan sebesar 92.7022% atau model dapat memprediksi 470 data dengan benar dari total 507 data. Hasil ini ditunjukkan oleh gambar 6.8.

```
C:\Users\Rani\Documents\NetBeansProjects\svm\libsvm-3.21\windows>svm-predict re
view.test.data review.model4.data review.result4.data
Accuracy = 92.7022% (470/507) (classification)
```

Gambar 6. 12 Percobaan *Testing* Keempat

6.5. Analisis Hasil

Setelah didapatkan hasil penelitian, langkah selanjutnya adalah melakukan analisis dari hasil yang didapat. Analisis dilakukan menjadi dua tahapan yaitu analisis penilaian dan pengurutan menggunakan *weighted sum* dan analisis klasifikasi menggunakan *support vector machine*.

6.5.1. Hasil Penilaian dan Pengurutan *Weighted Sum*

Setelah dilakukan penilaian dan pengurutan ulasan dengan *weighted sum* dari 2513 data, penilaian yang paling tinggi sebesar 0,736 dengan id 491. Adapun tujuh ulasan dengan penilaian tertinggi ditunjukkan pada tabel 6.12.

Tabel 6. 12 Tujuh Ulasan dengan Penilaian Tertinggi

Urutan	1
Nilai	0.735597252
id	491
Helpful	1 of 1 people found the following review helpful
Rating	5.0 out of 5 stars

Detail	i have to say that i had my doubts about this product, and did a lot of research before purchase it, but i'm really impressed about it, it can recognize my voice practically from everywhere inside my home, i can even send commands from my first floor i have the echo on the third floor, of course when house is quiet, i'm pretty sure it will more development on this device.
Urutan	2
Nilai	0.728022245
id	384
Helpful	1 of 1 people found the following review helpful
Rating	5.0 out of 5 stars
Detail	I got this as a birthday gift to myself best thing I bought so far I recommend this to everyone I love this product Amazon has come up with a great way to listen to music and other things just by voice command I say thank you to the one who created this as for the set up of the echo was very easy and very easy to handle all the features as well
Urutan	3
Nilai	0.71695946
id	120
Helpful	7 of 15 people found the following review helpful
Rating	5.0 out of 5 stars
Detail	I love the device I was skeptical at first because of the price but it is so worth it i've seen reviews on those that have purchased the original dot who are upset with amazon because of the release of a 2nd generation just two months later after there purchase and they are prime users I find that disheartening myself but we do live in an unfair world full of selfishness and money hungry corporations so what do you expect some of the complaints are towards the dot not being able to understand you from far away but hey thats why you purchase the echo right or it would be unfair to us since we paid almost two hundred dollars sounds like someones trying to get a better

	device for a cheaper price
Urutan	4
Nilai	0.699465966
id	1746
Helpful	3 of 3 people found the following review helpful
Rating	5.0 out of 5 stars
Detail	I cant believe how cool this device is and I am in my 60s so I am not very up on devices that need additional steps but I was able to get the Echo connected and listen to beautiful music and the sound is amazing ! I will enjoy this each and every day from updates on weather and answers to questions I want to know I did not think I wanted or needed this Echo but Now I am over the moon with all the Echo can do I am listening to more music and turning off the TV which is a good thing and I am dancing by myself in the kitchen enjoying my Echo Thank you so much Amazon as I rock on!
Urutan	5
Nilai	0.69590019
id	2438
Helpful	1 of 1 people found the following review helpful
Rating	5.0 out of 5 stars
Detail	Best thing to have around the house if you are just one of the people that wish you were tony stark and have jarvis around the house well Alexa is just like that jarvis in a speaker so ya if you are just one of those people then this product is worth the price and worth having around the house
Urutan	6
Nilai	0.692456939
id	1523
Helpful	2 of 2 people found the following review helpful
Rating	5.0 out of 5 stars
Detail	Everyone in our family has played with it all weekend and had fun Great sound Amazing sound It's worth much more than I paid But as a shopping service

	Amazon is a genius I knew when I took it out of the box based on weight it would be high quality Fast Smart Fun She knows every song Very impressive
Urutan	7
Nilai	0.690648778
id	780
Helpful	1 of 1 people found the following review helpful
Rating	5.0 out of 5 stars
Detail	Love my Echo, all I can say is she was an easy setup, Alexa is smart and fun to chat with, my Grand kids love to communicate with her, it does connects with my Nest Thermostat and works great, she is a must have in every household, planning to purchase other gadgets that will connect to her

Sedangkan ulasan dengan nilai terendah memiliki penilaian sebesar -0.104 dengan id 2513. Adapun tujuh ulasan dengan penilaian terendah dapat ditunjukkan pada tabel 6.13.

Tabel 6. 13 Tujuh Ulasan dengan Penilaian Terendah

Urutan	2507
Nilai	-0.04315625
id	63
Helpful	1 of 5 people found the following review helpful
Rating	4.4 out of 5 stars
Detail	Cool device, still would recommendbut...Pretty good, has some glitches, dropped wi-fi a few times already, not all apps work as planned, can't answer many simple questions I thought it should be able, product still needs work but I do like it.
Urutan	2508
Nilai	-0.061084915
id	1659
Helpful	3 of 25 people found the following review helpful
Rating	4.4 out of 5 stars

