

SKRIPSI

**PENAMBANGAN PESAN TWITTER UNTUK MEMAHAMI
SENTIMEN MASYARAKAT TERHADAP PERUBAHAN
HARGA PANGAN DI INDONESIA**



Ivan

NPM: 2014730026

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2018**

UNDERGRADUATE THESIS

**MINING TWITTER MESSAGES TO UNDERSTAND PUBLIC
SENTIMENTS OF FOOD PRICE CHANGES IN INDONESIA**



Ivan

NPM: 2014730026

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2018**

LEMBAR PENGESAHAN



**PENAMBANGAN PESAN TWITTER UNTUK MEMAHAMI
SENTIMEN MASYARAKAT TERHADAP PERUBAHAN
HARGA PANGAN DI INDONESIA**

Ivan

NPM: 2014730026

Bandung, 23 Mei 2018

Menyetujui,

Pembimbing

Dr. Veronica Sri Moertini

Ketua Tim Penguji

Kristopher David Harjono, M.T.

Anggota Tim Penguji

Mariskha Tri Adithia, P.D.Eng

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng



PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

PENAMBANGAN PESAN TWITTER UNTUK MEMAHAMI SENTIMEN MASYARAKAT TERHADAP PERUBAHAN HARGA PANGAN DI INDONESIA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 23 Mei 2018



Ivan

NPM: 2014730026

ABSTRAK

Twitter merupakan salah satu media sosial mikroblog dimana pengguna dapat memperbarui status tentang apa yang sedang dipikirkan dan dilakukan. Twitter dapat digunakan sebagai sarana untuk menyampaikan opini atau sentimen tentang berbagai topik atau isu-isu yang sedang terjadi. Pesan Twitter dapat digunakan sebagai sumber data untuk menilai opini atau sentimen masyarakat. Pada penelitian ini pesan yang menjadi sumber data yaitu pesan mengenai perubahan harga pangan.

Penelitian ini bertujuan untuk membangun sebuah perangkat lunak yang dapat mengumpulkan pesan Twitter mengenai perubahan harga pangan dan menganalisis pesan tersebut sehingga didapatkan kategori (positif, negatif atau netral) dari pesan tersebut. Dalam penelitian ini dilakukan analisis sentimen dengan menggunakan lima algoritma, yaitu algoritma *Naive Bayes*, *Decision Tree (C4.5)*, *k-Nearest Neighbor (k-NN)*, *Support Vector Machine (SVM)* dan metode statistik. Langkah pertama yaitu mengumpulkan pesan yang mengandung *keyword* mengenai pangan dengan menggunakan API Twitter (proses *crawling*). Langkah kedua yaitu tahap prapengolahan data. Pada tahap ini pesan akan dipersiapkan agar dapat dianalisis dengan menggunakan algoritma yang telah disebutkan. Persiapan yang dimaksud yaitu mengambil fitur-fitur (kata-kata) penting yang terkandung dalam pesan. Langkah ketiga yaitu tahap analisis. Pada tahap ini pesan yang sudah melalui tahap prapengolahan akan dikategorikan ke dalam kategori positif, negatif atau netral.

Pada penelitian ini berhasil dibangun sebuah perangkat lunak yang dapat mengumpulkan pesan Twitter berdasarkan *keyword* mengenai pangan. Perangkat lunak juga dapat melakukan proses prapengolahan terhadap pesan yang telah dikumpulkan. Terakhir perangkat lunak dapat mengelompokkan pesan ke dalam kategori yang ada (positif, negatif atau netral). Hasil dari eksperimen yang telah dilakukan yaitu algoritma *Naive Bayes* memberikan tingkat akurasi tertinggi dibandingkan algoritma lainnya.

