



TUGAS AKHIR- KM184801

**IMPLEMENTASI ANALISIS SENTIMEN DALAM
KLASIFIKASI ARTIKEL BERITA DARING
MENGUNAKAN ALGORITMA HYBRID
CONVOLUTIONAL NEURAL NETWORK -
HIDDEN MARKOV MODELS (HYBRID CNN-
HMM)**

**ATHYAH DANNI SURYA GAMA
NRP 061115 4000 0061**

**Dosen Pembimbing :
Dr. Imam Mukhlash, S.Si., M.T.
Dr. Darmaji, S.Si., M.T.**

**DEPARTEMEN MATEMATIKA
Fakultas Matematika Komputasi dan Sains Data
Institut Teknologi Sepuluh Nopember
Surabaya 2019**



FINAL PROJECT- KM184801

**SENTIMENT ANALYSIS IMPLEMENTATION
FOR ONLINE NEWS ARTICLE
CLASSIFICATION USING HYBRID
CONVOLUTIONAL NEURAL NETWORK -
HIDDEN MARKOV MODELS (HYBRID CNN-
HMM)**

**ATHYAH DANNI SURYA GAMA
NRP 061115 4000 0061**

**Supervisors :
Dr. Imam Mukhlash, S.Si., M.T.
Dr. Darmaji, S.Si., M.T.**

**DEPARTEMENT OF MATHEMATICS
Faculty of Mathematics, Computation, and Data Science
Institut Teknologi Sepuluh Nopember
Surabaya 2019**

LEMBAR PENGESAHAN

**IMPLEMENTASI ANALISIS SENTIMEN DALAM KLASIFIKASI
ARTIKEL BERITA DARING MENGGUNAKAN ALGORITMA
HYBRID CONVOLUTIONAL NEURAL NETWORK – HIDDEN
MARKOV MODELS (HYBRID CNN-HMM)**

***SENTIMENT ANALYSIS IMPLEMENTATION FOR ONLINE NEWS
ARTICLE CLASSIFICATION USING HYBRID CONVOLUTIONAL
NEURAL NETWORK – HIDDEN MARKOV MODELS (HYBRID CNN-
HMM)***

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Matematika
Pada Bidang Studi Ilmu Komputer
Program Studi S-1 Departemen Matematika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember Surabaya

Oleh :

ATHYAH DANNI SURYA GAMA
NRP. 0611154000061

Menyetujui,

Dosen Pembimbing II,

Dosen Pembimbing I,

Dr. Darmaji, S.Si., M.T.

NIP. 19691015 199412 1 001

Dr. Imam Mukhlash, S.Si., M.T.

NIP. 19700831 199403 1 003

Mengetahui,
Kepala Departemen Matematika
FMKSD ITS

Dr. Imam Mukhlash, S.Si., M.T.

NIP. 19700831 199403 1 003

Surabaya, Juli 2019

**IMPLEMENTASI ANALISIS SENTIMEN DALAM
KLASIFIKASI ARTIKEL BERITA DARING MENGGUNAKAN
ALGORITMA HYBRID CONVOLUTIONAL NEURAL
NETWORK – HIDDEN MARKOV MODELS (HYBRID CNN-
HMM)**

Nama : Athyah Danni Surya Gama
NRP : 061115 4000 0061
Departemen : Matematika
Dosen Pembimbing : 1. Dr. Imam Mukhlash, S.Si., M.T.
2. Dr. Darmaji, S.Si., M.T.

ABSTRAK

Era Big Data dan Internet of Things telah merambah hampir pada semua bidang, di antaranya informasi komunikasi, pendidikan, industri bahkan dalam bidang keuangan. Seperti yang telah diketahui, bidang informasi dan komunikasi merupakan bidang yang menjadi sumber paling berpengaruh pada bidang keuangan. Bidang ini akan sangat dipengaruhi oleh informasi yang muncul dan berkembang di masyarakat. Informasi negatif maupun positif yang muncul melalui berita akan mempengaruhi saham. Untuk mengetahui hal tersebut, diperlukan sebuah metode yang mampu melakukan hal tersebut. Salah satu metode yang dapat digunakan adalah analisis sentimen. Pada penelitian ini akan dikembangkan algoritma untuk mencari sentimen/opini yang berkembang melalui pemberitaan di media berita *online* / daring (dalam jaringan). Untuk mendapatkan berita dari portal berita tersebut, terlebih dahulu akan dilakukan pengambilan data menggunakan metode web crawling. Kemudian untuk melakukan pengenalan sentimen/opini akan digunakan algoritma Hybrid Convolutional Neural Network-Hidden Markov Models (Hybrid CNN-HMM). Hasil dari metode tersebut menunjukkan performansi pembelajaran 99.92% dan performansi pengujian sekitar 70,45 %.

Kata kunci: web crawling, sentiment analysis, Hybrid CNN-HMM, data mining, ekonomi dan keuangan.

**SENTIMENT ANALYSIS IMPLEMENTATION FOR
ONLINE NEWS ARTICLE CLASSIFICATION USING
HYBRID CONVOLUTIONAL NEURAL NETWORK –
HIDDEN MARKOV MODELS (HYBRID CNN-HMM)**

Name : Athyah Danni Surya Gama
NRP : 061115 4000 0061
Department : Mathematics
Supervisor : 1. Dr. Imam Mukhlash, S.Si., M.T.
2. Dr. Darmaji, S.Si., M.T.

ABSTRACT

The Big Data era and the Internet of Things have penetrated almost all fields, including information on communication, education, industry and even in the financial sector. As is well known, the field of information and communication is a field that has become the most influential source of finance. This field will be greatly influenced by information that arises and develops in the community. Negative and positive information that appears through the news will affect shares. To find out this, a method is needed to do this. One method that can be used is sentiment analysis. In this study, algorithms will be developed to find sentiments/opinions that develop through reporting in online news media. To get news from the total news, first, it will take data retrieval using the web crawling method. Then to identify sentiments/opinions will be used Hybrid Convolutional Neural Network algorithm - Hidden Markov Models (Hybrid CNN-HMM). The results of the method show a learning performance of around 99.92% and a testing performance of around 70.45%.

Keywords: *web crawling, sentiment analysis, CNN-HMM Hybrid, data mining, economics and finance.*

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillahilahihirobil'aalamiin, segala puji dan syukur bagi Allah SWT yang memiliki apa yang ada di langit dan di bumi dan yang telah memberikan limpahan rahmat, petunjuk serta hidayahNya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul

“IMPLEMENTASI ANALISIS SENTIMEN DALAM KLASIFIKASI ARTIKEL BERITA DARING MENGGUNAKAN ALGORITMA HYBRID CONVOLUTIONAL NEURAL NETWORK – HIDDEN MARKOV MODELS (HYBRID CNN-HMM)”

sebagai salah satu syarat kelulusan Program Sarjana Departemen Matematika FMKSD Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

Tugas Akhir ini dapat terselesaikan dengan baik berkat bantuan dari berbagai pihak. Suatu kebahagiaan dan kewajiban bagi penulis untuk menyampaikan terima kasih kepada semua pihak yang telah membantu dan mendukung atas terselesainya Tugas Akhir:

1. Kepala Departemen, Sekretaris Departemen, Ketua Program Studi Sarjana, dan Sekretaris Program Studi Sarjana Matematika FMKSD ITS.
2. Bapak Dr. Imam Mukhlash, S.Si., M.T. dan Dr. Darmaji, S.Si, M.T. selaku dosen pembimbing Tugas Akhir.
3. Bapak dan Ibu dosen penguji atas semua saran dan masukan yang diberikan.
4. Bapak Dr. Mahmud Yunus, M.Si dan Bapak Drs. Suhud Wahyudi, M.Si selaku dosen wali penulis.
5. Bapak dan Ibu dosen serta para staf Departemen Matematika FMKSD ITS.

6. Kedua orang tua penulis, Bapak Anwar Daramo dan Ibu jujuk Eka Ria Dewi serta keluarga atas dukungan dan semangat yang telah diberikan
7. Capt.Aldo, Agung komting, Nida nadi, Inayah inay, nenek AyuN, Riko senpai, svmihar, li',Ciko cik, Ijoel, Bapak kus, dan WJ, selaku orang yang berjasa selama pengerjaan.
8. DOHMAIn, sahabat - sahabat, dan teman - teman penulis yang telah memberikan bantuan dan motivasi.
9. De Kepeed, PENDOSA, Alan Turing, Mubadzir, yang telah menjadi GAP penulis.
10. Cindy Chelia yang telah memberikan motivasi, dukungan, omelan dan semangat.
11. Toilet, Lab.Ilkom, dan Jalanan Manukan-ITS yang telah menjadi tempat mencari inspirasi selama pengerjaan.
12. Semua pihak yang belum disebutkan yang telah membantu dalam penyusunan Tugas Akhir ini.

Penulis menyadari bahwa laporan Tugas Akhir ini masih jauh dari sempurna, sehingga penulis mengharapkan kritik dan saran yang sangat diharapkan dari semua pihak demi kesempurnaan laporan Tugas Akhir ini. Semoga laporan Tugas Akhir ini dapat bermanfaat bagi penulis dan bagi pembaca pada umumnya.

Surabaya, 26 Juli 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	v
ABSTRAK	vii
<i>ABSTRACT</i>	ix
KATA PENGANTAR	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xix
BAB I	1
1. 1 Latar Belakang	1
1. 2 Rumusan Masalah	4
1. 3 Batasan Masalah	4
1. 4 Tujuan	4
1. 5 Manfaat	5
1. 6 Sistematika Penulisan Tugas Akhir	5
BAB II	7
2. 1 Penelitian Terdahulu	7
2. 2 Landasan Teori	7
2.2.1 <i>Web Crawling</i>	7
2.2.2 <i>Text Mining</i>	8
2.2.3 <i>Sentiment Analysis</i>	8
2.2.4 <i>Word2vec</i>	9
2.2.5 <i>Convolutional Neural Network</i>	9
2.2.6 <i>Hidden Markov Model</i>	14
2.2.7 <i>Hybrid CNN-HMM</i>	16
2.2.8 Evaluasi Sistem	17
BAB III	19
3. 1 Studi literatur	19
3. 2 Pengumpulan Data	19
3. 3 Praproses Data	19
3. 4 Ekstraksi Fitur	20
3. 5 Implementasi Algoritma Hybrid CNN-HMM	21
3. 6 Analisa Hasil dan Penarikan Simpulan	21
3. 7 Validasi Sistem	21
3. 8 Penyusunan Laporan	22

3. 9	Diagram Alir Metodologi Penelitian	22
BAB IV	25
4. 1	Analisis Implementasi Sistem	25
4.1.1	<i>Use Case Diagram</i>	25
4.1.2	<i>Activity Diagram</i>	26
4.1.3	Analisis Kebutuhan Fungsional.....	28
4.1.4	Analisis Kebutuhan Non-Fungsional.....	28
4. 2	Perancangan Sistem.....	29
4.2.1	Perancangan Data	29
4.2.2	Perancangan Proses	31
4.2.3	Perancangan Antarmuka.....	52
4. 3	Penggunaan Library.....	53
4. 4	Implementasi Sistem	53
4.4.1	Implementasi Pra-proses Data	53
4.4.2	Implementasi Algoritma Hybrid CNN-HMM ..	62
4.4.3	Implementasi Proses <i>Web Crawling</i>	65
BAB V	69
5. 1	Uji Coba Algoritma Hybrid CNN-HMM	69
5.1.1	Data Uji Coba Algoritma.....	69
5.1.2	Hasil Uji Coba Algoritma.....	70
5. 2	Uji Coba Metode <i>Web Crawling</i>	77
5.2.1	Data Uji Coba Metode <i>Web Crawling</i>	77
5.2.2	Hasil Uji Coba Metode <i>Web Crawling</i>	79
5. 3	Hasil Uji Coba Prediksi <i>Sentiment Analysis</i>	83
BAB VI	84
6. 1	Simpulan.....	85
6. 2	Saran.....	85
DAFTAR PUSTAKA	87

DAFTAR GAMBAR

Gambar 2. 1. Arsitektur MLP Sederhana	9
Gambar 2. 2. Arsitektur model dengan dua channel untuk sebuah kalimat	10
Gambar 2. 3. Jaringan dengan 3 convolutional layer	11
Gambar 2. 4. Contoh proses konvolusi contoh kalimat.....	11
Gambar 2. 5. Contoh penggunaan HMM pada klasifikasi teks	16
Gambar 2. 6. Contoh penggunaan Hybrid CNN-HMM pada image processing.....	17
Gambar 3. 1. Diagram Alir penelitian (1)	22
Gambar 3. 2. Diagram alir penelitian (2)	23
Gambar 4. 1. <i>Usecase diagram</i> sistem	26
Gambar 4. 2. <i>Activity diagram</i> sistem	27
Gambar 4. 3. Data monitoring harian berita.....	32
Gambar 4. 4. Flowchart proses persiapan data masukan.....	33
Gambar 4. 5. Flowchart proses casefolding	35
Gambar 4. 6. Flowchart proses tokenisasi.....	36
Gambar 4. 7. Flowchart proses filterisasi	38
Gambar 4. 8. Flowchart proses stemming	39
Gambar 4. 9. Flowchart proses ekstraksi fitur.....	40
Gambar 4. 10. Arsitektur jaringan Hybrid CNN-HMM.....	42
Gambar 4. 11. Contoh mekanisme pada convolutional layer dengan 2 window dimana terdapat 2 filter tiap windownya	45
Gambar 4. 12. Ilustrasi mekanisme pada pooling layer	46
Gambar 4. 13. Contoh gambaran jaringan pada output layer...	47
Gambar 4. 14. Flowchart proses pemilihan situs web.....	48
Gambar 4. 15. Situs web kompas	48
Gambar 4. 16. Situs web bisnis	48
Gambar 4. 17. Flowchart proses penentuan tanggal pencarian	49
Gambar 4. 18. Kemungkinan pemisahan tanggal inputan untuk masing-masing situs web	49

Gambar 4. 19. URL situs web bisnis setelah diberi tanggal pencarian	49
Gambar 4. 20. URL situs web Kompas setelah diberi tanggal pencarian	49
Gambar 4. 21. Flowchart proses halaman induk	50
Gambar 4. 22. Dokumen web situs web berita online.....	50
Gambar 4. 23. Flowchart proses pengunjung dan akuisisi data	51
Gambar 4. 24. URL dari halaman induk situs berita online yang akan dikunjungi	51
Gambar 4. 25. Contoh halaman yang telah dikunjungi dan data yang akan diambil pada halaman tersebut.....	52
Gambar 4. 26. Rancangan desain antarmuka program.....	52
Gambar 4. 27. Data mentah dari Bank Indonesia.....	54
Gambar 4. 28. Data yang telah dilakukan pemilihan komponen serta perubahan format	54
Gambar 4. 29. Data yang akan dimasukkan ke dalam program	55
Gambar 4. 30. Tanda pemisah data antar kolom pada file CSV	55
Gambar 4. 31. Tampilan data yang telah dimasukkan dalam program.....	56
Gambar 4. 32. Contoh data label sebelum diubah.....	61
Gambar 4. 33. Contoh data label setelah diubah	62
Gambar 4. 34. Contoh hasil akuisisi data	67
Gambar 5. 1. Grafik pergerakan akurasi model pada percobaan pertama.....	71
Gambar 5. 2. Grafik pergerakan akurasi model pada percobaan kedua.....	72
Gambar 5. 3. Grafik pergerakan akurasi model pada percobaan ketiga.....	73
Gambar 5. 4. Grafik pergerakan akurasi model pada percobaan keempat.....	74
Gambar 5. 5. Grafik pergerakan akurasi model pada percobaan kelima.....	75
Gambar 5. 6. Hasil <i>web crawling</i> Kompas.com.....	81

Gambar 5. 7. Hasil <i>web crawling</i> bisnis.com.....	81
Gambar 5. 8. Halaman artikel berita yang gagal dilakukan <i>web crawling</i>	82
Gambar 5. 9. Halaman artikel berita yang gagal dilakukan <i>web crawling</i>	83
Gambar 5. 10. Hasil prediksi <i>sentiment</i> dari tiap artikel berita	83

DAFTAR TABEL

Tabel 4. 1. Lingkungan hardware dan software	28
Tabel 4. 2. Contoh artikel berita dengan tone positif, netral, dan negatif.....	33
Tabel 4. 3. Contoh data inputan	34
Tabel 4. 4. Tabel input dan output proses Case folding	36
Tabel 4. 5. Tabel input dan output proses tokenisasi	37
Tabel 4. 6. Tabel input dan output proses filterisasi.....	37
Tabel 4. 7. Tabel input dan output proses stemming	38
Tabel 4. 8. Hasil dari ekstraksi fitur	41
Tabel 4. 9. Arsitektur jaringan CNN	44
Tabel 5. 1. Rincian data.....	69
Tabel 5. 2. Data yang digunakan dalam proses komputasi	70
Tabel 5. 3. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan pertama).....	72
Tabel 5. 4. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan kedua)	73
Tabel 5. 5. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan ketiga).....	74
Tabel 5. 6. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan keempat)	75
Tabel 5. 7. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan kelima).....	76
Tabel 5. 8. Nilai <i>precision</i> dan <i>recall</i> pada tiap percobaan.....	76
Tabel 5. 9. Rincian jumlah data artikel yang tersedia selama bulan Juni	78
Tabel 5. 10. Rincian jumlah data artikel yang terkumpul selama bulan Juni	79

BAB I PENDAHULUAN

Pada bab ini menjelaskan gambaran umum dari penulisan Tugas Akhir yang terdiri atas latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.

1.1 Latar Belakang

Era *Big Data* dan *Internet of Things* telah merambah hampir pada semua bidang, di antaranya informasi komunikasi, pendidikan, industri bahkan dalam bidang keuangan. Menurut Qing Li, bidang informasi (dalam hal ini artikel berita) dapat memberikan dampak pada bidang ekonomi (dalam hal ini *return* saham) [1]. Bidang ini akan sangat dipengaruhi oleh informasi yang muncul dan berkembang di masyarakat. Sebagai contoh, informasi negatif maupun positif yang muncul melalui berita akan mempengaruhi nilai saham. Oleh karena itu, untuk mendapatkan informasi yang muncul di masyarakat, khususnya melalui pemberitaan maka diperlukan seperangkat metode (algoritma/aplikasi) yang mampu melakukan analisis secara cepat dan akurat. Salah satu bidang ilmu yang dapat melakukan hal tersebut adalah *data mining*. *Data mining* adalah proses yang menggunakan teknik matematika dan statistik untuk mendapatkan informasi dari *database* yang besar. Algoritma-algoritma *data mining* mampu menemukan pola yang baru, *valid* dan menarik dari sejumlah besar data sumber.

Perkembangan teknologi telah mengubah pola penyampaian berita, penyampaian berita yang dahulu menggunakan media cetak telah berkembang dengan adanya internet. Kini penyampaian berita kepada masyarakat luas dapat diberikan melalui portal-portal berita *online* / daring (dalam jaringan) tersedia di internet. Adanya hal tersebut memberikan kemudahan pada pembaca untuk dapat mendapatkan berita dengan cepat dan gratis. Dengan kemudahan yang telah tersedia

ini, diharapkan penyampaian berita dapat segera diketahui oleh masyarakat. Salah satu topik yang berpengaruh terhadap masyarakat secara luas adalah topik tentang ekonomi dan keuangan. Hal ini dikarenakan ekonomi dan keuangan menjadi sebuah kebutuhan primer bagi masyarakat secara luas. Dalam sebuah berita pasti memiliki kecenderungan sebagai berita positif ataupun berita negatif terhadap suatu hal. Namun, juga tidak menutup kemungkinan bahwa dalam suatu berita tidak memiliki kecenderungan positif ataupun negatif dengan kata lain netral. Untuk mampu mengetahui kecenderungan yang terdapat dalam sebuah artikel berita daring, maka dibutuhkan sebuah teknologi untuk mengumpulkan artikel-artikel berita tersebut dan melakukan pengolahan terhadapnya. Untuk melakukan pengumpulan data, telah ada sebuah cara dengan mengambil data dari beberapa website yang menjadi sumber data. Proses pengambilan informasi-informasi dari beberapa *website* disebut dengan *web crawling* [2]. Dengan adanya hal tersebut tentunya akan dapat membantu melakukan suatu pencarian data dalam beberapa *website* untuk mendapatkan sebuah informasi didalamnya. Hal inilah yang akan digunakan untuk mengumpulkan artikel-artikel berita sebelum nantinya akan dilakukan proses lebih lanjut.

Menurut Kamus Besar Bahasa Indonesia, berita adalah cerita atau keterangan mengenai kejadian atau peristiwa yang hangat . Untuk mengetahui kecenderungan yang terkandung di dalamnya , telah dikenal sebuah proses yaitu Analisis sentiment (*sentiment analysis*). *Sentiment analysis* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis agar didapatkan informasi sentimen yang terkandung di dalam sebuah kalimat opini [3]. *Sentiment analysis* dilakukan untuk melihat pendapat ataupun kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah orang tersebut cenderung memiliki pandangan yang bersifat positif atau negatif [3]. Proses inilah yang akan digunakan untuk mengetahui apakah dalam sebuah artikel berita sebenarnya memiliki unsur positif, negatif, atau netral didalamnya. Qing Li dkk menjelaskan

bagaimana pengaruh berita dan *mood* dari masyarakat terhadap keputusan investor [1]. Hal serupa juga telah dikemukakan oleh Thien Hai Nguyen dkk dan M.S. Checkley yang meneliti bagaimana analisis sentimen berpengaruh pada pergerakan saham [4] [5]. Ditambah dengan penelitian yang dilakukan oleh Kelly dan Khursid yang kali ini mencari pengaruh analisis sentimen pada sebuah asset finansial [6]. *Sentiment analysis* sendiri bisa dianggap sebagai kombinasi dari *text mining* dan *natural language processing(NLP)*. *Text mining* merupakan cabang ilmu dari *data mining* yang berfokus pada penggalian informasi pada data berbentuk teks, sedangkan NLP menurut James Putstějovský merupakan penerapan ilmu komputer untuk mengkaji interaksi antara komputer dengan bahasa manusia. Salah satu metode dari *text mining* yang bisa digunakan untuk menyelesaikan masalah *sentiment analysis* adalah algoritma *Convolutional Neural Network* (CNN). Pada umumnya CNN digunakan pada pengolahan citra digital sebagai alat untuk klasifikasi maupun kluster. Dengan inovasi dari Kim pada tahun 2014 yaitu penerapan model CNN pada NLP khususnya dalam klasifikasi kalimat [7]. Kim mengusulkan konsep baru dalam penggunaan CNN pada pengolahan teks yang menunjukkan bahwa CNN merupakan metode yang unggul dalam pengolahan teks. Terbukti dengan adanya penelitian tugas akhir yang dilakukan oleh M. Fakhur Rozi yang menggunakan metode CNN-L2-SVM untuk analisis sentimen pada ulasan buku serta penelitian yang dilakukan oleh Imam Mukhlash dkk yaitu analisis sentimen pada ulasan buku dengan menggunakan metode CNN-LSTM [8] [9]. Selain CNN, terdapat pula metode *Hidden Markov Models* (HMM) yang juga mampu melakukan klasifikasi dalam pengolahan teks seperti halnya pada penelitian yang dilakukan oleh Mangi Kang dkk [10]. Mengacu pada hal tersebut, maka penulis akan melakukan kombinasi antara metode CNN dan HMM dengan merujuk jurnal yang ditulis oleh Qiang Guo dkk tentang penggunaan metode Hybrid CNN-HMM untuk *scene number recognition* [11]. Maka dari itu dalam tugas akhir ini metode yang digunakan oleh penulis adalah Hybrid CNN-

HMM. Diharapkan dengan adanya kombinasi metode tersebut serta mengaplikasikannya pada klasifikasi teks, dapat memberikan hasil dengan akurasi yang baik sehingga akan dapat memudahkan apabila ingin dilakukan pengembangan dari hasil yang didapatkan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disajikan sebelumnya, masalah yang akan dibahas dalam penelitian Tugas akhir ini adalah sebagai berikut:

1. Bagaimana menggunakan metode Hybrid CNN-HMM untuk melakukan *sentiment analysis* pada artikel berita
2. Bagaimana menggunakan metode *web crawling* untuk mendapatkan informasi dari portal berita daring
3. Bagaimana hasil *Sentiment analysis* dari artikel berita daring yang didapatkan dengan metode Hybrid CNN-HMM

1.3 Batasan Masalah

Dalam tugas akhir ini penulis membatasi masalah sebagai berikut:

1. Portal berita yang digunakan adalah Kompas dan Bisnis dengan topik berita mengenai ekonomi dan keuangan
2. Proses *web crawling* akan dilakukan pada berita bulan Juni 2019
3. *Web Crawling* dan *Sentiment analysis* akan menggunakan Bahasa pemrograman Python
4. Algoritma yang akan digunakan adalah Hybrid CNN-HMM

1.4 Tujuan

Tujuan dari penelitian tugas akhir ini adalah sebagai berikut:

1. Mengetahui implementasi metode Hybrid CNN-HMM dalam melakukan *sentiment analysis* pada artikel berita

2. Mengetahui implementasi metode *web crawling* dalam mendapatkan informasi dari portal berita daring
3. Mengetahui hasil *sentiment analysis* dari artikel berita daring menggunakan metode Hybrid CNN-HMM

1.5 Manfaat

Manfaat dari tugas akhir ini adalah sebagai berikut:

1. Menerapkan teori-teori matematika dalam bidang ekonomi dan keuangan
2. Memberikan informasi mengenai *sentiment analysis* artikel berita bidang ekonomi dan keuangan
3. Sebagai literatur penunjang bagi penelitian selanjutnya

1.6 Sistematika Penulisan Tugas Akhir

Sistematika penulisan dalam tugas akhir ini adalah sebagai berikut:

1. BAB I : PENDAHULUAN

Pada bab ini menjelaskan gambaran umum dari penulisan Tugas Akhir yang terdiri atas latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.