Detail	What an idiotic device. Amazon makes it sound so simple...."just plug it in and launch the Amazon Echo App..."except that it crashed my network connection to my iPad twice. Would never let me connect to my Apple Extreme network...finally could only connect via my older Netgear router, and then it wouldn't let me register it. IN MY OPINION - THIS DEVICE IS NOT READY FOR PRIME TIME. I will return mine, and suggest you save yourself the trouble. UPDATE: after the initial setup difficulties - which the amazon echo folks had no explanation for - apparently there may be some conflict issues for people using an apple extreme router and/or multiple routers - i now love this device. The voice recognition is virtually flawless and quick - very natural. Even though the activities she can perform are limited - they are useful. And i find the sound quality to be excelent.
Urutan	2509
Nilai	-0.06948804
id	11
Helpful	1 of 13 people found the following review helpful
Rating	4.4 out of 5 stars
Detail	Worth the constant audio surveillance? Probably, if convenience is your preference. Easy to interconnect devices, apps, other online services. Occasionally comes in handy in unexpected ways, such as reading books that are not stored as audiobooks in your digital libraries, music discovery, and handsfree calculations on the fly. Can Echo serve as a personal assistant? Not yet. And does Echo automatically connect to smart TVs? Nope. So certainly some functions that could improve Echo in the future.
Urutan	2510
Nilai	-0.072968835
id	8
Helpful	1 of 13 people found the following review helpful
Rating	4.4 out of 5 stars
Detail	I love my Echo except for two things: I can't hook it

	<p>up to a more powerful speaker. Don't get me wrong, it has a good speaker but I want a superb speaker especially now that I have Amazon Music Unlimited. I can't understand why aux in and out ports weren't included with the Echo. The other complaint is that I can't link my Dots to my Echo so that when I move from room to room my music (or whatever I'm listening to) follows me. I imagine they did this so people in other rooms can select their own uses for the dots but it would be nice to have a link command. I would suggest, that if you are going to use Alexa to mainly provide basic services that you buy the Dots as they do have the ability to hook up to a Bluetooth speaker or system. I hope Amazon develops a way to link my Echo and Dots as that alone would solve my speaker wants. If anyone has a figured out a way to do that I hope you post it here.</p>
Urutan	2511
Nilai	-0.076422954
id	21
Helpful	1 of 10 people found the following review helpful
Rating	4.4 out of 5 stars
Detail	Great sound Alexa needs more options and skills already loaded but she is a blast. I replaced a sound system and timer in the kitchen. I liked washing dishes and getting the news update and picking literally any music there is.
Urutan	2512
Nilai	-0.095208887
id	41
Helpful	1 of 15 people found the following review helpful
Rating	4.4 out of 5 stars
Detail	You gotta love the echo. I have it integrated with my home automation software CQC (www.charmedquark.com). I can ask Alexa to turn on lights, fans, start my roomba, even control the TV. The echo can generally hear me from across the room without issue. The one thing I absolutely HATE is that

	Amazon's commercials tend to make my Echo start playing music. They really need to talk to their marketing department.... But I love the Echo itself. I will probably order a couple more.
Urutan	2513
Nilai	-0.104420057
id	39
Helpful	1 of 15 people found the following review helpful
Rating	4.4 out of 5 stars
Detail	I love my Amazon Echo. I use it all the time and also use it to freak out my friends. The only problem I have is that it MUST have an internet connection to work. If the internet is of, or out, it will not work to turn on lights or anything else that would only be on your LAN.

Jika dilihat dari sepuluh ulasan yang memiliki penilaian tertinggi, kesepuluh ulasan tersebut memiliki opini positif terhadap produk. Sedangkan sepuluh ulasan yang memiliki penilaian terendah, enam dari sepuluh ulasan memiliki opini negatif terhadap produk dan empat ulasan lainnya memiliki opini positif. Hal ini bukan menjadi tolak ukur apabila opini positif akan memiliki kualitas yang lebih bagus dibandingkan opini negatif. Namun hal ini dapat dijadikan salah satu faktor yang mempengaruhi kualitas suatu ulasan. Berdasarkan hasil yang diperoleh, opini positif mendapatkan penilaian yang lebih tinggi dari segi kualitas dibandingkan opini negatif. Selain itu, sepuluh ulasan yang memiliki penilaian tertinggi memiliki ulasan yang cukup ringkas dibandingkan sepuluh ulasan yang memiliki penilaian terendah.

Lalu apabila dilihat dari masing-masing karakteristik, karakteristik yang paling berpengaruh adalah karakteristik keterbacaan dengan atribut terkait yaitu *Flesch Kincaid* dan *Automated Readability Index*. Hal ini dapat dilihat dari

perbedaan hasil uji keterbacaan yang cukup jauh antara sepuluh ulasan teratas dan sepuluh ulasan terbawah. Rata-rata hasil perhitungan uji keterbacaan *Flesch Kincaid* untuk sepuluh ulasan teratas adalah sebesar 30,26 dan rata-rata hasil perhitungan uji keterbacaan *Automated Readability Index* adalah sebesar 39,1. Sedangkan rata-rata hasil perhitungan uji keterbacaan *Flesch Kincaid* untuk sepuluh ulasan terbawah adalah sebesar 8,36 dan rata-rata hasil perhitungan uji keterbacaan *Automated Readability Index* adalah sebesar 11,5. Semakin besar hasil uji keterbacaan, maka semakin besar pula nilai yang didapatkan.

Hasil uji keterbacaan *Flesch Kincaid* menggambarkan kerumitan suatu teks sedangkan uji keterbacaan *Automated Readability Index* menggambarkan kekomprehensifan bacaan. Sehingga sepuluh ulasan dengan nilai tertinggi tidak lebih sulit dan mudah dibaca dibandingkan sepuluh ulasan dengan nilai terendah.

