

**ANALISIS SENTIMEN MEDIA SOSIAL (TWITTER) TERHADAP
LAYANAN *PROVIDER* TELEKOMUNIKASI (TELKOMSEL)
MENGUNAKAN METODE *MULTINOMIAL
NAÏVE BAYES***

SKRIPSI



Oleh :

WAHDATUN NOVI YANTI

11650006

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2018**

HALAMAN PENGAJUAN

**ANALISIS SENTIMEN MEDIA SOSIAL (TWITTER) TERHADAP
LAYANAN PROVIDER TELEKOMUNIKASI (TELKOMSEL)
MENGUNAKAN METODE *MULTINOMIAL
NAÏVE BAYES***

SKRIPSI

Diajukan Kepada :
Dekan Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memperoleh Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
WAHDATUN NOVI YANTI
NIM. 11650006

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2018**

LEMBAR PERSETUJUAN
ANALISIS SENTIMEN MEDIA SOSIAL (TWITTER) TERHADAP
LAYANAN PROVIDER TELEKOMUNIKASI (TELKOMSEL)
MENGGUNAKAN METODE MULTINOMIAL
NAÏVE BAYES

SKRIPSI

Oleh:
WAHDATUN NOVI YANTI
NIM. 11650006

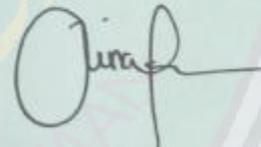
Telah Disetujui untuk diuji
Malang, 18 Mei 2018

Dosen Pembimbing I

Dosen Pembimbing II



Fachrul Kurniawan, M.MT
NIP. 19771020 200901 1 001



Ainatul Mardhiah, M.Cs
NIDT. 19860330 20160801 2 075

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



De Cahyo Crysdian
NIP. 19740424 200901 1 008

LEMBAR PENGESAHAN
ANALISIS SENTIMEN MEDIA SOSIAL (TWITTER) TERHADAP
LAYANAN PROVIDER TELEKOMUNIKASI (TELKOMSEL)
MENGGUNAKAN METODE MULTINOMIAL
NAÏVE BAYES

SKRIPSI

Oleh :

WAHDATUN NOVI YANTI
NIM. 11650006

Telah Dipertahankan Di Depan Dewan Penguji Skripsi
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)

Tanggal : 6 Mei 2018

Susunan Dewan Penguji

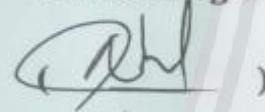
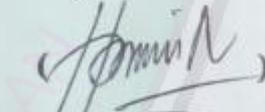
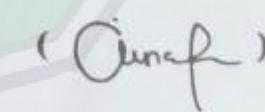
Tanda Tangan

Penguji Utama : Fressy Nugroho, M.T
NIP. 19710722 201101 1 001

Ketua Penguji : Hani Nurhayati, M.T
NIP. 19780625 200801 2 006

Sekretaris Penguji : Fachrul Kurniawan, M.MT
NIP. 19771020 200901 1 001

Anggota Penguji : Ainatul Mardhiyah, M.Cs
NIDT. 19860330 20160801 2 075

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Ganjo Crysdian
NIP. 19740424 200901 1 008

HALAMAN PERSEMBAHAN

Segala puji syukur atas limpahan rahmat, taufik, dan hidayah kepada Allah SWT atas terselesaikannya skripsi ini. Tak luput ucapan terima kasih sebanyak-banyaknya kepada :

- a. Bapak Muhammad Asrofi dan Ibu Kufayah, Orang tua saya yang selalu memberi dorongan mulai dari materi dan doa yang tiada henti-hentinya untuk saya
- b. Mbak Nur Kholidah, Mbak Liana, Acak Purwanto dan Mbak Titik, Mbak Munazati, Wak Kumayah , Mas Eko dan Segenap Keluarga yang selalu menyemangati dan memberikan dukungan dan doa
- c. Segenap dosen dan civitas di Jurusan Teknik Informatika yang membimbing saya selama menjalani Studi dengan penuh ketulusan
- d. Teman-teman seperjuangan Integer 2011, khususnya mereka; Emil, Masiti, Ulfa, Faaris, Alvian, Hafidz Nur, Hafidz Maulana, Riki dan seluruh teman dan sahabat saya yang tidak dapat saya sebutkan satu persatu yang selalu memberikan dukungan hingga skripsi saya dapat diselesaikan dengan baik.

HALAMAN PERNYATAAN ORISINALITAS PENELITIAN

Saya yang bertanda tangan dibawah ini :

Nama : Wahdatun Novi Yanti

NIM : 11650006

Fakultas/ Jurusan : Sains dan Teknologi/ Teknik Informatika

Judul Penelitian : Analisis Sentimen Media Sosial (Twitter) Terhadap
Layanan Provider Telekomunikasi (Telkomsel)
Menggunakan Metode *Multinomial Naïve Bayes*

Menyatakan dengan sebenar-benarnya bahwa hasil penelitian saya ini tidak terdapat unsur-unsur penjiplakan karya penelitian atau karya ilmiah yang pernah dilakukan atau dibuat oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata hasil penelitian ini terbukti terdapat unsur-unsur jiplakan, maka saya bersedia untuk mempertanggungjawabkan, serta diproses sesuai peraturan yang berlaku.

Malang, 18 Mei 2018



Wahdatun Novi Yanti
11650006

MOTTO

*"Manusia tidak merancang untuk gagal,
mereka gagal untuk merancang."
(William J. Siegel)*



KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillahrabbi'Alamin penulis haturkan kehadiran Allah SWT yang telah memberikan rahmat, hidayah, dan ridha-Nya, sehingga penulis dapat menyelesaikan studi di jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan skripsi ini dengan baik dan lancar.

Selanjutnya penulis menghaturkan ucapan terima kasih kepada semua pihak yang telah memberikan doa, harapan, dan semangat untuk terselesaikannya skripsi ini. Ucapan terima kasih, penulis sampaikan kepada :

1. Prof. Dr. Abd. Haris, M. Ag. Selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang, yang telah banyak memberikan pengetahuan dan pengalaman yang berharga.
2. Dr. Sri Harini, M.Si selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Dr. Cahyo Crysdiyan selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang
4. Fachrul Kurniawan, M.MT dan Ainatul Mardhiyah, M.CS selaku dosen pembimbing skripsi, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
5. Ibu Ririn Kusumawati, M.Kom, selaku dosen wali yang juga selalu memberi nasihat akademik kepada penulis selama masa studi.
6. Segenap civitas akademika jurusa Teknik Informatika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya.
7. Semua pihak yang ikut membantu dalam menyelesaikan skripsi ini baik berupa materiil maupun moril.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. Amiin Ya Rabbal Alamiin.

Wasaalamu'alaikum Wr.Wb.

Malang, 18 Mei 2018
Penulis

DAFTAR ISI

HALAMAN PENGANTAR	ii
LEMBAR PERSETUJUAN	iii
LEMBAR PENGESAHAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN PERNYATAAN	vi
MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
ABSTRACT	xiv
المخلص	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.4 Manfaat Penelitian	6
1.5 Batasan Penelitian	6
BAB II KAJIAN PUSTAKA	8
2.1 Penelitian Terdahulu	8
2.2 Rancangan Bangun Web Base Application	10
2.3 Analisis Sentimen	10
2.4 Media Sosial (<i>Twitter</i>)	12
2.5 Porter Stemmer	16
2.5.1 Tentang porter Stemmer	16
2.5.2 <i>Porter Stemmer</i> Bahasa Indonesia	16
2.6 Operator Seluler (Provider) GSM (Global System for Mobile communications)	21
2.7 Pembobotan TF-IDF	22
2.8 <i>Naïve Bayes</i>	24
2.9 <i>Multinomial Naïve Bayes</i>	25
BAB III DESAIN DAN IMPLEMENTASI	27
3.1 Gambaran Penelitian	27
3.2 Gambaran Umum	27
3.3 Perancangan Desain Sistem	30
3.3.1 Perancangan Alur Sistem	31
3.3.2 Flowchart	33
3.4 Alur Manual Analisis Sentimen	35
3.5 Implementasi Algoritma	36
3.5.1 Data	36
3.5.2 Metode Pengolahan Data	36
3.5.3 <i>Multinomial Naïve Bayes</i>	37
BAB IV HASIL DAN PEMBAHASAN	44
4.1 Deskripsi Aplikasi	44
4.2 Implementasi	44
4.2.1 Pengambilan data <i>Twitter</i>	44
4.2.2 Ekstraksi Fitur	45

4.2.3	Pembobotan kata	46
4.2.4	Multinomial <i>Naïve Bayes</i>	47
4.2.5	Tampilan antarmuka	49
4.3	Pengujian	50
4.3.1	Multinomial <i>Naïve Bayes</i>	51
4.3.2	Presisi dan <i>Recall</i>	52
4.3.3	Kecepatan	54
4.4	Pembahasan	55
BAB V PENUTUP		64
5.1	Kesimpulan	64
5.2	Saran	65
DAFTAR PUSTAKA		66
LAMPIRAN 1		68



DAFTAR TABEL

Tabel 2. 1 Aturan 1 <i>Inflection Particle</i>	18
Tabel 2. 2 <i>Inflection Possesive Pronouns</i>	18
Tabel 2. 3 <i>First Order of Derivational Prefixs</i>	19
Tabel 2. 4 <i>Second Order of Derivational Prefixs</i>	20
Tabel 2. 5 <i>Derivational suffixes</i>	20
Tabel 2. 6 Suku Kata dalam Bahasa Indonesia	21
Tabel 3. 1 Contoh Kasus Data Training	37
Tabel 3. 2 Hasil Perhitungan Prior Probabilities Untuk Setiap Kelas.....	38
Tabel 3. 3 Contoh term pada dokumen	38
Tabel 3. 4 Model Perhitungan Klasifikasi.....	40
Tabel 3. 5 Data Testing	41
Tabel 3. 6 Hasil Nilai Conditional Probabilities	41
Tabel 3. 7 Hasil Perkalian Nilai Conditional Probabilities dengan Term Frequency-nya	42
Tabel 4. 1 Kelas Positif	51
Tabel 4. 2 Kelas Positif Relevan	51
Tabel 4. 3 Kelas Negatif.....	52
Tabel 4. 4 Presisi dan <i>Recall</i> Kelas	53
Tabel 4. 5 Presisi dan <i>Recall</i> Kelas Negatif	53
Tabel 4. 6 Presisi dan <i>Recall</i> Kelas Netral	53

DAFTAR GAMBAR

Gambar 2. 1 Alur Port Stemmer.....	17
Gambar 3. 1 Alir Sistem.....	28
Gambar 3. 2 Alur proses besar aplikasi.....	30
Gambar 3. 3 Perancangan Alur proses besar aplikasi	31
Gambar 3. 4 Desain entity	33
Gambar 3. 5 Alir pre-poses	34
Gambar 3. 6 Alur Preproses	34
Gambar 3. 7 Alur Proses Klasifikasi.....	35
Gambar 4. 1 Fitur pencarian Twitter	45
Gambar 4. 2 Input data	50
Gambar 4. 3 Hasil Pie Chart.....	50
Gambar 4. 4 Contoh Data Latih	56
Gambar 4. 5 Hasil sentimen terhadap telkomsel.....	61



ABSTRAK

Novi yanti, Wahdatun. 2018., **Analisis Sentimen Media Sosial (Twitter) Terhadap Layanan Provider Telekomunikasi (Telkomsel) Menggunakan Metode Multinomial Naïve Bayes** Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) Fachrul Kurniawan, M.MT (II), Ainatul Mardiyah, M.Cs

Kata Kunci: Twitter, Analisis Sentimen, Multinomial *Naïve Bayes*

Mikroblog Twitter dewasa ini menjadi salah satu rujukan bagi para netizen untuk memberikan kesan dan pendapatnya, baik opini yang bersifat positif ataupun bersifat negatif. Para produsen barang dan jasa di Negara barat memanfaatkannya sebagai bahan koreksi terhadap barang dibuat apakah para konsumen merasa puas atau tidak. Di Indonesia hal tersebut dapat di implementasikan juga sebagai tolak ukur kinerja layanan dari produk, karena sebagian besar dari mereka merupakan pengguna twitter.

Analisis Sentimen merupakan teknik yang digunakan untuk melakukan evaluasi dan mengidentifikasi emosi dan opini baik positif maupun negatif. Penelitian ini mencoba memanfaatkan dengan menganalisis tweet berbahasa Indonesia yang membicarakan tentang *merk* atau *brand* (provider seluler). *Merk* dan *brand* tersebut dianalisis merupakan yang memiliki nilai popularitas tinggi.

Oleh karena itu, perlu dibuat yang mampu mengetahui respon sentiment *netizen* terhadap objek yang ditentukan baik bersifat positif maupun bersifat negatif menggunakan metode Multinomial *Naïve Bayes* yang berfungsi sebagai penentu relevansi dari hasil pengambilan data tweet di twitter (*crawling*) terhadap tema yang diinginkan dan kemudian metode Multinomial *Naïve Bayes* dipakai sebagai pengklasifikasian hasil relevansi terhadap dua kategori yaitu positif dan negative. Berdasarkan hasil pengujian dapat disimpulkan bahwa metode Multinomial *Naïve Bayes* dapat memberikan hasil relevansi dan klasifikasi akurat terhadap analisis sentiment.