2. BAB II : TINJAUAN PUSTAKA

Pada bab ini dijelaskan mengenai penelitian terdahulu serta beberapa teori dasar yang mendukung dalam pengerjaan Tugas Akhir ini.

3. BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tahap-tahap pengerjaan dalam menyelesaikan Tugas Akhir ini sehingga penelitian dapat dirancang secara sistematis dan diatur dengan sebaik-baiknya.

4. BAB IV : PERANCANGAN DAN IMPLEMENTASI

Pada bab ini menjelaskan tentang model dan desain dari sistem yang akan dibentuk. Hal-hal tersebut meliputi

visualisasi data, tahap pra pengolahan data, transformasi data dengan *Word2vec*, pembuatan model Hybrid CNN-HMM sebagai acuan dalam mengimplementasikan sistem.

5. **BAB V : UJI COBA DAN EVALUASI SISTEM**

Pada bab ini membahas tentang pengujian sistem yang telah terimplementasi dengan melakukan proses verifikasi dan validasi beserta pengujian kinerja dari sistem yang telah dibuat.

6. **BAB VI : SIMPULAN DAN SARAN**

Pada bab ini berisi Simpulan dari Tugas Akhir yang diperoleh dari bab uji coba dan evaluasi serta saran untuk pengembangan penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

Pada bab ini dijelaskan mengenai penelitian terdahulu serta beberapa teori dasar yang mendukung dalam pengerjaan Tugas Akhir ini.

2.1 Penelitian Terdahulu

Dalam hal domain permasalahan, penulis merujuk pada 4 makalah yaitu yang pertama jurnal oleh Qing Li dkk tentang pengaruh berita dan *mood* masyarakat terhadap keputusan investor [1]. Kemudian jurnal dari Thien Hai Nguyen dkk dan M.S. Checkley yang meneliti bagaimana analisis sentimen berpengaruh pada pergerakan saham [4] [5]. Serta yang terakhir oleh Kelly dan Khursid mengenai dampak dari berita terhadap asset finansial [6].

Kemudian dalam hal metode/algorithm pada penelitian ini, penulis merujuk pada jurnal yang ditulis oleh Qiang Guo dkk yang berjudul “Convolutional feature learning and Hybrid CNN-HMM for scene number recognition” yang menerapkan algoritma Hybrid CNN-HMM pada *image processing* [11]. Terdapat penelitian Tugas Akhir yang dilakukan oleh Muhammad Fakhru Rozi (2017) yang berjudul ‘Opinion Mining On Book Review Using CNN-L2-SVM Algorithm’ [8]. Terdapat pula jurnal yang disusun oleh Imam Mukhlash dkk dengan judul “Opinion Mining on Book Review using CNN-LSTM” [9]. Serta jurnal yang dikarang oleh Mangi kang dkk dengan judul “Opinion mining using ensemble text Hidden Markov Models for text classification” [10].

2.2 Landasan Teori

2.2.1 Web Crawling

Web crawling adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman *web* dalam Bahasa *markup* seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data

tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain [2].

2.2.2 Text Mining

Text mining, pada proses mengambil informasi dari teks. Informasi biasanya diperoleh melalui peramalan pola dan kecenderungan pembelajaran pola statistik. Text mining yaitu parsing, bersama dengan penambahan beberapa fitur linguistik turunan dan penghilangan beberapa diantaranya, dan penyisipan subsequent ke dalam database, menentukan poladalam data terstruktur, dan akhirnya mengevaluasi dan menginterpretasi output, text mining biasanya mengacu ke beberapa kombinasi relevansi, kebaruan, dan interestingness. Proses text mining yang khas meliputi kategorisasi teks, text clustering, ekstraksi konsep/entitas, produksi taksonomi granular, sentiment analysis, penyimpulan dokumen, dan pemodelan relasi entitas yaitu, pembelajaran hubungan antara entitas [12].

2.2.3 Sentiment Analysis

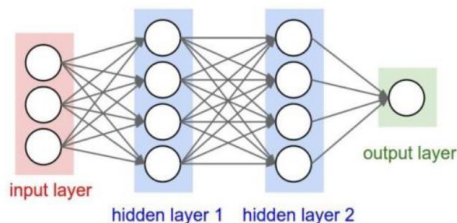
Sentiment analysis didefinisikan sebagai tugas menemukan opini dari penulis tentang entitas tertentu. Menurut Tang dalam Haddi, *sentiment analysis* pada *review* adalah proses menyelidiki *review* produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan [13]. Menurut Thelwall dalam Haddi, analisa sentimen diperlakukan sebagai suatu tugas klasifikasi yang mengklasifikasikan orientasi suatu teks ke dalam positif atau negatif. Menurut Mejova dalam Basari, tujuan dari *sentiment analysis* adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan (bagaimana penulis ingin mempengaruhi pembaca) [14] [15].

2.2.4 *Word2vec*

Word2vec merupakan cara untuk membuat suatu bentuk representasi kata dalam suatu ruang vektor. Vektor tersebut dapat membantu algoritma pembelajaran untuk mencapai performansi yang lebih baik dalam NLP dengan mengelompokkan kata-kata yang mirip atau sama. Ada beberapa macam cara dalam membuat vektor kata.

2.2.5 *Convolutional Neural Network*

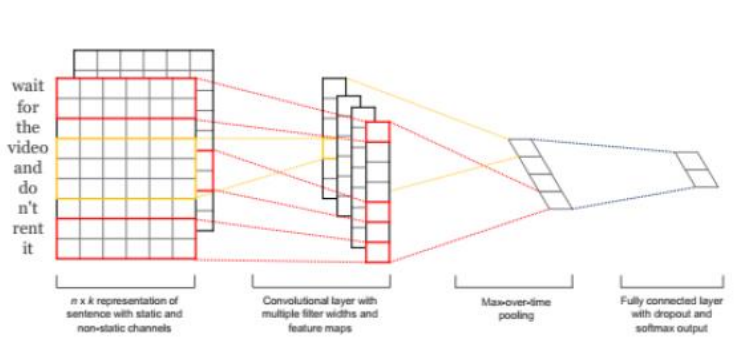
Convolutional Neural Network (CNN) adalah pengembangan dari *Multilayer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. CNN termasuk dalam jenis *Deep Neural Network* karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra. Cara kerja CNN memiliki kesamaan pada MLP, namun dalam CNN setiap neuron dipresentasikan dalam bentuk dua dimensi, tidak seperti MLP yang setiap neuron hanya berukuran satu dimensi.



Gambar 2. 1. Arsitektur MLP Sederhana

Sebuah MLP seperti pada Gambar 2.1. memiliki i layer (kotak merah dan biru) dengan masing-masing layer berisi ji neuron (lingkaran putih). MLP menerima input data satu dimensi dan mempropagasikan data tersebut pada jaringan hingga menghasilkan output. Setiap hubungan antar *neuron* pada dua layer yang bersebelahan memiliki parameter bobot satu dimensi yang menentukan kualitas mode. Disetiap data input pada layer dilakukan operasi linear dengan nilai bobot yang ada, kemudian

hasil komputasi akan ditransformasi menggunakan operasi nonlinear yang disebut sebagai fungsi aktivasi.

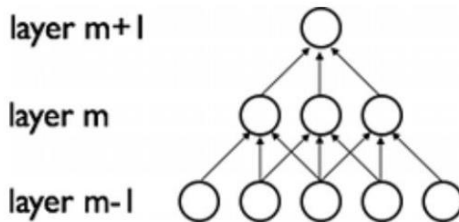


Gambar 2. 2 Arsitektur model dengan dua channel untuk sebuah kalimat

Sekarang CNN dapat diterapkan pada dua hal yaitu pada *computer vision* dan NLP. Gambar 2.2 merupakan arsitektur model dari CNN pada NLP. Bagaimanapun juga terdapat dua komponen utama dari CNN yaitu *convolution layer* dan *pooling layer*.

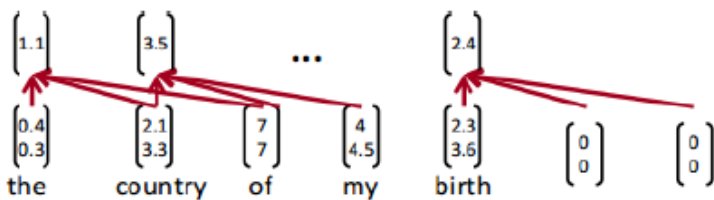
1. Convolutional Layer

Pada tiap *convolutional layer*, filter konvolusi menggunakan korelasi lokal dengan memaksakan pola konektivitas lokal diantara layer yang berdekatan. Layer yang lebih atas m telah didapat dari sebuah subhimpunan satuan dari layer yang lebih rendah $m-1$. Dibandingkan dengan MLP, keuntungan lainnya dari *convolution layer* adalah jumlah parameter yang berkurang signifikan disebabkan oleh parameter bagi. Konektivitas tersebut digambarkan dalam Gambar 2.3



Gambar 2. 3. Jaringan dengan 3 convolutional layer

Lapisan ini merupakan lapisan pertama dalam arsitektur jaringan Hybrid CNN-HMM. Pada lapisan ini dilakukan proses konvolusi untuk semua vektor kata yang ada pada artikel. Perlu diingat bahwa Gambar 2.3 hanya ilustrasi semata, sehingga pada layer m+1 tidak wajib menghasilkan satu *output*, akan tetapi yang wajib dilakukan adalah m+1 selalu menghasilkan output kurang dari sama dengan m. Gambar 2.4 menunjukkan contoh dari proses konvolusi (penjelasan dari proses konvolusi akan disajikan pada Gambar 4.12) pada suatu kalimat dimana untuk kata yang terakhir vektor kata setelahnya yang berada di window dibuat suatu vektor kata yang berupa vektor nol.



Gambar 2. 4 Contoh proses konvolusi contoh kalimat

Jika vektor-vektor kata digabungkan, maka artikel dapat direpresentasikan dalam *array* sebagai berikut

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \end{bmatrix} \quad (1)$$

Keterangan:

X : *array* kumpulan kata dalam sebuah artikel berita
 i : jumlah kata terbanyak dalam satu artikel
 x_i : vektor kata yang berada pada \mathbb{R}^{50} dengan indeks ke- i

Secara umum x_i merupakan *array* yang berisi vektor-vektor representasi kata sebagai berikut.

$$x_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad \dots \quad x_{i50}] \quad (2)$$

Dalam x_i terdapat *array* vektor yang merupakan representasi dari kata x_i . Dalam Tugas Akhir ini, jumlah vektor adalah 50, sehingga dalam indeks vektor terakhir dalam x_i adalah x_{i50} .

Operasi konvolusi menggunakan sebuah filter untuk menghasilkan suatu nilai *feature* maka dapat dirumuskan sebagai berikut

$$c_i = f(w \cdot x_{i:1:50} + b) \quad (3)$$

Keterangan:

c_i : nilai *feature map* pada indeks ke- i
 h : ukuran *window* kata
 w : filter yang berada pada \mathbb{R}^{50}
 b : parameter bias

Pada Tugas akhir ini fungsi non linear yang digunakan merupakan fungsi *Rectified Linear Unit (ReLU)*. Keluaran dari fungsi tersebut memberikan batasan keluaran bernilai positif. Fungsi tersebut dapat dituliskan sebagai berikut:

$$ReLU(x) = \max\{0, x\} \quad (4)$$

Sehingga persamaan menjadi seperti berikut:

$$c_i = \text{ReLu}(\text{net}) \quad (5)$$

Filter w di terapkan untuk tiap *window* kata yang mungkin di dalam kalimat artikel sehingga dihasilkan sebuah *feature map* seperti berikut:

$$c = [c_1, c_2, c_3, \dots, c_n] \quad (6)$$

Dengan c merupakan *feature map*.

Untuk mendapatkan hasil yang maksimal maka filter yang akan digunakan pada Tugas Akhir ini lebih dari satu filter dan begitupun juga dengan *window* kata.

2. *Pooling Layer*

Komponen penting kedua CNN adalah *pooling*. Fungsi *layer pooling* sebagai *down-sampling* yang non linear. Algoritma *pooling* hanya fokus pada sub-daerah tertentu dari *feature map*. Tujuan dari *pooling* adalah mencoba mendapatkan informasi yang paling bermanfaat dari daerah tertentu. Algoritma ini sangat penting karena beberapa hal antara lain:

- Dengan mengeliminasi nilai yang tidak maksimal, mengurangi komputasi dari layer yang di atasnya
- Menyediakan sebuah bentuk dari tranlasi invariant
- Beberapa kasus, *pooling* membantu untuk merangkum perbedaan Panjang fitur menjadi dimensi yang sama

Setelah didapatkan *feature map* yang berasal dari *convolution layer*, maka masing-masing dari *feature map* tersebut dilakukan operasi *pooling* pada *pooling layer*. Pada

lapisan ini diambil nilai-nilai yang penting dari *feature map* dengan mengambil nilai yang paling maksimum di tiap *feature map*. Secara matematis, operasi *pooling* dapat dirumuskan sebagai berikut:

$$\hat{c} = \max\{c\} \quad (7)$$

\hat{c} merupakan nilai maksimum dari *feature map* c yang berada pada Persamaan 6.

Dikarenakan terdapat sebanyak m filter, maka hasil dari *pooling layer* merupakan suatu vektor yang terdiri atas nilai maksimum tiap *feature map* dan berjumlah m elemen. Sehingga dapat diperoleh

$$z = [\hat{c}_1, \hat{c}_2, \hat{c}_3, \dots, \hat{c}_m] \quad (8)$$

z merupakan vektor hasil dari *pooling layer* yang akan masuk kelapisan berikutnya.

3. *Output layer*

Lapisan terakhir arsitektur jaringan pada Tugas Akhir ini adalah *output layer*. Pada lapisan ini akan digunakan jaringan syaraf tiruan yang bertujuan untuk mengklasifikasikan artikel ke kelas-kelas yang ditentukan antara artikel positif, netral, atau negatif. Hasil dari *pooling layer* yang berupa vektor *feature* akan melewati hubungan penuh jaringan *Hidden Markov Model*(HMM) untuk mendapatkan skor artikel.

2.2.6 *Hidden Markov Model*

Hidden Markov Models (HMM) adalah sebuah model stokastik dan pada dasarnya merupakan perluasan dari rantai

Markov. Dalam HMM, ada dua tipe state: state yang dapat diobservasi dan state yang tersembunyi [16]. Dikutip dari jurnal Zoubin Ghahramani yang berjudul “*An introduction to hidden markov models and bayesian network*”, HMM merupakan sebuah alat untuk merepresentasikan distribusi peluang dari pengamatan keterurutan [17]. Misalkan akan tunjukkan sebuah pengamatan pada waktu t dan dengan variable Y_t . Hal ini dapat menjadi sebuah simbol dari sebuah alfabet diskrit, variabel nilai riil, sebuah bilangan bulat, atau objek lainnya, selama dapat didefinisikan menjadi sebuah peluang distribusi dari hal tersebut. Akan diasumsikan bahwa pengamatan tersebut adalah sampel dari diskrit, dengan interval waktu yang sama, sehingga t dapat menjadi sebuah indeks waktu bernilai bulat.

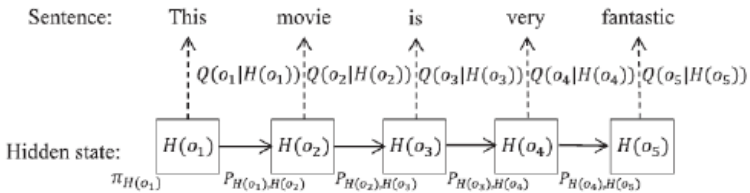
Hidden Markov Model mendapatkan nama dari dua properti yang menentukan. Pertama, akan diasumsikan bahwa pengamatan pada waktu t dibentuk oleh beberapa proses dimana konsisi S_t disembunyikan dari pengamat. Kedua, akan diasumsikan bahwa kondisi dari proses *hidden* merujuk *Markov property*. Hal itu, memberikan nilai S_{t-1} , kondisi saat ini S_t independen dari semua kondisi sebelumnya pada $t-1$. Dengan kata lain kondisi pada beberapa waktu merangkum semua apa yang ingin diketahui tentang sejarah dari proses sebagai prediksi masa depan dari proses. Keluaran juga merujuk pada sifat Markov terhubung dengan kondisi: diberikan S_t , Y_t adalah independent dari sebuah keadaan dan pengamatan pada semua indeks waktu lainnya.

Diambil secara bersamaan, sifat-sifat markov berikut ini berarti bahwa distribusi gabungan dari *hidden state* dan *observation state* dapat dituliskan seperti pada Persamaan(9):

$$P(S_{1:T}, Y_{1:T}) = P(S_1)P(Y_1|S_1) \prod_{t=2}^T P(S_t|S_{t-1})P(Y_t|S_t) \quad (9)$$

Rumus dari probabilitas gabungan dapat digambarkan seperti pada Persamaan (9). Model ini, diketahui sebagai

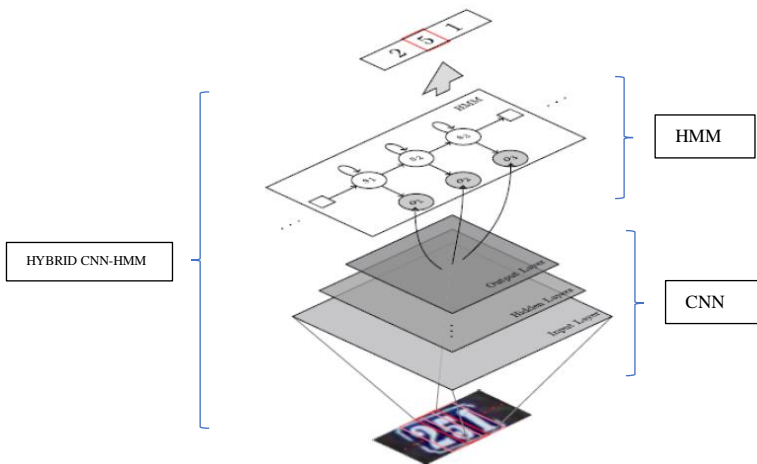
jaringan Bayesian, jaringan kepercayaan, grafik model kepercayaan, atau jaringan probabilitas independent.



Gambar 2. 5 Contoh penggunaan HMM pada klasifikasi teks

2.2.7 Hybrid CNN-HMM

Metode Hybrid CNN-HMM merupakan metode gabungan antara algoritma CNN dan HMM. CNN akan melakukan proses dimana akan dicari probabilitas keterhubungan antar kata yang muncul. Berdasarkan probabilitas dari CNN ini akan dilakukan klasifikasi menggunakan HMM. HMM digunakan sebab memiliki kemampuan untuk mendeteksi pola keterhubungan antar objek berdasarkan probabilitas keterhubungannya.



Gambar 2. 6 Contoh penggunaan Hybrid CNN-HMM pada *image processing*

2.2.8 Evaluasi Sistem

Sistem yang telah dibentuk dengan menggunakan algoritma Hybrid CNN-HMM akan dilakukan evaluasi untuk mengetahui performansi dari sistem. Terdapat tiga cara evaluasi yang akan digunakan dalam Tugas Akhir ini yaitu akurasi (accuracy), presisi (*precision*), dan *recall*.

Akurasi merupakan prosentase artikel yang diklasifikasikan benar ke dalam kelas sesuai dengan kelas target sebenarnya. Presisi merupakan prosentase artikel positif yang diklasifikasikan secara benar ke kelas target positif. Sedangkan *recall* adalah prosentase artikel positif yang diklasifikasikan secara benar sesuai kelas targetnya. Rumus yang digunakan untuk mengukur akurasi, presisi, dan *recall* adalah sebagai berikut:

$$accuracy = \frac{TP + TN}{P + N} \quad (3)$$

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{TP + FN} \quad (5)$$

Dengan:

TP : Jumlah artikel yang diprediksi positif dan kelas target sebenarnya adalah positif.

TN : Jumlah artikel yang diprediksi non-positif (negatif dan netral) dan kelas target sebenarnya adalah non-positif.

FP : Jumlah artikel yang diprediksi positif sedangkan kelas target sebenarnya adalah non-positif

FN : Jumlah artikel yang diprediksi non-positif sedangkan kelas target sebenarnya adalah positif

P : Jumlah artikel yang sebenarnya bernilai positif

N : Jumlah artikel yang sebenarnya bernilai non-positif

BAB III

METODOLOGI PENELITIAN

Bab ini menjelaskan tentang tahap pengerjaan dalam menyelesaikan Tugas Akhir ini sehingga penelitian ini dapat dirancang sistematis dan diatur dengan sebaik-baiknya.

3.1 Studi literatur

Studi literatur dilakukan dengan mengumpulkan bahan referensi dan rujukan tentang dasar-dasar teori yang digunakan dalam analisa sentiment (*opinion mining*) sebagai pendukung dalam kegiatan pengerjaan tugas akhir ini.

3.2 Pengumpulan Data

Pada tahap awal dilakukan pengumpulan data sebagai bahan masukan untuk tugas akhir ini. Data yang didapat merupakan data ringkasan artikel berita dengan tipe data berupa teks dan bobot.

