



**Klasifikasi Pengaduan Pelayanan Publik di Kota
Semarang dari Data di *Twitter* Menggunakan Metode
*Naive Bayes Classifier***

Skripsi

disusun sebagai salah satu syarat
untuk memperoleh gelar Sarjana Komputer
Program Studi Teknik Informatika

oleh

Muhammad Abdurrokhim

4611412007

UNNES
UNIVERSITAS NEGERI SEMARANG

JURUSAN ILMU KOMPUTER

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

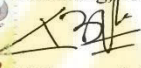
2017

PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.



Semarang, 15 Juni 2017


Muhammad Abdurrokhim
4611412007

UNNES
UNIVERSITAS NEGERI SEMARANG

PERSETUJUAN PEMBIMBING

Nama : Muhammad Abdurrokhim

Nim : 4611412007

Program Studi : S-1 Teknik Informatika

Judul Skripsi : Klasifikasi Pengaduan Pelayanan Publik di Kota Semarang dari Data di *Twitter* Menggunakan Metode *Naive Bayes Classifier*

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian skripsi Program Studi Teknik Informatika FMIPA UNNES.

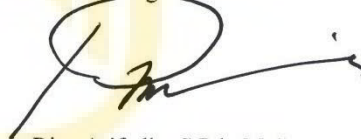
Semarang, 15 Juni 2017

Pembimbing 1



Isa Akhlis S.Si., M.Si.
NIP. 197001021999031002

Pembimbing 2



Riza Arifudin, S.Pd., M.Cs.
NIP. 198005252005011001

UNNES
UNIVERSITAS NEGERI SEMARANG

PENGESAHAN

Skripsi yang berjudul

Klasifikasi Pengaduan Pelayanan Publik di Kota Semarang dari Data di
Twitter Menggunakan Metode *Naive Bayes Classifier*

disusun oleh

Muhammad Abdurokhim
4611412007

telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES
pada tanggal 20 Juni 2017.



Prof. Dr. Zaenuri, S.E., M.Si., Akt.
NIP. 196412231988031001

Sekretaris

Endang Sugiharti, S.Si., M.Kom.
NIP. 197401071999032001

Ketua Penguji

Much Aziz Muslim, S.Kom., M.kom.
NIP. 197404202008121001

Anggota Penguji/
Pembimbing I

Isa Akhlis S.Si., M.Si.
NIP. 197001021999031002

Anggota Penguji/
Pembimbing II

Riza Arifudin, S.Pd., M.Cs.
NIP. 198005252005011001

MOTTO DAN PERSEMBAHAN

MOTTO

- Tidak ada hal yang tidak bisa kita lakukan semua itu bergantung pada niat dan semangat kita
- Sesungguhnya jiwa itu bagaikan kaca, dan akal pikiran, bagaikan lampunya dan hikmah (kebijakan) Allah bagaikan minyaknya, dan jika ia padam kamu menjadi mati (Ibnu Sina)

PERSEMBAHAN

Skripsi ini ku persembahkan kepada:

1. Orang tua tercinta terimakasih atas doa, dukungan dan kasih sayang yang tiada hentinya engkau berikan.
2. Saudara saya yang saya selalu memotivasi dan senantiasa menasehati serta memberikan saran dan masukan.
3. Sahabat terdekat yang telah memberikan kesan dalam penulisan skripsi ini.
4. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu hingga terselesaikannya penulisan skripsi ini.
5. Almamaterku UNNES.

PRAKATA

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan segala rahmat dan hidayah-Nya dalam penyusunan skripsi, sehingga penulis dapat menyelesaikan skripsi dengan judul “**Klasifikasi Pengaduan Pelayanan Publik di Kota Semarang dari Data di Twitter Menggunakan Metode Naive Bayes Classifier**”.

Skripsi ini dapat diselesaikan karena adanya kerjasama, bantuan dan motivasi dari berbagai pihak. Ucapan terima kasih ini penulis tujukan kepada yang terhormat:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang, yang telah memberikan kesempatan kepada penulis untuk dapat berkuliah di Jurusan Ilmu Komputer Program Studi Teknik Informatika FMIPA UNNES.
2. Prof. Dr. Zaenuri, S.E., M.Si., Akt., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang yang telah memberikan izin kepada penulis untuk menyusun skripsi.
3. Endang Sugiharti, S.Si., M.Kom., Ketua Jurusan Ilmu Komputer yang telah banyak memberi bimbingan kepada penulis.
4. Much Aziz Muslim, S.Kom., M.Kom., selaku ketua penguji, yang telah memberikan ijin kepada penulis untuk menyusun skripsi, serta memberikan banyak masukan, kritik dan saran dalam penyelesaian skripsi ini.

5. Isa Akhlis S.Si., M.Si., Dosen Pembimbing I yang telah meluangkan banyak waktu, membantu, membimbing, dan mengarahkan untuk memberikan bimbingan pada penulis dalam menyelesaikan skripsi.
6. Riza Arifudin, S.Pd, M.Cs., Dosen Pembimbing II yang telah meluangkan banyak waktu, membantu, membimbing, dan mengarahkan untuk memberikan bimbingan pada penulis dalam menyelesaikan skripsi.
7. Bapak Muhtadi dan Ibu Suwarni tercinta, yang telah memberikan do'a dan dorongan baik secara moril, materil maupun spiritual dalam menyelesaikan skripsi.
8. Sahabat-sahabat k9ndo *family* Kiky, Whisnu, Bayu, Ageng, Oky, Odi, Eka Ar, Hardo & Zahra.
9. Sahabat-sahabat KKF Uli, Yuniati, Yuniari, Yahya, Angga, Noval, Diah, Anis dan Rivfan dan mahasiswa Ilmu komputer 2012 dan rekan – rekan palatikom yang tidak bisa disebutkan satu persatu.
10. Teman – teman penghuni lab jarkom (Ukhti, Eka lis, Elham, Rovi, Mirqoh, Juli, dsb).
11. Seluruh staf dosen di Universitas Negeri Semarang.

Semoga bantuan yang telah diberikan kepada kepada penulis mendapatkan imbalan dari Allah Yang Maha Pengasih.

Semarang, 20 Juni 2017

Muhammad Abdurrokhim

ABSTRAK

Abdurrokhim, Muhammad. 2017. *Klasifikasi Pengaduan Pelayanan Publik di Kota Semarang dari Data di Twitter Menggunakan Metode Naive Bayes Classifier*. Skripsi, Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama Isa Akhlis S.Si., M.Si. dan Pembimbing Pendamping Riza Arifudin, S.Pd, M.Cs.

Kata kunci: *Twitter, Naive bayes, klasifikasi*.

Perkembangan media sosial sekarang ini tumbuh sangat pesat seperti hanya *twitter*. Setiap hari server *twitter* menerima data tweet dengan jumlah yang sangat besar, dengan demikian, kita dapat melakukan data dari *twitter* untuk tujuan tertentu. Salah satunya adalah untuk visualisasi pengaduan masyarakat di sebuah kota.

Naive bayes classifier adalah pendekatan yang mengacu pada teorema bayes, dengan mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Sehingga merupakan salah satu algoritma *klasifikasi* yang sederhana namun memiliki akurasi tinggi. Untuk itu, dalam penelitian ini akan membuktikan kemampuan *naive bayes* untuk mengklasifikasikan tweet yang berisi informasi dari pengaduan masyarakat terhadap pelayang atau kondisi sosial yang ada di kota Semarang. Hal ini dilakukan oleh pemerintah kota Semarang bertujuan untuk menghimpun atau mendapatkan informasi secara langsung dari masyarakat dengan mengetweet berhashtagkan *laporhendi*.