ABSTRACT

Novi yanti, Wahdatun. 2018. **Social Media Sentiment Analysis (Twitter) Against Telecommunication Provider Service (Telkomsel) Using Multinomial Method Naïve Bayes Thesis**. Department of Informatics Faculty of Science and Technology State Islamic University Maulana Malik Ibrahim Malang.

Counselor: (I) Fachrul Kurniawan, M.MT (II), Ainatul Mardhiyah, M.Cs

Keywords: Twitter, Sentiment Analysis, Multinomial Naïve Bayes

Twitter microblogs today become one of the references for netizens to give the impression and opinion, either opinion that is positive or negative. The producers of goods and services in the western State use it as a corrective material to the goods made whether the consumers are satisfied or not. In Indonesia it can be implemented also as a benchmark of service performance of the product, because most of them are users of twitter.

Sentiment Analysis is a technique used to evaluate and identify both positive and negative emotions and opinions. This research tries to make use of by analyzing Indonesian tweets that talk about brand or brand (cellular provider). Brand and brand are analyzed is that has a high popularity value.

Therefore, it is necessary to be made able to know the response of netizen sentiment toward the specified object either positive or negative using Multinomial Naïve Bayes method which serves as determinant of relevance of tweet data retrieval result on the desired theme and then Multinomial method Naïve Bayes is used as a classification of relevance results in two categories: positive and negative. Based on the test results can be concluded that Multinomial Naïve Bayes method can provide accurate relevance and classification of sentiment analysis.

المخلص

نوفي يانتي ، وحدة ٢٠١٨ ، تحليل المشاعر الاجتماعية وسائل الإعلام) تويتر (ضد خدمة مزود خدمة الاتصالات) تكمسل (باستخدام طريقة متعدد الحدود .قسم المعلوماتية كلية العلوم والتكنولوجيا الجامعة الإسلامية - مولانامالك إبراهيم مالانج.

المستشار (I) :فشرل كورنياوان، م.مت(II) ، اينتل مرضية، م.قس
كلمات البحث :تويتر ، تحليل المشاعر ، والخلفيات متعددة السداجة

أصبحت المدونات الصغيرة على تويتر اليوم واحدة من الإشارات لمستخدمي الإنترنت لإعطاء الانطباع والرأي ، إما رأي إيجابي أو سلبي .منتجي السلع والخدمات في الدول الغربية استخدامه بمثابة تصحيح ليكون على ما إذا كانت راضية السلع الاستهلاكية أم لا .في إندونيسيا ، يمكن تطبيقه أيضًا كميّار لأداء الخدمة للمنتج ، لأن معظمهم من مستخدمي تويتر .
تحليل المشاعر هو أسلوب يستخدم لتقييم وتحديد كل من المشاعر والآراء الإيجابية والسلبية .يحاول هذا البحث الاستفادة من تحليل التغريدات الإندونيسية التي تتحدث عن العلامة التجارية أو العلامة التجارية (المزود الخليوي) .يتم تحليل العلامة التجارية والعلامة التجارية التي لديها قيمة شعبية عالية .
لذلك ، فمن الضروري أن تكون قادرة على معرفة استجابة مشاعر مستخدمي الإنترنت تجاه الكائن المحدد إما إيجابية أو سلبية باستخدام طريقة بايز مولاييد متعدد الاستخدامات التي تعمل كمحدد لمدى صلة نتيجة استرجاع بيانات على الموضوع المطلوب ثم الأسلوب متعدد الحدود يستخدم بايز يمكن أن كتصنيف لنتائج الملاءمة في فئتين :الإيجابية والسلبية .استنادا إلى نتائج الاختبار يمكن الاستنتاج بأن طريقة ملتتميل نيفي بايز يمكن أن توفر نتائج دقيقة وتصنيف لتحليل المشاعر .



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini situs *microblogging* telah menjadi alat komunikasi yang sangat populer dikalangan pengguna internet. Dimana jutaan pesan muncul setiap hari di situs *web* populer yang menyediakan layanan *microblogging* seperti *Twitter*, *Tumblr*, dan *Facebook*.

Hal tersebut menyebabkan semakin banyak pengguna yang melakukan unggahan tentang suatu produk dan layanan yang mereka gunakan, atau mengekspresikan pandangan mereka terhadap sesuatu. *Twitter* sebagai salah satu situs *microblogging* dengan pengguna lebih dari 500 juta dan 400 juta *tweet* perhari (Farber, 2012), memungkinkan pengguna untuk berbagi pesan menggunakan teks pendek disebut *Tweet* (Twitter, 2013). *Twitter* dapat menjadi sumber data pendapat dan sentimen masyarakat. Data tersebut dapat digunakan secara efisien untuk pemasaran atau studi sosial (Pak & Paroubek, 2010).

Sentimen analysis atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Pang & Lee, 2008).

Saat ini, dua situs jejaring sosial *Facebook* (yang diciptakan tahun 2004) dan *Twitter* (yang diciptakan tahun 2006) sebagai jejaring sosial ciptaan terbaru

yang menawarkan fitur-fitur yang lebih *fresh* dalam berkomunikasi memperoleh perhatian yang dimata khalayak dunia, termasuk Indonesia yang perkembangan pengguna situs jejaring sosial *Facebook* dan *Twitter* semakin tahun meningkat dengan cepat. Selain sebagai alat komunikasi, media sosial biasa dimanfaatkan banyak orang untuk wadah promosi, iklan pertokoan, bahkan sampai kampanye politik. Atau wadah penyapaian pendapat terkait opini-opini publik, kritik dan saran yang disampaikan kepada tokoh tertentu.

Hal ini berkaitan dengan hadist Rasulullah SAW, sebagai berikut:

“ فَلَمْ يَزَلْ عُمَرُ يُرَاجِعُنِي فِيهِ حَتَّى شَرَحَ اللَّهُ لِي ذَلِكَ صَدْرِي، وَرَأَيْتُ الَّذِي رَأَى عُمَرُ

Artinya “Umar senantiasa membujukku untuk mengevaluasi pendapatku dalam permasalahan itu hingga Allah melapangkan hatiku dan akupun berpendapat sebagaimana pendapat Umar” [HR. Bukhari]. Seseorang dapat terbantu untuk mengevaluasi diri dengan bermusyawarah bersama dengan niat untuk mencari kebenaran. Imam Bukhari mengeluarkan suatu riwayat yang menceritakan usul Umar kepada Abu Bakr radhiallahu anhu untuk mengumpulkan Al-Quran. Tatkala saat itu Abu Bakr menolak usul tersebut, tetapi Umar terus mendesak Beliau dan mengatakan bahwa hal itu merupakan kebaikan. Abu Bakr tidak bersikukuh dengan pendapatnya dan ketika terdapat usulan yang baik dan kedudukan Beliau lebih tinggi tidak menghalangi untuk menerima kebenaran pihak yang memiliki pendapat yang berbeda. Hadist tersebut mengajarkan kita bahwa tidak ada manusia yang sempurna yang tak luput dari salah dan lupa sehingga kita sebagai sesama umat manusia wajib untuk saling mengingatkan.

Oleh karena itu, masyarakat sangat menyambut baik hal tersebut, tidak terkecuali para tokoh masyarakat dan pemilik usaha yang sering menyatakan opini dan inovasi, promosi melalui media sosial *twitter*. *Twitter* dianggap lebih diminati oleh pengguna karena dirasa lebih mudah dan *simple* dalam mempresentasikan opininya. Seringkali *tweet* seseorang dijadikan tolak ukur pendapat oleh acara-acara televisi. Sehingga citra yang berusaha direpresentasikan tersebut disisi lain terkadang melampaui realitas dalam kehidupan manusia atau dalam bahasa Jean Baudrillard terjadi *hyperreality*. Sehingga pada titik tertentu, masyarakat modern menerima realitas dengan beraneka macam bentuk citra dihadapkan pada mereka, kemudian serta merta meyakini kebenaran yang diberikan atau dipresentasikan dari citranya. Dengan demikian media sosial mampu membentuk gambaran dengan tujuan mempengaruhi perilaku politik masyarakat.(Debrix, 2010).

Penelitian ini mencoba memanfaatkan dengan menganalisis *tweet* berbahasa Indonesia yang membicarakan tentang merek atau *brand* (Provider Seluler). Merek atau *Brand* dianalisis merupakan yang memiliki nilai popularitas tinggi.

Analisis sentimen merupakan teknik yang digunakan untuk melakukan *evaluasi* dan mengidentifikasi emosi dan opini baik positif maupun negatif (Wilson *et al.* 2009). Penelitian analisis setimen telah banyak dilakukan pada media dokumen, *Twitter* sebagai salah satu media sosial populer dimana penggunaanya dapat megekspresikan opini yang objektif tentang topik yang berbeda (Coletta *et al.* 2014).

Pengguna telekomunikasi di Indonesia dari tahun ke tahun terus tumbuh pesat. Hal ini seiring dengan kebutuhan publik komunikasi yang terus meningkat

baik melalui SMS (*Short Message Service*), telepon maupun data. Dengan demikian terjadi persaingan para *provider* telekomunikasi Indonesia untuk menarik atau mempertahankan pelanggannya. Berbagai opini yang dikemukakan oleh pelanggan tentang *provider* telekomunikasi dapat diketahui melalui media sosial *Twitter*. Berbagai opini sentimen konsumen terhadap produk dalam media sosial *Twitter* merupakan bentuk dari kualitas produk. Begitupun opini sentimen konsumen terhadap *provider* telekomunikasi merupakan bentuk dari kualitas pelayanan *provider* terhadap konsumennya.

Analisis sentimen pada *Twitter* terdapat kelemahan dalam kata-kata yang terdapat pada kalimat yang diposting oleh pengguna situs tersebut. *Twitter* hanya memungkinkan pengguna menulis sebanyak 140 karakter, hal ini menyebabkan para pengguna sering menggunakan singkatan kata dan ejaan kata yang salah.

Dalam pengerjaannya penulis sangat berhati-hati didalam pengerjaannya karena opinion maining tidak dapat diartikan satu persatu kata. Sebagaimana yang diajarkan Nabi Muhammad SAW. Dalam hadist riwayat Al-Turmudzi:

ابن عن جُبَيْرِ بْنِ سَعِيدٍ عَنِ الْأَعْلَى عَبْدِ عَن سُفْيَانَ حَدَّثَنَا السَّرِيُّ بْنُ بِشْرٍ حَدَّثَنَا عَلِيَّانُ بْنُ مَحْمُودٍ حَدَّثَنَا
فَلْيَتَّبِعُوا عِلْمَ بَعْضِ الْقُرْآنِ فِي قَوْلِ مَنْ وَسَلَّمَ عَلَيْهِ اللَّهُ صَلَّى اللَّهُ رَسُولُ قَالَ قَالَ عَنْهُمَا اللَّهُمَّ رَضِيَ عَنَّا
صَحِيحٌ حَسَنٌ حَدِيثٌ هَذَا عَيْسَى أَبُو قَالَ النَّارِ مِنْ مَفْعَدَهُ

Artinya : (Al-Turmuzi berkata): Mahmud bin gailan telah menceritakan kepada kami, (Mahmud berkata): Bisyr bin al-Syariyy menceritakan kepada kami, (Bisyr berkata): Sufyan menceritakan kepada kami dari ‘Abd al-A-‘la dari Sa’id bin Jubair dari Ibn ‘Abbas Ra, ia berkata : Rasulullah SAW bersabda : siapa yang mengatakan tentang (isi) al-Qur’an dengan tanpa landasan pengetahuan, maka

hendaklah ia menempati tempat duduknya dari api neraka” Abu “isa (al-Turmuzi) berkata: hadist ini hasan sahih.

Pada penelitian lain sering digunakan *Naïve Bayes* sebagai klasifikasi opini maupun untuk pengklasifikasian. Dalam penelitian Imam Fahrur Rozi, dkk yang menggunakan *Naïve Bayes* sebagai metode klasifikasi untuk analisis sentimen untuk data ekstraksi opini perguruan tinggi menyebutkan *Naïve Bayes* mempunyai presisi data 0.92 dan recall 0.93 dari seluruh total data klasifikasi. Meskipun dalam penelitian tersebut metode yang digunakan tidak hanya *Naïve Bayes* tetapi juga menggunakan metode lain untuk ekstraksi opini. Tapi pada *Naïve Bayes* yang digunakan tidak bergantung pada metode lain.

Meskipun begitu untuk mempercepat proses perhitungan digunakan algoritma *Multinomial Naïve Bayes*. Algoritma *Multinomial Naïve Bayes* merupakan pengembangan dari *Naïve Bayes* sehingga mungkin digunakan pada pembuatan analisis sentiment dikarenakan algoritma ini bertujuan sebagai mode klasifikasi kedalam kategori positif dan negatif.