3.3 Praproses Data

Selanjutnya data yang telah didapatkan terlebih dahulu dilakukan pra-proses data. Tahap ini bertujuan untuk mendapatkan bentuk data yang diinginkan sebelum masuk ke tahap implementasi. Terdapat beberapa tahapan pada praproses ini, antara lain:

a. Case Folding

Seluruh kata pada artikel berita akan diubah menjadi huruf kecil (*lowercase*). Hal ini bertujuan untuk memperkecil kemungkinan terjadinya multi data pada kata yang sama. Multi data adalah keadaan suatu kata dianggap memiliki arti yang berbeda karena terdapat

perbedaan penulisan huruf kapital (contoh: Jakarta dan jakarta).

b. Tokenisasi

Pembentukan *array* dari kata-kata yang ada di dalam artikel berita. Hal ini bertujuan agar data dapat dijadikan sebagai masukan pada algoritma CNN

c. Filterisasi

Penghilangan angka, kata yang dianggap tidak bertendensi yang biasa disebut *stopword*. Sebelum berlanjut pada proses selanjutnya, kata-kata pada artikel berita tersebut disaring dari kata *stopword* tersebut.

d. Stemming

Masing-masing *array* kata akan diubah menjadi kata dasar. Dengan kata lain pada tahap ini akan dihilangkan prefix dan suffix dari sebuah kata. Hal ini akan memudahkan proses pengenalan kata, sehingga kata yang memiliki kata dasar sama akan dianggap sama

3.4 Ekstraksi Fitur

Data yang sudah dilakukan tahap pra-proses data, selanjutnya dilakukan tahapan ekstraksi fitur. Ekstraksi fitur dilakukan untuk mendapatkan ciri dari suatu data atau kata. Pada tahap ini setiap kata diubah menjadi representasi vektor yang mewakili polaritas suatu kata. Salah satu metode yang sedang berkembang sekarang yaitu metode *Word2vec*. Metode ini bertugas untuk merubah sebuah kata menjadi suatu vektor. Vektor yang dihasilkan merupakan konteks dari kata tersebut yang memperhatikan peluang kata yang muncul disekitarnya. Pada perkembangannya *Word2vec* telah dijadikan sebagai dasar dari penelitian NLP. *Word2vec* ini juga memiliki beberapa jenis,

namun pada Tugas Akhir ini peneliti menggunakan jenis Gensim. Dengan vektor ini, data dapat diolah untuk proses selanjutnya

3.5 Implementasi Algoritma Hybrid CNN-HMM

Pada tahap ini dilakukan proses klasifikasi data yang telah melalui tahap ekstraksi fitur. Output dari ekstraksi fitur nantinya berupa matriks yang berupa kumpulan vektor-vektor kata dalam sebuah artikel. Matriks tersebut akan masuk ke tahapan selanjutnya. Terdapat dua tahapan utama dalam proses ini yaitu tahap CNN dan HMM. Pada tahap CNN di penelitian Tugas Akhir ini terdiri dari dua tahapan lagi yaitu proses konvolusi dan proses *pooling*. Hasil dari CNN selanjutnya akan digunakan sebagai inputan pada HMM. Pada Proses Implementasi Algoritma ini terdiri dari tahap uji dan latih. Tahap uji (*training*) digunakan untuk mendapatkan parameter yang sesuai dan optimal, sedangkan tahap uji (*testing*) untuk melihat keberhasilan dari parameter yang telah didapatkan setelah penelitian. Data akan diklasifikasikan menjadi tiga kelas utama yaitu positif, netral, dan negatif dengan menggunakan HMM.

3.6 Analisa Hasil dan Penarikan Simpulan

Pada tahap ini akan dilakukan Analisa terhadap hasil yang didapatkan selama pengerjaan Tugas Akhir ini apakah hasil yang dilakukan oleh program telah sesuai ataukah tidak. Kemudian akan dilakukan pengambilan Simpulan berdasarkan hasil yang didapat tadi.

3.7 Validasi Sistem

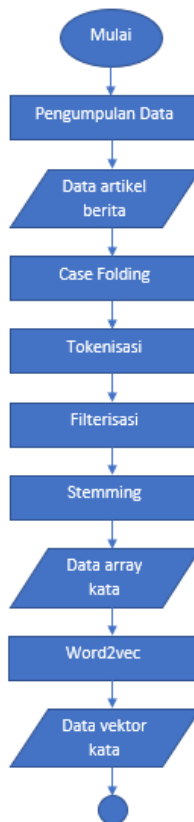
Pada tahap ini akan dilakukan validasi terhadap berjalannya program. Untuk mengukur performansi program dalam menghasilkan *output*.

3.8 Penyusunan Laporan

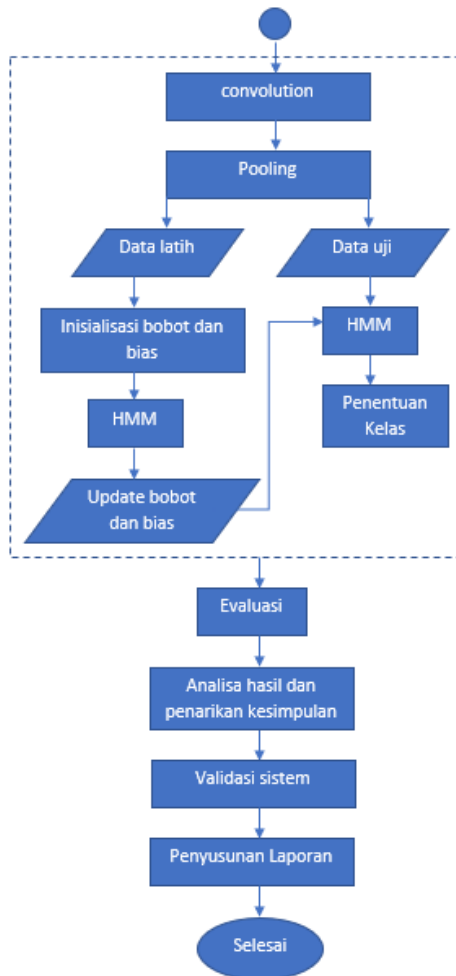
Tahap akhir setelah semua tahap dijalankan adalah penulisan laporan tugas akhir. Konten laporan tugas akhir sesuai dengan hasil yang telah didapatkan dari proses awal pengumpulan data sampai dengan penarikan Simpulan.

3.9 Diagram Alir Metodologi Penelitian

Gambaran tahap-tahap dalam penelitian pada Tugas Akhir ini disajikan dalam diagram alir berikut



Gambar 3. 1. Diagram Alir penelitian (1)



Gambar 3. 2 Diagram alir penelitian (2)

BAB IV

PERANCANGAN DAN IMPLEMENTASI

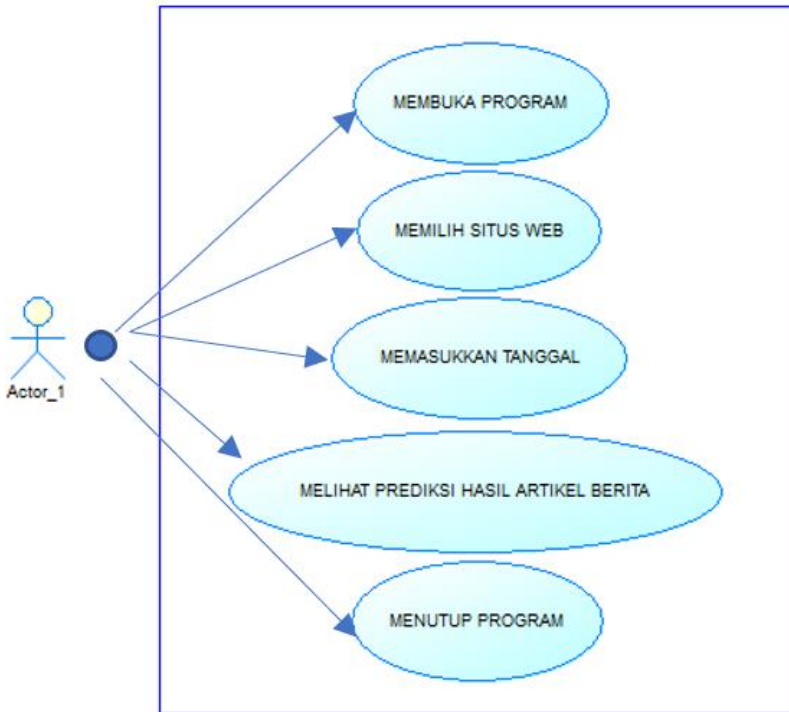
Pada bab ini menjelaskan perancangan desain sistem yang digunakan sebagai acuan untuk implementasi sistem. Perancangan sistem menggambarkan proses rancang bangun secara terperinci dari awal tahap pengumpulan data hingga pembuatan model Hybrid CNN-HMM. Langkah-langkah dalam pengimplementasian sistem berdasarkan desain sistem yang sudah dibentuk

4.1 Analisis Implementasi Sistem

Dalam proses implementasi suatu sistem diperlukan suatu analisis terhadap sistem sehingga sistem dapat dibuat secara optimal. Ada banyak cara dalam menganalisis suatu sistem yaitu menggunakan *usecase diagram* dan *activity diagram*.

4.1.1 Use Case Diagram

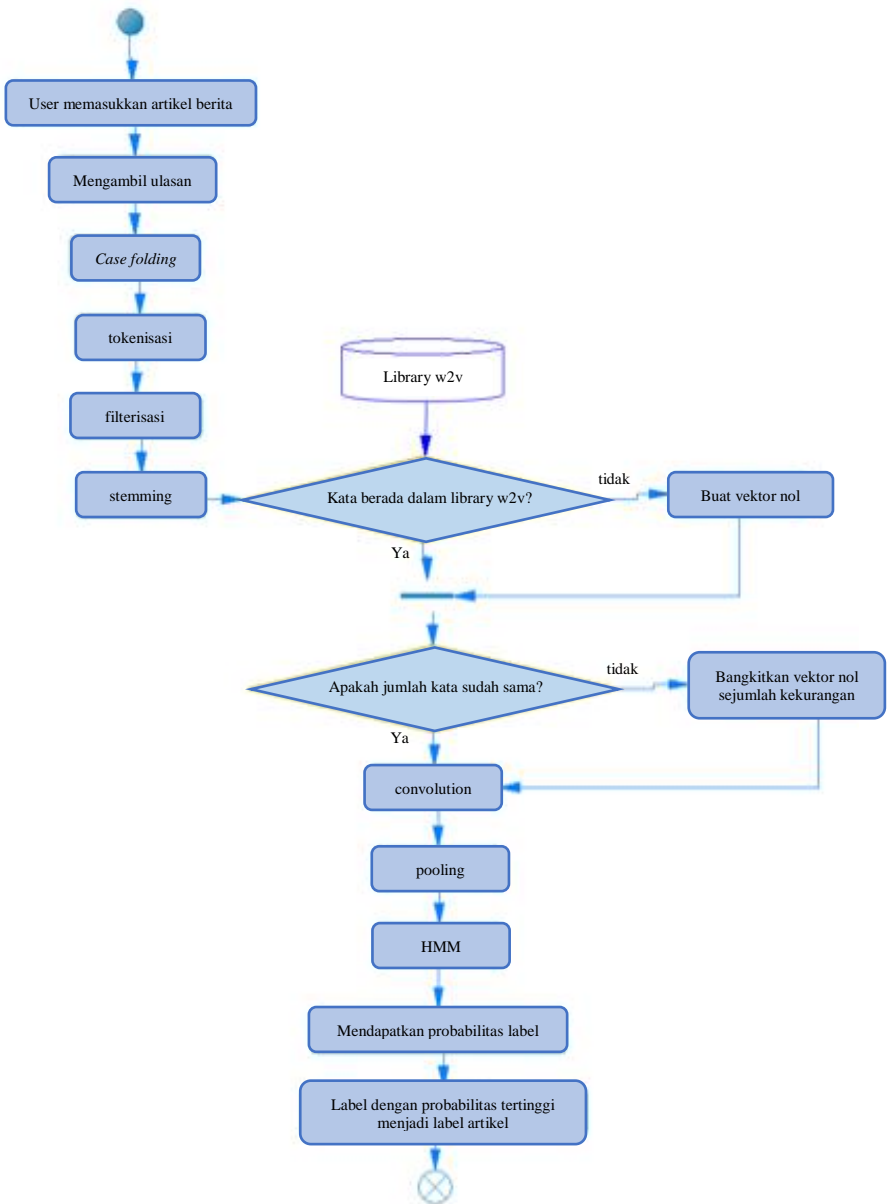
Use case diagram merupakan suatu diagram yang menjelaskan interaksi yang terjadi pada suatu sistem dengan aktor sistem. Adapun *use case diagram* dari sistem Tugas Akhir ini seperti pada Gambar 4.1.



Gambar 4. 1. Use case diagram sistem

4.1.2 Activity Diagram

Untuk menjelaskan proses berjalannya suatu program dari awal sampai akhir, dapat digunakan activity diagram. Adapun activity diagram pada Tugas Akhir ini seperti pada Gambar 4.2 berikut.



Gambar 4. 2. Activity diagram sistem

4.1.3 Analisis Kebutuhan Fungsional

Berdasarkan hasil analisis sistem menggunakan *use case diagram* dan *activity diagram* diatas, sistem diharapkan mampu melakukan hal sebagai berikut:

1. Sistem mampu melakukan proses pembelajaran dengan menggunakan algoritma *Hybrid CNN-HMM* pada data latih.
2. Sistem mampu menerapkan proses pembelajaran dengan menggunakan algoritma *Hybrid CNN-HMM* pada data uji dan melakukan klasifikasi.
3. Sistem mampu melakukan *web crawling* terhadap situs yang telah ditentukan
4. Sistem mampu mengumpulkan artikel berita dari situs yang telah dikunjungi
5. Sistem mampu melakukan klasifikasi terhadap artikel yang telah dikumpulkan
6. Sistem mampu menyimpan data hasil klasifikasi artikel yang telah dikumpulkan

4.1.4 Analisis Kebutuhan Non-Fungsional

Lingkungan perancangan sistem dibangun dari dua lingkungan *software* dan lingkungan *hardware*. Spesifikasi lingkungan perancangan sistem secara lengkap dapat dilihat di Tabel 4.1

Tabel 4. 1. Lingkungan *hardware* dan *software*

Lingkungan	Spesifikasi	
<i>Hardware</i>	<i>Processor</i>	Intel Core I3
	RAM	4GB
<i>Software</i>	Sistem Operasi	Windows 10 Education 64-bit
	<i>Tools</i>	<ul style="list-style-type: none"> • Python 3.5 • Google Colab

4.2 Perancangan Sistem

Dari hasil analisis sistem akan dirancang program yang diharapkan sesuai dengan tujuan penelitian Tugas Akhir ini.

4.2.1 Perancangan Data

Data pada sistem dibagi menjadi tiga yaitu data masukan (*input*), data proses (sistem), data luaran (*output*).

1. Data masukan

Terdapat dua data masukan pada sistem yaitu data mentah dan data hasil *web crawling*. Data mentah yaitu berupa kumpulan artikel berita dari Bank Indonesia berformat (.xlsx) yang akan digunakan sebagai data uji dan latih. Data hasil *web crawling* merupakan kumpulan data artikel yang diperoleh melalui proses *web crawling* berformat (.csv).

2. Data proses

Data proses merupakan data yang digunakan dalam proses pengolahan data masukan hingga hasil dari proses tersebut. Data proses sistem ini terbagi menjadi beberapa tahapan sebagai berikut:

a) Persiapan data masukan

Masukan : Data mentah yaitu berupa kumpulan artikel berita dari Bank Indonesia berformat (.xlsx)

Luaran : Data awal yaitu data berupa artikel dan label

b) *Case folding*

Masukan : List artikel berita dari data awal

Luaran : List artikel berita dengan format *lowercase*

c) *Tokenisasi*

Masukan : List artikel berita dengan format *lowercase*

- Luaran : list artikel yang telah dipisah per-kata
- d) *Filtering*
- Masukan : List artikel yang telah dipisah per-kata
 Luaran : List artikel berita tanpa ada angka dan tanda baca
- e) *Stemming*
- Masukan : List artikel berita tanpa ada angka dan tanda baca
 Luaran : list artikel berita dengan hanya kata dasar
- f) Ekstraksi Fitur
- Masukan : list artikel berita dengan kata dasar
 Luaran : array vektor representasi kata dari artikel
- g) *Hybrid CNN-HMM*
- Masukan : array vektor representasi kata dari artikel
 Luaran : probabilitas prediksi kelas dari artikel berita
- h) Penentuan situs web
- Masukan : pilihan radio button
 Keluaran : URL situs berita online
- i) Penentuan tanggal pencarian
- Masukan : tanggal berupa teks
 Keluaran : format tanggal untuk tambahan masukan pada URL
- j) Halaman induk
- Masukan : URL halaman situs web lengkap (sudah tertera tanggal)
 Luaran : URL halaman tiap artikel
- k) Pengunjungan dan akuisisi data
- Masukan : URL halaman tiap artikel
 Luaran : Data artikel berita

3. Data luaran

Terdapat dua data luaran sistem ini yaitu akurasi pembelajaran menggunakan algoritma Hybrid CNN-HMM dan berupa kumpulan artikel berita yang telah dilakukan *web crawling* dan telah dilakukan klasifikasi kelas pada tiap artikel.

4.2.2 Perancangan Proses

Perancangan proses dimulai dari perancangan pra-proses data, perancangan proses algoritma Hybrid CNN-HMM, dan perancangan proses *web crawling*. Pada pra-proses data akan dilakukan proses persiapan awal guna menyiapkan data agar dapat menjadi masukan pada proses algoritma. Pada proses algoritma akan dilakukan proses pembelajaran berdasarkan data latih sehingga sistem mampu melakukan klasifikasi pada data uji maupun data hasil *web crawling*. Pada proses *web crawling* akan dilakukan pengambilan data artikel berita dari situs berita *online* yang akan dilakukan klasifikasi *sentiment* pada tiap artikel berita. Berikut ini adalah perancangan proses dari sistem.

4.2.2.1 Perancangan Pra-proses Data

Dalam perancangan pra-proses data akan dilakukan enam proses guna menyiapkan data sebagai masukan pada proses selanjutnya. Berikut merupakan proses dalam pra-proses data.

1. Perancangan Proses Persiapan Data Masukan

Dalam penerapan Tugas Akhir ini dibutuhkan data sebagai bahan masukan dari algoritma yang akan dijalankan. Data yang digunakan dalam Tugas Akhir kali ini menggunakan dataset dari Bank Indonesia yaitu Data Monitoring harian berita.

	A	B	C	D	E	F	G	H
1								
2	Tanggal	Nama Media	Ringkasan Berita	Kaitan Berita	Ruang Lingkup	Klasifikasi Berita	Label Berita	
3								
4	01-Jan-17	Surya	(Surabaya) Harga Beras Medium Distabilkan Rp. 9.350 per Kil	Komoditas	Lokal	Ekonomi	2	
5		Kompas	(Jakarta) Konsolidasi Aset Setelah Libur Panjang. Pasar keuan	Investasi	Nasional	Ekonomi	3	
6	01-Jul-17	Radar	(Surabaya) Aperindo: Penjualan Lampu LED Masih Rendah. A	Industri	Lokal	Ekonomi	1	
7	01-Jul-17	Radar	(Kebomas) Bisnis Ritel Tahun Ini Melambat. Ramadhan tahun	Industri	Lokal	Ekonomi	1	
8	01-Jul-17	Radar	(Kebomas) Bahan Naik, Indospring Lakukan Efisiensi. Naiknya	Ekspor/Impor	Lokal	Ekonomi	2	
9	01-Jul-17	Jawa Pos	(Jakarta) Harga Sulit Turun Pasca-Lebaran. Memasuki masa-m	Komoditas	Nasional	Ekonomi	1	
10	01-Jul-17	Jawa Pos	(Surabaya) Tarif Penerbangan Terealisasi Separo. Perkembang	Inflasi/Harga	Lokal	Ekonomi	3	
11	01-Jul-17	Jawa Pos	(Surabaya) Mandiri Gencarkan Transaksi Cashless. Bank Manc	GNNT	Lokal	sistem Pembayaran	2	
12	01-Jul-17	Jawa Pos	(Jakarta) Bank Swasta Klaim Semakin Efisien. Bank swasta mei	GNNT	Nasional	sistem Pembayaran	2	
13	01-Jul-17	Jawa Pos	(Jakarta) Menilai Prospek Mata Uang Digital. Dalam beberapa	GNNT	Nasional	sistem Pembayaran	2	
14	01-Jul-17	Jawa Pos	(Jakarta) Suku Bunga Bertahan Sampai Akhir Tahun. Bank Indi	days Repo Rat	Nasional	Ekonomi	2	
15	01-Jul-17	Jawa Pos	(Surabaya) Hospitality Dongkrak Kinerja Industri Kertas. Indu	Komoditas	Lokal	Ekonomi	3	
16	01-Jul-17	Jawa Pos	(Surabaya) Sekar Laut Ekspansi ke Eropa Timur. Produsen mak	Ekspor/Impor	Lokal	Ekonomi	3	
17	01-Jul-17	Kompas	(Jakarta) Ketersediaan Beras Tahun Ini Perlu Menjadi Perhatian	Komoditas	Nasional	Ekonomi	3	
18	03-Jul-17	Radar	(Surabaya) MNC Bank Perluas Pasar melalui Aplikasi. Gellat b	Perbankan	Lokal	SSK	3	
19	03-Jul-17	Radar	(Surabaya) Libur Lebaran Kerek Konsumsi Gasoline 24%. PT Pi	Komoditas	Lokal	Ekonomi	3	
20	03-Jul-17	Kompas	(Jakarta) Penyelundupan Benih Lobster Terus Mingkat. Hingg	Komoditas	Nasional	Ekonomi	1	
21	03-Jul-17	Kompas	(Jakarta) Dendapan Kurang Optimal. Meskipun ntromoni dai	Keruangan	Nasional	SSK	1	

Gambar 4. 3. Data monitoring harian berita

Komponen data yang didapat dari Data Monitoring tersebut antara lain:

- Tanggal
- Nama media
- Isi Berita
- Kaitan Berita
- Ruang lingkup
- Klasifikasi berita
- Label Berita (Kecenderungan berita berunsur positif, negatif, atau netral)
- Pembicara BI

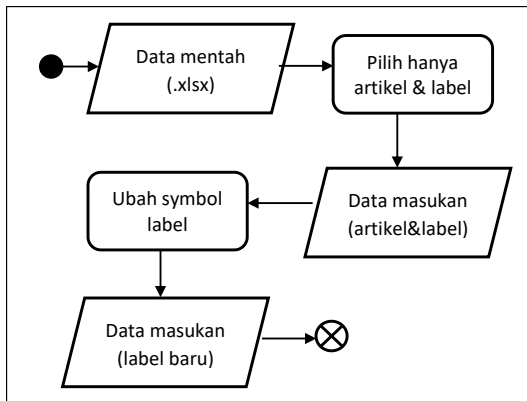
Berdasarkan data tersebut, penulis melakukan filtering secara manual dengan cara mengambil data berupa isi berita dan tone berita. Kemudian dari data tersebut penulis membuat dataset baru dengan isi yaitu isi berita dan tone berita. Untuk tone berita, penulis melakukan perubahan dengan cara *replace* angka yang terdapat pada data sumber. Dimana data awal berisi:

- a. 1 (merah) = negatif
- b. 2 (kuning) = netral
- c. 3 (hijau) = positif

Akan diubah menjadi:

- a. -1 = negatif
- b. 0 = netral
- c. 1 = positif

Hal ini dilakukan agar dapat mudah melakukan proses pada program sehingga *tone* berita akan diubah menjadi -1, 0, dan 1. Contoh potongan artikel berita pada Tugas Akhir ini dapat dilihat pada Tabel 4.2 berikut:



Gambar 4. 4. Flowchart proses persiapan data masukan

Tabel 4. 2 Contoh artikel berita dengan *tone* positif, netral, dan negatif

1	<i>Tone</i> positif	Tambah Produksi Tambang Emas. Aktivitas di tambang emas Tumpang Pitu (Tujuh Bukit) di Banyuwangi, Jawa Timur. Tambang Pitu merupakan tambang ramah lingkungan yang tidak menghasilkan limbah batuan atau tanah halus...
----------	--------------------------------	---

2	Tone netral	Suku Bunga Bertahan Sampai Akhir Tahun. Bank Indonesia (BI) masih mempertahankan suku bunga acuan, yakni BI 7-day reverse repo rate di posisi 4,75% sejak Oktober 2016. Posisi itu diprediksi bertahan hingga penghujung tahun ini...
3	Tone negatif	Ketersediaan Beras Tahun Ini Perlu Menjadi Perhatian. Stok beras diperkirakan berkurang karena terjadi gagal panen akibat serangan hama di lahan sekitar 500000 hektar pada tahun ini, sehingga pemerintah perlu mewaspadai ketersediaan beras untuk tahun ini...