Kata-kata kunci: Pangan, Twitter, Sentimen, *Crawling* Pesan Twitter, Prapengolahan Data, Analisis Sentimen, Algoritma *Naive Bayes*, Algoritma *Decision Tree (C4.5)*, Algoritma *k-Nearest Neighbor (k-NN)*, Algoritma *Support Vector Machine (SVM)*, Metode Statistik

ABSTRACT

Twitter is one of the microblog social media where users can update the status of what is being thought and done. Twitter can be used as a means to convey opinions or sentiments about various topics or issues that are going on. Twitter messages can be used as a source of data to assess public opinion or sentiments. In this research the message that becomes the source of data is message about the change of food price.

This study aims to build a software that can collect Twitter messages about changes in food prices and analyze the message to get the category (positive, negative or neutral) of the message. In this research, sentiment analysis performed with using five algorithms that are Naive Bayes algorithm, Decision Tree (C4.5), k-Nearest Neighbor (k-NN), Support Vector Machine (SVM) and statistical methods. The first step is to collect messages containing keywords about food using the Twitter API (crawling process). The second step is preprocessing data. At this stage the message will be prepared to be analyzed using the algorithm already mentioned. Preparation in question is taking the important features (words) contained in the message. The third step is the analysis phase. At this stage messages that have been through the preprocessing phase will be categorized into positive, negative or neutral categories.

In this study successfully built a software that can collect Twitter messages based on keywords about food. The software can also process the preprocessing of messages that have been collected. Finally software can group messages into existing categories (positive, negative or neutral). The results of the experiments that have been done that is Naive Bayes algorithm provides the highest level of accuracy than other algorithms.

Keywords: Food, Twitter, Sentiment, Crawling Twitter Messages, Preprocessing Data, Sentiment Analysis, Naive Bayes Algorithm, Decision Tree (C4.5) Algorithm, k-Nearest Neighbor (k-NN) Algorithm, Support Vector Machine Algorithm (SVM), Statistical Methods

Skripsi ini dipersembahkan kepada keluarga tercinta dan orang-orang yang memberikan kontribusi dalam penyusunan skripsi ini.

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa karena dengan rahmat dan karuniaNya, penulis dapat menyelesaikan skripsi berjudul Penambangan Pesan Twitter untuk Memahami Sentimen Masyarakat terhadap Perubahan Harga Pangan di Indonesia ini dengan baik meskipun masih banyak kekurangan didalamnya. Dalam kesempatan ini, penulis ingin menyampaikan rasa terima kasih kepada:

1. Keluarga tercinta yang selalu mendoakan dan memberikan dukungan kepada penulis.
2. Ibu Dr. Veronica Sri Moertini selaku dosen pembimbing dan dosen wali yang telah memberikan kritik, saran dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Kristopher David Harjono, M.T. dan Ibu Mariskha Tri Adithia, P.D.Eng selaku dosen penguji.
4. Daniel, Kevin, Kresna, Fedrian, Stillmen dan Stanley selaku teman seperjuangan yang selalu memberikan bantuan dan dukungan.
5. Ancella Hendrika selaku teman yang selalu mendoakan dan memberikan dukungan kepada penulis.
6. Dosen-dosen yang mengajar pada program studi Teknik Informatika UNPAR.
7. Teman-teman Teknik Informatika UNPAR.
8. Seluruh pihak yang tidak dapat disebutkan namanya satu persatu yang telah memberikan kontribusi dalam penyusunan skripsi ini.

Penulis berharap skripsi ini dapat berguna dalam rangka menambah wawasan serta pengetahuan mengenai penambangan data pada pesan Twitter. Penulis menyadari bahwa di dalam makalah skripsi ini terdapat kekurangan dan jauh dari kata sempurna. Oleh sebab itu, penulis berharap adanya kritik, saran dan usulan yang membangun demi perbaikan makalah ini di waktu yang akan datang. Terima kasih.