Oleh karena itu penulis merasa perlu untuk mengajukan sebuah penelitian dengan judul Analisis sentiment dengan menggunakan metode *Multinomial Naïve Bayes* dengan study kasus untuk mengukur respon masyarakat dunia maya kepada proses kinerja suatu provider telekomunikasi (telkomsel).

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalahnya adalah:

- a. Bagaimana akurasi *Multinomial Naïve Bayes* dalam analisis sentiment pada media sosial (*twitter*) tentang layanan provider telekomunikasi.

- b. Bagaimana klasifikasi opini di media social sehingga bisa memvisualisasikan terhadap layanan provider GSM (*Global System for Mobile communications*)

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

- a. Mengetahui akurasi *Multinomial Naïve Bayes* dalam analisis sentiment pada media sosial (*twitter*) tentang layanan provider telekomunikasi.
- b. Mengklasifikasi opini di media social sehingga bisa memvisualisasikan terhadap layanan provider GSM (*Global System for Mobile communications*)

1.4 Manfaat Penelitian

Manfaat aplikasi ini adalah sebagai berikut:

- a. Memberikan aplikasi tentang layanan telekomunikasi provider GSM (*Global System for Mobile communications*) dengan analisis sentiment
- b. Mencari informasi tentang suatu produk, provider GSM (*Global System for Mobile communications*) dapat menganalisis sentiment tercapainya kepuasan konsumen terhadap berbagai keluhan
- c. Memperoleh hasil analisis sentiment terhadap layanan provider dengan multinomial *Naïve Bayes*

1.5 Batasan Penelitian

Agar penyusunan tugas akhir ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup lingkungan pembahasan dibatasi pada:

- a. Algoritma yang digunakan dalam pengklasifikasian ini adalah *Multinomial Naïve Bayes* dan tidak menggunakan membanding dengan algoritma lain.
- b. Data yang digunakan terdiri dari tweet Provider telekomunikasi (Telkomsel) berbahasa Indonesia saja.



BAB II KAJIAN PUSTAKA

2.1 Penelitian Terdahulu

Peneliti mengenai klasifikasi sentiment telah dilakukan oleh Bo Pang (2002). Pada papernya, Bo Pang melakukan klasifikasi sentiment terhadap review film dengan menggunakan berbagai teknik pembelajaran mesin. Teknik pembelajaran mesin yang digunakan yaitu *Naïve Bayes*, *Maximum Entropy*, dan *Support Vector Machine* (SVM). Pada penelitian ini juga digunakan beberapa pendekatan untuk melakukan ekstraksi fitur, yaitu *unigram*, *unigram+bigram*, *unigram+Part of Speech* (POS), *adjective*, dan *unigram+posisi*. Hasil dari eksperimen yang dilakukan dipenelitian ini menemukan bahwa SVM menjadi metode ketika dikombinasikan dengan unigram dengan akurasi 82.9% (Pang, et. Al. 2002)

Penelitian terhadap sentimen opini *twitter* menggunakan algoritma *Naïve Bayes* untuk mendeteksi polaritas *tweet* bahasa Inggris menunjukkan kinerja terbaik menggunakan *classifier biner* antara dua kategori polaritas tajam positif dan negatif (Gamallo dan Garcia 2014). Analisis kualitas layanan *provider* telepon seluler menggunakan media sosial *Twitter* menggunakan *Naïve Bayes* menunjukkan *provider* dengan tingkat kepuasan pelanggan tertinggi (Setiawan 2014). Penelitian lain mengenai analisis sentimen *Twitter* mengenai operator seluler Indonesia dengan menggunakan metode *Naïve Bayes* diperoleh akurasi 72,22 % (Wijaya et al. 2013). Analisis kualitas layanan *provider* telepon seluler

pada media sosial *Twitter* menggunakan *Naïve Bayes* menunjukkan *provider* dengan tingkat kepuasan pelanggan tertinggi (Calvin dan Setiawan 2014).

Penelitian yang serupa juga dilakukan oleh Paulina Aliandu (2013) dengan judul *Twitter Used by Indonesian President: An Sentimen Analysis of Timeline*. Penelitian ini melakukan eksperimen untuk melakukan klasifikasi sentimen terhadap data yang diperoleh dari *Twitter* dengan mengambil *tweet* akun Presiden RI @SBYudhoyono baik sentimen positif, negatif ataupun netral. Aliandu menerapkan *Naïve Bayes* untuk klasifikasi sentimen tersebut dan dapat mengklasifikasi dengan baik dengan akurasi 79,42% (Aliandu, 2013).

Pada penelitian Riszki Wijayatun Pratiwi (2016) *Naïve Bayes* digunakan untuk melakukan proses prediksi rating film. Sumber data yang diambil dari www.imdb.com sebagai sumber data. Penelitian ini menggunakan banyak variable sebagai data masukkan. Penelitian ini bertujuan mengetahui minat perfilman melalui penilaian film tersebut. Penelitian berhasil mengetahui pemniat film dari penilaian film menggunakan variable input dengan algoritma *Naïve Bayes*.

Penelitian yang dilakukan oleh Imam Fahrur Rozi (2012) yang meneliti data opini publik pada Perguruan Tinggi meneliti tentang ekstraksi data opini publik menggunakan tiga buah subproses dari opinion mining yaitu, document subjectivity, opinion orientation dan target detection. Selain itu pada subproses document subjectivity dan target detection digunakan Part-of-Speech (POS) Tagging menggunakan Hidden Markov Model (HMM). Hasil POS diklasifikasikan menggunakan *Naïve Bayes*. Hasil yang diharapkan adalah pengenalan dokumen yang dikenali sebagai opini selanjutnya diklasifikasikan ke dalam opini negatif

dan positif (subproses opinion orientation). Dari penelitian tersebut berhasil didapatkan data precision dan recall yang besar. Sehingga dapat dikatakan penelitian tersebut berhasil dengan baik. (Imam Fahrur Rozi, 2012)

2.2 Rancangan Bangun Web Base Application

Web base application ialah suatu aplikasi yang dapat berjalan dengan menggunakan basis teknologi *web* atau *browser*. Aplikasi ini dapat diakses dimana saja asalkan terdapat koneksi internet yang mendukung tanpa melakukan penginstallan dikomputer masing-masing seperti pada aplikasi *desktop*.

Keunggulan dari *web base application*

1. Dapat menjalankan aplikasi dimanapun dan kapanpun tanpa melakukan penginstallan
2. Tidak memerlukan lisensi hak cipta untu menggunakan aplikasi, sebab aplikasi itu sudah menjadi tanggung jawab web penyedia aplikasi
3. Dapat dijalankan sistem operasi manapun
4. Dapat diakses melalui beberapa media semisal komputer, laptop dan smartphone

2.3 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisi sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini positif atau negatif.

Sentiment analysis atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistic dan text mining yang

bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenan dengan suatu topik, produk, layanan, organisasi, individu ataupun kegiatan tertentu (Liu, 2012).

Penelitian di bidang opinion mining mulai marak pada tahun 2002. Turney pada tahun 2002 melakukan penelitian dengan tema opinion mining dengan menggunakan data berupa data review konsumen atas suatu produk. Metode yang digunakan adalah Semantic Orientation (Orientasi Semantik) menggunakan Pointwise Mutual Information (SO-PMI). Hasil terbaik yang dicapai adalah 84% akurasi terhadap data review kendaraan bermotor dan 66% untuk data review film Pang et.al. pada tahun 2002 mengklasifikasikan review dari film pada level dokumen yang memiliki pendapat positif atau negatif dengan menggunakan teknik supervised learning. Sekumpulan dari review film yang sebelumnya telah ditentukan menjadi baik positif ataupun negatif digunakan sebagai data latihan untuk beberapa algoritma machine learning yang sudah ada.

Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral. Sentimen analis juga dapat menyatakan perasaan emosional sedih, gembira atau marah.

Opinion mining bisa dianggap sebagai kombinasi antara text mining dan natural language processing. Salah satu metode dari text mining yang bisa digunakan untuk menyelesaikan masalah opinion mining adalah Naïve Bayes Classifier (NBC). NBC bisa digunakan untuk mengklasifikasikan opini ke dalam opini positif dan negatif. NBC bisa berfungsi dengan baik sebagai metode

pengklasifikasi teks. Penelitian tentang penggunaan NBC sebagai metode pengklasifikasi teks telah dilakukan oleh SM Kamaruzzaman dan Chowdury Mofizur Rahman (Kamaruzzaman, 2014) serta Ashraf M Kibriya et.al (Kibriya Ashraf M., 2014).pada tahun 2004. Dari proses pengujian secara kualitatif disebutkan bahwa teks bisa diklasifikasikan dengan akurasi yang tinggi.

Ekspresi atau sentimen mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda. Dari sentiment kita dapat mencari pendapat tentang produk-produk, merek atau orang-orang dan menentukan apakah mereka dilihat positif atau negatif di *web* (Saraswati, 2011). Hal ini memungkinkan kita untuk mencari informasi tentang:

- a. Detektif Flame (rants buruk)
- b. Persepsi produk baru
- c. Persepsi merek
- d. Manajemen reputasi

Ekspresi atau sentimen mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda. Oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining (Barber, 2010).

2.4 Media Sosial (*Twitter*)

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh Twitter Inc., yang menawarkan jaringan sosial berupa *microblog* sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan *Tweets*

(Twitter, 2013). *Microblog* adalah salah satu jenis alat komunikasi *online* dimana pengguna dapat memperbarui status tentang mereka yang sedang memikirkan dan melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu. *Tweets* adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. *Tweets* bisa dilihat secara publik, namun pengiriman dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Pengguna dapat melihat *tweets* pengguna lain yang dikenal dengan sebutan pengikut (*follower*). Pesan pada awalnya diatur hanya mempunyai batasan 240 karakter disesuaikan dengan kompatibilitas dengan pesan SMS, perkenalkan singkatan notasi dan slang yang biasa digunakan dalam pesan SMS. Batas karakter 240 juga meningkatkan penggunaan layanan mempendek URL seperti bit.ly, goo.gl, dan tr.im, dan jasa *hosting* konten seperti *Twitpic*, *Tweephoto*, *memozu.com* dan *NotePub* untuk mengakomodasi multimedia isi dan teks yang lebih panjang daripada 240 karakter (Twitter, 2017).

Tidak seperti *Facebook*, *LinkedIn* dan *MySpace*, *Twitter* merupakan sebuah jejaring sosial yang dapat digambarkan sebagai sebuah *graph* berarah (Wang, 2010), yang berarti bahwa pengguna dapat mengikuti pengguna lain, namun pengguna kedua tidak diperlukan untuk mengikutinya kembali. Kebanyakan akun berstatus publik dapat diikuti tanpa memerlukan persetujuan pemilik.

Semua pengguna dapat mengirim dan menerima *tweet* melalui situs *Twitter*, aplikasi eksternal yang kompatibel (telepon seluler), atau dengan pesan singkat (SMS) yang tersedia di negara-negara tertentu (Twitter, 2013). , apa pendapat mereka tentang suatu objek atau fenomena tertentu. *Tweets* adalah teks

tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. *Tweets* bisa dilihat secara publik, namun pengiriman dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda # (*hashtag*). Sedangkan untuk menyebutkan atau membalas pesan dari pengguna lain bisa menggunakan tanda @.

Pesan pada awalnya diatur hanya mempunyai batasan 240 karakter disesuaikan dengan kompatibilitas dengan pesan SMS, perkenalkan singkatan notasi dan slang yang biasa digunakan dalam pesan SMS. Batas karakter 240 juga meningkatkan penggunaan layanan memperpendek URL seperti *bit.ly*, *goo.gl*, dan *tr.im*, dan jasa *hosting* konten seperti *Twitpic*, *Tweephoto*, *memozu.com* dan *NotePub* untuk mengakomodasi multimedia isi dan teks yang lebih panjang daripada 240 karakter (Twitter, 2017). Twitter menggunakan *bit.ly* untuk memperpendek otomatis *url* yang dikirim-tampil. Fitur yang terdapat dalam Twitter antara lain:

1. Laman utama (*Home*)

Pada halaman utama kita bias melihat *tweets* yang dikirim oleh orang-orang yang menjadi teman kita atau yang kita ikuti (*following*)

2. Profil (*Profile*)

Pada halaman ini yang akan dilihat oleh seluruh orang mengenai profil atau data diri serta *tweets* yang sudah pernah kita buat

3. Followers

Pengikut adalah pengguna lain yang ingin menjadikan kita sebagai teman. Bila pengguna lain menjadi pengikut akun seseorang, maka *tweets* seseorang yang ia ikuti tersebut akan masuk ke dalam halaman utama

4. Following

Kebalikan dari pengikut, *following* adalah akun seseorang yang mengikuti akun pengguna lain agar *tweets* yang dikirim oleh orang yang diikuti tersebut masuk ke dalam halaman utama

5. Mentions

Biasanya konten ini merupakan balasan dari percakapan agar sesama pengguna bisa langsung menandai orang yang diajak bicara

6. Favorite

Tweets ditandai sebagai favorit agar tidak hilang oleh halaman sebelumnya.