Tujuan penelitian ini adalah untuk mendapatkan informasi dari data *twitter* pengaduan masyarakat yang berhashtag *lapor hendi* dengan mendapatkan info laporan berupa masing – masing kelasnya. Hal ini dilakukan bertujuan agar dapat secara mudah untuk melakukan tindakan terkait. Pada penelitian ini dengan menerapkan algoritma *naive bayes* dalam melakukan prses klasifikasi data tweet yang sudah diambil dari *twitter*. Proses klasifikasi nantinya akan mendapatkan 3 kelas yaitu insiden, infrastruktur dan pelayanan.

Hasil dari penelitian ini mendapatkan hasil dari data yang diperoleh dengan *klasifikasi* masing – masing kelas yang sudah ditentukan nilainya. Perhitungan *klasifikasi* dihitung melalui nilai probabilitas masing – masing kata yang sudah memiliki nilainya. Kata yang sudah didapat menjadi kata kunci untuk memperoleh hasil *klasifikasi* tiap kelasnya. Berdasarkan hasil uji coba data dari *twitter* diperoleh hasil akurasi sebesar 77% nialai ini didapat dari perhitungan penjumlahan seluruh data yang terprediksi benar dibagi dengan seluruh data uji sedangkan *error rate* sebesar 23% didapatkan dari perhitungan penjumlahan nilai 1 dikurangi dari hasil akurasi.

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
PERNYATAAN	ii
PERSETUJUAN PEMBIMBING.....	iii
HALAMAN PENGESAHAN.....	iv
MOTTO DAN PERSEMBAHAN	v
PRAKATA	vi
ABSTRAK	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
BAB	
1. PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
2. TINJAUAN PUSTAKA.....	7
2.1 Konsep dasar sistem	7

2.1.1	Definisi Sistem	7
2.1.2	Karakteristik Sistem	7
2.2	Konsep Dasar Informasi	10
2.2.1	Definisi Data	10
2.2.2	Definisi Informasi	10
2.3	Data Mining	10
2.3.1	Fungsi <i>Data Mining</i>	12
2.4	Konsep Dasar Sistem Informasi	16
2.5	Klasifikasi	17
2.5.1	Konsep Klasifikasi	17
2.5.2	Model Klasifikasi	18
2.6	<i>Naive Bayes Classifier</i>	20
2.6.1	Teorema Bayes	20
2.6.2	<i>Naive Bayes</i> Untuk Klasifikasi	21
2.6.3	Karakteristik <i>Naive Bayes</i>	23
2.7	Evaluasi Model	24
2.7.1	<i>Confusion Matrix</i>	24
2.8	<i>Twitter API</i>	25
2.8.1	Definisi <i>Twitter</i>	25
2.8.2	Fitur – fitur <i>API Twitter</i>	25
2.9	Penelitian Terkait	26
3.	METODE PENELITIAN	29
3.1	Studi Pendahuluan	29

3.2 Tahap Pengambilan Dan Pengumpulan Data	30
3.3 Tahap Analisis Data	31
3.3.1 Pembagian Data <i>Traning</i> dan Data <i>Testing</i>	31
3.3.2 <i>Preprocessing</i>	32
3.3.3 Klasifikasi <i>Naive Bayes</i>	33
3.4 Perancangan Sistem.....	34
3.4.1 <i>Data Flow Diagram</i> (DFD)	34
3.4.2 <i>Entity Relationship Diagram</i> (ERD).....	36
4. HASIL DAN PEMBAHASAN.....	38
4.1 Hasil Penelitian.....	38
4.1.1 Tahap Pengambilan Data	38
4.1.2 Pemisahan Data <i>Training</i> dan Data <i>Testing</i>	40
4.1.3 Tahap Pengolahan Data atau <i>Preprocessing</i>	41
4.1.4 Tahap Pengklasifikasian.....	43
4.1.5 Tahap Penghitungan Probabilitas	46
4.2 Tahap Implementasi Sistem.....	51
4.3 Pembahasan	56
5. PENUTUP	59
5.1 Simpulan.....	59
5.2 Saran.....	59
DAFTAR PUSTAKA	60
LAMPIRAN	62

DAFTAR TABEL

Tabel	Halaman
2.1 Tabel <i>Confusion Matrix</i>	24
4.1 Database yang akan diolah.....	41
4.2 Data <i>training</i> yang sudah memiliki kategori.....	46
4.3 Menjelaskan tabel kemunculan kata.....	47
4.4 Kemunculan <i>term</i> dalam kalimat	48
4.5 Tabel probabilitas hasil dari dokumen 5	48
4.6 Tabel probabilitas hasil dari dokumen 5	50
4.7 Evaluasi Data Hasil Klasifikasi.....	57



DAFTAR GAMBAR

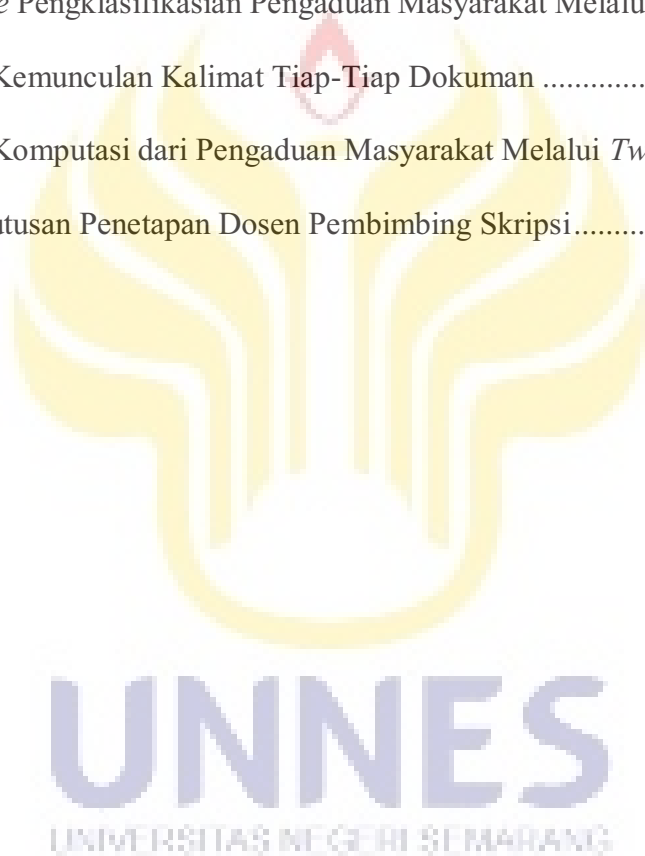
Gambar	Halaman
2.1 Proses Klasifikasi	19
3.1 Model Sekuensial Linier (<i>waterfall</i>)	29
3.2 Alur Penelitian Sebuah Sistem.....	31
3.3 <i>Flowchart Naive Bayes</i>	33
3.4 DFD Level 0.....	34
3.5 DFD Level 1.....	35
3.6 <i>ERD</i> Klasifikasi Pengaduan Pelayanan Publik	36
4.1 API Twitter.....	39
4.2 Tampilan Data Yang Sudah Di Ambil.....	40
4.3 <i>Source code</i> untuk menjalankan proses <i>tokenisasi</i>	42
4.4 <i>Source Code</i> untuk menjalankan proses <i>filtering</i>	42
4.5 <i>Source Code</i> untuk menjalankan proses <i>stemming</i>	43
4.6 <i>Source Code</i> menghitung probabilitas dokumen data <i>training</i>	44
4.7 <i>Source Code</i> menghitung nilai <i>term</i> dokumen data <i>testing</i>	45
4.8 <i>Source code</i> menghitung nilai <i>term</i> pada dokumen metode <i>naive bayes</i> ...	45
4.9 Tampilan awal.....	52
4.10 Tampilan menu sistem	52
4.11 Tampilan menu <i>stopword</i>	53
4.12 Tampilan <i>Term</i>	54
4.13 Tampilan data <i>training</i>	54