Bandung, Mei 2018

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR TABEL	xxv
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	2
1.5 Metodologi	2
1.6 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 Data Mining	5
2.1.1 Pengertian Data Mining	5
2.1.2 Knowledge Discovery and Data Mining (KDD)	6
2.1.3 Tugas dan Teknik Data Mining	7
2.1.4 Teknik Klasifikasi	8
2.2 Algoritma Klasifikasi	10
2.2.1 Algoritma Naive Bayes	10
2.2.2 Algoritma Decision Tree (C4.5)	11
2.2.3 Algoritma k-Nearest Neighbor	13
2.2.4 Algoritma Support Vector Machine	14
2.3 Sentiment Analysis	17
2.4 Data Tidak Terstruktur	17
2.5 Text Mining	18
2.5.1 Definisi dan Tujuan Text Mining	18
2.5.2 Proses Text Mining	18
2.5.3 Pengolahan Teks	19
2.6 Morfologi Teks pada Bahasa Indonesia	20
2.6.1 Prefiks	21
2.6.2 Infiks	23
2.6.3 Sufiks	23
2.6.4 Konfiks	24
2.7 Stemming	24
2.7.1 Algoritma Stemming Nazief dan Adriani	24
2.8 Crawling	25
2.9 Twitter	26
2.9.1 API Twitter	27

2.10	Weka	27
2.11	Evaluasi Teknik Klasifikasi	27
3	STUDI DAN EKSPLORASI PERANGKAT LUNAK WEKA	31
3.1	Package Perangkat Lunak Weka	31
3.2	Diagram Kelas Perangkat Lunak Weka	32
3.3	Input Perangkat Lunak Weka	35
3.4	Output Perangkat Lunak Weka	36
3.5	Penggunaan Weka pada Java Netbeans	38
3.5.1	Konfigurasi Weka dengan Netbeans	38
3.5.2	Fungsi-Fungsi Weka pada Java	38
4	ANALISIS	43
4.1	Data Twitter Berbahasa Indonesia Mengenai Perubahan Harga Pangan	43
4.2	Analisis Masalah	44
4.3	Analisis Data	45
4.4	Rancangan Solusi	46
4.5	Modul-modul Perangkat Lunak	47
4.5.1	Modul Pengambil Pesan Twitter	47
4.5.2	Modul Prapengolahan Data	48
4.5.3	Modul Analisis Sentimen	50
5	ANALISIS DAN PERANCANGAN	59
5.1	Diagram Use Case	59
5.1.1	Diagram Use Case	59
5.1.2	Skenario	60
5.2	Perancangan Fisik Basis Data	65
5.3	Perancangan Kelas dan Method Perangkat Lunak	67
5.3.1	Diagram Kelas	67
5.3.2	Deskripsi Kelas dan Method	72
5.4	Perancangan Tampilan Antarmuka	87
6	IMPLEMENTASI, PENGUJIAN DAN EKSPERIMEN	97
6.1	Implementasi Modul-modul	97
6.1.1	Implementasi Modul Pengambil Pesan Twitter	97
6.1.2	Implementasi Modul Prapengolahan Data	98
6.1.3	Implementasi Modul Analisis Sentimen	100
6.1.4	Implementasi Modul Analisis Sentimen dengan Metode Statistik	104
6.2	Pengujian Fungsional Perangkat Lunak	104
6.2.1	Skenario Pengujian	105
6.2.2	Hasil Pengujian	105
6.3	Eksperimen Perangkat Lunak	122
6.3.1	Skenario Eksperimen	122
6.3.2	Hasil Eksperimen	122
6.3.3	Kesimpulan Hasil Eksperimen	127
7	KESIMPULAN DAN SARAN	131
7.1	Kesimpulan	131
7.2	Saran	131
	DAFTAR REFERENSI	133
	A KODE PROGRAM	135

B HASIL EKSPERIMEN	173
C DAFTAR KATA STOPWORD	177
D DAFTAR KATA KUNCI	181
E DAFTAR KATA SLANG	183
F DAFTAR SEMBAKO DAN PROVINSI	185