Jika dilihat dari karakteristik metadata dengan atribut terkait yaitu *helpfulness* tidak begitu mempengaruhi kualitas ulasan. Dapat dilihat dari tidak ada perbedaan yang signifikan antara sepuluh ulasan dengan penilaian tertinggi dan sepuluh ulasan dengan penilaian terendah. Semua ulasan tersebut memiliki nilai yang cukup sama, sepuluh ulasan dengan penilaian tertinggi memiliki nilai *helpfulness* sebesar satu sedangkan enam dari sepuluh ulasan dengan penilaian terendah memiliki nilai *helpfulness* sebesar satu. Sedangkan untuk atribut *star rating* dan *review extremity*, sepuluh ulasan dengan nilai tertinggi memiliki nilai *star rating* dan *review extremity* yang paling tinggi yaitu lima dan 0,53. Sedangkan sepuluh ulasan dengan nilai terendah memiliki nilai *star rating* dan *review extremity* yang paling rendah yaitu 4,4 dan -1,84. Sehingga

dapat disimpulkan apabila suatu ulasan mempunyai nilai *review extremity* yang cukup rendah (seberapa besar perbedaan *star rating* yang diberikan pelanggan dengan *star rating* rata-rata ulasan produk tersebut), maka kemungkinan besar bahwa ulasan yang ditulis memiliki kualitas yang tidak lebih baik dibandingkan ulasan dengan *review extremity* yang mendekati angka 0.

Jika dilihat dari karakteristik struktural yang meliputi atribut jumlah karakter, kalimat, dan kata, terlihat bahwa semakin banyaknya jumlah karakter maka semakin tidak baik juga kualitas ulasan tersebut. Hal ini terlihat dari perbandingan rata-rata jumlah karakter, kalimat, dan kata yang lebih sedikit untuk sepuluh ulasan dengan nilai tertinggi dengan sepuluh ulasan dengan nilai terendah.

6.5.2. Hasil Klasifikasi *Support Vector Machine*

Hasil klasifikasi SVM yang dilakukan ditunjukkan oleh tabel 6.14. Berdasarkan hasil dari empat percobaan menggunakan tipe kernel yang berbeda, didapatkan bahwa tipe kernel polynomial menghasilkan keakuratan tertinggi. Sedangkan tipe kernel sigmoid menghasilkan keakuratan terendah. Namun, nilai keakuratan dari keempat uji coba tidak berbeda signifikan.

Tabel 6. 14 Hasil Klasifikasi SVM

Percobaan	Kernel	Keakuratan
1	Linear	94.4773%
2	Polynomial	94.6746%
3	Radial Basis Function	94.4773%
4	Sigmoid	92.7022%

Tipe kernel linear merupakan tipe kernel dasar yang paling sederhana diantara tipe kernel yang lain. Tipe linear membagi kelas dengan membuat garis lurus yang menjadi *hyperplane* sebagai pemisah antar kelas. Sehingga, tipe kernel ini tidak membutuhkan waktu yang banyak dalam prosesnya. Hasil keakuratan yang didapat pun cukup bagus yaitu sebesar 94.4773% atau berhasil memprediksi 479 data dengan benar dari 507 data yang ada.

Tipe kernel polynomial merupakan tipe kernel lanjutan yang lebih rumit dibandingkan kernel linear. Tipe kernel polynomial menggunakan persamaan nonlinear untuk memperkirakan posisi *hyperplane* sebagai pemisah antar kelas. Sehingga, tipe kernel ini memerlukan waktu paling banyak dalam prosesnya diantara tipe kernel lain. Namun hasil keakuratannya dapat diandalkan dan mendapatkan nilai tertinggi dibandingkan tipe kernel lainnya yaitu sebesar 94.6746% atau berhasil memprediksi 480 data dengan benar dari 507 data yang ada.

Tipe kernel *radial basis function* (RBF) merupakan tipe kernel yang paling populer dan banyak digunakan peneliti. Tipe kernel RBF memetakan sampel ke ruang dimensi yang lebih tinggi secara nonlinear, sehingga tipe ini dapat menangani kasus dimana label dan atribut kelasnya terkait secara nonlinear. Tipe kernel ini merupakan tipe *default* yang dipilih LibSVM. Hasil keakuratannya cukup bagus yaitu sebesar 94.4773% atau berhasil memprediksi 479 data dengan benar dari 507 data yang ada. Hasil ini sama dengan hasil yang didapatkan saat menggunakan tipe kernel linear.

Tipe kernel sigmoid merupakan adaptasi dari penggunaan sigmoid pada *neural networks* namun tipe kernel ini mempunyai beberapa properti yang masih belum dikaji lebih

dalam. Hasil keakuratannya merupakan yang paling rendah diantara kernel lainnya yaitu sebesar 92.7022% atau berhasil memprediksi 470 data dengan benar dari 507 data yang ada. Tipe kernel sigmoid ini tidak lebih baik dari tipe kernel RBF, sehingga tipe kernel ini kurang populer digunakan untuk kasus SVM.

Model ini hanya berlaku untuk satu produk dalam satu kategori saja yaitu produk “Amazon Echo Black” dalam kategori *home automation hubs & controllers*. Model ini belum dilakukan *testing* untuk ulasan produk dan kategori lainnya, sehingga bisa dikatakan bahwa model ini *overfitting* yaitu keadaan dimana model sangat sesuai dengan data namun tidak sesuai untuk data secara umum.

Halaman ini sengaja dikosongkan

BAB VII

KESIMPULAN DAN SARAN

Bab ini menjelaskan kesimpulan dari hasil penelitian beserta saran yang bermanfaat untuk penelitian selanjutnya.