4.14 Tampilan data <i>testing</i>	55
4.15 Gambar tambah data <i>testing</i>	55



DAFTAR LAMPIRAN

Lampiran	Halaman
1 Data <i>Training</i> Dalam Sistem	63
2 Data <i>Testing</i> Dalam Sistem	80
3 <i>Sourcecode</i> Pengklasifikasian Pengaduan Masyarakat Melalui <i>Twitter</i>	86
4 Tampilan Kemunculan Kalimat Tiap-Tiap Dokumen	94
5 Tampilan Komputasi dari Pengaduan Masyarakat Melalui <i>Twitter</i>	95
6 Surat Keputusan Penetapan Dosen Pembimbing Skripsi.....	97



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengguna internet di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, 95% menggunakan internet untuk mengakses jejaring sosial. Situs jejaring sosial yang paling banyak diakses adalah facebook dan twitter. Indonesia menempati peringkat 5 pengguna twitter terbesar di dunia. Posisi Indonesia hanya kalah dari USA, Brazil, Jepang dan Inggris. Pengguna twitter, berdasarkan data PT Bakrie Telecom, memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global. Twitter menjadi salah satu jejaring sosial paling besar di dunia sehingga mampu meraup keuntungan mencapai USD 145 juta. Kebanyakan pengguna twitter di Indonesia adalah konsumen, yaitu yang tidak memiliki blog atau tidak pernah mengupload video di youtube namun sering *update* status di twitter dan facebook. Sangat disayangkan apabila perkembangan dan kemajuan teknologi internet ini hanya digunakan untuk sekadar *update* status atau juga saling menimpali komentar atau foto yang diunggah ke facebook dan twitter (Kominfo: 2013).

Pengguna media sosial semakin banyak yang ingin berbagi pendapat pribadi dengan orang lain di media sosial, seperti ulasan produk, analisi ekonomi, jejak pendapat politik dan sebagainya. Ada minat yang tumbuh dalam mencari tahu pendapat orang atau sikap tunduk terhadap beberapa objek dari media sosial, yang dapat membantu pengguna yang lain untuk memberikan keputusan yang membuat

keputusan dan mendapat umpan balik yang berharga bagi dirinya. Penelitian tersebut mencari tahu pendapat orang lain yang memiliki beberapa ekstrasi pendapat, klasifikasi sentimen, dan ringkasan pendapat (Pang & Lee, 2008: 120).

Banyak lembaga pemerintahan telah mulai berbagi video melalui youtube, tersedia di facebook dan jaringan sosial lainnya dan menawarkan informasi melalui blog dan mikroblog seperti twitter (Kavanaugh, dkk, 2012: 480-491). Secara umum diharapkan bahwa alat media sosial adalah untuk meningkatkan transparansi dan akuntabilitas di sektor publik, meningkatkan pelayanan publik, meningkatkan pembuatan kebijakan dengan memungkinkan masyarakat untuk mengambil bagian dalam proses pengambilan keputusan, mendorong kerjasama lintas instansi dan co-produksi antara mitra, dan untuk berkontribusi manajemen pengetahuan (Bertot, dkk, 2010: 53-59).

Untuk melakukan proses mencari tahu pendapat dari twitter diperlukan proses *data mining*. Untuk pengumpulan data pertama kali diperlukan kata kunci dan penggunaan API (Chae, 2014: 247-259). *Data mining* adalah sebuah proses dari *knowledge discovery* (penemuan pengetahuan) dari data yang sangat besar. Dari *data mining* itu terdapat *text mining*, yang merupakan bidang *data mining* bertujuan untuk mengumpulkan informasi *data mining* bertujuan untuk mengumpulkan informasi yang berguna dari data teks dilakukan *natural language processing* dan kemudian mengekstrak informasi yang berguna untuk tujuan tertentu (Noh, dkk, 2015: 4348-4360).

Di kalangan birokrasi pemerintahan, pelayanan publik memiliki aneka ciri dalam setiap masa. Salah satu faktor penentu keragaman layanan tersebut ialah

perkembangan teknologi informasi dan komunikasi. Satu dekade lalu, terdapat proses adopsi teknologi informasi dan komunikasi yang biasa dikenal sebagai *e-Government*. *Open Government* merupakan suatu konsep dalam pemerintahan yang bertujuan untuk menguatkan demokrasi serta menciptakan pemerintahan yang lebih efektif dan efisien dengan meningkatkan kualitas pemerintahan melalui peningkatan akses masyarakat terhadap kinerja pemerintah serta perluasan peran dan partisipasi publik dalam pembangunan. Terdapat dua pilar utama dalam perwujudan *Open Government* yaitu peningkatan akses masyarakat terhadap kinerja pemerintah yaitu transparansi, serta perluasan peran dan partisipasi publik dalam pembangunan atau partisipasi. (Obama, 2009; Rogers & Lindsey, 2012; Lathrop & Ruma, 2013).

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah maka dapat dirumuskan perumusan masalah yaitu,

- a) Bagaimana membangun sebuah sistem informasi yang dapat membantu mengklasifikasikan pelayanan publik di Kota Semarang dari data twitter dengan menentukan tema pengaduan masyarakat.
- b) Bagaimana tingkat keakurasian metode *naive bayes classifier* mengklasifikasikan pelayanan publik guna mendapatkan informasi.

1.3 Batasan Masalah

Pada penelitian ini diperlukan batasan-batasan agar tujuan penelitian dapat tercapai. Adapun batasan masalah yang dibahas pada penelitian ini adalah

- a) Penelitian ini menggunakan twitter untuk pengambilan data pengaduan masyarakat dengan total 300 data.
- b) Menggunakan algoritma *naive bayes* untuk pengklasifikasian pengaduan masyarakat terhadap pelayanan publik.
- c) Penelitian ini hanya mengklasifikasi teks dari twitter yang berhashtag #laporhendi

1.4 Tujuan Penelitian

Tujuan perancangan dan pembangunan aplikasi berbasis *web* ini adalah sebagai berikut:

- a) Membangun sebuah sistem informasi berbasis *web* yang dapat membantu mengklasifikasikan tweet dari masyarakat dan melakukan penentuan akhir dari hasil klasifikasi tersebut.
- b) Untuk mengetahui tingkat keakurasian sistem informasi dengan menggunakan metode klasifikasi *naive bayes classifier* pada sistem tersebut.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut,

- a) Mempermudah dan mempercepat mendapatkan informasi tentang pengaduan masyarakat yang sudah diklasifikasikan.
- b) Mempermudah para petugas dinas untuk mengambil tindakan - tindakan dari hasil klasifikasi di setiap kelasnya.