DAFTAR GAMBAR

2.1	Process Knowledge Discovery	6
2.2	Tahap Pelatihan	9
2.3	Tahap Pengujian	9
2.4	Contoh Support Vector Machine	14
2.5	Cara Kerja Support Vector Machine	16
2.6	Margin Terbaik	16
2.7	Transformasi Data Non-Linear	17
2.8	Confussion Matrix	28
2.9	Confussion Matrix	28
2.10	Confussion Matrix	29
2.11	Confussion Matrix	29
3.1	Diagram Package	31
3.2	Diagram Kelas Weka	32
3.3	Diagram Kelas Weka	33
3.4	Diagram Kelas Weka	33
3.5	Diagram Kelas Weka	34
3.6	Contoh File ARFF	35
3.7	Contoh File ARFF	36
3.8	Contoh Output Weka	37
3.9	Contoh Output Weka	37
3.10	Contoh Output Weka	37
3.11	Contoh Output Weka	38
4.1	Arsitektur Crawling Twitter	47
4.2	Flowchart Tahap Prapengolahan Data	49
4.3	Flowchart Tahap Analisis Sentimen	50
4.4	Contoh Set Data Pelatihan	51
4.5	Flowchart Analisis dengan Metode Statistik	53
4.6	Contoh Set Data Pelatihan	54
4.7	Hasil Pemilihan Atribut	56
4.8	Hasil Akhir Pohon Keputusan	56
4.9	Flowchart k-NN	57
4.10	Flowchart k-NN	58
5.1	Diagram Use Case	59
5.2	Diagram Kelas Package Crawling Twitter	67
5.3	Diagram Kelas Package Prapengolahan Data	68
5.4	Diagram Kelas Package Analisis	69
5.5	Diagram Kelas Package GUI	70
5.6	Diagram Kelas Package Data	71
5.7	Diagram Kelas Package Analisis	71
5.8	Rancangan Halaman Utama	88

5.9	Rancangan Halaman Data Master	89
5.10	Rancangan Halaman Data Master - Data Training	89
5.11	Rancangan Halaman Crawling Data Twitter	90
5.12	Rancangan Halaman Prapengolahan Data - Hasil Prapengolahan	91
5.13	Rancangan Halaman Pelabelan Data	92
5.14	Rancangan Halaman Input Data Tweet	92
5.15	Rancangan Halaman Perhitungan Model	93
5.16	Rancangan Halaman Pengujian Model	94
5.17	Rancangan Halaman Analisis Sentimen Baru - Crawling Data Baru	95
5.18	Rancangan Halaman Analisis Sentimen Baru - Pemilihan Algoritma	95
5.19	Rancangan Halaman Analisis Sentimen Baru - Hasil Dalam Bentuk Grafis	96
5.20	Rancangan Halaman Analisis Sentimen Baru - Crawling Data Baru	96
6.1	Tampilan Halaman Crawling Data Tweet	98
6.2	Tampilan Halaman Prapengolahan Data Tweet	99
6.3	Tampilan Halaman Hasil Prapengolahan Data Tweet	100
6.4	Tampilan Halaman Prediksi Sentimen Baru	101
6.5	Tampilan Halaman Prediksi Sentimen Baru	102
6.6	Tampilan Halaman Analisis Sentimen Baru	103
6.7	Tampilan Halaman Analisis Sentimen Baru	103
6.8	Tampilan Halaman Prediksi Sentimen Baru	104
6.9	Gambar Sebagian Data Master	106
6.10	Gambar Sebagian Data Master	107
6.11	Hasil Pengambilan Pesan Twitter	108
6.12	Hasil Pengambilan Pesan Twitter	108
6.13	Data Uji Pelabelan	111
6.14	Hasil Prapengolahan Data Uji	111
6.15	Hasil Pelabelan	111
6.16	Pengujian Input Data Tweet	112
6.17	Pengujian Input Data Tweet	112
6.18	Pengujian Input Data Tweet	113
6.19	Pengujian Perhitungan Model	113
6.20	Pengujian Perhitungan Model	114
6.21	Pengujian Perhitungan Model	114
6.22	Pengujian Perhitungan Model	115
6.23	Pengujian Fitur Pengujian Model	116
6.24	Pengujian Fitur Pengujian Model	116
6.25	Pengujian Fitur Pengujian Model	117
6.26	Pengujian Fitur Pengujian Model	117
6.27	Pengujian Analisis Sentimen Baru	119
6.28	Pengujian Analisis Sentimen Baru	119
6.29	Pengujian Analisis Sentimen Baru	120
6.30	Pengujian Analisis Sentimen Baru	120
6.31	Pengujian Analisis Sentimen Baru	121
6.32	Hasil Eksperimen Set Data 1	123
6.33	Hasil Eksperimen Set Data 2	123
6.34	Hasil Eksperimen Set Data 3	124
6.35	Hasil Eksperimen Set Data 4	125
6.36	Hasil Eksperimen Set Data 5	125
6.37	Hasil Eksperimen Set Data $k = 3$	126
6.38	Hasil Eksperimen Set Data $k = 5$	127
6.39	Hasil Eksperimen Set Data $k = 10$	127