7.1. Kesimpulan

1. Data yang digunakan pada penelitian ini adalah data ulasan produk “*Amazon Echo Black*” dari website Amazon.com sebanyak 2513 data. Dari sepuluh atribut yang diambil, dipilih tiga atribut yang digunakan untuk penelitian ini yaitu atribut *review_rating*, *review_helpful*, dan *review_detail*.
2. Pada tahapan ekstraksi fitur struktural dihasilkan data total kalimat, total kata, total karakter, dan total suku kata untuk masing-masing ulasan. Pada tahapan ekstraksi fitur metadata dihasilkan data rasio kegunaan, *rating*, dan *review extremity* untuk masing-masing ulasan. Dan pada tahapan ekstraksi fitur keterbacaan dihasilkan data uji keterbacaan *Flesch Kincaid* dan *Automated Readability Index* untuk masing-masing ulasan. Data yang sudah diekstrak, selanjutnya dilakukan normalisasi dan konversi untuk dapat digunakan di tahap selanjutnya.
3. Pembobotan masing-masing kriteria dilakukan menggunakan metode *rank order centroid*, metode ini memperhitungkan urutan kriteria mulai dari yang terpenting hingga yang kurang penting. Urutan kriterianya adalah dimulai dari rasio kegunaan, hasil uji *Flesch Kincaid*, hasil uji *Automated Readability Index*, *rating*, *extremity*, panjang ulasan, total kalimat dan total kata. Bobot tertinggi sebesar 0,34 dan bobot terendah sebesar 0,02 dari skala 1.

4. Berdasarkan perhitungan dengan *weighted sum*, diperoleh nilai kualitas ulasan tertinggi sebesar 0,736 dari skala 1 dan nilai kualitas ulasan terendah sebesar -0.104. Kategori yang paling mempengaruhi penilaian kualitas ulasan adalah nilai keterbacaan *automated readability index*, sedangkan kategori kegunaan tidak begitu mempengaruhi penilaian. Sehingga dapat disimpulkan bahwa ulasan dengan nilai kegunaan tinggi belum tentu memiliki kualitas ulasan yang bagus, sedangkan ulasan dengan nilai *automated readability index* yang tinggi kemungkinan besar ulasan tersebut memiliki kualitas yang bagus.
5. Berdasarkan hasil klasifikasi menggunakan *support vector machine* untuk memprediksi kegunaan suatu ulasan, diperoleh nilai keakuratan paling tinggi dengan menggunakan kernel polynomial dengan nilai sebesar 94.4773% atau berhasil memprediksi 479 data dengan benar dari 507 data yang ada. Hal ini disebabkan karena tipe kernel polynomial sesuai untuk masalah dimana semua data *training* telah dinormalisasi. Sehingga dapat disimpulkan bahwa tipe kernel polynomial paling cocok untuk memprediksi kegunaan suatu ulasan.

7.2. Saran

1. Menggunakan data yang lebih banyak dan bervariasi dari berbagai sumber pasar *online* agar hasil yang didapatkan lebih akurat dan lebih relevan sehingga tidak terjadi *overfitting*.
2. Menambahkan kategori lain sebagai faktor penentu kualitas ulasan pelanggan karena pada penelitian ini hanya memperhitungkan tiga kategori saja yaitu struktural, metadata, dan keterbacaan sedangkan masih ada faktor

lain penentu kualitas ulasan seperti kata kunci yang digunakan atau analisis sentimen. Sehingga semakin banyaknya faktor penentu kualitas ulasan pelanggan, maka hasilnya semakin dapat diandalkan untuk dijadikan acuan untuk menentukan kualitas ulasan bukan hanya dari nilai kegunaan yang didapatkan ulasan tersebut.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

- [1] J. P. Singh, S. Irani, N. P. Rana and dll, "Predicting the "helpfulness" of online consumer reviews," *Journal of Business Research*, 2016.
- [2] J. Rodak, M. Xiao and S. Longoria, "Predicting Helpfulness Ratings of Amazon Product Review," Stanford University, Stanford, 2014.
- [3] H.-. Y. Hsieh and S.-H. Wu, "Ranking Online Customer Reviews with the SVR Model," *IEEE 16th International Conference on Information Reuse and Integration*, 2015.
- [4] N. Korfiatis, E. Garcia-Bariocanal and S. Sanchez-Alonso, "Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content," *Electronic Commerce Research and Applications*, vol. 11, pp. 205-217, 2012.
- [5] S. M. Mudambi and D. Schuff, "What makes a helpful online review? A study of customer reviews on Amazon.com," *MIS Quarterly*, vol. 34, pp. 185-200, 2010.
- [6] B. Li, E. Ch'ng, A. Y.-L. Chong and H. Bao, "Predicting online e-marketplace sales performances: A big data approach," *Computers & Industrial Engineering*, 2016.
- [7] K. Arline, "What Is E-Commerce?," 26 Februari 2015. [Online]. Available: <http://www.businessnewsdaily.com/4872-what-is-e->

- commerce.html. [Accessed 28 April 2016].
- [8] Ezako, "Online Marketplace vs E-Commerce," February 2015. [Online]. Available: <http://ezako.com/en/onlime-marketplace-vs-e-commerce/>. [Accessed 26 January 2017].
- [9] Martin, "Top 5 Famous Online Marketplaces," CLEVERISM, 15 April 2016. [Online]. Available: <https://www.cleverism.com/top-5-famous-online-marketplaces/>. [Accessed 26 January 2017].
- [10] J. D'Onfro, "10 Quirky THings You Didn't Know About Amazon," 10 May 2014. [Online]. Available: <http://www.businessinsider.co.id/amazon-jeff-bezos-facts-story-history-2014-5/?r=US&IR=T#gjuKRKU8w8VQKzci.97>. [Accessed 02 January 2017].
- [11] B. Wire, "Social Shopping Study Reveals Changes in Customers' Online Shopping Habits and Usage of Cutomer Reviews," 03 May 2010. [Online]. Available: <http://www.businesswire.com/news/home/20100503005110/en/2010-Social-Shopping-Study-Reveals-Consumers%E2%80%99-Online>. [Accessed 02 January 2017].
- [12] D. Riyaldi, "Rancang Bangun Rest Web Service untuk Perbandingan Harga Pengiriman dengan Metode Web Scraping dan Pemanfaatan API," *Sekolah Tinggi Manajemen Informatika dan Komputer Amikom Yogyakarta*, 2013.
- [13] C. Advisor, "Web Site Scraper The Most Effective Tool for Web Data Extraction," 2015. [Online]. Available:

<http://www.thecomputeradvisor.net/web-site-scraper-the-most-effective-tool-for-web-data-extraction/>.
[Accessed 02 January 2017].