7. Pesan Langsung (*Direct Message*)

Fungsi pesan langsung lebih bisa disebut SMS karena pengirim pesan langsung diantara pengguna.

8. Hashtag

Hashtag “#” yang ditulis didepan topik tertentu agar pengguna lain bisa mencari topik yang sejenis yang ditulis oleh orang lain juga.

9. List

Pengguna *twitter* dapat mengelompokkan ikutan mereka ke dalam satu group sehingga memudahkan untuk dapat melihat secara keseluruhan para nama pengguna (*username*) yang mereka ikuti (*follow*)

10. Topik terkini (*Trending Topic*)

Topik yang sedang banyak dibicarakan banyak pengguna dalam suatu waktu bersamaan.

2.5 Porter Stemmer

2.5.1 Tentang porter Stemmer

Porter stemmer (Porter, 2006) merupakan sebuah algoritma yang digunakan untuk melakukan proses menghapus akhiran infleksional dan morfologi yang umum pada Bahasa Inggris. Terutama, penggunaan porter stemmer sebagai proses normalisasi pada kata yang digunakan pada saat membangun sebuah pemerolehan informasi (*Information retrieval*).

Algoritma porter stemmer didasarkan pada ide akhiran dalam Bahasa Inggris, secara umum terdiri dari kombinasi akhiran yang sederhana. Proses pada algoritma porter stemmer melalui beberapa tahapan, simulasi proses infleksional dan derivasional pada sebuah kata.

Pada tahapan tersebut, sebuah akhiran dihapus melalui aturan substitusi, aturan substitusi diterapkan pada suatu kondisi harus terpenuhi secara keseluruhan. Dalam suatu kondisi kata dasar yang diakhiri dengan huruf mati maupun huruf hidup, maka panjang pada kata dasar yang telah dihasilkan disebut dengan measure.

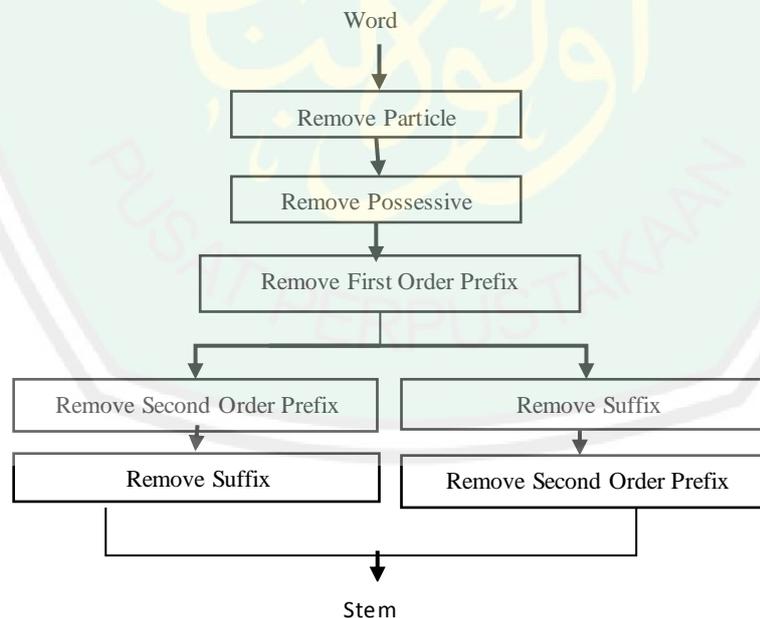
2.5.2 Porter Stemmer Bahasa Indonesia

Dalam Bahasa Indonesia terdapat sebuah kombinasi pada struktur yaitu infleksional dan derivasional yang sederhana. Tahap-tahap dalam porter stemmer adalah simulasi proses pada infleksional dan derivasional dalam Bahasa Inggris cocok pada Bahasa Indonesia.

Penggunaan bahasa yang berbeda tersebut adalah sangat berbeda sehingga terdapat aturan tersendiri pada penggunaan algoritma porter stemmer untuk

Bahasa Indonesia. Aturan untuk Bahasa Indonesia diantaranya, menghilangkan awalan, konflik, dan penyesuaian ejaan pada kasus, Dalam Bahasa Indonesia terdapat sebuah kombinasi pada struktur yaitu infleksional dan derivasional yang sederhana. Pada tahapan tersebut, sebuah akhiran dihapus melalui aturan substitusi, aturan substitusi menerapkan pada suatu kondisi harus terpenuhi secara keseluruhan. Dalam suatu kondisi kata dasar yang diakhiri dengan huruf mati maupun huruf hidup, maka panjang pada kata dasar yang telah dihasilkan diesbut dengan measure.

Tahap-tahap dalam porter stemmer adalah simulasi proses pada infleksional dan derivasional dalam Bahasa Inggris cocok pada Bahasa Indonesia dimana pada karakter awal pada sebuah kata dasar dapat berubah pada saat penambahan awalan dilakukan.



Gambar 2.1 Alur Port Stemmer

Aturan untuk melakukan pengakaran mengikuti rumus yang sudah dilakukan penelitian sebelumnya. Untuk aturan disajikan pada tabel dibawah.

Aturan 1 *Inflection Particle* digunakan pada akhiran yang menunjuk pada sebuah partikel.

Tabel 2.1 Aturan 1 *Inflection Particle*

Akhiran	Replacement	Measure Conditional	Additional Conditional	
-kah	NULL	2	NULL	Pensilkah -> pensil
-lah	NULL	2	NULL	Janganlah -> jangan
-pun	NULL	2	NULL	Pensilpun -> pensil

Inflection Particle untuk menunjukkan kata ganti orang.

Tabel 2.2 *Inflection Possesive Pronouns*

Akhiran	Replacement	Measure Conditional	Additional Conditional	
-ku	NULL	2	NULL	Pensilku -> pensil
-mu	NULL	2	NULL	bukumu -> buku
-nya	NULL	2	NULL	Pensilnya -> pensil

Aturan untuk mencari affix me- yang diikuti kata kerja yang tidak mengubah bentuk kata asal.

Tabel 2.3 First Order of Derivational Prefixs

Akhiran	Replacement	Measure Conditional	Additional Conditional	
meng-	NULL	2	NULL	menggali -> gali
meny-	NULL	2	NULL	Menyapa -> apa
men-	NULL	2	NULL	Mencuri -> curi
mem-	NULL	2	NULL	Membaca -> baca
me-	NULL	2	NULL	Menulis -> tulis
peng-	NULL	2	NULL	Menggali -> gali
peny-	NULL	2	NULL	Penyayang -> sayang
pen-	NULL	2	NULL	Pencuri -> curi
pem-	NULL	2	NULL	Pembawa -> bawa
di-	NULL	2	NULL	Dicuri -> curi
ter-	NULL	2	NULL	Tercuri -> curi
ke-	NULL	2	NULL	Kepada -> pada

Aturan untuk turunan kata kedua

Tabel 2.4 *Second Order of Derivational Prefixs*

Akhiran	Replacement	Measure Conditional	Additional Conditional	
Ber-	NULL	2	NULL	Berlutut -> lutut
Bel-	NULL	2	NULL	Belajar -> ajar
Be-	NULL	2	NULL	Bekerja -> kerja
Per-	NULL	2	NULL	Perjelas -> jelas
Pel-	NULL	2	NULL	Pelajar -> ajar
Pe-	NULL	2	NULL	Penulis -> tulis

Aturan untuk suffix dan tidak merubah kata asal.

Tabel 2.5 *Derivational suffixes*

Akhiran	Replacement	Measure Conditional	Additional Conditional	
-kan	NULL	2	Awalan E {ke, Peng}	Tuliskan -> tulis Carikan -> cari
-an	NULL	2	Awalan E {di, meng, ter}	(per)janjian -> janji Minuman -> minum

-i	NULL	2	V K.... c ₁ c ₁ , c ₁ ≠ s, c ₂ ≠ s, dan awalan € {ber, ke, peng}	Curangi -> curang (men)jalani -> jalan
----	------	---	--	---

Suku kata tersebut untuk memisahkan kata yang sudah berubah dari bentuk asal.

Tabel 2.6 Suku Kata dalam Bahasa Indonesia

Akhiran	Replacement	Meansure Conditional
0	Kh, ng, ny	Kh, ng, ny
1	Ma, af, nya, nga	Ma, af, nya, nga
2	Maaf, kami, rumpun, kompleks	Ma-af, ka-mi, pum-pun, kom-pleks
3	Mengapa, menggunung, tanda	Meng-apa, meng-gu-nung, tan-da-i

2.6 Operator Seluler (Provider) GSM (Global System for Mobile communications)

Telekomunikasi telah menjadi kebutuhan bagi masyarakat luas yang tidak dapat dihindari. Berkembangnya pengguna teknologi komunikasi di Indonesia mengakibatkan perkembangan teknologi informasi mampu menggeser media komunikasi dari kebutuhan sekunder atau tersier menjadi kebutuhan primer, dimana dulu telepon seluler menjadi barang mewah bagi kelas menengah ke atas, sekarang hampir seluruh elemen kelas masyarakat telah memiliki telepon seluler sebagai bagian dari kebutuhan dan gaya hidup. Seiring dengan perkembangan

kebutuhan akan telepon seluler disertai pula dengan kehadiran teknologi operator seluler (*Provider*) yang menyediakan dan memasarkan produk SIM Card (*Subscriber Identity Module Card*) yang diidentifikasi sebagai kartu pintar yang ditaruh ditelepon seluler yang menyimpan kunci mengenal jasa telekomunikasi. Kartu komunikasi tersebut dalam sistem GSM (*Global System for Mobile communications*) dan CDMA.

Melalui kemajuan teknologi komunikasi mampu menipiskan bahkan meniadakan jarak geografis melalui media komunikasi *virtual* yang diperlukan dalam kehidupan sehari-hari, sehingga tidak mengherankan jika bisnis telekomunikasi mengalami perkembangan yang sangat pesat dalam beberapa tahun terakhir ini. Di Indonesia sendiri terdapat beberapa perusahaan operator seluler GSM antara lain: Telkomsel (PT. Telekomunikasi Seluler), Indosat Ooredoo (PT. Satelit Indonesia / Satelindo), XL Axiata (PT XL Axiata Tbk), Hutchison (PT Hutchison CP Telecommunications Indonesia / HCPT). Masing-masing *provider* mempunyai produk yang berbeda-beda, misalnya Telkomsel (Simpati, AS dan Halo), Indosat Ooredoo (Im3 dan Mentari), XL Axiata (XL dan Axis), Hutchison (Tri).

2.7 Pembobotan TF-IDF

Term Frequency (tf) faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.

Inverse Document Frequency (idf) faktor yaitu pengurangan dominasi *term* yang sering muncul diberbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul diberbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kejarang-munculan kata (*term scarcity*) dalam koleksi dokumen harus diperhatikan dalam pemberian bobot. Menurut Mandala (dalam Witten, 1999) kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (*uncommon term*) dari pada kata yang muncul pada banyak dokumen. Pembobotan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*inverse document frequency*). Hal ini merupakan usulan dari George Zipf. Zipf mengamati bahwa frekuensi dari sesuatu cenderung kebalikan secara proposional dengan urutannya.

Metode TF-IDF merupakan metode pembobotan *term* yang banyak digunakan sebagai metode pembandingan terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot *term* t dalam sebuah dokumen dilakukan dengan mengalihkan nilai *term frequency* dengan *Inverse Document Frequency*. Pada *Term Frequency* (tf), terdapat beberapa jenis rumusan yang dapat digunakan yaitu (mandala, 2004):

- tf biner (*binery tf*), hanya memperhatikan apakah suatu kata ada atau tidak dalam dokumen, jika ada diberi nilai satu, jika tidak diberi nilai nol
- tf murni (*raw tf*), nilai tf diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya, jika muncul lima kali maka kata tersebut akan bernilai lima.

2.8 *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat (Feldman & Sanger 2001). Dalam penelitian ini yang menjadi data uji adalah dokumen tweets. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Dalam algoritma *Naïve Bayes* setiap dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori Tweet. Pada saat klasifikasi algoritma mencari probabilitas tertinggi dari semua.

Metode yang digunakan adalah *naïve bayes* yaitu metode yang mempunyai perhitungan matematik dasar yang sangat kuat serta dalam efisiensi klasifikasinya juga stabil, namun kekurangannya adalah parameter model *naïve bayes* perlu diperkirakan dan kurang peka terhadap data yang sudah hilang. Model *naïve bayes* memiliki tingkat kesalahan yang sangat minimum jika dibandingkan dengan algoritma klasifikasi lainnya [4]. Metode *naïve bayes* ini merupakan salah satu metode yang populer untuk pengkategorian teks dengan frekuensi kata sebagai fitur. Hal ini dapat disimpulkan bahwa fitur-fitur yang independen dapat dibuktikan dalam algoritma klasifikasi menjadi lebih efektif [5]

Naïve Bayes (Tan & Kumar, 2006) merupakan salah satu metode yang digunakan untuk pengklasifikasian sebuah data dengan berdasarkan *teorema bayes* dengan mengasumsikan bahwa data memiliki sifat tidak saling terkait antar

satu dengan yang lain atau disebut *independen*. Teknik penggunaan *Naïve Bayes* sangat sederhana dan cepat dengan penggunaan probabilitas. Namun normalnya *Naïve Bayes* memiliki input dalam bentuk binomial yakni hanya 1 dan 0. Mengatasi hal tersebut digunakan *Multinomial Naïve Bayes* yang memiliki input angka desimal. Memaksimalkan penggunaan dari metode *Naïve Bayes* maka digunakan metode *Multinomial Naïve Bayes*.