6.40 Hasil Eksperimen	128
6.41 Hasil Eksperimen	129
B.1 Hasil Eksperimen Set Data 1	173
B.2 Hasil Eksperimen Set Data 2	173
B.3 Hasil Eksperimen Set Data 3	174
B.4 Hasil Eksperimen Set Data 4	174
B.5 Hasil Eksperimen Set Data 8	174
B.6 Hasil Eksperimen	175
B.7 Hasil Eksperimen	175
B.8 Hasil Eksperimen	175
B.9 Hasil Eksperimen	176
B.10 Hasil Eksperimen	176

DAFTAR TABEL

2.1	Kombinasi Awalan-Akhiran yang Tidak Diizinkan [1]	25
5.1	Skenario Melihat <i>Data Master</i>	60
5.2	<i>Alternate Flow</i> Skenario Melihat <i>Data Master</i> - Melihat <i>Data Tweet</i>	60
5.3	<i>Alternate Flow</i> Skenario Melihat <i>Data Master</i> - Melihat Data Kata Master	60
5.4	<i>Alternate Flow</i> Skenario Melihat <i>Data Master</i> - Melihat Data Provinsi	60
5.5	<i>Alternate Flow</i> Skenario Melihat <i>Data Master</i> - Melihat Data Sembako	61
5.6	Skenario <i>Crawling</i> Data Twitter	61
5.7	<i>Alternate Flow</i> Skenario <i>Crawling</i> Data Twitter	61
5.8	Skenario Prapengolahan Data	62
5.9	Skenario Pelabelan <i>Data Tweet</i>	62
5.10	Skenario <i>Input Data Tweet</i>	63
5.11	Skenario Perhitungan Model	63
5.12	Skenario Pengujian Model	64
5.13	Skenario Analisis Sentimen Baru	64
5.14	<i>Alternate Flow</i> Skenario Analisis Sentimen Baru	65
5.15	Rancangan Fisik Basis Data	65
5.16	Rancangan Fisik Basis Data (Lanjutan)	66
6.1	Data Uji	109
6.2	<i>Output</i> Hasil Prapengolahan sampai Tahap <i>Stemming</i>	109
6.3	Data Uji	118
C.1	Daftar Kata Stopword dari Tala	177
D.1	Daftar Kata Kunci Orientasi Positif	181
D.2	Daftar Kata Kunci Orientasi Negatif	181
D.3	Daftar Kata Kunci Orientasi Netral	181
E.1	Daftar Kata Slang dan Kata Baku	183
F.1	Daftar Sembako, Buah dan Sayur	185
F.2	Daftar Provinsi di Indonesia	186