- [14] D. Kalyanmoy, "Multi-Objective Optimization Using Evolutionary Algorithms," Wiley, 2001.
- [15] E. Roszkowska, "Rank Ordering Criteria Weighting Methods - A Comparative Overview," *Optimum. Studia Ekonomiczne*, vol. 5, p. 65, 2013.
- [16] P. Flach, "Machine Learning: The Art and Science of Algorithms that Make Sense of Data," Cambridge University Press, 2012.
- [17] BrightLocal, "Local Consumer Review Survey," September 2016. [Online]. Available: <https://www.brightlocal.com/learn/local-consumer-review-survey/>. [Accessed 2 January 2017].
- [18] C. Forman, A. Ghose and B. Wiesenfeld, "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, vol. 19, no. 3, pp. 291-313, 2008.
- [19] E. Clemons, G. Gao and L. Hitt, "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, vol. 23, no. 2, pp. 149-171, 2006.
- [20] P. Chen, S. Dhanasobhon and M. Smith, "All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews on Sales on Amazon.com," Carnegie Mellon

University, 2008.

- [21] N. Kumar and I. Benbasat, "The Influence of Recommendations on Customer Reviews on Evaluations of Websites," *Information System Research*, vol. 17, no. 4, pp. 425-439, 2006.

BIODATA PENULIS



Penulis lahir di Jakarta pada tanggal 15 Oktober 1995. Merupakan anak pertama dari 2 bersaudara. Penulis telah menempuh pendidikan formal di SD Negeri Kenari 07 Jakarta, SMP Negeri 49 Jakarta, dan SMA Negeri 8 Jakarta.

Pada tahun 2013, penulis melanjutkan pendidikan di Jurusan Sistem Informasi – Institut Teknologi Sepuluh Nopember melalui jalur SBMPTN (Tulis) dengan NRP 5213100122. Selama kuliah, penulis aktif mengikuti kegiatan akademik dan non akademik. Penulis aktif menjadi asisten dosen untuk mata kuliah desain basis data, perancangan sumber daya perusahaan, dan riset operasi. Penulis juga aktif berpartisipasi dalam kegiatan organisasi dengan menjadi ketua divisi eksternal departemen hubungan luar BEM FTIf periode 2015/2016 dan menjadi kakak pendamping dalam kaderisasi HMSI periode 2014/2015. Selain itu, penulis juga berpartisipasi dalam kegiatan sosial dengan menjadi pengajar di program ITS Mengajar for Indonesia di Pulau Mandangin tahun 2014 dan menjadi volunteer dalam acara Young Environmental Leadership Program Kementerian Sosial Masyarakat BEM ITS periode 2014/2015.

Pada tahun keempat semester 8, penulis memilih lab Sistem Enterprise sebagai fokus tugas akhir. Di semester 8 juga, penulis mengikuti pertukaran pelajar di Bialystok University of Technology Polandia dengan beasiswa Erasmus+ sebagai mahasiswa Fakultas Manajemen. Penulis menyelesaikan tugas akhir selama mengikuti program pertukaran pelajar ini.

LAMPIRAN A

Hasil Penilaian dan Perhitungan dengan *Weighted Sum*

Tabel A. 1 Data Sebelum Dinormalisasi

rank	id	helpful	rating	fk	ari	karakter	kata	kalimat
1	491	1	5	30.54788732	38.88042254	374	71	1
2	384	1	5	29.37540541	37.59243243	346	74	1
3	120	1	5	54.46796992	69.89488722	701	133	1
4	1746	1	5	24.71333333	31.86309524	583	126	2
5	1747	1	5	24.71333333	31.86309524	583	126	2
6	2438	1	5	24.52666667	31.649	294	60	1
7	1523	1	5	23.498	31.9242	301	50	1
8	780	1	5	23.52857143	31.04517857	291	56	1
9	2307	0.981481481	5	7.744475754	11.27277414	9100	1635	126
10	706	1	5	20.44226415	26.66490566	243	53	1
11	919	0.927823692	5	8.044101844	12.02714296	9541	1622	141

12	920	0.927823692	5	8.044101844	12.02714296	9541	1622	141
13	1226	1	5	19.82097701	25.27758621	461	87	2
14	954	0.928571429	5	9.892886792	13.81309293	8858	1475	106
15	1440	1	5	19.21096995	24.11991803	881	183	4
16	2273	1	5	18.49352941	23.71455882	1219	204	6
17	154	1	5	18.14796748	25.07073171	679	123	3
18	1309	1	5	19.29162791	24.49627907	223	43	1
19	722	1	5	19.05914894	24.31723404	222	47	1
20	1003	1	5	6.396386893	10.22867954	6710	1151	137
21	1150	1	5	18.24	23.37544928	626	115	3
22	1151	1	5	18.24	23.37544928	626	115	3
23	1930	1	5	18.19866197	22.65619718	397	71	2
24	932	1	5	10.85961918	14.35750198	5015	865	51
25	1266	1	5	16.07150407	20.74429539	1668	328	9
26	2074	0.980769231	5	12.21927681	15.7436847	4579	802	39
27	669	1	5	16.12828426	21.55700508	974	197	5