2.9 *Multinomial Naïve Bayes*

Multinomial Naïve Bayes merupakan proses pengambilan jumlah kata yang muncul dalam setiap dokumen, dengan mengasumsikan dokumen memiliki beberapa kejadian dalam kata dengan panjang tidak tergantung dari kelasnya dalam dokumen. Menurut (Manning, Raghavan, & Schutze, 2008), probabilitas sebuah dokumen d berada dikelas c , kondisi berikut dapat dinyatakan dengan rumus:

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k | c)$$

$P(t_k | c)$ merupakan conditional probabilitas dari kata t_k yang terdapat dalam sebuah dokumen dari kelas c . $P(c)$ merupakan prior probabilitas dari sebuah dokumen yang terdapat dalam kelas c . $(t_1, t_2, \dots, t_{n_d})$ merupakan token dalam dokumen d yang merupakan bagian dari *vocabulary* yang digunakan sebagai klasifikasi dan merupakan jumlah *token* dalam dokumen d .

Untuk memperkirakan prior probabilitas $P(c)$ dinyatakan dengan rumus:

$$P(c) = \frac{N_c}{N}$$

N_c = Jumlah dokumen *training* dalam kelas c

N = Jumlah keseluruhan dokumen *training* dari seluruh kelas

Untuk memperkirakan *conditional probability* $P(t|c)$ dinyatakan dengan rumus :

$$P(t|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}}$$

T_{ct} = Jumlah kemunculan kata t dalam sebuah dokumen *training* pada kelas c

$\frac{T_{ct}}{\sum_{t' \in V} T_{ct'}}$ = Jumlah total keseluruhan kata dalam dokumen *training* pada kelas c

t' = Jumlah total kata dalam dokumen *training*

Untuk menghilangkan nilai nol pada sebuah dokumen, digunakan *laplace smoothing* sebagai proses penambahan nilai 1 pada setiap nilai T_{ct} pada perhitungan *conditional* probabilitas dan dinyatakan dengan rumus :

$$P(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} T_{ct'} + B'}$$

B' = total kata unik pada keseluruhan kelas dalam dokumen *training*

Untuk memperoleh nilai probabilitas yang tinggi setiap kata digunakan *laplace smoothing* atau *add-one*, *laplace smoothing* digunakan agar nilai dari probabilitas masing-masing kata dapat memenuhi syarat yaitu tidak sama dengan 0. Jika nilai dari probabilitas kata adalah 0 maka data tidak baik *training* maupun *testing* tidak akan pernah cukup untuk mewakili *frekuensi* saat terdapat kejadian langkah.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Gambaran Penelitian

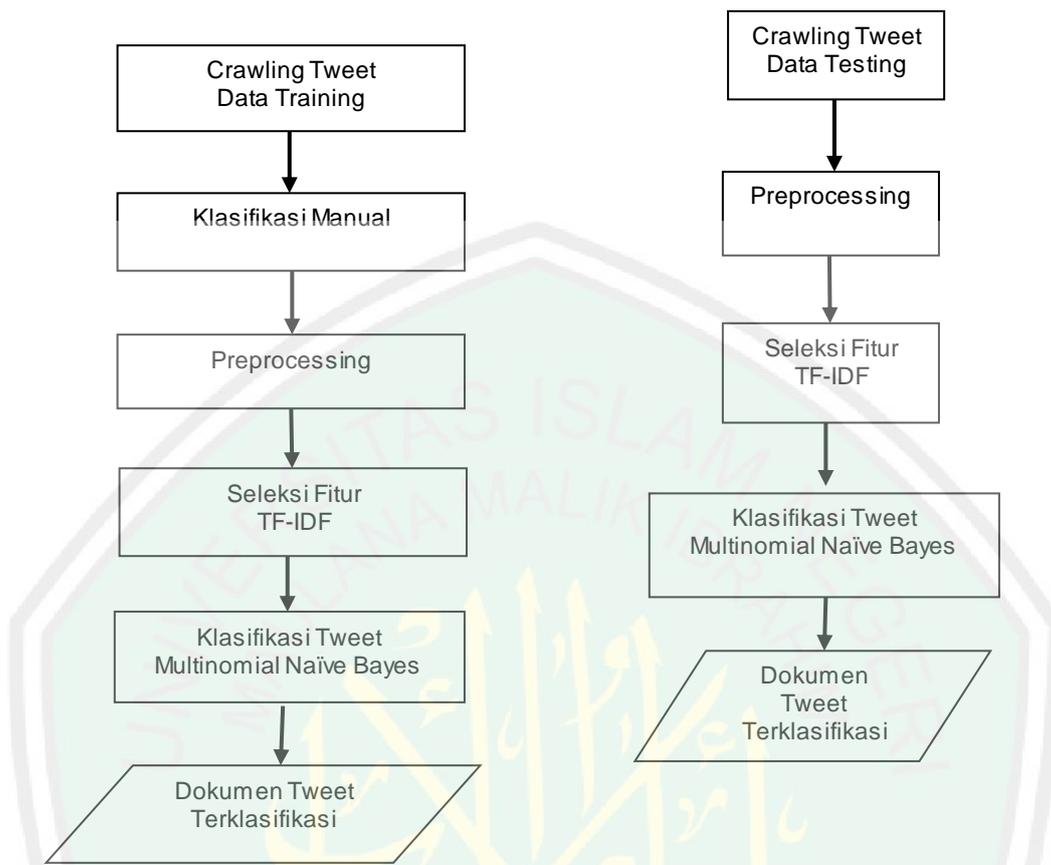
Sistem yang dibangun dalam penelitian ini adalah sistem pengujian akurasi dari penggunaan metode *Multinomial Naïve Bayes* pada klasifikasi data teks *tweet*. *Tweet* yang digunakan adalah yang berasal dari *Twitter* yang di unggah oleh masyarakat pada umumnya untuk menilai sentimen pada pelayanan publik. Proses klasifikasi yang dilakukan pada sistem ini menggunakan metode *Multinomial Naïve Bayes* sebagai klasifikasi.

Sebagai pelengkap juga dijelaskan proses perolehan data *Twitter* secara singkat karena hal tersebut tidak termasuk dalam penelitian ini. Teknik untuk mendapatkan data *Twitter* terdapat berbagai macam salah satunya yang disebutkan dalam penelitian ini yaitu melalui API yang telah disediakan oleh *Twitter*.

Penelitian ini menggunakan melakukan pengujian data dengan presisi dan *recall*. Pengujian tersebut digunakan untuk mengukur data yang terhimpun dari data yang dimasukkan kedalam sistem.

3.2 Gambaran Umum

Gambaran umum penelitian ini mengikuti diagram dijelaskan pada gambar 3.1. Pada gambar 3.1 dijelaskan alur dari *testing* dan *training*. Alur tersebut juga menggambarkan proses yang dilalui dari kedua sisi. Setiap proses akan dijelaskan lebih rinci pada halaman selanjutnya.



Gambar 3.1 Alir Sistem

Sistem ini dimulai dengan proses menginput data *tweet* dengan cara *crawling*. Proses *crawling* dibedakan menjadi dua, yaitu *crawling* data *training* dengan cara manual menggunakan *web browser* dan *fitur inspect element*, dan *crawling* data menggunakan *API Twitter*. *Output* dari sistem ini merupakan data *testing* dengan nilai *output* berupa opini positif atau negatif yang diklasifikasikan oleh sistem berdasarkan pembelajaran data *training*.

Jika pengguna ingin mendapatkan data *Twitter* yang sangat lampau akan digunakan teknik *crawling* data dari *frontend Twitter*. Dari data yang didapatkan tersebut akan disimpan dalam excel. Data tersebut terlebih dahulu diberi tanda positif, negatif atau netral pada setiap *tweet*.

Untuk melakukan proses pencarian data dibutuhkan *query*. *Query* ini dapat dicari melalui *frontend twitter* agar pencarian memperoleh lebih banyak data. Selain itu juga untuk mempersempit pengklasifikasian topik opini yang akan dilakukan.

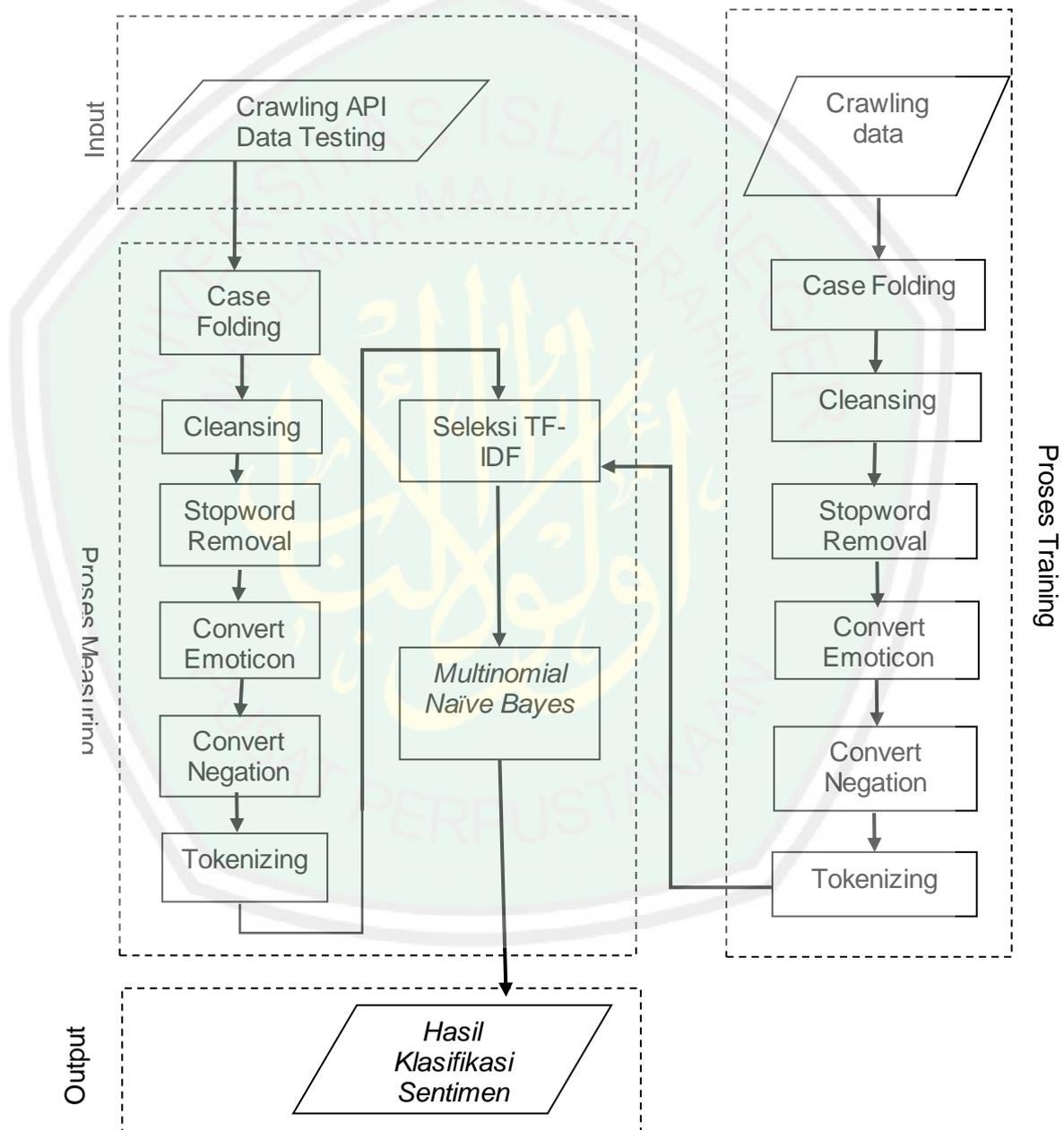
Selanjutnya adalah proses input data ke dalam sistem. Input ini dapat berupa file *excel* atau hasil *crawling* dari API *Twitter*. *Input* yang dimasukkan ke dalam sistem adalah dokumen yang berupa *tweet* dari akun *twitter* yang berupa opini ataupun keluhan. Data *tweet* tersebut didapat dengan memanfaatkan API (*Application Programming Interface*) yang disediakan *Twitter*. Dokumen yang dimaksud merupakan dokumen berbahasa Indonesia.