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Media sosial adalah sarana yang digunakan oleh orang-orang untuk berinteraksi satu sama lain dengan cara menciptakan, berbagi serta bertukar informasi dan gagasan dalam sebuah jaringan dan komunitas virtual¹. Pada saat ini media sosial digunakan sebagai sarana untuk menyampaikan opini atau sentimen masyarakat tentang berbagai topik atau membahas isu-isu yang sedang terjadi. Salah satu media sosial yang sering digunakan yaitu Twitter. Menurut Direktur Pelayanan Informasi Internasional Ditjen Informasi dan Komunikasi Publik (IKP), Selamatta Sembiring, media sosial yang paling banyak diakses oleh masyarakat Indonesia adalah Facebook dan Twitter². Indonesia menempati peringkat empat pengguna Facebook dan peringkat lima pengguna Twitter terbesar di dunia². Berdasarkan data dari PT Bakrie Telecom, Indonesia memiliki 19,5 juta pengguna Twitter dari total 500 juta pengguna global².

Twitter merupakan sebuah situs *web* yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan *Tweets*. [2] Mikroblog adalah salah satu jenis alat komunikasi *online* dimana pengguna dapat memperbarui status tentang apa yang sedang dipikirkan dan dilakukan. [2] *Tweets* adalah teks yang memiliki batas maksimum 280 karakter. *Tweets* menunjukkan pesan yang disampaikan oleh pengguna.

Pesan-pesan Twitter dapat digunakan sebagai sumber data untuk mengetahui opini atau sentimen dari masyarakat. Pada skripsi ini pesan yang menjadi sumber data yaitu pesan mengenai perubahan harga pangan. Opini yang disampaikan masyarakat dapat dijadikan masukan bagi pemerintah dalam menentukan kebijakan mengenai pangan. Analisis sentimen (*sentiment analysis*) dapat digunakan untuk memperoleh gambaran umum dari persepsi masyarakat terhadap perubahan harga pangan, apakah cenderung positif, negatif atau netral. Analisis sentimen dapat dilakukan dengan menggunakan teknik *data mining*. *Data mining* bertujuan untuk mengekstrak informasi atau pengetahuan dari sekumpulan data. [3] *Text mining* merupakan salah satu jenis dari *data mining* yang bertujuan untuk menganalisis data tekstual yang tidak mudah untuk diproses karena tidak terstruktur. [2]

Pesan-pesan Twitter yang akan dianalisis perlu dikumpulkan terlebih dahulu. Proses pengumpulan pesan-pesan tersebut dinamakan *crawling*. Pada proses *crawling* akan dikumpulkan pesan-pesan yang relevan dengan apa yang ingin dicari oleh pengguna. Pesan-pesan yang sudah dikumpulkan akan memasuki tahap prapengolahan. Tahap prapengolahan ini bertujuan untuk mempersiapkan pesan agar dapat dianalisis dengan menggunakan algoritma *data mining*. Pada skripsi ini algoritma yang akan digunakan yaitu algoritma klasifikasi. Algoritma klasifikasi akan mengelompokkan data ke dalam kelas atau kategori yang sudah ada. [3] Terdapat beberapa algoritma klasifikasi yang akan

¹<https://pakarkomunikasi.com/pengertian-media-sosial-menurut-para-ahli>

²https://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+Indonesia+63+Juta+0rang/0/berita_satker

digunakan diantaranya, yaitu algoritma *Naive Bayes*, algoritma *Decision Tree (C4.5)*, algoritma *k-Nearest Neighbor* (k-NN) dan algoritma *Support Vector Machine* (SVM). Selain algoritma klasifikasi akan digunakan metode statistik untuk mengelompokkan pesan Twitter.

1.2 Rumusan Masalah

Berdasarkan hal-hal yang dibahas pada latar belakang, terdapat beberapa permasalahan yang akan diselesaikan yaitu sebagai berikut.

1. Bagaimana cara mengumpulkan pesan Twitter berbahasa Indonesia mengenai perubahan harga pangan?
2. Bagaimana cara menyiapkan pesan Twitter berbahasa Indonesia mengenai perubahan harga pangan sehingga diperoleh data yang siap untuk dianalisis?
3. Bagaimana cara menganalisis pesan Twitter yang telah disiapkan di atas sehingga diperoleh hasil analisis atau informasi yang bermakna?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan dari tugas akhir ini adalah sebagai berikut.