Tabel A. 2 Data Setelah Normalisasi Hasil Penilaian dan Perhitungan *Weighted Sum*

rank	id	helpful	fk	ari	rating	extremity	karakter	kalimat	kata	penilaian
1	491	1	0.561	0.542	1	1	0.039	0.007	0.042	0.735597
2	384	1	0.539	0.524	1	1	0.036	0.007	0.038	0.728022
3	120	0.4667	1.000	0.975	1	1	0.073	0.007	0.078	0.716959
4	1746	1	0.454	0.444	1	1	0.061	0.014	0.064	0.699466
5	1747	1	0.454	0.444	1	1	0.061	0.014	0.064	0.699466
6	2438	1	0.450	0.441	1	1	0.031	0.007	0.032	0.695900
7	1523	1	0.431	0.445	1	1	0.032	0.007	0.032	0.692457
8	780	1	0.432	0.433	1	1	0.030	0.007	0.031	0.690649
9	2307	0.9815	0.142	0.157	1	1	0.954	0.894	0.962	0.674537
10	706	1	0.375	0.372	1	1	0.025	0.007	0.026	0.668833
11	919	0.9278	0.148	0.168	1	1	1	1	1	0.665757
12	920	0.9278	0.148	0.168	1	1	1	1	1	0.665757
13	1226	1	0.364	0.352	1	1	0.048	0.014	0.052	0.665316

14	954	0.9286	0.182	0.193	1	1	0.928	0.752	0.953	0.664149
15	1440	1	0.353	0.336	1	1	0.092	0.028	0.100	0.664073
16	2273	1	0.340	0.331	1	1	0.128	0.043	0.137	0.663359
17	154	1	0.333	0.350	1	1	0.071	0.021	0.070	0.660050
18	1309	1	0.354	0.342	1	1	0.023	0.007	0.025	0.659562
19	722	1	0.350	0.339	1	1	0.023	0.007	0.025	0.658254
20	1003	1	0.117	0.143	1	1	0.703	0.972	0.693	0.658107
21	1150	1	0.335	0.326	1	1	0.066	0.021	0.070	0.656507
22	1151	1	0.335	0.326	1	1	0.066	0.021	0.070	0.656507
23	1930	1	0.334	0.316	1	1	0.042	0.014	0.046	0.652896
24	932	1	0.199	0.200	1	1	0.526	0.362	0.552	0.652193
25	1266	1	0.295	0.289	1	1	0.175	0.064	0.184	0.651517
26	2074	0.9808	0.224	0.220	1	1	0.480	0.277	0.511	0.648
27	669	1	0.296	0.301	1	1	0.102	0.035	0.104	0.647

LAMPIRAN B

Data Hasil Penilaian dan Pengurutan dengan Weighted Sum

Tabel B. 1 Data *Review* Setelah Dinilai dan Diurutkan

rank	id	helpfulness	rating	review
1	491	1 of 1 people found the following review helpful	5.0 out of 5 stars	i have to say that i had my doubts about this product, and did a lot of research before purchase it, but i'm really impressed about it, it can recognize my voice practically from everywhere inside my home, i can even send commands from my first floor i have the echo on the third floor, of course when house is quiet, i'm pretty sure it will more development on this device.
2	384	1 of 1 people found the following review helpful	5.0 out of 5 stars	I got this as a birthday gift to myself best thing I bought so far I recommend this to everyone I love this product Amazon has come up with a great way to listen to music and other things just by voice command I say thank you to the one who created this as for the set up of the echo was very easy and very easy to handle all the features as well

3	120	7 of 15 people found the following review helpful	5.0 out of 5 stars	I love the device I was skeptical at first because of the price but it is so worth it i've seen reviews on those that have purchased the original dot who are upset with amazon because of the release of a 2nd generation just two months later after there purchase and they are prime users I find that disheartening myself but we do live in an unfair world full of selfishness and money hungry corporations so what do you expect some of the complaints are towards the dot not being able to understand you from far away but hey thats why you purchase the echo right or it would be unfair to us since we paid almost two hundred dollars sounds like someones trying to get a better device for a cheaper price
4	1746	3 of 3 people found the following review helpful	5.0 out of 5 stars	I cant believe how cool this device is and I am in my 60s so I am not very up on devices that need additional steps but I was able to get the Echo connected and listen to beautiful music and the sound is amazing ! I will enjoy this each and every day from updates on weather and answers to questions I want to know I did not think I wanted or needed this Echo

				but Now I am over the moon with all the Echo can do I am listening to more music and turning off the TV which is a good thing and I am dancing by myself in the kitchen enjoying my Echo Thank you so much Amazon as I rock on!
5	1747	3 of 3 people found the following review helpful	5.0 out of 5 stars	I cant believe how cool this device is and I am in my 60s so I am not very up on devices that need additional steps but I was able to get the Echo connected and listen to beautiful music and the sound is amazing ! I will enjoy this each and every day from updates on weather and answers to questions I want to know I did not think I wanted or needed this Echo but Now I am over the moon with all the Echo can do I am listening to more music and turning off the TV which is a good thing and I am dancing by myself in the kitchen enjoying my Echo Thank you so much Amazon as I rock on!
6	2438	1 of 1 people found the following	5.0 out of 5 stars	Best thing to have around the house if you are just one of the people that wish you were tony stark and have jarvis around the house well Alexa is just like that jarvis in a speaker so ya if you are just one of