Total data artikel yang diperoleh adalah 1760 data yang didapat dari sumber yang telah dijelaskan diatas. Untuk melakukan tahapan uji dan latih, data tersebut akan dilakukan *split data* dimana 80% akan digunakan sebagai data latih dan 20% digunakan sebagai data uji. Tidak ada perhitungan khusus dalam penetapan proporsi penggunaan data uji dan latih.

Berikut merupakan contoh data yang akan diinputkan untuk dilakukan proses pembelajaran untuk mendapatkan model.

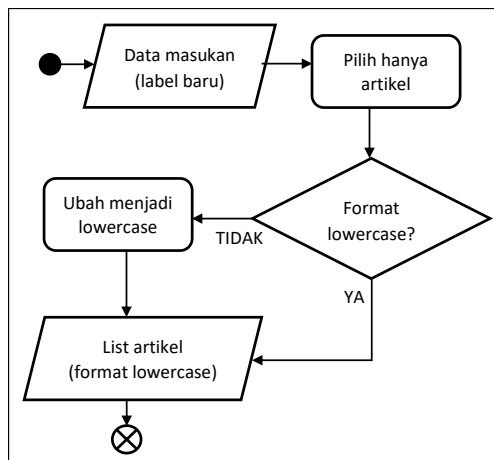
Tabel 4. 3. Contoh data inputan

Artikel	Label
Produksi Semen Melimpah. Produksi semen di Indonesia menunjukkan angka surplus jika dibandingkan dengan kebutuhan di pasar domestik. Berdasarkan data Asosiasi Semen Indonesia (ASI), kapasitas produksi semen nasional tahun ini tercatat 100 juta ton...	1
Penjualan Domestik Turun, Jual via Online. PT Semen Indonesia Tbk (SMGR) pada semester pertama ini mencapai 13,3 juta ton atau tumbuh 3 persen bila dibandingkan dengan periode yang sama 2017. Tetapi, jika di-breakdown, penjualan domestiknya turun 1,2 persen daripada tahun sebelumnya...	-1

Terbitkan Lagi SBI. Mengantisipasi pelemahan rupiah, BI memutuskan mengaktifkan kembali sertifikat BI (SBI) dengan tenor 9 bulan dan 12 bulan. Penerbitan kembali SBI tersebut ditujukan menarik dana asing agar mampu menstabilkan nilai tukar rupiah terhadap dolar AS.	1
Pemkab Terus Perbaiki Jalur Pedestrian. Sejumlah jalan protokol di Kota Delta belum dilengkapi dengan jalur pedestrian. Yakni, ruang nyaman untuk kegiatan pejalan kaki. Pemkab pun berjanji terus memperbaiki jalur pedestrian.	0
Peluang Domestik Terabaikan. Pemerintah dinilai kurang fokus mendorong pergerakan wisatawan nusantara untuk menggenjot pendapatan negara, padahal pelancong lokal menghabiskan uang lebih banyak dari turis asing saat berpelesir...	-1

2. Perancangan proses *Case Folding*

Pada tahap *case folding* akan dilakukan penyamaan bentuk penulisan. Dalam Tugas akhir ini, seluruh kata dalam artikel akan diubah menjadi bentuk penulisan *lowercase*. Hal ini bertujuan agar tidak terjadi data ganda apabila ada penulisan dengan huruf kapital ataupun tidak.



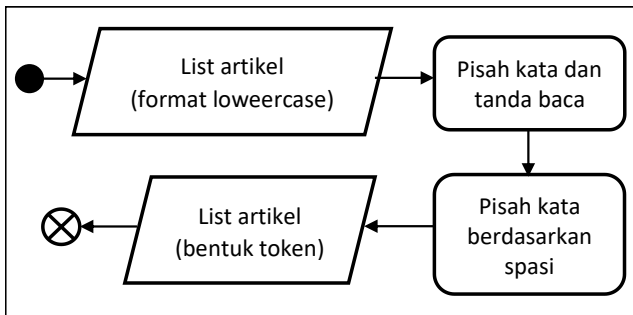
Gambar 4. 5. Flowchart proses *casefolding*

Tabel 4. 4. Tabel input dan output proses *Case folding*

Input	Tambah Produksi Tambang Emas. Aktivitas di tambang emas Tumpang Pitu (Tujuh Bukit) di Banyuwangi, Jawa Timur. Tambang Pitu merupakan tambang ramah lingkungan...
Output	tambah produksi tambang emas. aktivitas di tambang emas tumpang pitu (tujuh bukit) di banyuwangi, jawa timur. tambang pitu merupakan tambang ramah lingkungan...

3. Perancangan Proses Tokenisasi

Pada tahap tokenisasi akan dilakukan pemisahan kata pada artikel, pemisahan ini akan dilakukan dengan meninjau ‘unsur’ pemisah atau yang biasa disebut *delimiter*. Dalam Tugas Akhir ini pemisahan kata menggunakan *delimiter* yaitu ‘spasi’. Kemudian, masing-masing kata yang telah dipisah akan dimasukkan kedalam sebuah list. Hal ini bertujuan agar masing-masing kata dapat diamati maknanya.



Gambar 4. 6. Flowchart proses tokenisasi

Hasil dari proses tokenisasi dapat dilihat pada Tabel 4.5 berikut:

Tabel 4. 5. Tabel input dan output proses tokenisasi

Input	tambah produksi tambang emas. aktivitas di tambang emas tumpang pitu (tujuh bukit) di banyuwangi, jawa timur. tambang pitu merupakan tambang ramah lingkungan...
Output	'tambah', 'produksi', 'tambang', 'emas', '.', 'aktivitas', 'di', 'tambang', 'emas', 'tumpang', 'pitu', '(', 'tujuh', 'bukit', ')', 'di', 'banyuwangi', ',', 'jawa', 'timur', '.', 'tambang', 'pitu', 'merupakan', 'tambang', 'ramah', 'lingkungan', ...

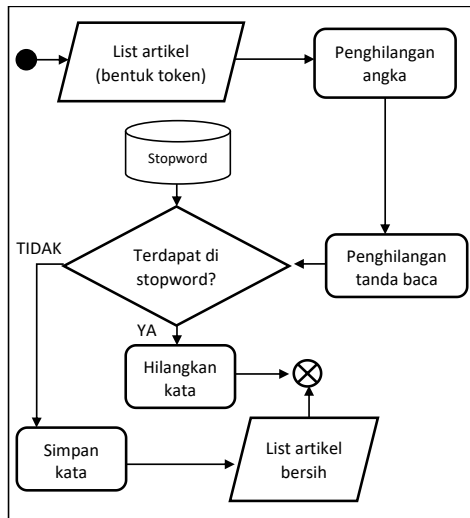
4. Perancangan Proses Filterisasi

Pada tahap filterisasi akan dilakukan penghilangan kata atau entitas yang dirasa tidak penting dalam penentuan sentimen suatu artikel berita. Dalam Tugas Akhir ini penulis menggunakan *library stopwords* milik nltk untuk melakukan filterisasi pada artikel yang telah berbentuk token. *Library* yang digunakan dalam proses filterisasi ini sudah tersedia untuk Bahasa Indonesia. Selain penghilangan *stopwords*, pada tahap ini penulis juga menghilangkan tanda baca dan angka dari list diatas.

Hasil dari proses filterisasi dapat dilihat pada Tabel 4.6 berikut

Tabel 4. 6. Tabel input dan output proses filterisasi

Input	'tambah', 'produksi', 'tambang', 'emas', '.', 'aktivitas', 'di', 'tambang', 'emas', 'tumpang', 'pitu', '(', 'tujuh', 'bukit', ')', 'di', 'banyuwangi', ',', 'jawa', 'timur', '.', 'tambang', 'pitu', 'merupakan', 'tambang', 'ramah', 'lingkungan', 'menghasilkan', ...
Output	'produksi', 'tambang', 'emas', 'aktivitas', 'tambang', 'emas', 'tumpang', 'pitu', 'tujuh', 'bukit', 'banyuwangi', 'jawa', 'timur', 'tambang', 'pitu', 'tambang', 'ramah', 'lingkungan', 'menghasilkan', ...



Gambar 4. 7. Flowchart proses filterisasi

5. Perancangan Proses *Stemming*

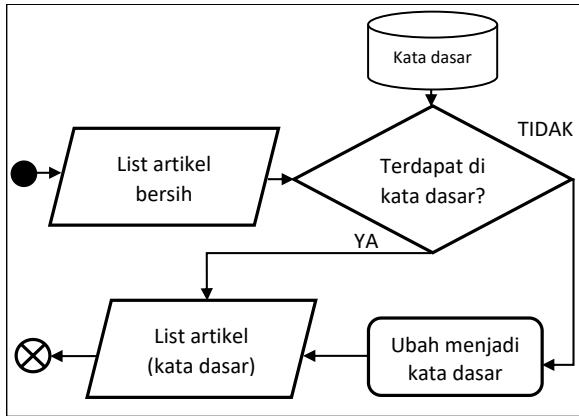
Pada tahap *stemming* akan dilakukan penghilangan prefix dan suffix dari sebuah kata sehingga list kata dari proses sebelumnya akan berubah menjadi kata dasar. Hal ini bertujuan untuk memudahkan program mengenali makna dari sebuah kata dalam artikel berita tersebut. Dalam Tugas Akhir ini penulis menggunakan *library sastrawi* untuk melakukan proses *stemming*.

Hasil dari proses *stemming* dapat dilihat pada Tabel 4.7 berikut

Tabel 4. 7. Tabel input dan output proses *stemming*

Input	'produksi', 'tambang', 'emas', 'aktivitas', 'tambang', 'emas', 'tumpang', 'pitu', 'tujuh', 'bukit', 'banyuwangi', 'jawa', 'timur', 'tambang', 'pitu', 'tambang', 'ramah', 'lingkungan', 'menghasilkan', ...
--------------	---

Output	'produksi', 'tambang', 'emas', 'aktivitas', 'tambang', 'emas', 'tumpang', 'pitu', 'tujuh', 'bukit', 'banyuwangi', 'jawa', 'timur', 'tambang', 'pitu', 'tambang', 'ramah', 'lingkung', 'hasil', ...
---------------	--



Gambar 4. 8. Flowchart proses stemming

6. Perancangan Proses Ekstraksi Fitur

Setelah diperoleh array kata-kata pada tiap artikel yang bebas dengan *stopwords*, maka dilakukan perubahan kata-kata tersebut menjadi representasi vektor kata. Vektor yang digunakan merupakan vektor *Word2vec* yang dibentuk secara manual dengan bantuan *tools* Gensim.

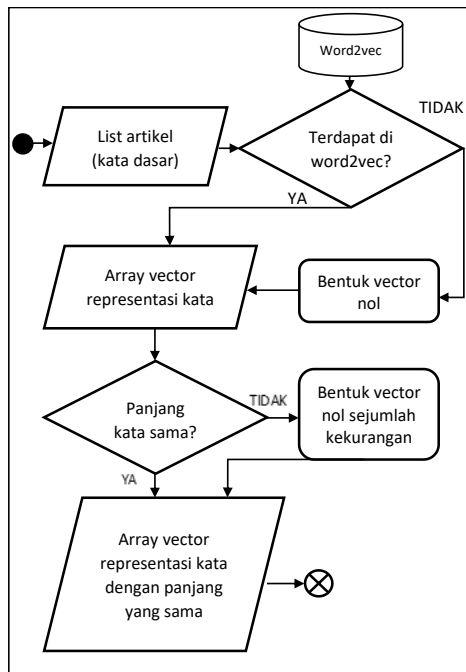
Pembuatan *library* ini akan diproses secara otomatis oleh gensim dengan mempertimbangkan jumlah kemunculan, dan keterkaitan antara kata satu dengan kata lainnya. Pada Tugas Akhir ini penulis memberikan Batasan minimum kemunculan kata agar dapat dibentuk ke dalam vektor adalah sebanyak 5. Hal ini dilakukan agar dapat menghindari apabila terjadi kesalahan penulisan atau *typo* sehingga kata-kata yang dibentuk ke dalam vektor adalah kata-kata yang memang dianggap penting, sering muncul, dan representatif bagi masing-masing artikel.

Sedangkan untuk ukuran dari vektor tiap kata yaitu 1×50 . Kemudian untuk kata yang tidak diubah ke dalam vektor dan dimasukkan ke dalam *library*, akan dibentuk kedalam vektor nol.

Sehingga hasil dari proses ekstraksi fitur ini adalah diperolehnya matriks representasi dari masing-masing artikel dengan ukuran $\text{max_len} \times 50$ dimana max_len merupakan ukuran terpanjang dari artikel dalam data tersebut. Kemudian untuk artikel dengan panjang $< \text{max_len}$ maka:

$$\text{add} = \text{max len} - \text{panjang artikel} \quad (6)$$

Pada indeks ke-*add* hingga indeks terakhir, artikel tersebut akan diisi dengan vektor nol.



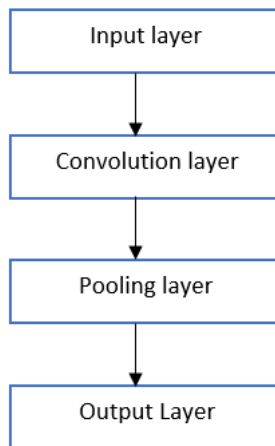
Gambar 4. 9. Flowchart proses ekstraksi fitur

Tabel 4. 8. Hasil dari ekstraksi fitur

No	Array vektor artikel
1	<pre>[-1.12407207e-02, -6.30088570e-03, -7.27833249e-03, ..., 1.24360221e-02, -4.49447567e-03, 8.02137703e-03], [8.13056249e-03, 2.46114377e-03, -8.43414757e-03, ..., 8.06766190e-03, -3.49627179e-03, -7.40810018e-03], [4.02027136e-03, -5.62059507e-03, -1.02112386e-02, ..., 1.60642981e-03, -1.92127621e-03, 7.69580901e-03], ..., [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00], [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00], [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00]]</pre>
2	<pre>[[-1.11767268e-02, 3.53055773e-03, -5.85780339e-03, ..., -4.83378721e-03, -8.72269273e-03, 7.81101966e-03], [-4.87548020e-03, -4.64063603e-03, -6.59062201e-03, ..., -3.24410666e-03, -9.81057761e-04, -6.59447350e-03], [-6.02664752e-03, -7.95502122e-03, 2.72068125e-03, ..., 1.09558913e-03, 3.37153999e-03, -3.51140555e-03], ..., [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00], [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00], [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00]]</pre>
3	<pre>[[-1.12407207e-02, -6.30088570e-03, -7.27833249e-03, ..., 1.24360221e-02, -4.49447567e-03, 8.02137703e-03], [-5.71386190e-03, -3.98984039e-03, 7.15243118e-03, ..., 1.40229624e-03, -8.61123612e-04, 3.27721960e-03], [-6.32367982e-03, -5.11622475e-03, -5.00117010e-03, ..., 1.19813706e-03, -6.47535734e-03, -6.66034780e-03], ..., [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00], [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00], [0.00000000e+00, 0.00000000e+00, 0.00000000e+00, ..., 0.00000000e+00, 0.00000000e+00, 0.00000000e+00]]</pre>

4.2.2.2 Perancangan Algoritma Hybrid CNN–HMM

Hybrid CNN-HMM tergolong ke dalam algoritma *Deep Learning*. Pada penerapan algoritma *Deep Learning*, kedalaman arsitektur merupakan hal yang membedakan algoritma tersebut dengan yang lainnya. Hal ini juga mempengaruhi tingkat kemampuan suatu jaringan tersebut. Pada tugas akhir ini arsitektur jaringan model *Hybrid CNN-HMM* terdiri atas 4 layer utama antara lain *input layer*, *convolutional layer*, *pooling layer*, dan *output layer*. Dengan inputan berupa matriks dari artikel dan suatu vektor kelas berdimensi tiga. Pada Tugas Akhir ini hanya menggunakan masing-masing satu *convolutional layer* dan *pooling layer*. Hal ini dikarenakan data yang akan diolah merupakan data teks yang tidak memiliki dimensi yang tinggi seperti data citra sehingga arsitektur ini sudah mampu mengklasifikasikan data teks artikel dengan baik.



Gambar 4. 10. Arsitektur jaringan Hybrid CNN-HMM

Pada lapisan pertama adalah *input layer* yang merupakan lapisan masukan. Pada lapisan ini terdapat matriks dari

gabungan vektor-vektor representasi kata dalam satu artikel. Pada Tugas Akhir ini dimensi yang digunakan pada vektor kata yaitu 50, sehingga pada *input layer* matriks masukkan berukuran $k \times 50$ dengan k merupakan jumlah kata terbanyak dalam sebuah artikel.

Setelah itu, matriks pada input layer masuk ke *convolution layer*. Pada lapisan ini dibentuk suatu vektor *feature map* sebanyak filter yang digunakan pada proses konvolusi. Tiap filter digunakan untuk semua *window* yang mungkin sehingga tiap *feature map* terdiri dari berbagai jenis *window*.

Hasil pada *convolution layer* masuk ke lapisan selanjutnya yaitu *pooling layer*. Pada lapisan ini diambil nilai *feature map* terbaik dari tiap filter yang digunakan. Sehingga didapatkan *feature* yang paling penting dari artikel. Hasil dari *pooling layer* berupa vektor berukuran banyaknya filter yang digunakan. Vektor tersebut yang akan digunakan untuk mendapatkan *output layer* dengan jaringan terhubung penuh. *Hidden Markov Models* (HMM) digunakan untuk mendapatkan skor tiap kelas sehingga dapat diklasifikasikan dalam kategori positif, negatif atau netral. Metode pelatihan yang digunakan pada Tugas Akhir ini yaitu *mini-batch gradient descent* dengan aturan pembaharuan *Adam* (*Adaptive Moment Estimation*).

Perancangan Proses Konstruksi Jaringan Algoritma Hybrid CNN-HMM

Konstruksi model jaringan mengacu pada arsitektur CNN yang sudah dibangun sebelumnya. Model berupa persamaan antara pola fitur dengan pola fitur lainnya. Berikut penjabaran model pada setiap lapisan arsitektur jaringan.

1. Input Layer

Input layer berisi matriks berukuran $k \times 50$ dengan k merupakan jumlah kata terbanyak dalam sebuah artikel. Matriks tersebut merupakan hasil dari ekstraksi fitur yaitu merubah dari

kumpulan kata menjadi bentuk vektor yang merepresentasikan masing-masing kata tersebut. Vektor tersebut yang nantinya akan diproses pada *layer-layer* selanjutnya.

2. Convolutional layer

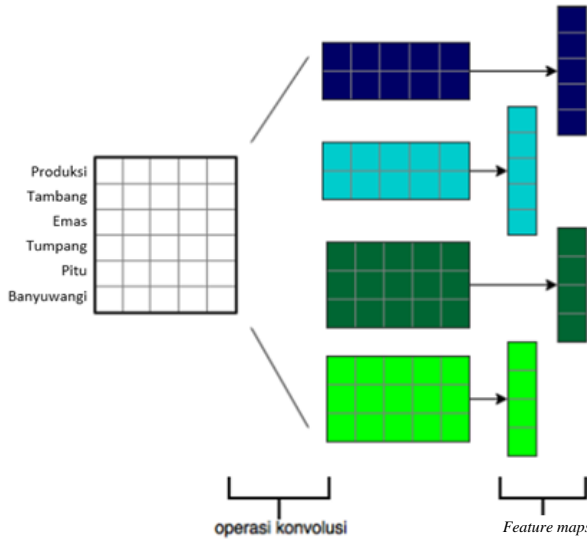
Pada layer pertama pada algoritma *convolutional neural network* dilakukan proses konvolusi untuk mendapatkan *feature map* dari inputan awal. Pada tabel 4.9 dijabarkan parameter-parameter yang digunakan pada Tugas Akhir ini:

Tabel 4. 9. Arsitektur jaringan CNN

No	Hyperparameter	Ukuran
1	Banyaknya filter	100
2	Ukuran <i>kernel</i>	5

Pemilihan ukuran *kernel* dan *filter* tidak terdapat aturan tertentu yang harus diikuti. Penulis memilih nilai tersebut berdasarkan referensi serta percobaan yang dilakukan selama melakukan penelitian.

Gambar 4.11 menjelaskan contoh mekanisme pada *convolutional layer*. Pada contoh tersebut dimisalkan satu kalimat artikel terdiri dari 6 kata dengan tiap kata direpresentasikan sebagai vektor kata. Dimensi dari vektor kata tersebut adalah 5. Vektor-vektor kata tersebut selanjutnya dilakukan proses konvolusi dengan ukuran *kernel* 2 dan 3. Tiap *kernel* diberlakukan 2 filter untuk mendapatkan *feature maps*, sehingga terdapat 4 filter yang berbeda yang digunakan. Oleh karena itu terbentuk *feature map* sebanyak 4 dimana tiap *kernel* menghasilkan 2 *feature map*.



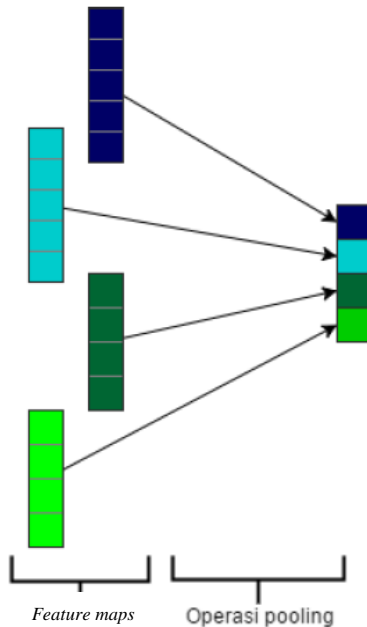
Gambar 4. 11. Contoh mekanisme pada *convolutional layer* dengan 2 *window* dimana terdapat 2 filter tiap windownya

Input bagi *convolutional layer* adalah matriks berukuran $k \times 50$ dan *output* yang akan dihasilkan adalah *feature map*. Dalam Gambar 4.11 terlihat bahwa ukuran *kernel* yang digunakan adalah 2 dan 3. Sehingga untuk kernel berukuran 2, *input* akan di proses dengan pola ('produksi' 'tambang', 'tambang' 'emas', 'emas' 'tumpang', 'tumpang' 'pitu', 'pitu' 'banyuwangi') yang diilustrasikan dengan kotak berwarna biru dengan ukuran 2×5 pada Gambar 4.11. Sedangkan untuk kernel berukuran 3, *input* akan di proses dengan pola ('produksi' 'tambang' 'emas', 'tambang' 'emas' 'tumpang', 'emas' 'tumpang' 'pitu', 'tumpang' 'pitu' 'banyuwangi') yang diilustrasikan dengan kotak berwarna hijau berukuran 3×5 pada Gambar 4.11.