Dataset berupa teks berbahasa Indonesia yang diambil dari *website* <https://twitter.com>. Ada beberapa data yang diambil dari *website* tersebut, dalam penelitian yang dilakukan data diambil menggunakan *query* ‘telkomsel’, ‘myTelkomsel’, ‘simpati’, ‘simpatiLoop’. *Query* merupakan akun resmi dari *Provider Telekomunikasi Telkomsel*.

Dataset dari hasil *crawling* dengan API *Twitter* ini akan dibagi menjadi dua yaitu data *training* dan data *testing*, hasil dari *crawling* ini berupa dokumen *tweet* yang tidak menyertakan atribut lainnya. Data *training* ini akan dimasukkan ke dalam *dataset* MySQL, dan diklasifikasikan secara manual dengan label sentimen positif dan negatif. Data *testing* yang diperoleh dari proses *crawling* ini akan disimpan di dalam *dataset* MySQL, yang nantinya akan diolah ke dalam sistem untuk menghasilkan *output* otomatis berupa sentimen positif atau negatif.

Keluaran yang dihasilkan berupa sentimen positif dan negatif serta netral dari *dataset testing*. Keluaran berupa diagram yang menggambarkan sebaran data kelas sentimen.

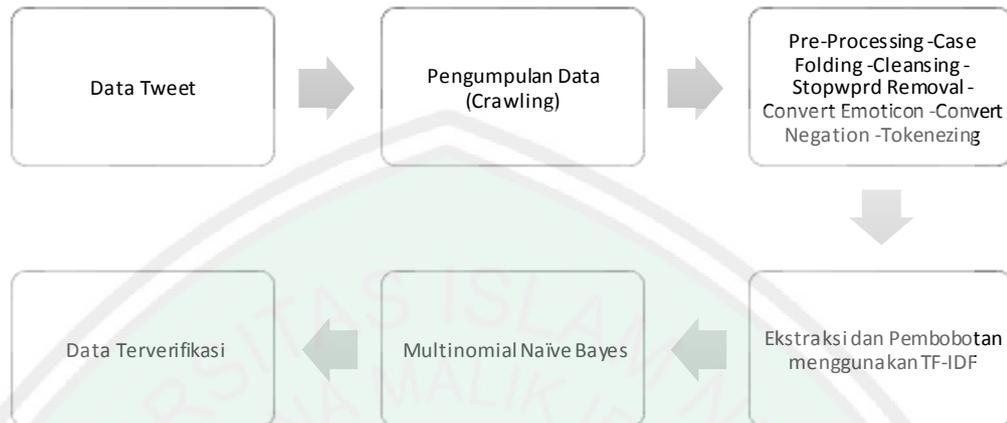
3.3 Perancangan Desain Sistem



Gambar 3.2 Alur proses besar aplikasi

3.3.1 Perancangan Alur Sistem

Berikut adalah rancangan sistem yang dibuat pada penelitian ini



Gambar 3.3 Perancangan Alur proses besar aplikasi

3.3.1.1 Crawling Data

Crawling data yang dimaksud adalah proses pencarian *tweet* pada *twitter* yang ditujukan pada satu akun *provider* seluler. *Crawling* data ini dilakukan dengan mengakses API dari *Twitter*. Teknik *crawling* yang digunakan sesuai dengan dokumentasi yang sudah disediakan oleh *Twitter*.

Ketentuan *crawling* data *Twitter* terdapat pada website resmi dari *Twitter*. Teknik *crawling* yang akan digunakan adalah melalui *Stream* data *Twitter* yang berformat JSON. Selain lebih mudah dalam mengelola data, format JSON juga ringan dalam ukuran.

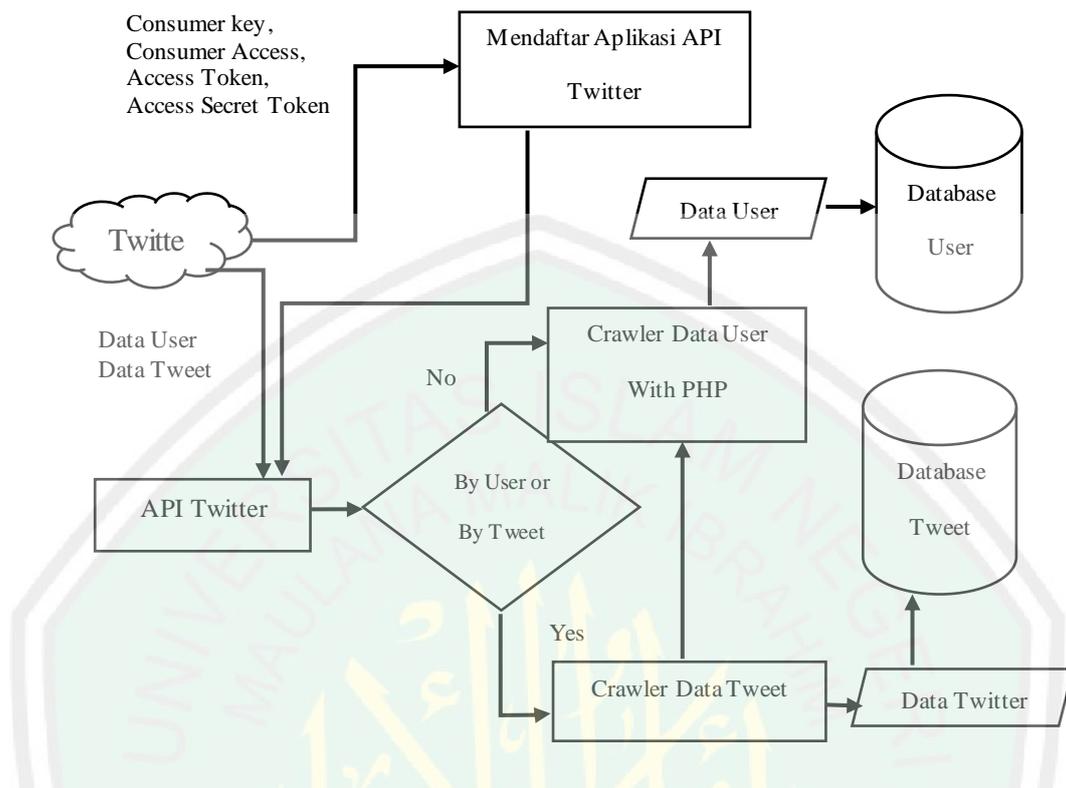
Crawling akan dilakukan selama 1 bulan untuk mendapatkan data yang sesuai dengan yang diinginkan. Selain itu juga untuk memperoleh data yang cukup banyak.

Arsitektur *crawling* tidak akan dibahas ini karena membutuhkan bahasan tersendiri.

3.3.1.2 Pre-processing

- *Case folding*, adalah penyeragaman bentuk huruf serta penghapusan angka dan tanda baca. Dalam hal ini yang digunakan hanya huruf latin antara a sampai dengan z
- *Cleansing*, adalah proses membersihkan dokumen dari kata yang tidak diperlukan untuk mengurangi *noise*. Kata yang dihilangkan adalah karakter HTML, kata kunci, ikon emosi, *hashtag* (#), *username* (@username), *url*
- *Stopword removal*, adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen (Dragut et al. 2009). Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan sebagainya. Sebelum proses *stopword removal* dilakukan, harus dibuat daftar *stopword* (*stoplist*). Jika termasuk didalam *stoplist* maka kata-kata tersebut n dihapus dari deskripsi dianggap sebagai kata-kata yang mencirikan isi dari suatu dokumen atau *keywords*. Daftar kata *stopword* dipenelitian ini bersumber dari Tala (2013)
- *Convert emoticon*, adalah proses seleksi bentuk *emoticon* senyum, sedih dan sebagainya
- *Convert negation*, adalah proses seleksi bentuk kalimat tidak baku ke dalam kalimat baku
- *Tokenizing*, adalah pemotongan kata dalam tiap kalimat.

Alur dari sistem yang dibangun seperti yang disajikan pada gambar Gambar 3.4 sebagai berikut:



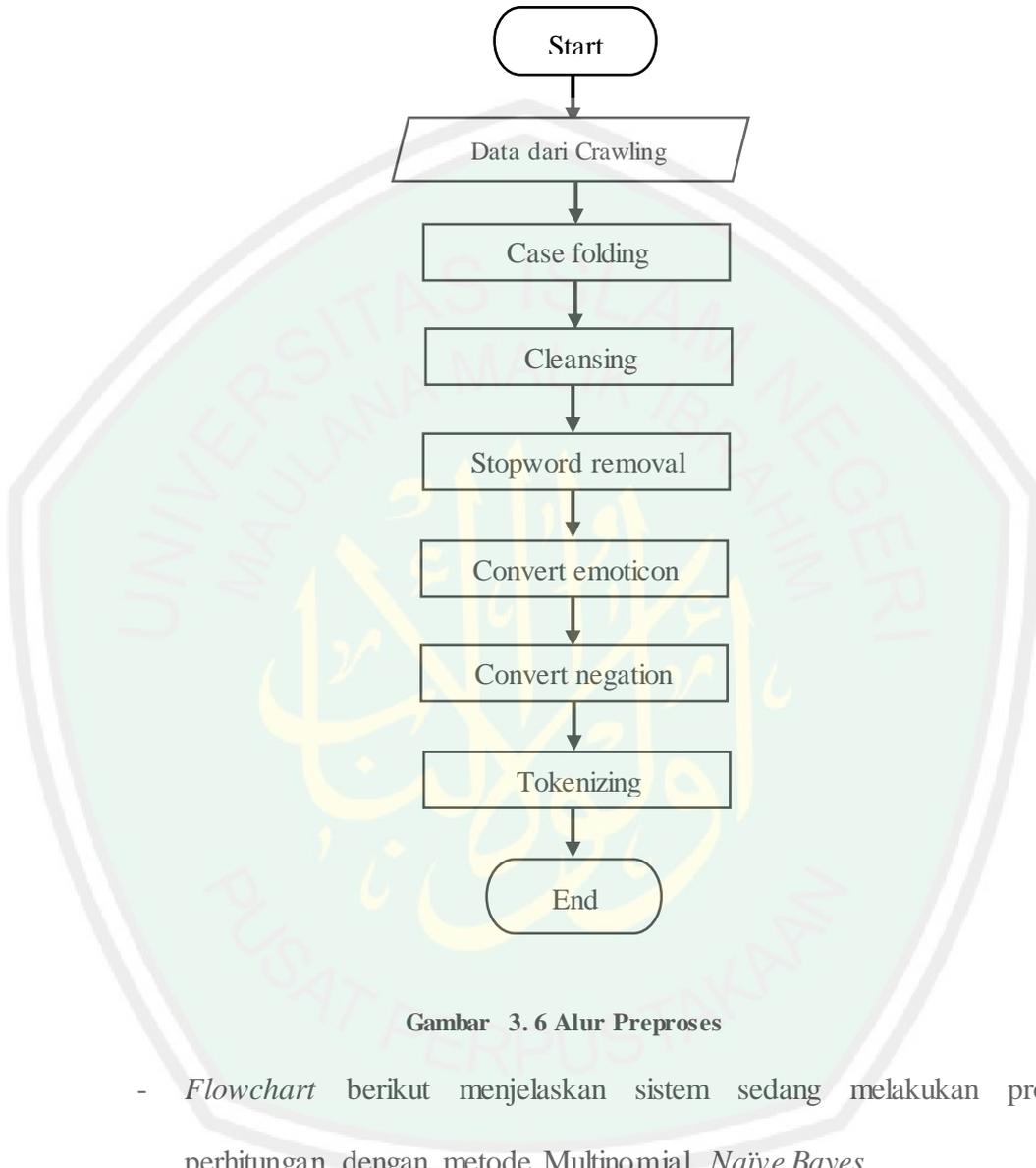
Gambar 3.4 Desain entity

API Twitter berfungsi sebagai penghubung antara sistem yang dibangun dengan twitter. *API Twitter* membutuhkan *consumer key*, *consumer access*, *access token* dan *access secret token* yang didapatkan dengan cara mendaftarkan aplikasi *API Twitter* di <http://dev.twitter.com>. Berikutnya pencarian data dari *twitter* dilakukan menggunakan bahasa pemrograman PHP dengan pencarian data berdasarkan *user* dan *keyword*. Terakhir data diunduh dan disimpan ke dalam database tweet dan user.