		review helpful		those people then this product is worth the price and worth having around the house
7	1523	2 of 2 people found the following review helpful	5.0 out of 5 stars	Everyone in our family has played with it all weekend and had fun Great sound Amazing sound It's worth much more than I paid But as a shopping service Amazon is a genius I knew when I took it out of the box based on weight it would be high quality Fast Smart Fun She knows every song Very impressive
8	780	1 of 1 people found the following review helpful	5.0 out of 5 stars	Love my Echo, all I can say is she was an easy setup, Alexa is smart and fun to chat with, my Grand kids love to communicate with her, it does connects with my Nest Thermostat and works great, she is a must have in every household, planning to purchase other gadgets that will connect to her
9	2307	5 of 6 people found the following review helpful	5.0 out of 5 stars	The Echo and Alexa not only are a quantum step forward in home automation, but she (yes its that good I'll call her she) has utterly and completely exploded the fun scale. I'm a late adopter of Echo; (obviously with over 3000 Echo reviews already posted). So I'll walk through a few ways Alexa's

			<p>grown up in the past few months and how I enjoy her every day. Firstly; I use the Wink app (not the hub, just the free app). I would use the hub, but I don't feel like I need one (yet). The SmartThings hub from Samsung seems to have more product support, but you can't use the app alone and it get's about 1.75 stars on iTunes. So the free Wink app is the winner. It's important to add that the Wink app is phenomenally simple, yet fairly robust, which makes it the perfect sister (or brother) to Alexa. I also use IFTTT. The simplest method (simple being relative I know) is to create an IFTTT that directs Alexa to a Wink Shortcut. Although this takes a few steps to set up, it makes it soooooo easy afterwards to work well with Alexa. All you say is, "Alexa, trigger X". X being things like, "Arm system", "Set Away", "Movienight" etc. You can customize the phrases pretty much however you want. I like, "Alexa, trigger Self Destruct". She doesn't actually explode, fyi. So with that being said, here's some ways Alexa makes life better. * Nest: Nest control isn't currently native</p>
--	--	--	---

				<p>with Alexa. There are two work arounds 1) The aforementioned use of IFTTT to trigger a Wink or using Reflect Connect (free skill in Alexa); then telling Alexa, "Talk to Reflect". Each serves a different purpose and you'll probably use both. With the IFTTT method, you can direct a specific set of pre set actions. I use this to "Trigger Away" which then set's the Nest to Away, turns on my drop cam and my Canary, turn off Hue lights etc. Also use it to "Trigger Home" which does the opposite. Reflect Connect, primes Alexa through a third party to interact more directly with the Nest. So you can say things like, "Set Nest to 70, Turn up Nest 2 degrees", "Run Fan for 15 minutes" etc. So while it's a little more clunky than you may prefer, once set up it works very well. Ecobee integrates more tightly with Alexa, but you still have a few workarounds for things like setting it to heat and cool mode at the same time. I'm sure in short time there will be a direct control from Alexa to Nest but for now, Google has it's vice grip on the system, so the workarounds are</p>
--	--	--	--	--

			<p>the way to go. All in all, I love this setup.</p> <p>*Dropcam/Nest Cam: Being a now Google owned product, you have some of the same challenges as above and I answer them the same way. I link the Dropcam to Nest and use IFTTT/Wink triggers to set away, set home, create a schedule and turn on my hue lights if it activates and I'm not at home. By packaging commands with Wink, I use one phrase like "Arm System" to activate all the components. *</p> <p>Canary: I've been testing a Canary, which I'll review separately. In short, I really like it; given what it is and how it's designed to work. IE it's not a full home security system, but it is a good solution and has major advantages over Nest cam (mainly you're aren't limited to alerts only once in 30 minutes). Similar to what I described above, when I say, "Alexa arm system" or "disarm system"; the Canary is set to Armed/Disarmed mode. Or if I say, "Alexa, trigger Away" she turns my Nest to Away, shuts off the lights, arms Canary and turns on the Dropcam. Arming and Disarming are about the only Echo</p>
--	--	--	---

			<p>features so far. Canary hasn't opened up the "trigger alarm" ability, except from the app. That would be a great emergency feature, "Alexa, alarm", but I haven't found a way to make that work. * Hue Lights (and other lights like Lix). Hue integration is very simple and seamless. Once you set up your Hue system, you just have Alexa "find" the devices. They all show up in the Alexa app and you're good to go. You can say, "Alexa turn on living room light" or group them together and she will. You can say, "Alexa, dim living room light 50%" and she will. Program color changes etc and Alexa will execute. Super useful and easy. I love the convenience before bed, in the morning etc. * Traffic: Alexa gives you a traffic report, but only on a preset route, i.e.; you have to have put the destination into the app. I was considering moving back towards the beach, which would increase my commute significantly. I set up the destination in the app and ask Alexa a few times throughout the morning, "Alexa, how's traffic?". She tells me the estimated time and route. I've been</p>
--	--	--	--

			<p>logging this to see if the new commute changes at different times and if it's worth the move. So far it is definitely not. Alexa can't move time and space, yet.</p> <p>* Music: I use Alexa mostly to play Spotify or Pandora. "Alexa, play Pandora, Of Monsters and Men". She's off and running. Of course you can ask, what is this song, who is this artist etc and she'll provide you an accurate fast answer. She's compatible with tons' of music service and the ones I've used work flawlessly. If you're an Amazon Prime member, and let's face it, you should be really unless you live under a rock, but probably mostly then; you just say, "Alexa, play Band of Horses" and she almost immediately fires off the tunes. Super easy and efficient. Sound quality is very good. It's not Sonos quality and you can't stream to multiple rooms (even with multiple echo's they will act independent) but it's very good. Better than most typical bluetooth speakers. Speaking of Sonos, I think you can use a Smartthings Hub and link Alexa to turn on your Sonos, but I'm not sure how effective it is, nor what</p>
--	--	--	---