Setelah dikenakan fungsi linear *ReLU*. *Feature map* siap untuk menjadi inputan bagi *pooling layer*.

3. Pooling layer

Feature map hasil dari *convolutional layer* selanjutnya akan diambil fitur yang paling penting dari vektor artikel yang menjadi *inputan*. Ilustrasi dari *pooling layer* dapat dilihat pada Gambar 4.12:



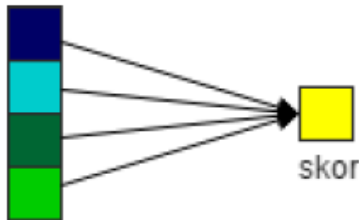
Gambar 4. 12. Ilustrasi mekanisme pada *pooling layer*

Berdasarkan *feature map* yang telah di bentuk, akan dicari nilai maksimum pada tiap *filter*. Dalam Gambar 4.12 diilustrasikan dengan tanda panah yang menunjukkan bahwa dari *feature maps* yang awalnya berjumlah 5, hanya akan diambil satu nilai saja yang maksimal. Oleh karena itu output dari proses *pooling* adalah satu nilai *feature maps* terbaik dari masing masing *filter*. Nilai terbaik tersebut akan digabungkan menjadi satu buah array yang dalam Gambar 4.12 diilustrasikan dengan 4

buah kotak (berwarna biru tua, biru muda, hijau tua, dan hijau muda). Hasil dari *pooling layer* tersebut akan masuk kelapisan berikutnya.

4. *Output layer*

Hasil dari *pooling layer* akan masuk ke dalam jaringan *output layer*. Dalam *output layer* akan dicari probabilitas terbesar dari sebuah kelas yang nantinya akan menentukan artikel tersebut masuk kedalam kelas positif, negatif, atau netral.



Feature map

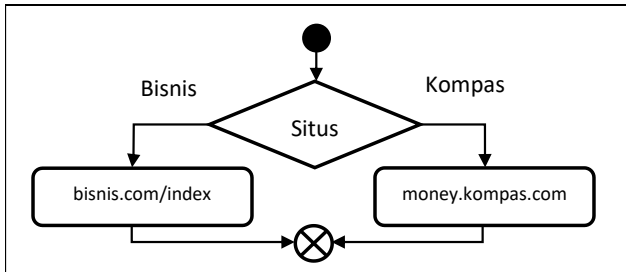
Gambar 4. 13. Contoh gambaran jaringan pada *output layer*

4.2.2.3 Perancangan Proses *Web Crawling*

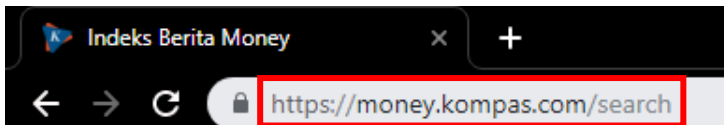
Dalam perancangan proses *web crawling* akan dilakukan empat proses untuk mendapatkan data artikel berita dari situs berita *online*. Data artikel berita tersebut akan diklasifikasikan berdasarkan *sentiment*. Berikut merupakan proses dalam *web crawling*.

1. Perancangan Proses Penentuan Situs Web

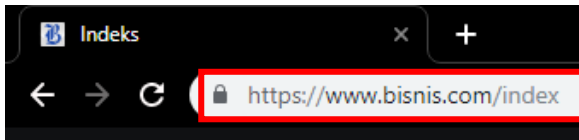
Dalam proses ini akan dilakukan pemilihan situs web mana yang akan dilakukan *web crawling* data. Pada Tugas Akhir ini terdapat tiga situs web yang akan dilakukan proses *web crawling* yaitu kompas dan bisnis.



Gambar 4. 14. Flowchart proses pemilihan situs web



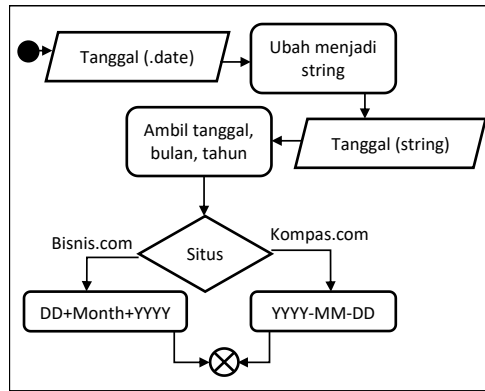
Gambar 4. 15. Situs web kompas



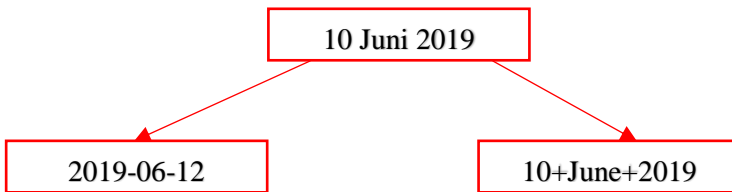
Gambar 4. 16. Situs web bisnis

2. Perancangan Proses Penentuan Tanggal Pencarian

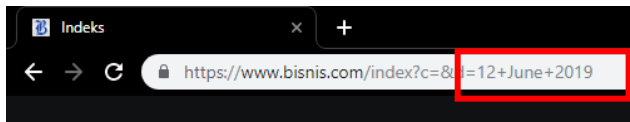
Dalam Proses ini akan dilakukan proses *input* tanggal berupa teks yang nantinya akan digunakan untuk URL pencarian seperti pada Gambar 4.18



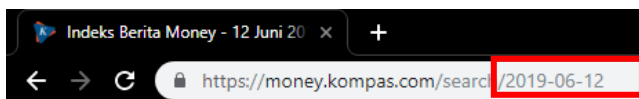
Gambar 4. 17. Flowchart proses penentuan tanggal pencarian



Gambar 4. 18. Kemungkinan pemisahan tanggal inputan untuk masing-masing situs web



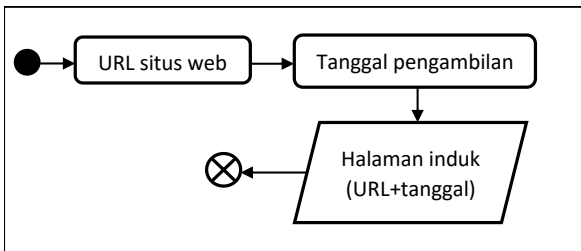
Gambar 4. 19. URL situs web bisnis setelah diberi tanggal pencarian



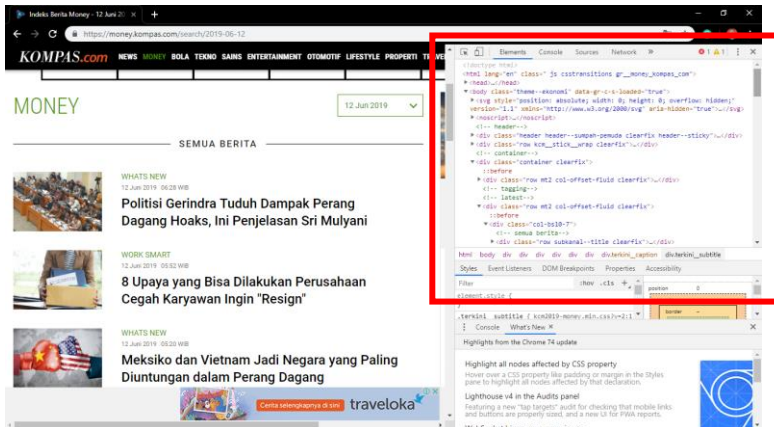
Gambar 4. 20. URL situs web kompas setelah diberi tanggal pencarian

3. Perancangan Proses Halaman Induk

Pada perancangan proses pengambilan halaman induk, URL yang sudah tersusun melalui proses pemilihan situs web dan proses penentuan tanggal pencarian akan dikunjungi. Dokumen web halaman artikel berita akan diambil. Selanjutnya, URL dari dokumen tersebut akan diambil. Dokumen web yang diambil berformat (.html) seperti pada gambar 4.22



Gambar 4. 21. Flowchart proses halaman induk

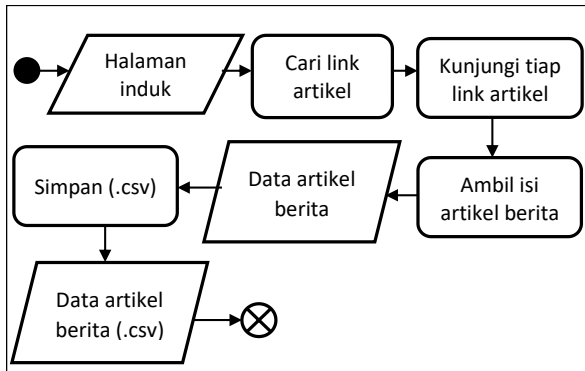


Gambar 4. 22. Dokumen web situs web berita online

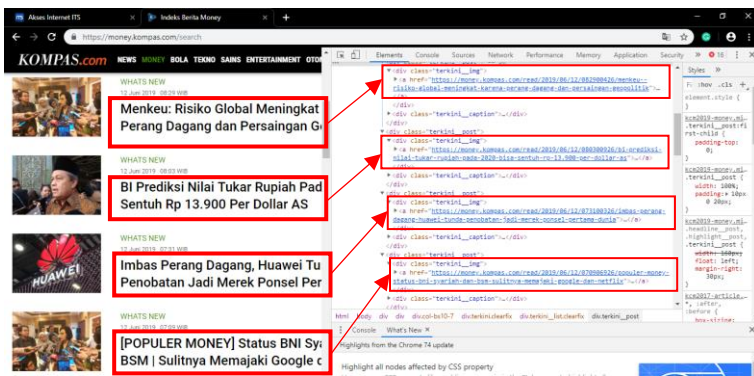
4. Perancangan Proses Pengunjungan dan Akuisisi Data

Pada perancangan proses pengunjungan dan akuisisi data akan dilakukan pengunjungan pada masing-masing halaman

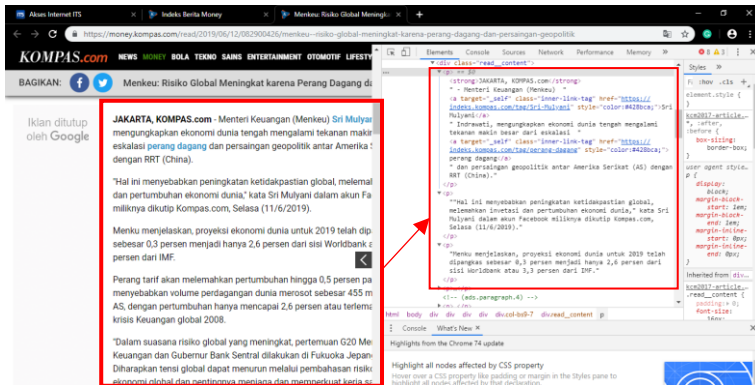
artikel berita *online*. Halaman tersebut kemudian akan dilakukan proses akuisisi data teks dari artikel berita yang telah dikunjungi.



Gambar 4. 23. Flowchart proses pengunjungan dan akuisisi data



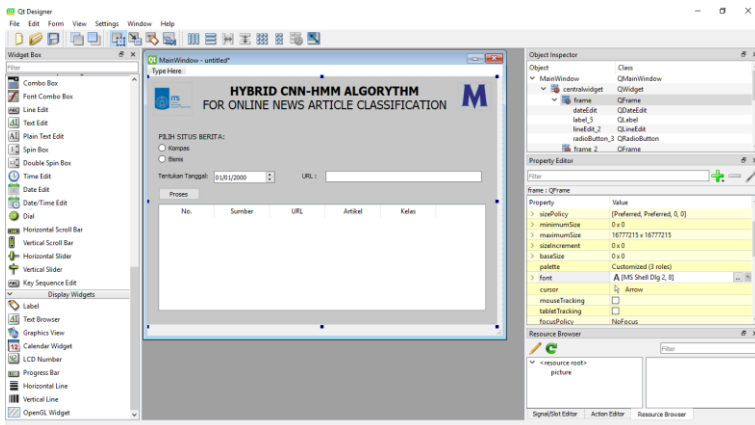
Gambar 4. 24. URL dari halaman induk situs berita *online* yang akan dikunjungi



Gambar 4. 25. Contoh halaman yang telah dikunjungi dan data yang akan diambil pada halaman tersebut

4.2.3 Perancangan Antarmuka

Desain antarmuka sistem dibutuhkan untuk memudahkan pengguna dalam mengoperasikan sistem yang dibangun. Berikut akan disajikan rancangan desain antarmuka sistem.



Gambar 4. 26. Rancangan desain antarmuka program

4.3 Penggunaan *Library*

Pada tugas akhir ini digunakan *library* yang ada pada Python sebagai pembentuk model *deep learning* yang akan dibangun. Terdapat tiga *library* utama yang digunakan pada Tugas Akhir ini, antara lain *Keras*, *Sckit-learn*, dan *Gensim*. *Library* tersebut menyediakan fungsi-fungsi untuk dapat membangun model *deep learning* maupun HMM dengan mudah.

Library Keras merupakan salah satu *library* yang khusus dibuat untuk mengolah data dengan menggunakan model *deep learning*. Selain itu untuk membentuk model jaringan HMM digunakan *library* bernama *Sckit-learn*. Sama halnya dengan *Keras*, *library* ini dibuat khusus untuk membentuk model jaringan syaraf tiruan dengan salah satu modelnya yaitu HMM. Untuk dapat merubah kata menjadi vektor maka digunakan *library Gensim*. *Library* tersebut menyediakan model *Word2vec* sehingga dapat digunakan dengan mudah untuk merubah kata menjadi representasi vektor kata.

4.4 Implementasi Sistem

Implementasi sistem merupakan bentuk realisasi dari tahap perancangan yang telah dilakukan sebelumnya. Dalam tahap ini terdapat 4 tahap implementasi yang akan dijelaskan yaitu implementasi pra-proses data, implementasi algoritma, implementasi *web crawling* dan implementasi desain antar muka.

4.4.1 Implementasi Pra-proses Data

Implementasi pra-proses data dilakukan untuk mendapatkan penerapan secara nyata dari perancangan pra-proses data pada program yang sedang dibangun.

4.4.1.1 Implementasi Persiapan Data Masukan

Implementasi data persiapan dari data masukan adalah tahap perubahan secara manual data mentah dari Bank Indonesia yang berformat (.xlsx) menjadi file berformat (.csv). Dalam tahap ini juga akan dilakukan pemilihan secara manual mengenai

komponen yang akan digunakan dalam Tugas Akhir ini yaitu artikel dan label.

Tanggal	Nama Media	Ringkasan Berita	Saitan Berita	Ruang Lingkup	Klasifikasi Berita	Label Berita
01-Jan-17	Surya	(Surabaya) Harga Beras Medium Distabilkan Rp. 9.350 per Kil	Komoditas	Local	Ekonomi	2
01-Jul-17	Kompas	(Jakarta) Konsolidasi Aset Setelah Libur Panjang. Pasar keuangan	Investasi	Nasional	Ekonomi	3
01-Jul-17	Radar	(Surabaya) AperIndo: Penjualan Lampu LED Masih Rendah. A	Industri	Local	Ekonomi	1
01-Jul-17	Radar	(Kebomas) Bisnis Ritel Tahun Ini Melambat. Ransudhan tahun	Industri	Local	Ekonomi	1
01-Jul-17	Radar	(Kebomas) Bahan Naik, Indogring Lakukan Efisiensi. Naiknya	Ekspor/Impor	Local	Ekonomi	2
01-Jul-17	Jawa Pos	(Jakarta) Harga Solut Turun Pasca-Lebaran. Memasuki masa-m	Komoditas	Nasional	Ekonomi	1
01-Jul-17	Jawa Pos	(Surabaya) Tarif Penerbangan Teredukasi Separo. Perikembang	Inflasi/harga	Local	Ekonomi	3
01-Jul-17	Jawa Pos	(Surabaya) Mandiri Gencarkan Transaksi Cashless. Bank Manc	GNNT	Local	sistem Pembayaran	2
01-Jul-17	Jawa Pos	(Jakarta) Bank Swasta Klaim Semakin Efisien. Bank swasta me	GNNT	Nasional	sistem Pembayaran	2
01-Jul-17	Jawa Pos	(Jakarta) Menilai Prospek Mata Uang Digital. Dalam beberapa	GNNT	Nasional	sistem Pembayaran	2
01-Jul-17	Jawa Pos	(Jakarta) Suku Bunga Bertahan Sampai Akhir Tahun. Bank Ind	ays Repo Rat	Nasional	sistem Pembayaran	2
01-Jul-17	Jawa Pos	(Surabaya) Hospitality Dongkrak Kinerja Industri Kertas. Indu	Komoditas	Local	Ekonomi	3
01-Jul-17	Jawa Pos	(Surabaya) Sekar Laut Ekspansi ke Eropa Timur. Produsen mak	Ekspor/Impor	Local	Ekonomi	3
01-Jul-17	Kompas	(Jakarta) Ketersediaan Beras Tahun Ini Perlu Menjadi Perhatian	Komoditas	Nasional	Ekonomi	1
03-Jul-17	Radar	(Surabaya) MNC Bank Perkuat Pasar melalui Aplikasi. Geliat b	Perbankan	Local	SSK	3
03-Jul-17	Radar	(Surabaya) Libur Lebaran Kerek Konsumsi Gasoline 24%. PT P	Komoditas	Local	Ekonomi	3
03-Jul-17	Kompas	(Jakarta) Penyelundupan Benih Lobster Terus Mingkat. Hingg	Komoditas	Nasional	Ekonomi	3
03-Jul-17	Kompas	(Jakarta) Penyelundupan Benih Lobster Terus Mingkat. Hingg	Komoditas	Nasional	SSK	3

Gambar 4. 27. Data mentah dari Bank Indonesia

A	B	C
1 Produksi Semen Melimpah. Produksi semen di Indonesia men	1	
2 Penjualan Domestik Turun, Jual via Online. PT Semen Indone	-1	
3 Terbitkan Lagi SBI. Mengantisipasi rupiah tidak melemah, BI r	1	
4 Inalun Utang Rp 55,4 T dari Luar Negeri. PT Indonesia Asahar	0	
5 Tambah Produksi Tambang Emas. Aktivitas di tambang emas	1	
6 Arus Logistik Terhambat. Penertiban persoalan truk overdim	-1	
7 Targetkan 100 Ribu Nasabah Baru. Perhelatan Asian Games 2	1	
8 Perluas ke Pasar Nontradisional. Kinerja produksi dan ekspor	1	
9 Pembak Terus Perbaiki Jalur Pedestrian. Sejumlah jalan proto	0	
10 APBD 2019 Diprediksi menurun Rp 300 Miliar. Tahun depan i	-1	
11 Rencana Investasi Hotel di Arab. Peningkatan fasilitas asrama	1	
12 Sampah yang memiliki nilai jual Itu adalah Plastik. Sampah ple	1	
13 Peluang Domestik Terabaikan. Pemerintah dinilai kurang foku	-1	
14 Pertumbuhan Ritel Harus Dipertahankan. Asosiasi Pengusaha	0	
15 Pelaksanaan Post-Tariff Jadi PR. Kesepakatan mengenai waki	0	
16 Strategi Omnichannel Bakal Makin Populer. Pelaku usaha pe	1	
17 Desain Lokal Mulai Diminati. Konsultan properti menilai saat	1	
18 Pengusaha Semen Minta Ditunda 1 Tahun. Asosiasi Semen li	0	

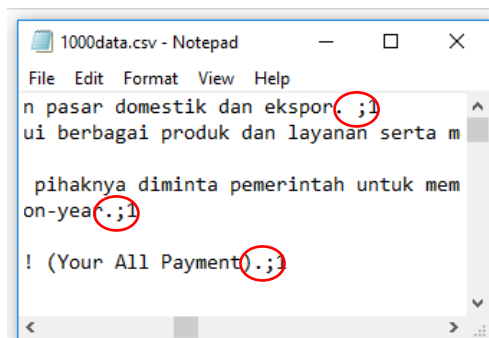
Gambar 4. 28. Data yang telah dilakukan pemilihan komponen serta peubahan format

Dalam implementasi persiapan data masukan, terdapat tahapan dimana akan dilakukan pemasukan data ke dalam program sebagai berikut.



Gambar 4. 29. Data yang akan dimasukkan ke dalam program

File yang akan digunakan dalam program ini berformat CSV (*Comma Separated Value*) dimana file jenis ini data antar kolomnya akan dipisahkan dengan tanda baca ‘;’. Nama file yang digunakan adalah ‘1000data.csv’.



Gambar 4. 30. Tanda pemisah data antar kolom pada file CSV

Code yang digunakan untuk memasukkan data ke dalam program adalah sebagai berikut:

```
data = pd.read_csv('1000data.csv',encoding = "ISO-8859-1",
sep=';', header=None)
```

Data yang telah dimasukkan ke dalam program akan berbentuk sebagai berikut:

```

0  (Jakarta) Produksi Semen Melimpah. Produksi se... 1
1  Penjualan Domestik Turun, Jual via Online. PT ... -1
2  (Jakarta) Terbitkan Lagi SBI. Mengantisipasi p... 1
3  (Jakarta) Inalut Utang Rp 55,4 T dari Luar Neg... 0
4  Tambah Produksi Tambang Emas. Aktivitas di ta... 1
5  (Jakarta) Arus Logistik Bisa Terhambat. Penert... -1
6  (Jakarta) Targetkan 100 Ribu Nasabah Baru. Per... -1
7  (Jakarta) Perluas ke Pasar Nontradisional. Kin... 1
8  (Sidoarjo) Pemkab Terus Perbaiki Jalur Pedestr... 0
9  (Sidoarjo) APBD 2019 Diprediksi Drop Rp 300 Mi... -1
10 (Surabaya) Rencana Investasi Hotel di Arab. Pe... 1
11 Sampah Mahal Itu adalah Plastik. Sampah plasti... 1
12 (Jakarta) Peluang Domestik Terabaikan. Pemerin... -1
13 (Jakarta) Pertumbuhan Ritel Harus Dipertahanka... 0
14 (Jakarta) Pelaksanaan Post-Tariff Jadi PR. Kes... 0
15 (Jakarta) Strategi Omnichannel Bakal Makin Pop... 1
16 (Jakarta) Desain Lokal Mulai Diminati. Konsult... 1
17 (Jakarta) Pengusaha Semen Minta Ditunda 1 Tahu... 0
18 (Jakarta) GPEI Minta Direct Call Diperbanyak. ... 0
19 Prahara di Tubuh Commuter Line. Hampir sebaga... -1
20 (Jakarta) Angkasa Pura II Tak Libatkan Pebisni... -1
..

```

Gambar 4. 31. Tampilan data yang telah dimasukkan dalam program

Code yang dimasukkan sebelumnya akan memisah antara data yang berisi artikel berita pada kolom ‘0’ dan data yang berisi label pada kolom ‘1’.

Data yang akan digunakan selanjutnya adalah data artikel, sehingga untuk memilih hanya data tersebut saja maka akan dimasukkan *code* sebagai berikut:

```
data_mentah=data[0]
```

Code diatas akan menyimpan data artikel ke dalam *list* bernama data_mentah.