1.6 Sistematika Skripsi

Sistematika penulisan untuk memudahkan dalam memahami alur pemikiran secara keseluruhan skripsi. Penulisan skripsi ini secara garis besar dibagi menjadi tiga bagian yaitu sebagai berikut:

a. Bagian Awal Skripsi

Bagian awal skripsi terdiri dari halaman judul, halaman pengesahan, halaman pernyataan, halaman motto dan persembahan, abstrak, kata pengantar, daftar isi, daftar gambar, daftar tabel, dan daftar lampiran.

b. Bagian Isi Skripsi

Bagian isi skripsi terdiri dari lima bab yaitu sebagai berikut:

1) Bab 1: Pendahuluan

Bab ini terdiri atas latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika skripsi.

2) Bab 2: Tinjauan Pustaka

Bab ini terdiri atas landasan teori, contoh kasus dan penelitian terkait.

3) Bab 3: Metode Penelitian

Bab ini terdiri atas studi pendahuluan, tahap pengumpulan dan pengumpulan data, studi pustaka, teknik analisis data, analisis kebutuhan, dan pengambilan kesimpulan.

4) Bab 4: Hasil dan Pembahasan

Bab ini terdiri atas hasil penelitian dan pembahasan penelitian.

5) Bab 5: Penutup

Bab ini terdiri atas simpulan dan saran.

c. Bagian Akhir Skripsi

Bagian akhir skripsi berisi daftar pustaka yang merupakan informasi mengenai buku-buku, sumber-sumber dan referensi yang digunakan penulis serta lampiran-lampiran yang mendukung dalam penulisan ini..

BAB II

TINJAUAN PUSTAKA

2.1 Konsep Dasar Sistem

2.1.1 Definisi Sistem

Sistem merupakan suatu bentuk integrasi antara satu komponen dengan komponen lain karena sistem memiliki sasaran yang berbeda untuk setiap kasus yang terjadi dalam sistem tersebut yang di jelaskan Sutabri (2012: 22). Menurut McLeod (2010: 34) dalam bukunya Sistem adalah sekelompok elemen-elemen yang terintegrasi dengan tujuan yang sama untuk mencapai tujuan. Berdasarkan beberapa pendapat yang di atas dapat penulis tarik kesimpulan bahwa sistem adalah kumpulan bagian-bagian atau subsistem-subsistem yang disatukan dan dirancang untuk mencapai suatu tujuan.

2.1.2 Karakteristik Sistem

Dalam bukunya Sutabri (2012: 20), menjelaskan bahwa sebuah sistem mempunyai karakteristik atau sifat-sifat tertentu yang mencirikan bahwa hal tersebut bisa dikatakan sebagai suatu sistem. Adapun karakteristik yang dimaksud adalah sebagai berikut:

a) **Komponen Sistem (*System Components*)**

Suatu sistem terdiri dari sejumlah komponen yang saling berinteraksi, artinya saling bekerja sama membentuk satu kesatuan. Komponen-komponen sistem tersebut dapat berupa suatu subsistem. Setiap subsistem

2.2 Konsep Dasar Informasi

2.2.1 Definisi Data

Data merupakan bentuk yang masih mentah yang belum dapat bercerita banyak. Karena itu, perlu diolah lebih lanjut. Data diolah melalui suatu model agar menjadi informasi menurut Sutabri (2012: 32).

2.2.2 Definisi Informasi

Menurut McLeod dalam bukunya Yakub (2012: 8), informasi adalah data yang diolah menjadi bentuk yang berguna dan berarti bagi yang menerimanya, informasi disebut juga data yang diproses atau data yang memiliki arti. Informasi adalah sebuah istilah yang tepat dalam pemakaian umum. Informasi dapat mengenai data mentah, data tersusun, kapasitas sebuah saluran komunikasi, dan lain sebagainya menurut Sutabri (2012: 29). Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa informasi adalah sebagai data yang sudah diolah, dibentuk, atau dimanipulasi sesuai dengan keperluan tertentu.

2.3 Data Mining

Data mining (Connolly dan Begg, 2010: 180) adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. *Data mining* (Segall ,dkk, 2008) biasa juga disebut dengan “Data atau *knowledge discovery*” atau menemukan pola tersembunyi pada data. *Data mining* adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna.

Data mining (Han dan Kamber, 2006: 5) didefinisikan sebagai proses mengekstrak atau menambang pengetahuan yang dibutuhkan dari sejumlah data besar.

Prasetyo (2014: 3-4) menyatakan bahwa *data mining* adalah pencarian otomatis pola dalam basis data besar, menggunakan teknik komputasional campuran dari statistik, pembelajaran mesin, dan pengenalan pola; mengekstraksi implisit non-trivial, yang sebelumnya belum diketahui secara potensial adalah informasi berguna dari data; ilmu pengestrakan informasi yang berguna dari *dataset*/basis data besar, eksplorasi otomatis/semiotomatis dan analisis data dalam jumlah besar, dengan tujuan untuk menemukan pola yang bermakna; proses penemuan informasi otomatis dengan mengidentifikasi pola dan hubungan tersembunyi dalam data.

Pada prosesnya *data mining* akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *Database System*, *Data Warehousing*, *Statistic*, *Machine Learning*, *Information Retrieval*, dan Komputasi Tingkat Tinggi. Selain itu *data mining* didukung oleh ilmu lain seperti *Neural Network*, Pengenalan Pola, *Spatial Data Analysis*, *Image Database*, *Signal Processing*. Beberapa *survey* tentang proses pemodelan dan metodologi menyatakan bahwa, “*Data mining* digunakan sebagai penunjuk, dimana *data mining* menyajikan intisari atas sejarah, deskripsi dan sebagai standar petunjuk mengenai masa depan dari sebuah proses model *data mining*”(Mariscal, dkk, 2010) Karakteristik *data mining* sebagai berikut:

- a) *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b) *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dapat dipercaya.
- c) *Data mining* berguna untuk membuat keputusan kritis.

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *Data Mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

2.3.1 Fungsi Data Mining

Teknik – teknik data mining telah digunakan untuk menemukan pola yang tersembunyi dan memprediksi tren masa depan. Keuntungan kompetitif dari data mining termasuk dengan meningkatnya pendapatan, berkurangnya pengeluaran, dan kemampuan pemasaran yang meningkat. (Pujari, dkk, 2012) *Data mining* dibagi menjadi dua kategori utama (Han dan Kamber, 2006 : 21- 29) yaitu,

- a) Prediktif

Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variable tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai *explanatory* atau *variable* bebas.

b) Deskriptif

Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, *trend*, *cluster*, teritori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas *data mining* deskriptif sering merupakan penyelidikan dan seringkali memerlukan teknik *post-processing* untuk validasi dan penjelasan hasil.

Fungsi dari *data mining* juga ada dalam dunia kesehatan, dimana data mining telah digunakan untuk meningkatkan diagnosis dan pengobatan atau lebih mengerti perilaku dari pasien. (Sandra, dkk,2009). *Data mining* juga memiliki beberapa fungsionalitas yaitu *Concept/Class Description: Characterization and Discrimination*, *Mining Frequent Patterns, Associations, and Correlations*, *Classification and Prediction*, *Cluster Analysis*, *Outlier analysis*, dan *Evolution analysis*. (Han dan Kamber, 2006 : 21 – 27) Berikut adalah penjelasan dari masing-masing fungsi diatas:

- *Concept/Class Description: Characterization and Discrimination*

Data characterization adalah ringkasan dari semua karakteristik atau fitur dari data yang telah diperoleh dari target kelas. Data yang sesuai dengan kelas yang telah ditentukan oleh pengguna biasanya dikumpulkan di dalam *database*. Misalnya, untuk mempelajari karakteristik produk perangkat lunak dimana pada tahun lalu seluruh penjualan telah meningkat sebesar 10%, data yang terkait dengan produk-produk tersebut dapat dikumpulkan dengan menjalankan sebuah *query SQL*. Sedangkan, *data discrimination* adalah perbandingan antara

fitur umum objek data target kelas dengan fitur umum objek dari satu atau satu set kelas lainnya. target diambil melalui *query database*. Misalnya, pengguna mungkin ingin membandingkan fitur umum dari produk perangkat lunak yang pada tahun lalu penjualannya meningkat sebesar 10% tetapi selama periode yang sama seluruh penjualan juga menurun setidaknya 30%.