			<p>music would play (you're last station maybe?). But if you're interested it's worth exploring further. That would be a big win to me, as I'd rather use Alexa to run my Sonos, vs opening another app. * Time, Alarms and Weather etc: this works like you'd expect and similar to Siri or Cortana etc. It's easy, fast and reliable. I use these type of commands almost daily and no need for me to rehash things like to do lists, alarms or timers. * Family Friendly: My wife is an anti tech person mildly convinced that eventually all the electronics will cause brain cancer and gluten will most likely kill us. In less than a week, she's now using Alexa to check the time, check the weather and stream music every morning. I've caught her asking Alexa random questions to test her. Although she initially approached Alexa baring her teeth like a protective mother wolf, she's now realized Alexa's not a threat and accepted her warmly into the pack. * A billion other things: There are so many things, questions, commands etc that you can give Alexa. This is one of my favorite sites and is really helpful if</p>
--	--	--	--

			<p>you are new to Alexa or considering her; [...]: search "The Ultimate Guide to Mastering Amazon Echo" they have over 300 easter eggs that are worth exploring and show Alexa's versatility. -Downsides: are there any? Yep, some, but none that make her less than worthy. As described above, to get full function you still need to maximize a combo of IFTTT, Wink and third party skills. But, they are simple to set up and then make for a flawless experience. Think of it, like setting up your TV and Surround sound so you can watch Game of Thrones (of The Voice if that's your thing) in immersive auditory bliss. It's like that. -Also, she isn't integrated with everything yet. Like Nest, Canary, Door locks, etc there are closed software platforms that she can't control. She's only so smart, there are many things you'll ask and get, "I'm sorry, I didn't understand that question". However, she's getting smarter every day and you'll be more surprised at the things she does know vs her limitations. -Portable; she has to be plugged in; oh wait, Amazon Tap you say? Yep, now there's a</p>
--	--	--	---

			<p>portable, recharging version if you need to move it around and free yourself from a cable. Big caveat though; she isn't always listening, you have to tap a button first, hence the name Tap because "button" sounds too cute. I'd still opt for plugged and always on, that's the magic of Alexa, throwing out random questions or commands, but too each their own. The new Echo dot looks like a great way to expand your system into other rooms and use your own speakers...more to follow. -Speaker quality: It's not as good as it could be. The money is in the technology and her speed of response, not in the speaker. So to keep price manageable, it was either a really smart product, or a really good speaker. It's smart and fairly good speaker. Don't expect it replace your Sonos Play 3. -Connectivity and syncing: It's too bad you can't use these to replace your Sonos. Put a few around the house and have them all play the same music? Maybe in near future; its completely software updatable, but right now...no you can't. I'll update as I find fun new uses for Alexa, but I'm already head over heels. This</p>
--	--	--	--

				is a fantastic new product that moves past gimmick and becomes essentially integrated to your life style. [...]. Thanks for reading.
10	706	1 of 1 people found the following review helpful	5.0 out of 5 stars	We love this thing I looked at these since they came out and hemmed and hawed about buying one well I finally bit the bullet and bought one I liked it so much I was taking it from room to room So I bought a Echo Dot and that works just as good

LAMPIRAN C

Sparse Format Data untuk SVM

Tabel C. 1 Sparse Format Data

2 1:0.542748735 2:0.542748735 3:-0.606825779 4:-0.5283659 5:-0.0335889 6:0.026816224 7:0.054581977
2 1:0.542748735 2:0.542748735 3:10.6289626 4:16.47440594 5:-0.752168329 6:-0.93730798 7:- 0.463019941
2 1:0.542748735 2:0.542748735 3:-2.529890705 4:-2.254258439 5:-0.929845245 6:-0.949748292 7:-0.808087887
2 1:0.542748735 2:0.542748735 3:-2.529890705 4:-2.254258439 5:-0.929845245 6:-0.949748292 7:-0.808087887
2 1:0.542748735 2:0.542748735 3:6.854282043 4:5.421498084 5:-0.935467932 6:-0.980849073 7:-0.808087887
2 1:0.542748735 2:0.542748735 3:-1.987648988 4:-1.158694331 5:-0.930969783 6:-0.955968448 7:-0.894354873
2 1:0.542748735 2:0.542748735 3:-0.639784051 4:-0.243292896 5:-0.622846522 6:-0.651180797 7:-0.549286928
2 1:-1.841740026 2:-1.841740026 3:0.480901762 4:0.297266336 5:0.057498633 6:0.138779035

7:-0.204218982
2 1:0.542748735 2:0.542748735 3:0.220602256 4:0.10801743 5:-0.149416257 6:-0.110027211 7:-0.290485968
2 1:0.542748735 2:0.542748735 3:-0.909158972 4:-0.730399483 5:-0.304602425 6:-0.333952833 7:0.22711595
2 1:-1.841740026 2:-1.841740026 3:0.829819033 4:0.618853731 5:-0.409184407 6:-0.520557517 7:-0.204218982
2 1:0.542748735 2:0.542748735 3:0.606364696 4:0.840129698 5:0.18682044 6:0.213420909 7:-0.204218982
2 1:0.542748735 2:0.542748735 3:-0.063761746 4:0.018654333 5:-0.489026566 6:-0.520557517 7:-0.463019941
2 1:0.542748735 2:0.542748735 3:-0.071690077 4:-0.004091443 5:-0.5531252 6:-0.576538923 7:-0.549286928
2 1:0.542748735 2:0.542748735 3:-1.016593517 4:-1.009869486 5:-0.209016742 6:-0.178448929 7:0.22711595