4.4.1.2 Implementasi Proses *Case Folding*

Proses *case folding* pada artikel akan dilakukan dalam program dengan menggunakan *code* sebagai berikut:

```
#Proses case folding
artikel_lower=[]
for artikel in data_mentah:
    artikel=artikel.lower()
    artikel_lower.append(artikel)
```

Code di atas akan mengubah masing-masing artikel dalam *list* *data_mentah* menjadi artikel berbentuk *lowercase* (semua huruf kecil) yang akan disimpan dalam *list* baru bernama *artikel_lower*. Contoh input dan output dari tahap ini dapat dilihat pada tabel 4.3.

4.4.1.3 Implementasi Proses Tokenisasi

Proses tokenisasi pada artikel akan dilakukan dalam program dengan menggunakan *code* sebagai berikut:

```
#proses tokenisasi
artikel_word_tokenize=[]
for artikel in artikel_lower:
    artikel=word_tokenize(artikel)
    artikel_word_tokenize.append(artikel)
```

Code di atas akan memisahkan masing-masing unsur dalam setiap artikel baik itu kata maupun tanda baca kedalam suatu bentuk *list*. Kumpulan *list* tersebut akan disimpan dalam *list* baru bernama *artikel_word_tokenize*. Contoh *input* dan *output* dari tahap ini dapat dilihat pada tabel 4.4.

4.4.1.4 Implementasi Proses Filterisasi

Proses filterisasi pada artikel dilakukan dalam program dengan menggunakan *code* sebagai berikut:

```
#proses filterisasi
listStopwords=set(['.',',','(',')','?','/','-','?']+(stopwords.words('indonesian')))
```

```

artikel_bersihStopwords=[]
for kalimat in artikel_word_tokenize:
    t_kal=[]
    for kata in kalimat:
        if kata not in listStopwords:
            t_kal.append(kata)
    artikel_bersihStopwords.append(t_kal)

```

Code di atas akan menghilangkan kata *stopwords* dan tanda baca dari masing-masing artikel berita. Proses di atas akan melakukan pengecekan, apabila *token* tidak terdapat pada *liststopwords* maka akan dimasukkan ke dalam *list t_kal*. Kumpulan dari seluruh *list* tersebut akan disimpan ke dalam *list* baru bernama *artikel_bersihStopwords*. Contoh *input* dan *output* dari tahap ini dapat dilihat pada tabel 4.5.

4.4.1.5 Implementasi Proses *Stemming*

Proses *stemming* pada artikel akan dilakukan dalam program dengan menggunakan *code* sebagai berikut:

```

#proses stemming
artikel_stemming=[]
for kalimat in artikel_bersihStopwords:
    tampung=[]
    for kata in kalimat:
        bersih=stemmer.stem(kata)
        tampung.append(bersih)
    artikel_stemming.append(tampung)

```

Code di atas akan mengubah *token* kata dalam artikel berita menjadi kata dasar. Hasil dari proses ini akan disimpan dalam *list* baru bernama *artikel_stemming*. Contoh *input* dan *output* dari tahap ini dapat dilihat pada tabel 4.6.

4.4.1.6 Implementasi Proses Ekstraksi Fitur

Proses ekstraksi fitur yaitu pengubahan kata menjadi bentuk vektor. Pertama akan dibuat algoritma yang berfungsi merubah kata menjadi vektor-vektor kata berdasarkan hubungan kertaikatan antar kata. Sumber basis data vektor kata yang digunakan merupakan hasil dari model Gensim dengan ukuran dimensi untuk tiap kata adalah 50.

```

from gensim.models.word2vec import Word2Vec
num_features = 50
min_word_count = 5
num_workers = 2
window_size = 7
subsampling = 1e-3

model = Word2Vec(
    artikel_stemming,
    workers=num_workers,
    size=num_features,
    min_count=min_word_count,
    window=window_size,
    sample=subsampling)

vocab_w2v = "my_domain_specific_word2vec_model"
model.save(vocab_w2v)

```

Code di atas bertujuan untuk membentuk *library* vektor kata yang kemudian akan disimpan ke dalam file bernama `vocab_w2v` dengan keterangan sebagai berikut:

```
Word2Vec(vocab=2300, size=50, alpha=0.025).
```

List kata yang telah dibentuk kemudian akan diubah ke dalam bentuk vektor dengan code sebagai berikut:

```
word_vectors = Word2Vec.load(vocab_w2v)
```

```

def vectorize(dataset, maxlen):
    vectorized_data=[]
    expected=[]
    for kalimat in dataset:
        sample_vecs=[]
        for kata in kalimat:
            if kata in word_vectors:
                sample_vecs.append(word_vectors[kata])
            else:
                sample_vecs.append(np.random.uniform(0.0,0.0,50))
        if len(sample_vecs) < maxlen:
            additional_elems = maxlen - len(sample_vecs)
            for _ in range(additional_elems):
                sample_vecs.append(np.random.uniform(0.0,0.0,50))
        sample_vecs=np.array(sample_vecs)
        vectorized_data.append(sample_vecs)
    vectorized_data=np.array(vectorized_data)
    return vectorized_data

```

Kata yang terdapat dalam *library word2vec* akan langsung diubah ke dalam bentuk vektor sesuai dengan nilai yang tertera pada *library*. Namun, untuk kata yang tidak terdapat dalam *library* akan tetap diubah ke dalam bentuk vektor nol berukuran 1×50

Code di atas juga akan melakukan pengecekan, apakah panjang *array* artikel tersebut sudah sesuai dengan panjang *array* artikel maksimal atau belum. Jika belum maka *code* di atas akan melakukan penambahan *array* nol berukuran 1×50 sebanyak kekurangannya.

4.4.1.7 Implementasi Proses *Splitting* Data

Data artikel dan data label akan dipisah ke dalam data latih dan data uji dengan perbandingan 80:20 (data latih : data uji). Tidak ada ketentuan khusus dalam pemilihan perbandingan data latih dan uji, perbandingan tersebut dipilih berdasarkan referensi dari buku acuan yang digunakan oleh penulis. *Code* yang digunakan untuk membagi data tersebut adalah sebagai berikut:


```

split_point = int(len(data)*.8)
x_train = vectorize_data[:split_point]
y_train = data[1][:split_point]
x_test = vectorize_data[split_point:]
y_test = data[1][split_point:]

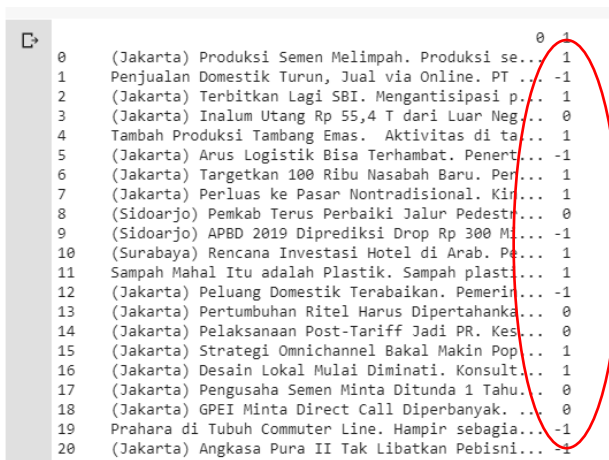
```

Data label yang telah dilakukan pembagian belum sepenuhnya siap untuk diproses ke dalam algoritma *Hybrid CNN-HMM*. Data label yang awalnya dalam bentuk angka -1 (negatif), 0 (netral), dan 1 (positif) akan diubah ke dalam bentuk array berukuran 1×3 agar dapat dilakukan pemrosesan dalam algoritma *Hybrid CNN-HMM*. *Code* yang digunakan untuk melakukan proses tersebut adalah sebagai berikut:

```

Y_train=pd.get_dummies(y_train).values
Y_test=pd.get_dummies(y_test).values

```



```

0 (Jakarta) Produksi Semen Melimpah. Produksi se... 0 1
1 Penjualan Domestik Turun, Jual via Online. PT... -1
2 (Jakarta) Terbitkan Lagi SBI. Mengantisipasi p... 1
3 (Jakarta) Inalum Utang Rp 55,4 T dari Luar Neg... 0
4 Tambah Produksi Tambang Emas. Aktivitas di ta... 1
5 (Jakarta) Arus Logistik Bisa Terhambat. Penert... -1
6 (Jakarta) Targetkan 100 Ribu Nasabah Baru. Per... 1
7 (Jakarta) Perluas ke Pasar Nontradisional. Kir... 1
8 (Sidoarjo) Pemkab Terus Perbaiki Jalur Pedestr... 0
9 (Sidoarjo) APBD 2019 Diprediksi Drop Rp 300 M... -1
10 (Surabaya) Rencana Investasi Hotel di Arab. Pa... 1
11 Sampah Mahal Itu adalah Plastik. Sampah plasti... 1
12 (Jakarta) Peluang Domestik Terabaikan. Pemerir... -1
13 (Jakarta) Pertumbuhan Ritel Harus Dipertahanka... 0
14 (Jakarta) Pelaksanaan Post-Tariff Jadi PR. Kes... 0
15 (Jakarta) Strategi Omnichannel Bakal Makin Pop... 1
16 (Jakarta) Desain Lokal Mulai Diminati. Konsult... 1
17 (Jakarta) Pengusaha Semen Minta Ditunda 1 Tahu... 0
18 (Jakarta) GPEI Minta Direct Call Diperbanyak... 0
19 Prahara di Tubuh Commuter Line. Hampir sebagia... -1
20 (Jakarta) Angkasa Pura II Tak Libatkan Pebisni... -1

```

Gambar 4. 32. Contoh data label sebelum diubah

```
array([[1, 0, 0],
       [0, 0, 1],
       [0, 0, 1],
       [0, 0, 1],
       [0, 1, 0],
       [0, 0, 1],
       [0, 1, 0],
       [1, 0, 0],
       [0, 0, 1],
       [0, 1, 0],
       [0, 0, 1],
       [0, 0, 1],
       [1, 0, 0],
       [0, 0, 1],
       [0, 0, 1],
       [0, 0, 1]])
```

Gambar 4. 33. Contoh data label setelah diubah

Data label yang telah diubah akan memiliki identitas yang berbeda dari sebelumnya, keterangan untuk data label setelah perubahan adalah sebagai berikut.

$[1, 0, 0] = \textit{negatif}$

$[0, 1, 0] = \textit{netral}$

$[0, 0, 1] = \textit{positif}$

Perubahan ini dilakukan untuk menyesuaikan kebutuhan dari *library* yang membutuhkan kelas target dalam bentuk *array*. Oleh karena itu, dalam Tugas Akhir ini, pada tahap *splitting* data dilakukan adaptasi terhadap label dari artikel berita.

4.4.2 Implementasi Algoritma Hybrid CNN-HMM

Pada tahap implementasi algoritma, terdapat proses pembentukan model dan tahap pembelajaran dengan dilakukan validasi untuk tiap iterasinya. Iterasi yang digunakan sebanyak 100 iterasi. Pembentukan model jaringan dapat diimplementasikan sebagai berikut:

```

model = Sequential()
model.add(Conv1D(filters=100,
                 strides=1,
                 kernel_size=5,
                 padding='valid',
                 activation='relu',
                 input_shape=(max_len, embedding_dims)))
model.add(MaxPooling1D())
model.add(Conv1D(filters=50,
                 strides=1,
                 kernel_size=5,
                 padding='valid',
                 activation='relu'))
model.add(MaxPooling1D())
model.add(Flatten())
model.add(Dense(int(hidden_dims), activation='relu'))
model.add(Dense(int(hidden_dims)/2), activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(3, activation='sigmoid'))
model.compile(loss='categorical_crossentropy', optimizer='adam',
              metrics=['accuracy'])

```

Selanjutnya dilakukan proses pembelajaran dengan menggunakan data yang sudah dibentuk. Implementasi tersebut dapat buat sebagai berikut:

```

model.fit(x_train, Y_train,
         batch_size=batch_size,
         epochs=1000,
         validation_data=(x_test, Y_test))

```

Inputan bagi *Layer output* dari proses CNN diatas akan dimasukkan ke dalam algoritma HMM dengan cara mengambil hasil dari layer tersebut. Implementasi dari proses tersebut adalah:

```

def get_kernels_output(model,
                       layer_index,
                       model_input,

```

```

training_flag=True):
get_outputs = K.function([model.layers[0].input,
                          K.learning_phase()],
                          [model.layers[layer_index].output])
kernel_output = get_outputs([model_input, training_flag])[0]
return kernel_output

```

```

output_train = get_kernels_output(model, layer_index=8,
model_input=x_train)
output_test = get_kernels_output(model, layer_index=8,
model_input=x_test)

```

Setelah didapatkan hasil dari layer tersebut, maka akan dimasukkan ke dalam modul utama HMM sebagai berikut:

```

from hmm_filter.hmm_filter import HMMFilter

hmmfilter = HMMFilter()
hmmfilter.fit(latih, session_column="session",
prediction_column="label")

df_uji = hmmfilter.predict(uji, session_column='session',
probabs_column="probabs", prediction_column='prediksi')
df_latih = hmmfilter.predict(latih, session_column='session',
probabs_column="probabs", prediction_column='prediksi')

# akurasi uji
hmm_accuracy_uji = len(df_uji[df_uji.label == df_uji.prediksi]) /
len(df_uji)
hmm_accuracy_uji

# akurasi latih
hmm_accuracy_latih = len(df_latih[df_latih.label ==
df_latih.prediksi]) / len(df_latih)
hmm_accuracy_latih

```

4.4.3 Implementasi Proses *Web Crawling*

Pada tahap implementasi *web crawling*, akan dilakukan proses pengambilan data dari situs berita *online*:

4.4.3.1 Penentuan Situs Web

Situs dari portal berita *online* yang akan dikunjungi akan dipilih terlebih dahulu dengan menggunakan *radio button* pada GUI. Situs berita yang tersedia yaitu kompas.com dan bisnis.com

4.4.3.2 Penentuan Tanggal Pencarian

Tanggal akan dimasukkan sebagai inputan untuk melengkapi URL situs web agar spesifik mencari data artikel berita pada tanggal tertentu. Dengan menggunakan *code* berikut proses tersebut dieksekusi.

```

tgl = datetime.datetime.strptime(tanggal, "%d/%m/%Y")
if tgl.day < 10:
    day='0'+str(tgl.day)
else:
    day=str(tgl.day)

if tgl.month < 10:
    month = '0'+str(tgl.month)
else:
    month = str(tgl.month)

year = str(tgl.year)

```

4.4.3.3 Halaman Induk

Format tanggal yang sudah sesuai kemudian akan dimasukkan ke dalam URL situs web yang belum lengkap. Sehingga setelah dimasukkan format tanggal, akan menjadi halaman induk dari portal berita yang akan dilakukan *web crawling*. *Code* yang digunakan adalah sebagai berikut.

```

key_tgl='https://money.kompas.com/search/{}-{}-{}'.format(year,
month, day)

```

4.4.3.4 Pengunjungan dan Akuisisi Data

URL halaman induk yang telah didapat, kemudian akan dilakukan pengaksesan yang nantinya akan dicara data yang sesuai dengan diharapkan. *Code* yang digunakan adalah sebagai berikut.

```
hasil=[]
req=requests.get(key_tgl)
soup=BeautifulSoup(req.text, 'lxml')
kumpulan_page=soup.find_all('div', class_='terkini__post')
link_page=[kumpulan_page[i].a['href'] for i in
range(len(kumpulan_page))]
for i in link_page:
    item={ }
    req_berita=requests.get(i)
    soup_berita=BeautifulSoup(req_berita.text, 'lxml')
    item['URL']=i
    item['judul']=soup_berita.find_all('title')[0].text

    berita=soup_berita.find('div', class_='read__content')
    try:
        test=berita.text.strip()
    except:
        test=""
    item['paragraf']=test

    hasil.append(item)
```

	URL	Judul	paragraf
0	https://money.kompas.com/read/2019/06/13/21300...	Bappenas: Indonesia Bakal Jadi Negara Berpenda...	JAKARTA, KOMPAS.com - Menteri Perencanaan Pemb...
1	https://money.kompas.com/read/2019/06/13/21080...	KPPU Putuskan Perkara Dugaan Persekongkolan Te...	JAKARTA, KOMPAS.com - Komisi Pengawas Persaing...
2	https://money.kompas.com/read/2019/06/13/20530...	Indonesia Titi Cara Australia untuk Bangun SDM	JAKARTA, KOMPAS.com - Tahun 2019, masalah sumb...
3	https://money.kompas.com/read/2019/06/13/20470...	Agar Pertumbuhan Ekonomi Tak Stagnan di 5,3 Pe...	JAKARTA, KOMPAS.com - Kementerian Perencanaan ...
4	https://money.kompas.com/read/2019/06/13/20320...	Kemenhub Bantah Kabar China Airlines Layani Pe...	JAKARTA, KOMPAS.com - Di media sosial beredar ...
5	https://money.kompas.com/read/2019/06/13/20164...	Peraturan "Insentif Pajak Super" untuk Pengusa...	JAKARTA, KOMPAS.com - Menteri Perindustrian A...
6	https://money.kompas.com/read/2019/06/13/20133...	Kemenhub: Jangan Sampai Ada "Predatory Pricing...	JAKARTA, KOMPAS.com - Kementerian Perhubungan...
7	https://money.kompas.com/read/2019/06/13/20020...	QRIM Express Buka Kemitraan Gerai, Ini Syarat...	JAKARTA, KOMPAS.com - Perusahaan logistik QRI...
8	https://money.kompas.com/read/2019/06/13/19310...	Bi: Dari Awal Tahun, Rupiah Menguat 1,02 Persen	JAKARTA, KOMPAS.com - Bank Indonesia menyatak...
9	https://money.kompas.com/read/2019/06/13/18380...	Fitur Selfie Diberlakukan, Grab Ungkap Tak Ada...	JAKARTA, KOMPAS.com - Sejak meluncurkan fitur...
10	https://money.kompas.com/read/2019/06/13/18175...	Tarif Baru Ojek Online Diterapkan di Seluruh I...	JAKARTA, KOMPAS.com - Peraturan Menteri Perhub...
11	https://money.kompas.com/read/2019/06/13/17445...	Pertumbuhan Ekonomi 5,6 Persen, Butuh Investas...	JAKARTA, KOMPAS.com - Pemerintah menargetkan ...
12	https://money.kompas.com/read/2019/06/13/17373...	QRIM Express Manfaatkan Tol Trans Jawa untuk T...	JAKARTA, KOMPAS.com - Mahalnya harga pengirim...
13	https://money.kompas.com/read/2019/06/13/17320...	Kemenhub Hapus Diskon Ojek Online. Ini Komenta...	JAKARTA, KOMPAS.com - Presiden Grab Indonesia...
14	https://money.kompas.com/read/2019/06/13/17090...	Instansi Pemerintah Pusat dan Daerah Diminta H...	JAKARTA, KOMPAS.com - Instansi di pemerintah p...
15	https://money.kompas.com/read/2019/06/13/16460...	Sebulan, Donasi Digital di 3 Platform Ini Tela...	JAKARTA, KOMPAS.com - Program donasi digital ...
16	https://money.kompas.com/read/2019/06/13/16315...	Kemenhub Batal Larang Diskon Tarif Ojek Online	JAKARTA, KOMPAS.com - Kementerian Perhubungan...
17	https://money.kompas.com/read/2019/06/13/16030...	Penggunaan Biodiesel B30 Bisa Hemat Impor Sola...	JAKARTA, KOMPAS.com - Kepala Badan Penelitian ...
18	https://money.kompas.com/read/2019/06/13/15344...	Diabut Punya Utang Rp 614 Triliun, Ini Penjel...	JAKARTA, KOMPAS.com - Lion Air Group memberik...
19	https://money.kompas.com/read/2019/06/13/15314...	Hanya 1 dari 8 CEO Indonesia yang Yakini Blaris...	JAKARTA, KOMPAS.com - Para pemimpin perusahaan...

Gambar 4. 34. Contoh hasil akuisisi data dengan web crawling

Gambar 4.34 menampilkan hasil akuisisi data untuk tanggal 13 Juni 2019 dari situs kompas.com. dan diperoleh hasil sebanyak 20 artikel terakuisisi oleh program yang ditunjukkan dengan banyaknya baris sejumlah 20. Gambar 4.34 memiliki 3 buah kolom utama yaitu URL, judul, dan paragraph. Kolom URL berisi seluruh url artikel berita yang tersedia pada tanggal 13 Juni 2019, kolom judul berisi seluruh judul artikel berita yang teredia pada tanggal 13 Juni 2019, dan kolom paragraph berisi seluruh isi artikel berita yang tersedia pada tanggal 13 Juni 2019.

BAB V

UJI COBA DAN EVALUASI SISTEM

Pada bab ini dijelaskan tahap-tahap uji coba berdasarkan implementasi sistem yang telah dibuat. Hasil uji coba akan dianalisis melalui proses verifikasi dan validasi sehingga dapat melakukan evaluasi pada sistem.

5.1 Uji Coba Algoritma Hybrid CNN-HMM

Berikut akan dibahas mengenai uji coba yang dilakukan terhadap algoritma Hybrid CNN-HMM dalam melakukan *sentiment analysis*.

5.1.1 Data Uji Coba Algoritma

Data artikel berita pada Tugas Akhir ini sebanyak 1.760 data. Dari data tersebut dapat dibagi menjadi tiga jenis data yaitu data positif, data negatif dan data netral. Data positif didasarkan pada *tone* artikel yang bernilai 3, data netral didasarkan pada *tone* artikel yang bernilai 2, dan untuk data negatif didasarkan pada *tone* artikel yang bernilai 1. Rincian data dapat dilihat pada Tabel 5.1:

Tabel 5. 1. Rincian data

Jenis Data	Jumlah data
Data artikel positif	770
Data artikel netral	580
Data artikel negatif	410
Total data	1.760

Data tersebut akan dibagi menjadi data latih dan data uji dengan perbandingan sebanyak 80:20. Rincian tentang pembagian data latih dan uji dapat dilihat pada Tabel 5.2:

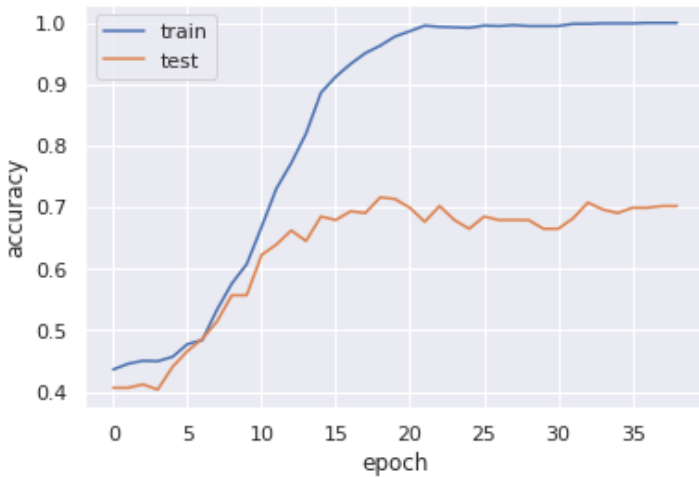
Tabel 5. 2. Data yang digunakan dalam proses komputasi

Jenis Data	Jumlah data
Data latih	1408
Data uji	352
Total data	1760

5.1.2 Hasil Uji Coba Algoritma

Setelah dilakukan proses pembelajaran maka didapatkan model yang telah menyesuaikan data yang dimasukkan. Model tersebut perlu diuji akurasi sehingga dapat disimpulkan bahwa model yang dihasilkan merupakan model yang sesuai atau tidak.