3.3.2 Flowchart

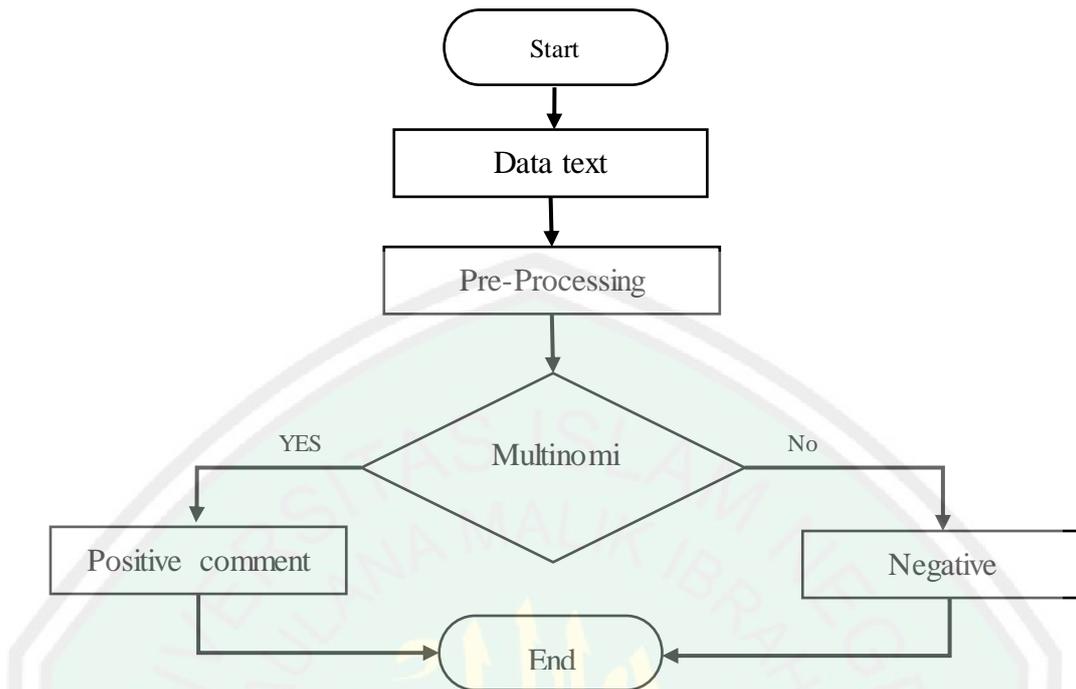
Flowchart berikut menjelaskan sistem sedang melakukan proses *Preprocessing*. *Flowchart* berikut adalah desain yang akan digunakan untuk penelitian ini. *Flowchart* yang digunakan peneliti telah banyak digunakan. Namun

peneliti telah memilih dan memodifikasi agar sesuai dengan penelitian yang dilakukan.



Gambar 3.6 Alur Preproses

- *Flowchart* berikut menjelaskan sistem sedang melakukan proses perhitungan dengan metode Multinomial *Naïve Bayes*



Gambar 3.7 Alur Proses Klasifikasi

3.4 Alur Manual Analisis Sentimen

Crawle dari Twitter

Dari hasil crawl twitter akun Telkomsel

RT @sahaL_AS: Ketika @telkomsel gak sedang modus, tp lemot :)

Hasil Case Folding

rt @sahaL_as: ketika @telkomsel gak sedang modus, tp lemot :)

Cleansing

ketika gak sedang modus, tp lemot :)

Stopword Removal

gak modus, tp lemot :)

Convert Emoticon

: gak modus, tp lemot emotsenang

Convert Negation

: gakmodus, tp lemot emotsenang

Tokenisasi

gak, modus, tp, lemot

3.5 Implementasi Algoritma

Implentasi Algoritma

3.5.1 Data

Pada penelitian ini digunakan data primer. Data primer yang digunakan dalam penelitian ini sebagai berikut:

a. Data training POS Tagging

Data training yang digunakan pada penelitian ini merupakan sekumpulan kamus data opini yang dibuat oleh peneliti. Pembuatan kamus data opini tersebut dengan mengumpulkan data opini melalui media data online. Kamus data opini yang dikumpulkan diklasifikasi menjadi kamus opini negatif dan positif. Untuk melakukan klarifikasi peneliti menggunakan bantuan daring online maupun offline untuk melakukan klarifikasi agar data kamus sesuai dan meminimalkan subjektifitas.

3.5.2 Metode Pengolahan Data

Sistem yang dikembangkan dengan menggunakan proses *opinion orientation*. *Opinion Orientation* merupakan proses penentuan kalimat termasuk

opini positif atau negatif. Penentuan tersebut digolongkan sebagai proses pengklasifikasian. Pengklasifikasian pada suatu dokumen termasuk kelas positif atau kelas negatif. Metode ini digunakan karena *Multinomial Naïve Bayes* merupakan metode klasifikasi.

3.5.3 *Multinomial Naïve Bayes*

Multinomial Naïve Bayes merupakan proses pengambilan jumlah kata yang muncul dalam setiap dokumen, dengan mengasumsikan dokumen memiliki beberapa kejadian dalam kata dengan panjang yang tidak tergantung dari kelasnya dalam dokumen.

4.1.1.1 Proses *Training*

Sebuah dokumen *training* sudah diklasifikasikan secara *manual* dan sudah dilakukan proses *pre-processing* seperti pada tabel 3.1 berikut:

Tabel 3.1 Contoh Kasus Data Training

Tweet	Fitur	Kelas
Tweet 1	kalau buka browser suka direct kesini terus ya terimakasih banyak	N
Tweet 2	pagi curhat dulu tentang costumer care telkomsel tugas akun twitter resmi mereka ah	P
Tweet 3	saya sudah bayar tagihan pasca bayar panggil keluar tidak bisa	N
Tweet 4	tumben nih telkomsel belum 4g di kampung myxl lte lagi	N
Tweet 5	masih ada saldo tcash gratis untuk telkomsel	P
Tweet 6	halo admin koneksi macam motor tahun 70an	N

	lambat kek kura2	
--	------------------	--

Dari data Tabel 3.1 dibuat sebuah model *prior probabilities* dengan mengacu pada persamaan

$$P(c) = \frac{Nc}{N}$$

Kelas positif muncul 2 kali dalam dokumen *training*. Total dokumen 6 adalah dokumen. Untuk itu jika dimasukkan dalam persamaan:

$$P(c) = \frac{2}{6}$$

Begitu pula untuk kelas negatif. Setiap dokumen yang muncul dibagi dengan total seluruh dokumen.

Tabel 3.2 Hasil Perhitungan Prior Probabilities Untuk Setiap Kelas

Atribut kelas	P(class)
Positif (P)	2/6
Negatif (N)	4/6

Model *confitional probabilities* berikut contoh perhitungan *conditional probabilities* untuk term “telkomsel” pada kelas positif:

$$P(\text{telkomsel} | \text{positif}) = \frac{2}{37}$$

Setiap term yang ditemukan akan dihitung probabilitas kemunculan untuk setiap kelas kemudian dibagi dengan total term yang ditemukan. Contoh term yang ditemukan terdapat pada tabel 3.3.

Tabel 3.3 Contoh term pada dokumen

Term	positif	Negatif
Ada	0/37	1/88
Akun	0/37	1/88

Care	1/37	0/88
costumer	1/37	0/88
Curhat	1/37	0/88
Dulu	0/37	1/88
Gratis	1/37	0/88
Masih	1/37	0/88
Mereka	0/37	1/88
Pagi	1/37	0/88
Resmi	1/37	0/88
Saldo	1/37	0/88
Tcash	1/37	0/88
telkomsel	2/37	0/88
Tentang	0/37	1/88
Tugas	1/37	0/88
Twitter	1/37	0/88
Untuk	1/37	0/88

Dalam proses perhitungan kemungkinan term tidak terdapat dalam suatu kelas sangat besar. Apabila term tersebut tidak terdapat pada kelas maka akan bernilai 0. Hal ini dapat menyebabkan seluruh perhitungan menjadi tidak sesuai. Untuk itu digunakan teknik *smoothing*. Teknik *smoothing* yang digunakan adalah *Laplace smoothing*.

Laplace smoothing adalah teknik yang menambahkan angka 1 untuk setiap term yang ditemukan. Dengan penambahan tersebut angka 0 dapat dihilangkan dan hasil perhitungan tidak rusak.

Laplace smoothing digunakan untuk menghilangkan nilai nol (*term* tanpa bobot) pada term dari tabel 3.3. *Laplace smoothing* merupakan tahap akhir dari proses *training* menggunakan algoritma *multinomial Naïve Bayes*. Hasil yang didapat dari proses ini menjadi model untuk melakukan klasifikasi.

Berikut adalah contoh perhitungan *probabilities* untuk term “telkomsel” pada kelas positif:

$$P(\text{telkomsel} | \text{positif}) = \frac{2 + 1}{37 + 125} = 0.22424$$

Pada persamaan diatas angka 2 mewakili term telkomsel yang ditemukan. Angka 37 mewakili total term pada kelas positif sedangkan 125 adalah total term pada seluruh dokumen. Total term tersebut didapat dari penjumlahan term pada tabel 3.3.

Tabel 3.4 Model Perhitungan Klasifikasi

Term	positif	negatif
Ada	0.07142857	0.4
Akun	0.07142857	0.4
Care	0.14285714	0.2
Costumer	0.14285714	0.2
Curhat	0.14285714	0.2
Dulu	0.07142857	0.4
Gratis	0.14285714	0.2
Masih	0.14285714	0.2
Mereka	0.07142857	0.4
Pagi	0.14285714	0.2
Resmi	0.14285714	0.2
Saldo	0.14285714	0.2
Tcash	0.14285714	0.2
Telkomsel	0.21428571	0.2
Tentang	0.07142857	0.4
Tugas	0.14285714	0.2
Twitter	0.14285714	0.2
Untuk	0.14285714	0.2

Pada tabel 3.4 merupakan hasil dari perhitungan setiap term yang sudah ditambahkan 1. Probabilitas tersebut nantinya dicari nilai yang paling besar dan dijadikan sebagai kelas untuk term tersebut. Misalnya saja telkomsel pada tabel 3.4 kelas positif menunjukkan 0.224 sedangkan kelas negatif 0.2. Dengan hasil tersebut didapatkan term telkomsel pada kelas positif karena nilai positif lebih besar.

3.5.1.1 Testing

Alur proses *testing* kurang lebih sama dengan alur *training*, Hanya saja setelah proses selesai data *testing* akan dihitung nilai probabilitas akhir. Data *testing* terdapat pada tabel 3.5. Data ini diambil dari tabel 3.4.

Tabel 3.5 Data Testing

Tweet	Fitur	Kelas
Tweet 7	kalah main game pake tidak guna koneksi makin lama makin butut sial	?
Tweet 8	maaf lha ini jaringan suka ilang timbul 4g nya padahal dulu lancar jaya	?
Tweet 9	minta rincih tagihan telkomsel tanpa ribet rasa kecewa	?

Perhitungan probabilitas pada bagian $\prod_{1 \leq k \leq n} P(t_k | c)$, persamaan tersebut dihitung terlebih dahulu untuk memudahkan kalkulasi selanjutnya. Untuk hasilnya terdapat tabel 3.6.

Untuk sebuah *term* yang kemunculannya lebih dari satu kali, nilai kemunculan tersebut menjadi pangkat dari nilai *conditional probabilities*-nya pada tabel 3.5 dengan *term frequency* masing-masing. Kemudian jumlahkan nilainya untuk masing-masing kelas.

Tabel 3.6 Hasil Nilai Conditional Probabilities

Term	tf	Positif	negatif
Kecewa	1	0.08	0.014388489
Minta	1	0.08	0.014388489
rasa	1	0.08	0.014388489
Ribet	1	0.08	0.014388489
Rincih	1	0.08	0.014388489
Tagihan	2	0.08	0.021582734
Tanpa	1	0.08	0.014388489

Telkomsel	4	0.04	0.007194245
-----------	---	------	-------------

Misalnya *term* 'telkomsel' memiliki kemunculan sebanyak 4 kali. Kemudian setiap probabilitas dari setiap kelas akan dipangkatkan dengan 4. Untuk nilai kemunculan yang sama agar proses kalkulasi tidak membebani maka proses tersebut dilakukan satu kali. Hal ini mempercepat proses yang pencarian probabilitas. Pada tabel 3.7 disebutkan hasil dari pemangkatan seluruh term yang ditemukan.

Tabel 3.7 Hasil Perkalian Nilai Conditional Probabilities dengan Term Frequency-nya

Term	Positif	Negatif
Kecewa	0.08	0.014388489
Minta	0.08	0.014388489
rasa	0.08	0.014388489
Ribet	0.08	0.014388489
Rincih	0.08	0.014388489
Tagihan	0.0064	0.000465814
Tanpa	0.08	0.014388489
Telkomsel	2.56E-06	2.68E-09
Total	0.48640256	0.086796751

Dari tabel 3.7 didapatkan nilai probabilitas setiap term. Untuk mendapatkan kelas yang diinginkan dari dokumen adalah dengan mengalikan probabilitas dokumen dengan probabilitas kelas yang didapat dari hasil perhitungan. Kemudian diambil nilai paling besar diantara kedua kelas tersebut. Sebagai contoh tweet 9. Didapatkan nilai probabilitas dari tweet 9 terhadap seluruh kelas dengan cara mengkalikan nilai *prior probabilities* dengan total nilai *conditional probabilities* (tabel 3.5) untuk masing-masing kelas

Probabilities dari tweet 9 terhadap kelas positif:

$$P(\text{positif} \mid \text{tweet 9}) = 2/6 * 0.48640256 = 0.0162134187$$

Probabilities dari tweet 9 terhadap kelas negatif:

$$P(\text{negatif} \mid \text{tweet 9}) = 2/6 * 0.086796751 = 0.0289322503$$

Dari hasil perhitungan probabilitas diatas diketahui bahwa *probabilities* tweet 9 terdapat kelas negatif memiliki nilai yang paling besar, sehingga masuk ke bisa diklasifikasikan kedalam kelas negatif.