- *Mining Frequent Patterns, Associations, and Correlations*

Frequent Patterns adalah pola yang sering terjadi di dalam data. Ada banyak jenis dari *frequent patterns*, termasuk di dalamnya pola, sekelompok *item set*, *sub-sequence*, dan sub-struktur. Sebuah *frequent patterns* biasanya mengacu pada satu set item yang sering muncul bersama-sama dalam suatu kumpulan data transaksional, misalnya seperti susu dan roti.

- *Associations Analysis* adalah pencarian aturan-aturan asosiasi yang menunjukkan kondisi-kondisi nilai atribut yang sering terjadi bersama-sama dalam sekumpulan data. Analisis asosiasi sering digunakan untuk menganalisa *Market Basket Analysis* dan data transaksi.

- *Classification and Prediction*

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari training data (yaitu objek data yang memiliki label kelas yang diketahui). Model

yang diturunkan dapat direpresentasikan dalam berbagai bentuk seperti *If-then* klasifikasi, *decision tree*, dan sebagainya.

- Teknik *classification* bekerja dengan mengelompokkan data berdasarkan *data training* dan nilai atribut klasifikasi. Aturan pengelompokan tersebut akan digunakan untuk klasifikasi data baru ke dalam kelompok yang ada. *Classification* dapat direpresentasikan dalam bentuk pohon keputusan (*decision tree*). Setiap *node* dalam pohon keputusan menyatakan suatu tes terhadap atribut *dataset*, sedangkan setiap cabang menyatakan hasil dari tes tersebut. Pohon keputusan yang terbentuk dapat diterjemahkan menjadi sekumpulan aturan dalam bentuk *IF condition THEN outcome*. (Mewati A, 2007 : 7).
- Dalam banyak kasus, pengguna ingin memprediksi nilai-nilai data yang tidak tersedia atau hilang (bukan label dari kelas). Dalam kasus ini nilai data yang akan diprediksi merupakan data *numeric*. Disamping itu, prediksi lebih menekankan pada identifikasi *trend* dari distribusi berdasarkan data yang tersedia.
- *Cluster Analysis*
Cluster adalah kumpulan objek data yang mirip satu sama lain dalam kelompok yang sama dan berbeda dengan objek data di kelompok lain. Sedangkan, *Clustering* atau Analisis *Custer* adalah proses pengelompokkan satu set benda-benda fisik atau abstrak kedalam kelas objek yang sama. Tujuannya adalah untuk menghasilkan

pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu *cluster* dan semakin besar perbedaan tiap *cluster* maka kualitas analisis *cluster* semakin baik.

2.4 Konsep Dasar Sistem Informasi

Sutabri (2012: 46) dalam bukunya menjelaskan, Sistem informasi adalah suatu sistem didalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian yang mendukung fungsi operasi organisasi yang bersifat manajerial dengan kegiatan strategi dari suatu organisasi untuk dapat menyediakan kepada pihak luar tertentu dengan laporan-laporan yang diperlukan. Sistem informasi adalah sistem yang dapat didefinisikan dengan mengumpulkan, memproses, menyimpan, menganalisis, menyebarkan informasi untuk tujuan tertentu. Seperti sistem lainnya, sebuah sistem informasi terdiri atas *input* (data, instruksi) dan *output* (laporan, kalkulasi). Menurut Sutarman (2012: 13) yang menyimpulkan pendapat di atas dapat ditarik kesimpulan bahwa Sistem informasi adalah sebuah sistem yang terdiri dari pengumpulan data, pemrosesan data, penyimpanan data, pengolahan data, pengendalian dan pelaporan sehingga tercapai sebuah informasi yang mendukung pengambilan keputusan didalam suatu organisasi untuk dapat mencapai sasaran dan tujuannya.

2.5 Klasifikasi

2.5.1 Konsep Klasifikasi

Hermawati (2013: 14) menyatakan bahwa klasifikasi adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori (atau kelas) yang telah didefinisikan sebelumnya. Disebut juga dengan ‘*supervised learning*’. Klasifikasi dapat didefinisikan secara detail sebagai suatu pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target f yang memetakan setiap vektor (set fitur) x ke dalam satu dari sejumlah label kelas y yang tersedia. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori (Prasetyo, 2014: 11). Klasifikasi menurut Han *et al.*, (2012: 18) adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep. Menurut Hermawati (2013: 56) ada dua jenis model klasifikasi yaitu:

- a) **Pemodelan Deskriptif** (*descriptive modelling*): Model klasifikasi yang dapat berfungsi sebagai suatu alat penjelasan untuk membedakan objek-objek dalam kelas-kelas yang berbeda.
- b) **Pemodelan Prediktif** (*predictive modelling*): Model klasifikasi yang dapat digunakan untuk memprediksi label kelas *record* yang tidak diketahui.

Teknik klasifikasi (*classifier*) merupakan suatu pendekatan sistematis untuk membangun model klasifikasi dari suatu himpunan data masukan. Tiap teknik menggunakan suatu algoritma pembelajaran (*learning algorithm*) untuk mendapatkan suatu model yang paling memenuhi hubungan antara himpunan atribut dan label kelas dalam data masukan. Tujuan dari algoritma pembelajaran

adalah untuk membangun model yang secara umum berkemampuan baik, yaitu model yang dapat memprediksi label kelas dari *record* yang tidak diketahui kelas sebelumnya dengan lebih akurat (Hermawati, 2013: 56).

Hermawati (2013: 56) menyatakan bahwa proses untuk membentuk model klasifikasi dengan suatu algoritma pembelajaran (*training*) berdasarkan himpunan data pembelajaran (*training set*) disebut dengan proses induksi (*induction*). Sedangkan proses penerapan model klasifikasi untuk memprediksikan kelas label dari data dalam himpunan data (*test set*) disebut dengan proses deduksi (*deduction*).

Menurut Hermawati (2013: 56-57), untuk mengevaluasi performa dari model yang dibangun, perlu dilakukan pengukuran performa, yaitu pengukuran akurasi (*accuracy*) atau tingkat kesalahan (*error rate*). Jika f_{ij} menotasikan jumlah *record* dari kelas j pada saat pengujian, maka pengukuran akurasi (*accuracy*) dapat dituliskan dengan Persamaan 1.

$$Accuracy = \frac{\text{jumlah_prediksi_yg_benar}}{\text{jumlah_prediksi_keseluruhan}} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}}$$

(1)

Sedangkan tingkat kesalahan (*error rate*) didefinisikan pada Persamaan 2.