Pengujian/validasi dilakukan dengan menggunakan data uji yang telah dikelompokkan sebelumnya. Data uji dan data latih merupakan data yang saling independen sehingga dapat dilihat kemampuan model untuk menangani data yang baru. Berdasarkan hasil implementasi yang dilakukan didapatkan keakurasian tiap iterasi untuk tiap jenis data. Dalam Tugas Akhir ini akan dilakukan 5 kali uji coba dengan jumlah iterasi maksimal sebanyak 500. Pada Gambar 5.1 menunjukkan akurasi model selama proses pembelajaran yang dilakukan pada percobaan pertama.



Gambar 5. 1. Grafik pergerakan akurasi model pada percobaan pertama

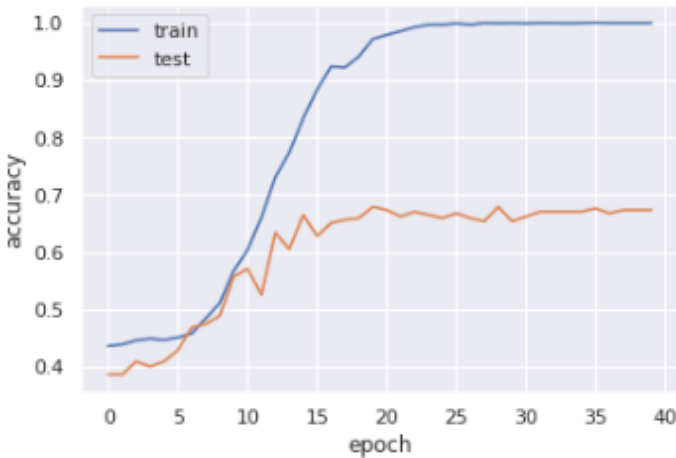
Akurasi model yang diperoleh adalah sebesar 99,14% untuk data latih dan 71,59% untuk data uji. Hasil tersebut diperoleh pada iterasi ke 20. Karena hingga iterasi ke 40 tidak terjadi penambahan akurasi pada data uji, sehingga pembelajaran dihentikan.

Hasil yang telah didapat tersebut, akan dilakukan klasifikasi ulang menggunakan algoritma HMM dan diperoleh hasil pada Tabel 5.3. Kolom Akurasi menunjukkan akurasi dari model setelah ditambahkan algoritma HMM, sedangkan kolom peningkatan menunjukkan peningkatan akurasi jika dibandingkan dengan model sebelum ditambahkan algoritma HMM.

Tabel 5. 3. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan pertama)

Jenis Data	Akurasi HMM	Peningkatan
Data latih	99,92%	0,78%
Data uji	71,87%	0,28%

Pada Gambar 5.2 menunjukkan akurasi model selama proses pembelajaran yang dilakukan pada percobaan kedua



Gambar 5. 2. Grafik pergerakan akurasi model pada percobaan kedua

Akurasi model yang diperoleh adalah sebesar 98,50% untuk data latih dan 67,04% untuk data uji. Hasil tersebut diperoleh pada iterasi ke 21. Karena hingga iterasi ke 41 tidak terjadi penambahan akurasi pada data uji, sehingga pembelajaran dihentikan.

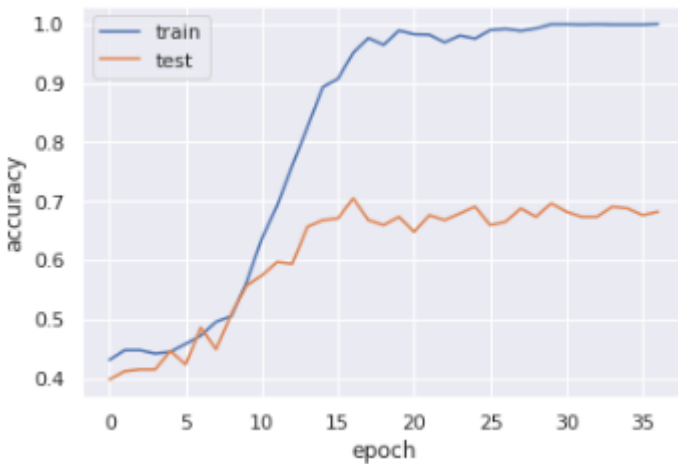
Hasil yang telah didapat tersebut, akan dilakukan klasifikasi ulang menggunakan algoritma HMM dan diperoleh hasil pada Tabel 5.4. Kolom Akurasi menunjukkan akurasi dari model setelah ditambahkan algoritma HMM, sedangkan kolom

peningkatan menunjukkan peningkatan akurasi jika dibandingkan dengan model sebelum ditambahkan algoritma HMM.

Tabel 5. 4. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan kedua)

Jenis Data	Akurasi HMM	Peningkatan
Data latih	99,92%	0,42%
Data uji	67,61%	0,47%

Pada Gambar 5.3 menunjukkan akurasi model selama proses pembelajaran yang dilakukan pada percobaan ketiga



Gambar 5. 3. Grafik pergerakan akurasi model pada percobaan ketiga

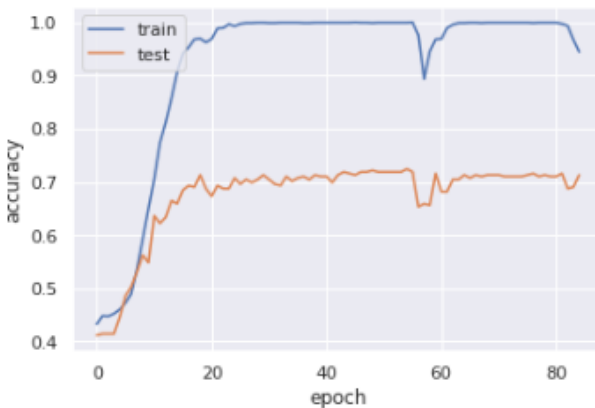
Akurasi model yang diperoleh adalah sebesar 98,93% untuk data latih dan 68,45% untuk data uji. Hasil tersebut diperoleh pada iterasi ke 17. Karena hingga iterasi ke 37 tidak terjadi penambahan akurasi pada data uji, sehingga pembelajaran dihentikan.

Hasil yang telah didapat tersebut, akan dilakukan klasifikasi ulang menggunakan algoritma HMM dan diperoleh hasil pada Tabel 5.5. Kolom Akurasi menunjukkan akurasi dari model setelah ditambahkan algoritma HMM, sedangkan kolom peningkatan menunjukkan peningkatan akurasi jika dibandingkan dengan model sebelum ditambahkan algoritma HMM.

Tabel 5. 5. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan ketiga)

Jenis Data	Akurasi HMM	Peningkatan
Data latih	99,78%	0,85%
Data uji	70,45%	2%

Pada Gambar 5.4 menunjukkan akurasi model selama proses pembelajaran yang dilakukan pada percobaan keempat.



Gambar 5. 4. Grafik pergerakan akurasi model pada percobaan keempat

Akurasi model yang diperoleh adalah sebesar 99,92% untuk data latih dan 72,44% untuk data uji. Hasil tersebut

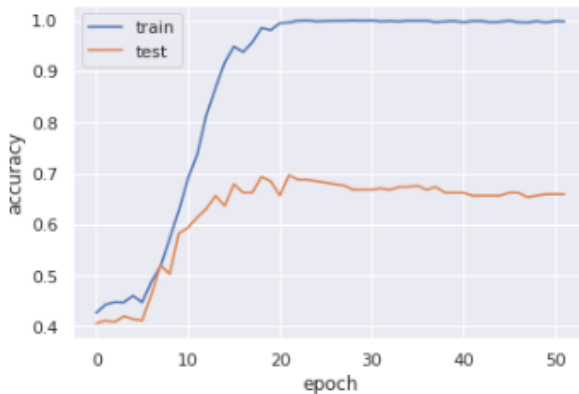
diperoleh pada iterasi ke 65. Karena hingga iterasi ke 85 tidak terjadi penambahan akurasi pada data uji, sehingga pembelajaran dihentikan.

Hasil yang telah didapat tersebut, akan dilakukan klasifikasi ulang menggunakan algoritma HMM dan diperoleh hasil pada Tabel 5.6. Kolom Akurasi menunjukkan akurasi dari model setelah ditambahkan algoritma HMM, sedangkan kolom peningkatan menunjukkan peningkatan akurasi jika dibandingkan dengan model sebelum ditambahkan algoritma HMM.

Tabel 5. 6. . Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan keempat)

Jenis Data	Akurasi HMM	Peningkatan
Data latih	100%	0,08%
Data uji	72,72%	0,28%

Pada Gambar 5.5 menunjukkan akurasi model selama proses pembelajaran yang dilakukan pada percobaan keempat.



Gambar 5. 5 Grafik pergerakan akurasi model pada percobaan kelima

Akurasi model yang diperoleh adalah sebesar 99,92% untuk data latih dan 69,60% untuk data uji. Hasil tersebut diperoleh pada iterasi ke 33. Karena hingga iterasi ke 53 tidak terjadi penambahan akurasi pada data uji, sehingga pembelajaran dihentikan.

Hasil yang telah didapat tersebut, akan dilakukan klasifikasi ulang menggunakan algoritma HMM dan diperoleh hasil pada Tabel 5.7. Kolom Akurasi menunjukkan akurasi dari model setelah ditambahkan algoritma HMM, sedangkan kolom peningkatan menunjukkan peningkatan akurasi jika dibandingkan dengan model sebelum ditambahkan algoritma HMM.

Tabel 5. 7. Hasil akurasi program setelah ditambahkan algoritma HMM (percobaan kelima)

Jenis Data	Akurasi HMM	Peningkatan
Data latih	100%	0,08%
Data uji	69,60%	0%

Berdasarkan 5 kali hasil percobaan, maka algoritma Hybrid CNN-HMM menghasilkan rata-rata akurasi sebesar 99,92% untuk data latih dan 70,45% untuk data uji.

Tabel 5.8 menunjukkan nilai *precision* dan *recall* pada masing-masing percobaan adalah sebagai berikut:

Tabel 5. 8. Nilai precision dan recall pada tiap percobaan

Percobaan Ke-	Precision	Recall
1	90,19%	86,79%
2	86,86%	76,78%
3	90%	77,14%
4	90,65%	84,34%
5	89,69%	94,56%

Jika dilihat hasil saat pembelajaran dan pengujian, dapat terlihat bahwa akurasi data uji tidak sebaik data latih. Hal tersebut menjadi ciri-ciri bahwa model mengalami *overfitting*. *Overfitting* merupakan suatu kejadian dimana model terlalu baik dalam mengklasifikasi data pembelajaran akan tetapi buruk dalam mengklasifikasi data uji. Walaupun model telah dibentuk dengan menggabungkan algoritma CNN dan algoritma HMM dengan harapan mendapatkan hasil yang lebih baik, namun *overfitting* tetap dialami oleh program. Permasalahan *overfitting* merupakan permasalahan yang umum terjadi pada klasifikasi menggunakan jaringan syaraf tiruan. Terdapat banyak hal yang menyebabkan *overfitting* antara lain jumlah data yang terlalu sedikit, jumlah variabel input yang terlalu sedikit atau kualitas data yang kurang (*outlier* data, *noise*, konsistensi, korelasi, *input* dan *output*, atau sebaran data).

5.2 Uji Coba Metode Web Crawling

Berikut akan dibahas mengenai uji coba yang dilakukan terhadap metode *web crawling* dalam melakukan pengambilan informasi dari situs portal berita *online*.

5.2.1 Data Uji Coba Metode Web Crawling

Data uji coba yang digunakan dalam Tugas Akhir ini adalah Seluruh artikel berita pada halaman pertama masing-masing situs portal berita *online* Kompas dan Bisnis selama bulan Juni dengan total data artikel berita yang tersedia sebanyak 836 data. Berikut rincian data per hari untuk masing-masing portal berita *online*. Pada Tabel 5.9, kolom jumlah berita kompas.com tersedia menunjukkan banyaknya artikel berita yang tersedia pada situs kompas.com bidang ekonomi dan keuangan pada tanggal yang telah ditentukan, sedangkan kolom jumlah berita bisnis.com tersedia menunjukkan banyaknya artikel berita yang tersedia pada situs bisnis.com bidang ekonomi dan keuangan pada tanggal yang telah ditentukan.

Tabel 5. 9. Rincian jumlah data artikel yang tersedia selama bulan Juni

Tanggal	Jumlah Berita Kompas.com tersedia	Jumlah Berita Bisnis.com tersedia
1	18	13
2	10	9
3	20	10
4	20	6
5	19	8
6	10	19
7	13	10
8	13	4
9	16	0
10	20	10
11	20	11
12	20	10
13	20	12
14	20	16
15	15	5
16	12	3
17	20	6
18	20	14
19	20	14
20	20	19
21	20	15
22	20	12
23	17	1
24	20	9
25	20	13
26	20	13
27	20	15
28	20	11
29	19	3
30	19	5

5.2.2 Hasil Uji Coba Metode *Web Crawling*

Dalam Tugas Akhir ini dilakukan uji coba pengambilan data pada situs portal berita *online* dengan menggunakan metode *web crawling*. Situs portal berita *online* yang dilakukan *web crawling* adalah situs *kompas.com* dan *bisnis.com*. Pengambilan data dilakukan dengan parameter hari dan halaman pertama pada tiap portal berita *online*. Parameter hari dipilih agar nantinya dapat diketahui bagaimana *sentiment analysis* muncul perharinya, sedangkan parameter halaman pertama dipilih karena terdapat perbedaan pola URL untuk masing-masing situs web, sehingga untuk memudahkan proses akuisisi data maka dalam Tugas Akhir ini hanya digunakan halaman pertama dari masing-masing situs web saja. Jumlah data yang berhasil dikumpulkan sebanyak 835 data. Berikut rincian data terkumpul per hari untuk masing-masing portal berita *online*. Pada Tabel 5.10, kolom jumlah berita *kompas.com* terakuisisi menunjukkan banyaknya artikel berita pada situs *kompas.com* bidang ekonomi dan keuangan yang berhasil diakuisisi oleh program pada tanggal yang telah ditentukan, sedangkan kolom jumlah berita *bisnis.com* terakuisisi menunjukkan banyaknya artikel berita yang tersedia pada situs *bisnis.com* bidang ekonomi dan keuangan yang berhasil diakuisisi oleh program pada tanggal yang telah ditentukan.

Tabel 5. 10. Rincian jumlah data artikel yang terkumpul selama bulan Juni

Tanggal	Jumlah Berita Kompas.com terakuisisi	Jumlah Berita Bisnis.com terakuisisi
1	18	13
2	10	9
3	20	10
4	20	6
5	19	8
6	10	19
7	13	10
8	13	4

9	16	0
10	20	10
11	20	11
12	19	10
13	20	12
14	20	16
15	15	5
16	12	3
17	20	6
18	20	14
19	20	14
20	20	19
21	20	15
22	20	12
23	17	1
24	20	9
25	20	13
26	20	13
27	20	15
28	20	11
29	19	3
30	19	5

Dengan menggunakan *web crawling*, program telah melakukan akuisisi seluruh data artikel berita yang tersedia pada situs kompas.com dan bisnis.com selama bulan Juni 2019. Data artikel yang berhasil diakuisisi oleh program adalah URL artikel berita, judul artikel berita, dan isi artikel berita. Hasil akuisisi data akan disajikan pada Gambar 5.6 dan Gambar 5.7. Proses *web crawling* berhasil mendapatkan 835 data dari 836 data yang tersedia, sehingga didapatkan ketercapaian dari metode tersebut sebesar 99,88%. Dengan hasil demikian, dapat disimpulkan bahwa metode *web crawling* efektif untuk melakukan pengambilan informasi dari situs portal berita *online*. Berikut akan disajikan contoh hasil dari *web crawling* pada masing-masing situs.

	URL	Judul	paragraf
0	https://money.kompas.com/read/2019/06/12/20585...	KKP Akhirnya Dapat WTP, Menteri Susi Akui Deg...	JAKARTA, KOMPAS.com - Laporan keuangan Kement...
1	https://money.kompas.com/read/2019/06/12/20540...	Pengusaha Yakini Pembangunan Infrastruktur Bisa...	JAKARTA, KOMPAS.com - Ketua Kamar Dagang dan I...
2	https://money.kompas.com/read/2019/06/12/20140...	Usai Impor 250 Ton, Kemendag Kembali Keluar...	JAKARTA, KOMPAS.com - Direktur Jenderal Perdag...
3	https://money.kompas.com/read/2019/06/12/20081...	Traveloka Serius Garap Bisnis Hiburan dan Gaya...	KOMPAS.com - Perusahaan teknologi penyedia jas...
4	https://money.kompas.com/read/2019/06/12/19452...	Sri Mulyani Yakini BI Bakal Turunkan Suku Bunga...	JAKARTA, KOMPAS.com - Pekan depan, Bank Indone...
5	https://money.kompas.com/read/2019/06/12/19250...	Tiket Mahal Pesawat adalah Puncak Gunung Es	JAKARTA, KOMPAS.com - Pengamat penerbangan se...
6	https://money.kompas.com/read/2019/06/12/19010...	Tiket Pesawat Mahal, Sri Mulyani Minta ASN Ham...	JAKARTA, KOMPAS.com - Tarif tiket pesawat yang...
7	https://money.kompas.com/read/2019/06/12/18321...	Chappy Hakim: Jangan Terlalu Cepat Undang Mask...	JAKARTA, KOMPAS.com - Ide mengundang maskapai...
8	https://money.kompas.com/jeotren-dan-tips-bis...	Tren dan Tips Bisnis Jastip Raup Rupiah - JEO ...	None
9	https://money.kompas.com/read/2019/06/12/17540...	Laporan Keuangan Pemerintah WTP Lagi, Sri Muly...	JAKARTA, KOMPAS.com - Badan Pemeriksa Keuangan...
10	https://money.kompas.com/read/2019/06/12/17433...	Lagi, Pemerintah Pulangkan Pekerja Migran Berm...	KOMPAS.com - Pemerintah Indonesia melalui KBRI...
11	https://money.kompas.com/read/2019/06/12/17380...	Soal Aturan Diskon Tarif Transportasi Online, ...	JAKARTA, KOMPAS.com - Pemerintah melalui Kamen...
12	https://money.kompas.com/read/2019/06/12/17232...	Pemerintah Disarankan Cari Pasar untuk Pelabuh...	JAKARTA, KOMPAS.com - Pemerintah disarankan un...
13	https://money.kompas.com/read/2019/06/12/17220...	Trik Atur Keuangan Biar Enggak "Tekor" Tengah ...	KOMPAS.com - Baru tengah bulan, namun Sinta su...
14	https://money.kompas.com/read/2019/06/12/16511...	Usai Diakuisisi, Bank Royal Akan Dimergeng deng...	JAKARTA, KOMPAS.com - PT Bank Central Asia Tbk...
15	https://money.kompas.com/read/2019/06/12/16200...	Soal Aturan Diskon Tarif, Ini Kata Pengemudi O...	JAKARTA, KOMPAS.com - Kementerian Perhubungan...
16	https://money.kompas.com/read/2019/06/12/15570...	Imbas Krisis 737 Max, Boeing Tak Dapat Order 2...	NEW YORK, KOMPAS.com - Pabrik pesawat Boein...
17	https://money.kompas.com/read/2019/06/12/15290...	Usai Lebaran, Harga Cabai Merah Masih Tinggi	JAKARTA, KOMPAS.com - Direktur Jenderal Perdag...
18	https://money.kompas.com/read/2019/06/12/15083...	Laporan Keuangan KKP Raih Opini Wajar Tanpa Pa...	JAKARTA, KOMPAS.com - Kementerian Kelautan dan...
19	https://money.kompas.com/read/2019/06/12/14160...	Ekspor Turun, Indonesia Cari Peluang di Tengah...	JAKARTA, KOMPAS.com - Defisit neraca perdagang...

Gambar 5. 6. Hasil web crawling kompas.com

	URL	Judul	paragraf
0	https://finansial.bisnis.com/read/20190612/215...	Soal Regulasi PAYDI, AAUI Minta Syarat Aktuari...	Bisnis.com, JAKARTA – Asosiasi Asuransi Umum I...
1	https://finansial.bisnis.com/read/20190612/215...	AAJ: Instrumen Investasi Jangka Panjang dapat...	Bisnis.com, JAKARTA — Penempatan investasi pad...
2	https://finansial.bisnis.com/read/20190612/215...	Regulasi PAYDI, Syarat Aktuaris Dinilai Wajib	Bisnis.com, JAKARTA – Kewajiban adanya ahli ak...
3	https://finansial.bisnis.com/read/20190612/215...	Kolaborasi Kunci Utama Bagi Insurtech	Bisnis.com, JAKARTA – Kolaborasi menjadi tren...
4	https://finansial.bisnis.com/read/20190612/90/...	Saham Bank Permata (BNLI) Masih Menjadi Incara...	Bisnis.com, JAKARTA – Saham PT Bank Permata Tb...
5	https://finansial.bisnis.com/read/20190612/90/...	Suku Bunga Berpotensi Turun, Ini Kata Bos BCA	Bisnis.com, JAKARTA - PT Bank Central Tbk. me...
6	https://finansial.bisnis.com/read/20190612/90/...	5 Berita Populer Finansial, BCA Kaji Akuisisi ...	1. Usai Caplok Bank Royal, BCA Incar Bank Keci...
7	https://finansial.bisnis.com/read/20190612/215...	Pasokan Aktuaris Asuransi Cukup	Bisnis.com, JAKARTA – Pasokan tenaga ahli aktu...
8	https://finansial.bisnis.com/read/20190612/215...	Transformasi Industri Asuransi, Langkah Mengga...	Bisnis.com, JAKARTA — Transformasi dan pembeng...
9	https://finansial.bisnis.com/read/20190612/90/...	CIMB Niaga Raih Penghargaan The Credit Risk Te...	Bisnis.com, JAKARTA - PT Bank CIMB Niaga Tbk. ...

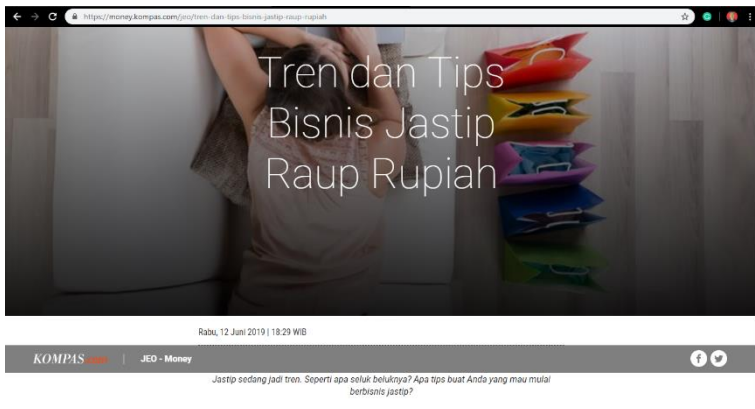
Gambar 5. 7. Hasil web crawling bisnis.com

Pada Gambar 5.6 dan Gambar 5.7, jumlah baris pada masing-masing gambar menunjukkan banyaknya artikel berita yang berhasil dilakukan *web crawling* oleh program. Terdapat pula tiga buah kolom yaitu “URL”, “Judul”, dan “Paragraf”. Berikut akan disajikan keterangan bagi masing-masing kolom.

URL : Kolom ini berisikan seluruh link dari setiap artikel berita yang tersedia pada halaman pertama situs berita *online* pada hari yang telah ditentukan.

- Judul** : Kolom ini berisikan seluruh judul artikel berita yang tersedia pada halaman pertama situs berita online dengan hari yang telah ditentukan.
- Paragraf** : Kolom ini berisikan seluruh isi artikel berita yang tersedia pada halaman pertama situs berita *online* dengan hari yang telah ditentukan.