1. Mengumpulkan dan mempelajari sampel pesan Twitter berbahasa Indonesia mengenai perubahan harga pangan yang dikumpulkan secara manual serta merancang dan mengimplementasikan teknik *crawling* Twitter dengan menggunakan API Twitter.
2. Mempelajari morfologi teks bahasa Indonesia serta merancang dan mengimplementasikan teknik prapengolahan data.
3. Mempelajari algoritma klasifikasi seperti *Naive Bayes*, *Decision Tree (C4.5)*, *k-Nearest Neighbor* (k-NN) dan *Support Vector Machine* (SVM) serta merancang dan mengimplementasikan algoritma tersebut.

1.4 Batasan Masalah

Batasan masalah dari tugas akhir ini:

1. Pesan Twitter yang dianalisis hanya dalam Bahasa Indonesia saja.
2. Pangan yang dicari dan dianalisis hanya sembako, buah dan sayur.
3. Pesan Twitter yang dapat dikumpulkan sebatas dalam waktu seminggu terakhir karena API Twitter membatasi pencarian dalam kurun waktu satu minggu terakhir.
4. Algoritma *Naive Bayes* dikembangkan sendiri oleh penulis sedangkan untuk algoritma *Decision Tree (C4.5)*, *k-Nearest Neighbor* (k-NN) dan *Support Vector Machine* (SVM) menggunakan *library* Weka.

1.5 Metodologi

Penyusunan tugas akhir ini menerapkan metodologi sebagai berikut.

1. Melakukan studi literatur mengenai konsep data tidak terstruktur, teks bahasa Indonesia, teknik-teknik prapengolahan data dan algoritma penambangan sentimen.

2. Melakukan studi literatur mengenai API Twitter.
3. Mengumpulkan sampel pesan Twitter berbahasa Indonesia mengenai perubahan harga pangan.
4. Mempelajari dan menganalisis permasalahan dalam mengolah pesan Twitter berbahasa Indonesia mengenai perubahan harga pangan.
5. Merancang teknik *crawling* Twitter, teknik prapengolahan data teks dan teknik analisis sentimen untuk pesan Twitter berbahasa Indonesia.
6. Mengimplementasikan teknik *crawling* Twitter, teknik prapengolahan data teks, dan teknik analisis sentimen untuk pesan Twitter berbahasa Indonesia.
7. Melakukan eksperimen untuk menguji *performance* perangkat lunak untuk menganalisis data.
8. Menulis dokumen skripsi.

1.6 Sistematika Pembahasan

Bab 1 Pendahuluan

Berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian dan sistematika pembahasan.

Bab 2 Dasar Teori

Berisi teori-teori mengenai *data mining*, penjelasan mengenai beberapa algoritma klasifikasi, *sentiment analysis*, data tidak terstruktur, *text mining*, morfologi teks bahasa Indonesia, *stemming*, *crawling*, Twitter, Weka dan pengujian algoritma klasifikasi.

Bab 3 Studi dan Eksplorasi Perangkat Lunak Weka

Berisi penjelasan *package* Weka, penjelasan diagram kelas Weka, format *input* Weka, format *output* Weka dan penggunaan Weka pada *Netbeans*.

Bab 4 Analisis

Berisi data Twitter berbahasa Indonesia mengenai perubahan harga pangan, analisis masalah, analisis data, rancangan solusi, modul-modul utama pada perangkat lunak.

Bab 5 Perancangan

Berisi diagram *use case*, perancangan fisik basis data, perancangan kelas dan *method* perangkat lunak dan perancangan tampilan antarmuka.

Bab 6 Implementasi, Pengujian dan Eksperimen

Berisi implementasi modul-modul, pengujian fungsional perangkat lunak dan eksperimen perangkat lunak.

Bab 7 Kesimpulan dan Saran

Berisi kesimpulan dan saran.