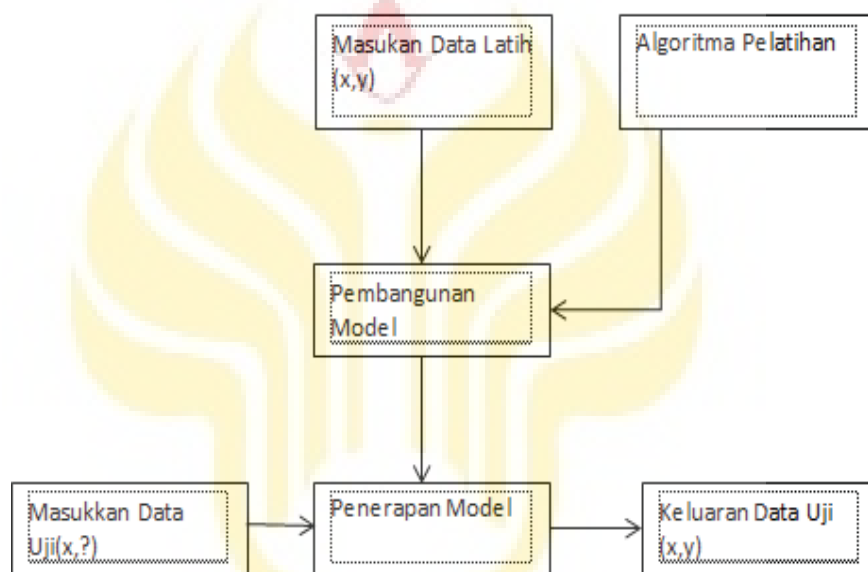
$$Accuracy = \frac{\text{jumlah_prediksi_yg_salah}}{\text{jumlah_prediksi_keseluruhan}} = \frac{f_{01}+f_{10}}{f_{11}+f_{10}+f_{01}+f_{00}}$$

(2)

2.5.2 Model Klasifikasi

Prasteyo (2012: 45) menjelaskan model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, dimana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut dan

memberikan jawaban sebagai keluaran dari hasil pemikirannya. Kerangka kerja (*framework*) klasifikasi ditunjukkan Gambar 2.1 pada gambar tersebut disediakan sejumlah data latih (x,y) untuk digunakan sebagai data pembangunan model. Model tersebut kemudian dipakai untuk memprediksi kelas dari data uji (x,y) sehingga diketahui kelas y yang sesungguhnya.



Gambar 2.1 Proses Klasifikasi

Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan memprediksi label kelas baru yang belum diketahui. Dalam pembangunan model selama proses pelatihan tersebut diperlukan suatu algoritma untuk membangunnya, yang disebut algoritma pelatihan (*learning algorithm*). Ada banyak algoritma pelatihan yang sudah dikembangkan oleh para peneliti, seperti *K-Nearest Neighbor*, *Artificial Neural Network*, *Support Vector Machine* dan lain sebagainya. Setiap algoritma mempunyai kelebihan dan kekurangan, tetapi semua algoritma berprinsip sama, yaitu melakukan suatu pelatihan sehingga di akhir pelatihan model dapat memetakan (memprediksi) setiap vektor masukan ke kelas dengan benar.

2.6 Naive Bayes Classifier

2.6.1 Teorema Bayes

Bayes merupakan teknik prediksi berbasis probalistik sederhana yang berdasar pada penerapan teorema bayes atau aturan *bayes* dengan asumsi independensi (ketidak tergantungan) yang kuat (*naive*). Dengan kata lain, *naive bayes*, model yang digunakan adalah model fitur independen (Prasetyo, 2012: 59). Dalam *bayes* (terutama *baive bayes*), maksud independendensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Prediksi *bayes* didasarkan pada teorema *bayes* dengan formula umum dengan Persamaan 3.

$$P(E|H) = \frac{P(E|H) P(H)}{P(E)} \quad (3)$$

Penjelasan formula diatas sebagai berikut:

Parameter	Keterangan
$P(H E)$	Probabilitas bebas bersyarat (<i>conditional probability</i>) suatu hipotesis H jika diberikan bukti (<i>Evidence</i>) E terjadi.
$P(E H)$	Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H
$P(H)$	Probabilitas awal (priori) hipotesis A terjadi tanpa memandang bukti apapun
$P(B)$	Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atas peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dalam aturan *bayes* tersebut yaitu,

- Sebuah probabilitas awal/prior A atau $P(A)$ adalah probabilitas suatu hipotesis sebelum bukti diamati.
- Sebuah probabilitas akhir B atau $P(A|B)$ adalah probabilitas suatu hipotesis setelah bukti diamati.

2.6.2 Naive Bayes Untuk Klasifikasi

Prasetyo (2012: 61) menjelaskan kaitan antara *naive bayes* dengan klasifikasi, kolerasi hipotesis dan bukti klasifikasi adalah bahwa hipotesis dalam teorema *bayes* merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadikan masukan dalam model klasifikasi. Jika X adalah vektor masukan yang berisi fitur dan Y adalah label kelas, *naive bayes* dituliskan dengan $P(X|Y)$. Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y, sedangkan $P(Y)$ disebut probabilitas awal (*prior probability*) Y.

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir $P(Y|X)$ pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari nilai Y' dengan memaksimalkan $P(X|Y)$ yang didapat. Formulasi *naive bayes* untuk klasifikasi yaitu pada Persamaan 4.

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (4)$$

$P(X|Y)$ adalah probabilitas data dengan vektor X pada kelas Y . $P(Y)$ adalah probabilitas awal kelas Y . $\prod_{i=1}^q P(X_i | Y)$ adalah probabilitas independen kelas Y dari semua fitur dalam vektor X . Nilai $P(X)$ selalu tepat sehingga dalam perhitungan prediksi nantinya kita tinggal menghitung bagian $P(Y) \prod_{i=1}^q P(X_i | Y)$ dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen $\prod_{i=1}^q P(X_i | Y)$ tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y , yang dinotasikan dengan Persamaan 5.

$$P(X|Y = y) = \prod_{i=1}^q P(X_i | Y = y) \quad (5)$$

Setiap set fitur $X = \{x_1, x_2, x_3, \dots, x_q\}$ terdiri atas q atribut (q dimensi). Umumnya, *bayes* mudah dihitung untuk fitur bertipe kategoris seperti pada kasus klasifikasi hewan dengan fitur “penutup kulit” dengan nilai {bulu, rambut, cangkang} atau kasus fitur “jenis kelamin” dengan nilai {pria, wanita}. Namun untuk fitur dengan tipe numerik (kontinu) ada perlakuan khusus sebelum dimasukkan dalam *naive bayes* dengan cara seperti berikut:

- a) Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasikan fitur kontinu ke dalam fitur ordinal.
- b) Mengasumsi bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data pelatihan. Distribusi Gaussian biasanya dipilih untuk merepresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas $P(X_i|Y)$, sedangkan distribusi Gaussian

dikarakteristikan dengan dua parameter: *mean*, μ dan *varian*, σ^2 . Untuk setiap kelas y_j , probabilitas bersyarat kelas y_j untuk fitur X_i adalah seperti pada Persamaan 6.

$$P(X_i = x_i | Y_j = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \frac{-(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \quad (6)$$

Parameter μ_{ij} bisa didapat dari *mean* sampel $X_i(\bar{x})$ dari semua data latih yang menjadi milik kelas y_j , sedangkan σ_{ij}^2 dapat diperkirakan dari *varian* sampel (s^2) dari data latih.

2.6.3 Karakteristik *Naive Bayes*

Klasifikasi dengan *naive bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *naive bayes* sebagai berikut:

1. Metode *naive bayes* teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*). *naive bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *naive bayes* karena asumsi independen atribut tersebut sudah tidak ada.