Satu artikel berita yang tidak berhasil dilakukan pengumpulan terjadi pada tanggal 12 Juni 2019. Hal ini disebabkan karena struktur halaman pada url yang tersedia terdapat perbedaan dibandingkan dengan lainnya. Artikel berita yang gagal dilakukan pengambilan disajikan pada Gambar 5.8 dan Gambar 5.9.



Gambar 5. 8. Halaman artikel berita yang gagal dilakukan web crawling



Gambar 5. 9. Halaman artikel berita yang gagal dilakukan *web crawling*

5. 3 Hasil Uji Coba Prediksi *Sentiment Analysis*

Dalam Tugas Akhir ini data hasil *web crawling* dari situs berita *online* yang telah didapat akan dilakukan pemrosesan sehingga akan ditampilkan prediksi hasil *sentiment analysis* dari masing-masing artikel berita tersebut

	URL	judul	paragraf	predict
0	https://finansial.bisnis.com/re	BPJS Kesehatan I	Bisnis.com, JAKA	-1
1	https://finansial.bisnis.com/re	BNI Syariah Targe	Bisnis.com, JAKA	0
2	https://finansial.bisnis.com/re	Mandiri Taspen J	Bisnis.com, DEN	1
3	https://finansial.bisnis.com/re	WOM Finance Ra	Bisnis.com, JAKA	1
4	https://finansial.bisnis.com/re	BIA 2019 : Ajang	Bisnis.com, JAKA	-1

Gambar 5. 10. Hasil prediksi *sentiment* dari tiap artikel berita

Gambar 5.10 menampilkan keluaran dari program dimana terdapat tambahan kolom jika dibandingkan dengan Gambar 5.6 dan Gambar 5.7 yaitu kolom “predict”. Kolom tersebut menunjukkan bagaimana *sentiment analysis* dari masing-masing artikel berita yang telah dilakukan *web crawling* oleh program. Dalam kolom tersebut angka -1 menunjukkan artikel memiliki *sentiment* negatif, angka 0 menunjukkan artikel memiliki *sentiment* netral, dan angka 1 menunjukkan artikel memiliki *sentiment* positif.

Dengan hasil demikian, maka program yang dibentuk dalam Tugas Akhir ini dapat dikatakan mampu melakukan prediksi *sentiment* dari tiap artikel berita dengan menggunakan Algoritma Hybrid CNN-HMM

BAB VI

SIMPULAN DAN SARAN

6.1 Simpulan

Berdasarkan hasil uji coba yang dilakukan maka dapat disimpulkan antara lain:

1. Performansi rata-rata yang didapatkan dari model Hybrid CNN-HMM untuk menentukan *sentiment* adalah sebesar 99,92% untuk data latih dan 70,45% untuk data uji.
2. Teknik *web crawling* mampu melakukan akuisisi data dari situs berita *online* kompas.com dan bisnis.com dengan ketercapaian sebesar 99,88%.
3. Konstruksi model Hybrid Convolutional Neural Network – Hidden Markov Models (Hybrid CNN-HMM) mampu menentukan *sentiment* dari artikel berita hasil dari akuisisi data situs berita *online* menggunakan Teknik *web crawling*

6.2 Saran

Saran yang diberikan untuk perbaikan pada penelitian selanjutnya antara lain:

1. Memberi jumlah data latih yang lebih banyak, agar sistem mampu lebih baik mengenali pola-pola yang muncul.
2. Memberi tambahan metode pada pra-proses agar didapatkan data masukan yang lebih optimal
3. Menambah metode untuk menghindari overfitting
4. Melakukan eksplorasi terhadap kedalaman jaringan dan jenis klasifikasi(dapat dilakukan dengan melakukan banyak percobaan dan penggantian terhadap parameter yang tersedia)
5. Menggunakan *library* peubah kata Bahasa Indonesia menjadi vektor yang lebih *up to date*.

DAFTAR PUSTAKA

- [1] Li,Q., Wang,T., Li,P., Liu,L., Gong,Q., Chen,Y. 2014. "The effect of news and public mood on stock movement". Elsevier. Vol. 278. Pg. 826-840.
- [2] Turlan, M. 2010. Php|architect's Guide to Web crawling. musketeers.me.
- [3] Fahrur Rozi, I, Hadi Pramono, S, Achmad Dahlan, E. 2012. Implementasi Opinion Mining (Analisa Sentimen) untuk ekstraksi data opini publik pada perguruan tinggi. Vol.6. Jurnal EECIS, Hal. 37-43.
- [4] Nguyen,T.H., Shirai,K., Velcin,J. 2015. "Sentiment analysis on social media for stock movement prediction". Elsevier. Vol. 42. Pg. 9603-9611.
- [5] Checkley,M.S., Higon,D.A., Alles,H. 2017. "The hasty wisdom of the mob: How market sentiment predicts stock market behaviour. Elsevier. Vol.77. Pg.256-263.
- [6] Kelly,S.,Ahmad,K., 2017. "Estimating the impact of domain-specific news sentiment on financial assets". School of computer science and statistic.
- [7] Kim, Y. 2014. Convolutional Neural Network for Sentence Classification. arXiv:1408.5882.
- [8] Rozi,MF, Mukhlash,I, Kimura, M. 2018. Opinion mining on book review using CNN-L2-SVM algorithm. Vol.974. Journal of Physics: Conference Series. Pg. 012004.
- [9] Mukhlash, I, Zamrudillah Arham, A, Rozi, F, Kimura, M, Adzkiya, D. 2018. Opinion Mining on Book Review using CNN-LSTM . Vol.8. International Journal of Machine Learning and Computing Pg. 437-441.
- [10] Kang,M, Ahn,J, Lee,K. 2018. Opinion mining using ensemble text hidden markov models for text classification. Vol.94. Elsevier Pg.218-227.
- [11] Guo, Q, Wang, F, Lei, J, Tu, D, Li, G. 2016. Convolutional feature learning and Hybrid CNN-HMM

- for scene number recognition. Vol.184. Elsevier Pg.78-90.
- [12] Bridge, Clara. 2011. What is Text Mining. <http://www.clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551>. Diakses 20 Oktober 2018.
- [13] Haddi, E, Liu, X, Shi, Y. 2013. The Role of Text Pre-processing in Sentiment Analysis. Vol.17. Procedia Cumputer Science Pg.26-32.
- [14] Basari, A.S.H. 2013. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Vol.53. Procedia Engineering Pg.453-462.
- [15] Feldman,R. 2013. Techniques and applications for sentiment analysis. Vol.56. Communications of the ACM Pg.82-89.
- [16] Tumilaar, K, Langi, Y, Rindegan, A. 2015. Hidden Markov Model. Vol.4. D'Cartesian Pg.1.
- [17] Ghahramani, Z. 2001. An Introduction to Hidden Markov Models and Bayesian Networks. Vol.15. Journal of Pattern Recognition and Artificial Intelegence Pg. 9-42.
- [18] Gorunescu, F. 2011. Data mining: Concepts, Models and Techniques. Springer: Berlin.

LAMPIRAN

	A	B	C	D	E	F	G
1							
2	Tanggal	Nama Media	Ringkasan Berita	Kaitan Berita	Ruang Lingkup	Klasifikasi Berita	Label Berita
3							
4	01-Jan-17	Surya	(Surabaya) Harga Beras Medium Distributions Rp. 9.350 per Kil	Komoditas	Lokal	Ekonomi	2
5		Kompas	(Jakarta) Konsolidasi Aset Serelah Libur Panjang, Pasar Keuangan	Investasi	Nasional	Ekonomi	3
6	01-Jul-17	Radar	(Surabaya) Aperiindo, Penjualan Lampu LED Masih Rendah, A	Industri	Lokal	Ekonomi	1
7	01-Jul-17	Radar	(Kebomas) Bisnis Ritel Tahun Ini Melambat, Ramadhan tahun	Industri	Lokal	Ekonomi	1
8	01-Jul-17	Radar	(Kebomas) Bahan Naik, Indospring Lakukan Efisiensi, Naiknya Ekspor/Impor	Ekspor/Impor	Lokal	Ekonomi	2
9	01-Jul-17	Jawa Pos	(Jakarta) Harga Sultir Turun Pasca-Lebaran, Memasuki masa-m	Komoditas	Nasional	Ekonomi	1
10	01-Jul-17	Jawa Pos	(Surabaya) Tarif Penanganan Tereuksi Separo, Perkembangan, Inflasi/Harga	Inflasi/Harga	Lokal	Ekonomi	3
11	01-Jul-17	Jawa Pos	(Surabaya) Mandiri Gencarkan Transaksi Cashless, Bank Manc	GNNIT	Lokal	sistem Pembayaran	2
12	01-Jul-17	Jawa Pos	(Jakarta) Bank Swasta Klaim Semakin Efisien, Bank swasta me	GNNIT	Nasional	sistem Pembayaran	2
13	01-Jul-17	Jawa Pos	(Jakarta) Mental Prospek Mata Uang Digital, Dalam beberapa	GNNIT	Nasional	sistem Pembayaran	2
14	01-Jul-17	Jawa Pos	(Jakarta) Suku Bunga Bertahan Sampai Akhir Tahun, Bank Indd days Repo Rat	GNNIT	Nasional	Ekonomi	2
15	01-Jul-17	Jawa Pos	(Surabaya) Hospitality Dongkrak Kinerja Industri Kertas, Indu	Komoditas	Lokal	Ekonomi	3
16	01-Jul-17	Jawa Pos	(Surabaya) Sekar Laut Ekspansi ke Eropa Timur, Prodiusen mak Ekspor/Impor	Komoditas	Lokal	Ekonomi	3
17	01-Jul-17	Kompas	(Jakarta) Ketersediaan Beras Tahun Ini Perlu Menjadi perhatian	Komoditas	Lokal	Ekonomi	1
18	03-Jul-17	Radar	(Surabaya) MNC Bank Perkuat Pasar melalui Aplikasi, Geliat b	Perbankan	Lokal	SSK	3
19	03-Jul-17	Radar	(Surabaya) Libur Lebaran Bereki Konsumsi Gasoline 24%, PT P	Komoditas	Lokal	Ekonomi	3
20	03-Jul-17	Kompas	(Jakarta) Penyelundupan Benih Lobster Terus Mingkat, Hingg	Komoditas	Nasional	Ekonomi	1
21	03-Jul-17	Kompas	(Jakarta) Pendanaan Kuran Optimal, Masalah rninggoti daei	Keuangan	Nasional	SSK	1

Gambar 1. Pembesaran Gambar 4.3

The image shows a browser window with a news feed and the Chrome DevTools developer interface. The news items are:

- Menkeu: Risiko Global Meningkatkan Perang Dagang dan Persaingan G** (12 Juni 2019, 08:29 WIB)
- BI Prediksi Nilai Tukar Rupiah Pad** (12 Juni 2019, 08:03 WIB)
- Imbas Perang Dagang, Huawei Tu** (12 Juni 2019, 07:31 WIB)
- [POPULER MONEY] Status BNI Sya** (12 Juni 2019, 07:09 WIB)

The developer tools show the following HTML structure for the news items:

```
<div class="terkini_img">
  <a href="https://money.kompas.com/read/2019/06/12/082900426/menkeu-
  risiko-global-meningkatkan-perang-dagang-dan-persaingan-g">
    <img alt="Thumbnail image for Menkeu article" data-bbox="668 458 788 488"/>
  </a>
</div>
<div class="terkini_caption"></div>
</div>
<div class="terkini_post">
  <div class="terkini_img">
    <a href="https://money.kompas.com/read/2019/06/12/080300256/BI-prediksi-
    nilai-tukar-rupiah-pada-2020-bisa-sentuh-rp-13.900-per-dolar-as">
      <img alt="Thumbnail image for BI article" data-bbox="448 458 568 488"/>
    </a>
  </div>
  <div class="terkini_caption"></div>
</div>
<div class="terkini_post">
  <div class="terkini_img">
    <a href="https://money.kompas.com/read/2019/06/12/073100326/Imbas-perang-
    dagang-huawei-kunda-penobatan-jadi-merk-ponsel-pertama-dunia">
      <img alt="Thumbnail image for Huawei article" data-bbox="328 458 448 488"/>
    </a>
  </div>
  <div class="terkini_caption"></div>
</div>
<div class="terkini_post">
  <div class="terkini_img">
    <a href="https://money.kompas.com/read/2019/06/12/070900256/populer-money-
    status-bni-syarikat-dan-bsm-sulitnya-memajaki-google-dan-netflix">
      <img alt="Thumbnail image for BNI/BSM article" data-bbox="208 458 328 488"/>
    </a>
  </div>
  <div class="terkini_caption"></div>
</div>
```

Red boxes highlight the news titles and the HTML elements: `<div class="terkini_img">`, ``, and ``. Red arrows point from the news items to these HTML elements.

Gambar 3. Pembesaran Gambar 4.24

Media Monitoring Juli 2018 Berudisk - Excel

ATHWAH DANNI SURIPA GAMAN537580

port Formulas Data Review View Help Nitro Pro 10 XLS/TXT Search

	C	D	E	F	G
a	Ringkasan Berita	Saluran Berita	Ruang Lingkup	Klasifikasi Berita	Label Berita
	(Surabaya) Harga Beras Medium Distabilkan Rp. 9.350 per Kil	Komoditas	Lokal	Ekonomi	2
	(Jakarta) Konsolidasi Aset Setelah Libur Panjang , Pasar Keuangan	Investasi	Nasional	Ekonomi	3
	(Surabaya) Apelindo: Penjualan Lampu LED Masih Rendah , A	Industri	Lokal	Ekonomi	1
	(Keboomas) Bisnis Ritel Tahun Ini Melambat , Ramadhan tahun	Industri	Lokal	Ekonomi	1
	(Keboomas) Bahan Naik, Indosping Lakukan Efisiensi , Naiknya	Ekspor/Impor	Lokal	Ekonomi	2
	(Jakarta) Harga Sulfur Turun Pasca-Lebaran , Memasuki masa m	Komoditas	Nasional	Ekonomi	1
	(Surabaya) Tarif Penerbangan Terealisasi Separo , Perkembangan	Infrastruktur/Harga	Lokal	Ekonomi	3
	(Surabaya) Mandiri Gencarkan Transaksi Cashless , Bank Manc	GNNIT	Lokal	sistem Pembayaran	2
	(Jakarta) Bank swasta Klam semakin Efisien , Bank swasta mei	GNNIT	Nasional	sistem Pembayaran	2
	(Jakarta) Menilai Prospek Mata Uang Digital , Dalam beberapa	GNNIT	Nasional	sistem Pembayaran	2
	(Jakarta) Suku Bunga BERTAHAN Sampai Akhir Tahun , Bank Ind	Days Repo Rat	Nasional	Ekonomi	2
	(Surabaya) Hospitality Dongkrak Kinerja Industri Kertas , Indu	Komoditas	Lokal	Ekonomi	3
	(Surabaya) Sektor Laut Ekspansi ke Eropa Timur , Produsen mak	Ekspor/Impor	Lokal	Ekonomi	3
	(Jakarta) Ketersediaan Beras Tahun Ini Perlu Mengalami Perubahan	Komoditas	Nasional	Ekonomi	1
	(Surabaya) MNC Bank Perkuat Pasar melalui Aplikasi , Gelat b	Perbankan	Lokal	SSK	3
	(Surabaya) Liburan Lebaran Kerak Konsumsi Gasoline 24% , PT P	Komoditas	Lokal	Ekonomi	3
	(Jakarta) Penyelundupan Benih Lobster Terus Miringkat , Hingg	Komoditas	Nasional	Ekonomi	1
		Kemampuan	Nasional	SSK	4

Gambar 5. Pembesaran Gambar 4.27

	URL	Judul	paragraf
0	https://money.kompas.com/read/2019/06/12/20595...	KPR Akhirnya Dapat WTP, Menteri Susi Alui Deg...	JAKARTA, KOMPAS.com - Laporan keuangan Kementre...
1	https://money.kompas.com/read/2019/06/12/20540...	Pengusaha Yakin Pembangunan Infrastruktur Bisa...	JAKARTA, KOMPAS.com - Ketua Kamar Dagang dan I...
2	https://money.kompas.com/read/2019/06/12/20140...	Usai Impor 250 Ton, Kemendag Kembali Keluarkan...	JAKARTA, KOMPAS.com - Direktur Jenderal Perdag...
3	https://money.kompas.com/read/2019/06/12/20081...	Traveloka Serius Garap Bisnis Hiburan dan Gaya...	KOMPAS.com - Perusahaan teknologi penyedia jas...
4	https://money.kompas.com/read/2019/06/12/19452...	Sri Mulyani Yakin BI Bakal Turunkan Suku Bunga...	JAKARTA, KOMPAS.com - Pekan depan, Bank Indone...
5	https://money.kompas.com/read/2019/06/12/19250...	Tiket Mahal Pesawat adalah Puncak Gunung Es	JAKARTA, KOMPAS.com - Pengamat penerbangan se...
6	https://money.kompas.com/read/2019/06/12/19010...	Tiket Pesawat Mahal, Sri Mulyani Minta ASN Hem...	JAKARTA, KOMPAS.com - Tarif tiket pesawat yang...
7	https://money.kompas.com/read/2019/06/12/18321...	Chappy Hakim: Jangan Terlalu Cepat Undang Mask...	JAKARTA, KOMPAS.com - Ide mengundang maskapai...
8	https://money.kompas.com/jeo/ren-dan-fips-his...	Tren dan Tips Bisnis Jastip Raup Rupiah - JEO	None
9	https://money.kompas.com/read/2019/06/12/17540...	Laporan Keuangan Pemerintah WTP Lagi, Sri Muly...	JAKARTA, KOMPAS.com - Badan Pemeriksa Keuangan...
10	https://money.kompas.com/read/2019/06/12/17433...	Lagi, Pemerintah Pulangkan Pekerja Migran Berm...	KOMPAS.com - Pemerintah Indonesia melalui KBRI...
11	https://money.kompas.com/read/2019/06/12/17380...	Soal Aturan Diskon Tarif Transportasi Online...	JAKARTA, KOMPAS.com - Pemerintah melalui Kementre...
12	https://money.kompas.com/read/2019/06/12/17232...	Pemerintah Disarankan Cari Pasar untuk Pelabu...	JAKARTA, KOMPAS.com - Pemerintah disarankan un...
13	https://money.kompas.com/read/2019/06/12/17220...	Trk Atur Keuangan Biar Enggak "Tekor" Tengah	KOMPAS.com - Baru tengah bulan, namun Sinta su...
14	https://money.kompas.com/read/2019/06/12/16511...	Usai Diakuisi, Bank Royal Akan Dimerger deng...	JAKARTA, KOMPAS.com - PT Bank Central Asia Tbk...
15	https://money.kompas.com/read/2019/06/12/16200...	Soal Aturan Diskon Tarif, Ini Kata Pengemudi O...	JAKARTA, KOMPAS.com - Kementerian Perhubungan...
16	https://money.kompas.com/read/2019/06/12/15570...	Impas Krisis 737 Max, Boeing Tak Dapat Order 2...	NEW YORK, KOMPAS.com - Pabrik pesawat Boein...
17	https://money.kompas.com/read/2019/06/12/15290...	Usai Labaran, Harga Gabai Merah Masih Tinggi	JAKARTA, KOMPAS.com - Direktur Jenderal Perdag...
18	https://money.kompas.com/read/2019/06/12/15083...	Laporan Keuangan KPR Raih Opini Wajar Tanpa Pe...	JAKARTA, KOMPAS.com - Kementerian Kelautan dan...
19	https://money.kompas.com/read/2019/06/12/14160...	Ekspor Turun, Indonesia Cari Peluang di Tengah	JAKARTA, KOMPAS.com - Defisit neraca perdagangan...

Gambar 6. Pembesaran Gambar 4.34 dan Gambar 5.6

	URL	Judul	paragraf
0	https://finansial.bisnis.com/read/20190612/215...	Soal Regulasi BKPD, AALU Minta Syarat Akumulasi	Bisnis.com, JAKARTA – Asosiasi Asuransi Umum I.
1	https://finansial.bisnis.com/read/20190612/215...	AALU: Instrumen Investasi Jangka Panjang Dapat	Bisnis.com, JAKARTA – Penempatan investasi pad.
2	https://finansial.bisnis.com/read/20190612/215...	Regulasi BKPD, Syarat Akumulasi Dana Wajib	Bisnis.com, JAKARTA – Kemungkinan adanya ahli ak.
3	https://finansial.bisnis.com/read/20190612/215...	Kolaborasi Kunci Utama Bagi Insurtech	Bisnis.com, JAKARTA – Kolaborasi menjadi tren.
4	https://finansial.bisnis.com/read/20190612/290...	Saran Bank Permodal (BNLI) Masih Menjadi Incarna	Bisnis.com, JAKARTA – Saham PT Bank Permodal To.
5	https://finansial.bisnis.com/read/20190612/290...	Sauu Bunga Berpendesi Tujuh, Ini Kata Bos BCA	Bisnis.com, JAKARTA - PT Bank Central Tbk. me.
6	https://finansial.bisnis.com/read/20190612/290...	5 Berita Populer Finansial, BCA Vajj Aduksi	1. Usai Ciptak Bank Royal, BCA Incar Bank Kecl.
7	https://finansial.bisnis.com/read/20190612/215...	Pasokan Akumulasi Asuransi Cukup	Bisnis.com, JAKARTA – Pasokan tenaga ahli sekur.
8	https://finansial.bisnis.com/read/20190612/215...	Transformasi Industri Asuransi, Langkah Mengena	Bisnis.com, JAKARTA – Transformasi dan pengemo.
9	https://finansial.bisnis.com/read/20190612/290...	CMB Negeri Raih Penghargaan The Credit Risk Te	Bisnis.com, JAKARTA - PT Bank CMB Negeri Tbk.

Gambar 7. Pembesaran Gambar 5.7

URL	judul	paragraf	predict
0 https://finansial.bisnis.com/	BPJS Kesehatan I	Bisnis.com, JAKA	-1
1 https://finansial.bisnis.com/	BNI Syariah Targ	Bisnis.com, JAKA	0
2 https://finansial.bisnis.com/	Mandiri Taspen J	Bisnis.com, DEN	1
3 https://finansial.bisnis.com/	WOM Finance Ra	Bisnis.com, JAKA	1
4 https://finansial.bisnis.com/	BIA 2019 : Ajang F	Bisnis.com, JAKA	-1

Gambar 8. Pembesaran Gambar 5.10

TENTANG PENULIS



Penulis memiliki nama lengkap Athyah Danni Surya Gama. Lahir di Surabaya pada tanggal 13 Juni 1997. Penulis menempuh pendidikan di SDN Babat Jerawat I Surabaya (2003-2009), SMP Negeri 26 Surabaya (2009-2012), dan SMA Negeri 11 Surabaya (2012-2015). Penulis memiliki hobi Olahraga (Voli, Billiard, Badminton). Pada tahun kedua perkuliahan, penulis aktif sebagai Staff Departemen *Student Research Development* HIMATIKA ITS 2016-2017 dalam Divisi Kaderisasi dan Staff Departemen *Human Research Development* ITS Billiard 2016-2017. Di tahun ketiga kuliah, penulis menjabat sebagai *Head of Student Research Development Department* HIMATIKA ITS 2017-2018. Di Departemen Matematika ini, penulis mengambil rumpun mata kuliah Ilmu Komputer untuk mengembangkan logika dan penerapan matematika pada perkembangan teknologi. Untuk informasi, kritik, atau saran lebih lanjut bisa disampaikan melalui *e-mail* penulis di athyah.danni@gmail.com.