2.7 Evaluasi Model

2.7.1 Confusion Matrix

(Deshpande, 2015: 257-258) Keakuratan dalam sebuah pengklasifikasian pada sebuah dataset yang diuji adalah tentang persentase dari sebuah tupel dalam data yang diuji, yang dikelompokkan secara benar oleh pengelompok. Dalam literature pengenalan pola, juga disebut sebagai presentase keseluruhan dari pengelompoknya, yaitu mencerminkan seberapa baik pengelompok tersebut mengelompokkan tupel dari berbagai kelas. Suatu *confusion matrix* merupakan alat yang berguna untuk menganalisis seberapa baik pengklasifikasi tersebut dapat mengenali tupel dalam kelas-kelas yang berbeda. Tabel 2.1 merupakan contoh confusion matrix :

Tabel 2.1 Tabel Confusion Matrix

		Predicted Class		
		Positif	Negatif	Netral
Actual Class	Positif	T_Pos	F_NegPos	F_NetPos
	Negatif	F_PosNeg	T_Neg	F_NetNeg
	Netral	F_PosNet	F_NegNeg	T_Net

Keterangan :

T_pos : True Positif
 F_pos : False Positif
 F_neg : False Negatif
 T_neg : True Negatif
 F_net : False Netral
 T_net : True Netral

Rumus untuk mencari nilai akurasi :

$$\text{Accuracy} = \frac{\text{Jumlah prediksi yang benar}}{\text{Jumlah prediksi benar dan salah}}$$

$$= \frac{T_{Pos} + T_{Neg} + T_{Net}}{T_{Pos} + F_{NegPos} + F_{NetPos} + F_{PosNeg} + T_{Neg} + F_{NetPos} + F_{NegNeg} + T_{Net}}$$

2.8 Twitter API

2.8.1 Definisi Twitter

Pada awalnya perusahaan Summize yang menyediakan fasilitas mencari data di twitter. Kemudian perusahaan Summize diakuisisi dan diganti mereka menjadi *twitter search* sehingga *search* API terpisah menjadi entitas sendiri Developer, (Twitter: 2012) twitter API terdiri dari 3 bagian yaitu,

- a) *Search* API : Dirancang untuk memudahkan *user* dalam mengolah *query search* di konten twitter. *User* dapat menggunakannya untuk mencari tweet keyword khusus atau mencari tweet lebih spesifik berdasarkan *username twitter*. *Search* api juga menyediakan akses data *Trending Topic*.
- b) *Representation* twitter (REST) API : Restelop api memperbolehkan *developer* untuk mengakses inti dari twitter seperti *timeline*, *status update* dan informasi *user*.
- c) *Streaming* API : *Streaming* API digunakan untuk *developer* untuk kebutuhan yang lebih intensif seperti melakukan penelitian dan analisis data.

2.8.2 Fitur – fitur API Twitter

Berikut beberapa fitur – fitur yang disajikan API twitter, diantaranya adalah

- a) *Tweet*

Tweet adalah dasar blok bangunan atom segala sesuatu twitter. *Tweet* , juga dikenal lebih umum sebagai update status. *Tweet* dapat embed, menjawab, menyukai, tidak menyukai dan menghapus.

b) *Limit*

Sebagai situs *mikroblogging*, twitter berusaha membatasi segala hal; jumlah karakter dari *tweet*, jumlah *tweet* dalam satuan waktu, jumlah melakukan follow dalam satuan waktu, jumlah API request, dan lain-lain.

c) *OAuth*

OAuth adalah cara twitter untuk meminta persetujuan *user* ketika ada aplikasi atau situs pihak ketiga ingin melakukan sesuatu dengan *account* twitter *user*, seperti melakukan kegiatan follow dan melakukan kegiatan tweet.

d) *Twitter Client*

Twitter client adalah situs atau aplikasi yang memudahkan *user* untuk melakukan kegiatan di twitter, baik melalui handphone (mobile) maupun komputer (PC). Jumlah dari twitter client ini banyak sekali dan semuanya mempunyai keunggulan dan kekurangannya masing-masing, di fitur maupun *bandwidth*.

2.9 Penelitian Terkait

Penelitian ini dikembangkan dari beberapa referensi penelitian terdahulu yang mempunyai keterkaitan dengan metode dan objek penelitian. Penggunaan referensi ini ditujukan untuk memberikan batasan-batasan terhadap metode yang nantinya akan dikembangkan lebih lanjut. Berikut adalah uraian dari penelitian dari peneliti sebelumnya.

- a) Rodiyansyah & Winarno (2012) dalam penelitiannya berjudul “Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan *Naive Bayesian Classification*” berisi tentang klasifikasi kemacetan yang ada di kota Bandung berawal mengambil data (*download*) *tweet* dari server Twitter dengan memanfaatkan API twitter. Data *tweet* yang diambil disimpan di database. Data *tweet* kemudian diolah dengan menggunakan *preprocessing*. Data bersih yang dihasilkan oleh *preprocessing* kemudian diolah dengan menggunakan *naive bayes classifier* sehingga membentuk model probabilitas klasifikasi. Model *probabilitas* klasifikasi ini kemudian digunakan untuk menentukan kelas pada *tweet* yang baru yang belum diketahui kelasnya.
- b) Ariadi & Fithriasari (2015) dalam penelitiannya yang berjudul “Klasifikasi Berita Indonesia menggunakan Metode *Naive Bayesian Classification* dan *Support Vector Machine* dengan *Confix Stripping Stemmer*” menggunakan algoritma klasifikasi *Naive Bayes Classifier* dan *Support Vector Machine* untuk mengkategorikan data berita yang berbentuk teks, dengan melalui proses *confix-stripping stemmer* untuk mendapatkan data kasar dari berita Indonesia. Hasil dari penelitian menunjukkan, perbandingan keduanya dalam klasifikasi data berita berdasarkan Akurasi, *Precision*, *Recall*, *F-Measure*. NBC menunjukkan 82.2%, 83.9%, 82.2%, 82.4%, sedangkan SVM menunjukkan 88.1%, 89.1%, 88.1 %, 88.3%. Dari hasil tersebut dapat dikatakan bahwa SVM melakukan lebih baik daripada NBC dalam mengklasifikasikan data berita di Indonesia.

- c) Chandra, Indawan dan Sukarajaya (2016) dalam penelitian berjudul “Klasifikasi Berita Lokal Radar Malang Menggunakan Metode *Naive Bayes* Dengan Fitur *N-Gram*” menggunakan *N-Gram* pada penelitian ini terbukti mampu menambah jenis kata sebelum masuk ke proses streaming. Dengan banyaknya tambahan jenis kata ini sangat membantu proses klasifikasi *naive bayes* menjadi lebih efektif dan akurat. Pada penelitian hasil akurasi maksimalnya adalah 78,66% untuk data uji berita ekonomi, *news*, edukasi, kesehatan, olahraga, *entertainment*, dan lain-lain dalam Bahasa Indonesia.
- d) Oman Somantri, Slamet Wiyono dan Dairoh (2016) dalam penelitian yang berjudul “Metode *K-Means* untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan *Support Vector Machine* (SVM)”. *K-Means Clustering* merupakan metode pengelompokan paling sederhana yang mengelompokkan data kedalam *k* kelompok berdasar pada *centroid* masing-masing kelompok. Optimasi klasifikasi tema tugas akhir mahasiswa menggunakan SVM dan *K-Means* untuk meningkatkan tingkat akurasi. Hasil yang diperoleh memiliki tingkat akurasi yang lebih baik yaitu 86,21%.
- e) Maulidia Rahmah Hidayah, Isa Akhlis dan Endang Sugiharti (2017) penelitian yang berjudul “Pengakuan Jumlah Pelat Kendaraan dengan Metode *Otsu* dan *K-Nearest Neighbour Classification*” Pengembangan program pengenalan plat dengan menggunakan metode *Otsu* dan klasifikasi KNN mengikuti langkah-langkah pengenalan pola, seperti input dan sensing, pre-processing, fitur ekstraksi metode biner, metode segmentasi, metode klasifikasi KNN dan post-processing oleh Menghitung tingkat

akurasi. Hasil penelitian menunjukkan bahwa program ini dapat mengenali 82% dari 100 plat uji dengan 93,75% akurasi pengenalan dan 91,92% akurasi pengenalan huruf.



BAB V

PENUTUP

5.1. Simpulan

Berdasarkan penelitian dan pembahasan terkait pengaduan masyarakat melalui jejaring sosial khususnya twitter dengan menggunakan metode *naive bayes* dapat ditarik kesimpulan sebagai berikut.

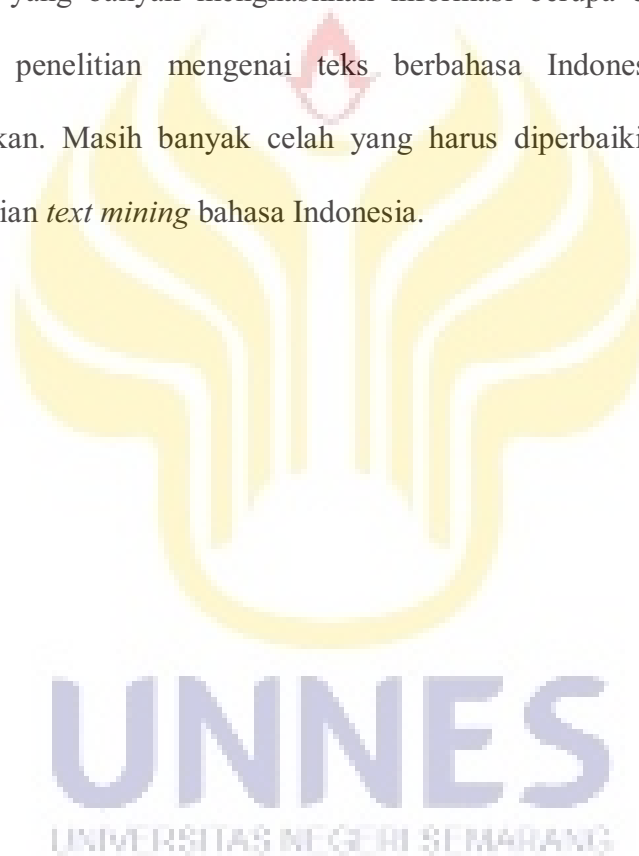
1. Penerapan membangun sistem mengklasifikasikan pelayanan pengaduan masyarakat dengan twitter menggunakan web sebagai alat untuk membuat sistem. Membangun sistem ini menggunakan beberapa tahap antara lain pengambilan data dengan menggunakan API twitter lalu proses *preprocessing* dan tahap terakhir ialah proses pengklasifikasi data uji dengan menghitung nilai probabilitas masing – masing teks.
2. Pada penelitian ini mendapatkan hasil akurasi dari sistem pengklasifikasi pengaduan masyarakat melalui twitter dengan menggunakan metode *naive bayes* yaitu sebesar 77%.

5.2. Saran

Saran yang dapat ditulis untuk penelitian ini adalah sebagai berikut.

- 1) Diharapkan dipenelitian mendatang akan ada bertambahnya kategori dalam proses klasifikasi dan menggunakan metode kombinasi untuk mendapatkan performansi yang lebih baik.

- 2) Belum ada proses pengembalian struktur kata yang disingkat dengan membandingkan dari kamus bahasa Indonesia membuat bahasa yang ada didalam sistem kurang baku.
- 3) Penelitian mengenai *text mining* merupakan salah satu penelitian yang sedang berkembang pesat saat ini seiring dengan berkembangnya teknologi digital yang banyak menghasilkan informasi berupa data tekstual. Akan tetapi, penelitian mengenai teks berbahasa Indonesia belum banyak dilakukan. Masih banyak celah yang harus diperbaiki dalam melakukan penelitian *text mining* bahasa Indonesia.



DAFTAR PUSTAKA

- Chae, B. 2014. Insights from hastag #supplyehain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal Production Economics*. Volume:165. Page:247-259.
- Deshpande, V. K.B. 2015. *Predictive Analytics and Data Mining*. USA: Morgan Kaufmam.
- Kavanaugh, A. L., et al. 2012. Social media use by government: From the routine to the critical. *Government Information Quarterly*, 29(4), 480–491.
- Lathrop, D. & Laurel, R. 2013. *Open Government*. California: O'Rilley.
- Manning, C.D., Raghavan, P., & Schütze, H. 2008, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge.
- Hidayah, M.R., Akhlis, I., & Sugiharti, E. 2017. Recognition Number of The Vehicle Plate Using Otsu Method and K-Nearest Neighbour Classification. *Scientific Journal of Informatics*. Vol 4(1):66-75.
- Munson, S., & Glaisyer, T. (2010). *Social media technology and government technology*. *Computer*, 43(11), 53–59.
- Noh, H., Jo, Y., & Lee, S. 2015. Keyword selection and procesing strategy for applying text mining to patent analysis. *Expret System With Application*, 42,4348-4360.
- Obama, B. 2009. President's Memorandum on *Transparency and Open Government - Interagency*. Washington DC: Executive Office of The President of United States.
- Pang, B. & Lee, I. 2008. Opinion Mining and Sentiment Analysis, *Foundation in Information Barivel*.1-135.
- Prasteyo E. 2012. *Buku Data Mining Konsep dan menggunakan MATLAB*, Yogyakarta: Andi
- Rish, I. 2006. An empirical study of The Naive Bayes Classifier, *International Joint Conference on Artificial Intelligence*, California.

- Rodiyansyah, F.S., & Winarko, E. 2012. Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan *Naive Bayesian Classification*. *Jurnal IJCCS*, Vol.6, No.1 , January 2012. 91-100.
- Rogers, P., & Lindsey, T.D. 2012. *Principle of Open Government: Transparency, Participation and Collaboration*. California Research Bureau.
- Sembiring, T. 2013. *Pengguna Internet di Indonesia 63 Juta Orang. Online*. Tersedia di <https://kominfo.go.id/>. [diakses 2-8-2016].
- Sibero, A.F.K. 2011. *Kitab Suci Web Programing*, Yogyakarta: MediaKom.
- Sobaci, M.Z., & Karkin, N. 2013. The use of twitter by mayors in Turkey: Tweets for better public services? *Government Information Quarterly*. 30, 417–425.
- Somantri, O., Wiyono, S., & Dairoh. 2016. Metode *K-Means* untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan *Support Vector Machine (SVM)*. *Scientific Journal of Informatics*. Vol 3(1):34-45.
- Sutarman. 2012. *Buku Pengantar Teknologi Informasi*. Jakarta: Bumi Aksara.
- Sutabri, T. 2012. *Analisis Sistem Informasi*. Yogyakarta: Andi.
- Tan, P. N., Steinbach, M., & Kumar, V. 2006, *Introduction to Data Mining*. Boston: Pearson Education.
- Twitter, 2012. *Twitter Api*. Tersedia di <https://dev.twitter.com/docs/api>. [diakses 10-8-2016].
- Yakub. 2012. *Pengantar Sistem Informasi*, Yogyakarta: Graha Ilmu.