BAB IV

HASIL DAN PEMBAHASAN

4.1 Deskripsi Aplikasi

Aplikasi yang dibuat ini termasuk dalam *opinion mining*. *Opinion mining* digunakan untuk mengetahui suatu teks termasuk positif, negatif atau netral. Proses yang digunakan untuk mencarinya menggunakan salah satu metode dari *text mining*. *Feature Extraction* digunakan untuk mencari hal yang penting dalam teks tersebut.

Setelah fitur tersebut pengelompokan dipilih menggunakan *Multinomial Naïve Bayes*. Pada model *multinomial* fitur vektor dokumen mengambil frekuensi dari kata-kata tidak hanya ada atau tidaknya kata tersebut (Shimodaira, 2015). Sehingga setiap kata merupakan fitur penting. *Naïve Bayes* merupakan *supervised learning* sehingga membutuhkan data training. Pada proses training pelabelan dilakukan manual oleh peneliti.

4.2 Implementasi

Implementasi pada penelitian ini menggunakan Bahasa *python* versi 2.7. Implementasi dikembangkan sendiri oleh peneliti. Untuk keperluan klasifikasi data ujilatih dibantu rekan dari disiplin ilmu yang terkait.

4.2.1 Pengambilan data Twitter

Twitter menyediakan API untuk mengambil datanya. Untuk pengambilan data melalui API harus melakukan registrasi aplikasi pada halaman yang telah disediakan twitter. Twitter menerapkan beberapa batasan dalam melakukan konsumsi pada API-nya. Salah satunya batasannya adalah hanya diberikan tweet

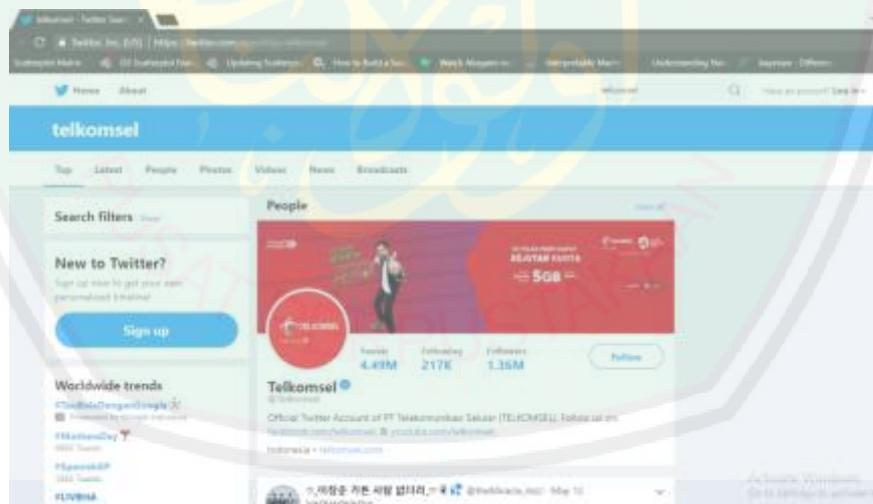
sampai batas 7 hari sebelum pengambilan data. Sehingga untuk tweet yang telah berumur 7 hari dari pengambilan data tidak bisa dilakukan. Batasan tersebut berlaku untuk melakukan pencarian dengan kata kunci tertentu. Untuk mengatasi hal itu teknik pengambilan data menggunakan sisi *front-end* dari Twitter. Sisi *front-end* ini masih terdapat tweet lama dan bisa dilakukan pencarian berdasarkan rentang waktu tertentu.

Berikut *pseudocode* untuk pengambilan data.

```

GetData (Term)
  1. T <- Input Term(Term)
  2. Q <- Search Query(T)
  3. For each result
  4. do tweet <- data
  5.   tweet <- data
End

```



Gambar 4.1 Fitur pencarian Twitter

4.2.2 Ekstraksi Fitur

Ekstraksi fitur ini termasuk diantaranya adalah melakukan proses pembersihan tweet, pembersihan dari *stopword*, proses pencarian kata dasar, dan lain-lain. Hal ini dilakukan untuk mengurangi gangguan (*noise*) yang bisa mengurangi keakuratan. Salah satu yang dilakukan dalam penelitian ini adalah

permembersihan tweet dari hastag, puntuasi, *url*, dan lain-lain. Proses pencarian kata dasar atau *stemming* juga bisa digunakan. Proses ekstraksi fitur ini menggunakan *Regex* untuk memilah data yang dibutuhkan atau tidak. Proses ekstraksi fitur tersebut sebagai berikut:

```
URLs <- re.sub(r'http[s]?://(?:[a-z]|[0-9]|[$-
_.&+]|[*\(\),]|(?:%[0-9a-f][0-9a-f]))+', '',) # URLs
misc <- re.sub(r'(?:\xe2\x80\x9c)', '', URLs)
misc <- re.sub(r'(?:\xe2\x80\xa6)', '', misc)
misc <- re.sub(r'(?:\xe2\x80\x99)', '', misc)
misc <- re.sub(r'(?:\xe2\x80\x9c)', '', misc)
misc <- re.sub(r'(?:\xe2\x80\x98)', '', misc)
misc <- re.sub(r'(?:\xe2\x80\x93)', '', misc)
misc <- re.sub(r'(?:\xe2\x80\x9d)', '', misc)
misc <- re.sub(r'(?:\xf0\x9f\x91\x87)', '', misc)
misc <- re.sub(r'(?:\xc2|\xa0|\xf0)', '', misc)
misc <- re.sub(r'^\x00-\x7F+', ' ', misc)
misc <- re.sub(r'(?<=\\w)\\. (?=\\w)', ' ', misc)
punct<-
re.sub(r'(?:!|\\?|\\. {3}|\\.\\.\\.D|\\.\\.s|\\.|[\\n]|\\(|\\)|\\:)', ' ',
misc)
retweet <- re.sub(r'(?:[A-
Z]+[\\s]+@[\\w_]+[:]|\\s[:])', '', punctuation) # Retweet
mention <- re.sub(r'(?:@[\\w_]+)', '', list_words[i]) # @-
mention
hashtags <- re.sub(r'(?:\\#[\\w_]+[\\w\\'_\\-
]*[\\w_]+[\\w:]|\\s[:])', '', list_words[i]) # hash-tags
html <- re.sub(r'<[^>]+>', '', data) #html tags
data <- re.sub(r'(?<=\\w) (?=\\w)', ' ', data)
data <- re.sub(r'(?<=\\w)\\.\\. (?=\\w)', ' ', data)
data <- re.sub(r'(?<=\\w)\\,\\. (?=\\w)', ' ', data)
casefold <- data.lower()
```

4.2.3 Pembobotan kata

Pembobotan kata yang digunakan menggunakan *TFIDF* (*term frequency Inverse Document Frequency*). Pembobotan kata untuk mencari penting atau tidaknya kata tersebut dalam dokumen tersebut. Pembobotan dilakukan dengan menghitung banyaknya kata yang muncul dalam dokumen dan dibandingkan dengan total kata dari seluruh dokumen. Untuk menghindari bobot kata yang berlebih digunakan sublinear. Skala sublinear dan mengecilkan ratio bias antara bobot kata mencegah terjadinya bobot kata yang terlalu berlebih (Haibing Wu, 2014). Bobot kata berlebih ini diakibatkan kata-kata yang tidak penting namun

muncul terlalu banyak. Sehingga menyebabkan bias. Untuk itu digunakan skala linear.

```

Similarity(documents)
1. tokenized_documents = [tokenize(d) for d in documents]
2. if vocabulary is None:
3. idf, vocabularies =
   inverse_document_frequencies(tokenized_documents)
4. else:
5. idf, vocabularies =
   inverse_document_frequencies(tokenized_documents,
   vocabulary)
6. tfidf_documents = []
7. for document in tokenized_documents:
8. doc_tfidf = []
9. for each term in idf.keys():
10.     tf = sublinear_term_frequency(term, document)
11.     doc_tfidf.append(tf * idf[term])
12.     tfidf_documents.append(doc_tfidf)
13.     return tfidf_documents, vocabularies
END

```

4.2.4 Multinomial Naïve Bayes

Multinomial Naïve Bayes merupakan metode pembelajaran probabilistik (Christopher D. Manning, 2008). Dalam hal ini yang dihitung adalah kemungkinan sebuah dokumen pada kelas tertentu. Untuk model dokumen untuk kelas akan membuat asumsi yang kuat. Misalnya saja $p(x|y)$ diasumsikan kondisi x independen terhadap y (Ng, 2017). Sehingga disebut *naive bayes assumption*. Hasilnya dalam berupa kelas yang bisa disebut sebagai klasifikasi *Naïve Bayes*.

Multinomial sendiri merupakan pengembangan dari *Naïve Bayes* agar bisa melakukan proses perhitungan data selain data biner. Data tersebut dapat berupa bilangan desimal maupun bilangan real. Salah satu contoh kasusnya adalah sentiment analisis. Sentimen analisis yang merupakan sekumpulan kata yang

menghitung setiap kata dan merepresentasikan kata tersebut berdasarkan total munculnya. Hal ini dipadukan dengan metode probabilitas dari *Naïve Bayes*.

Untuk menghitung probabilitas lebih dulu dilakukan pelabelan setiap data untuk pelatihan dilakukan oleh peneliti. Pelabelan yang dilakukan pada data latih berfungsi sebagai pengelompokan kelas. Proses pelabelan dilakukan pada proses pelatihan karena metode ini merupakan *supervised learning* yang membutuhkan data awal sebagai data latih. Proses perhitungan probabilitas menggunakan bobot kata pada dokumen. Perhitungan bobot kata dan probabilitas kata tersebut secara otomatis pada coding.

```

TRAINMULTINOMIALNB(C, D)
1  V ← EXTRACTVOCABULARY(D)
2  N ← COUNTDOCS(D)
3  for each c ∈ C
4  do Nc ← COUNTDOCSINCLASS(D, c)
5     prior[c] ← Nc/N
6     textc ← CONCATENATETEXTOFALLDOCSINCLASS(D, c)
7     for each t ∈ V
8     do Tct ← COUNTTOKENSOFTERM(textc, t)
9     for each t ∈ V
10    do condprob[t][c] ←  $\frac{T_{ct}+1}{\sum_{t'}(T_{ct'}+1)}$ 
11  return V, prior, condprob

APPLYMULTINOMIALNB(C, V, prior, condprob, d)
1  W ← EXTRACTTOKENSFROMDOC(V, d)
2  for each c ∈ C
3  do score[c] ← log prior[c]
4  for each t ∈ W
5  do score[c] += log condprob[t][c]
6  return arg maxc ∈ C score[c]

```

Gambar 4 1 Process Multinomial Naive Bayes

Proses selanjutnya adalah proses uji coba data. Setelah data latih dimasukkan maka data uji coba dimasukkan untuk dicari probabilitasnya. Dari proses didapatkan dua probabilitas. Probabilitas paling besar pada kelas tersebut dipilih sebagai kelas dokumen.

Data dibentuk menjadi matrik kata. Matrik kata terdiri dari seluruh kata yang ditemukan pada saat proses latih. Nilai dari matrik merupakan bobot kata. Hal ini menyebabkan terdapat data yang bernilai 0. Karena terdapat kata yang tidak terdapat dalam dokumen tersebut. Tentunya data tersebut bisa menyebabkan proses menjadi kacau. Proses tersebut sering terjadi dalam penelitian. Banyak metode yang digunakan untuk menghindari hal tersebut. Salah satunya adalah *Laplace smoothing*.

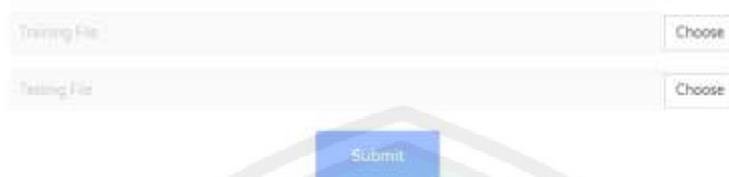
Laplace smoothing merupakan proses menambahkan seluruh data dengan angka 1. Sehingga tidak terdapat angka 0 yang dapat menjadikan proses penghitungan menjadi kacau. Metode ini digunakan pada proses perhitungan *Naïve Bayes* dan *tfidf*. Pada proses *Naïve Bayes* hal tersebut sangat penting dan digunakan pada proses latih maupun ujicoba.

4.2.5 Tampilan antarmuka

Tampilan antarmuka pada aplikasi ini menggunakan web. Selain lebih dinamis juga bisa dibuka dimana saja. Tampilan tersebut menampilkan input data latih dan data uji. Hasilnya yang ditampilkan dalam bentuk grafik.

Terdapat dua input yakni untuk data latih dan data uji. Data latih dalam file excel yang formatnya sudah ditentukan. Contoh format juga telahh sediakan pada halaman website.

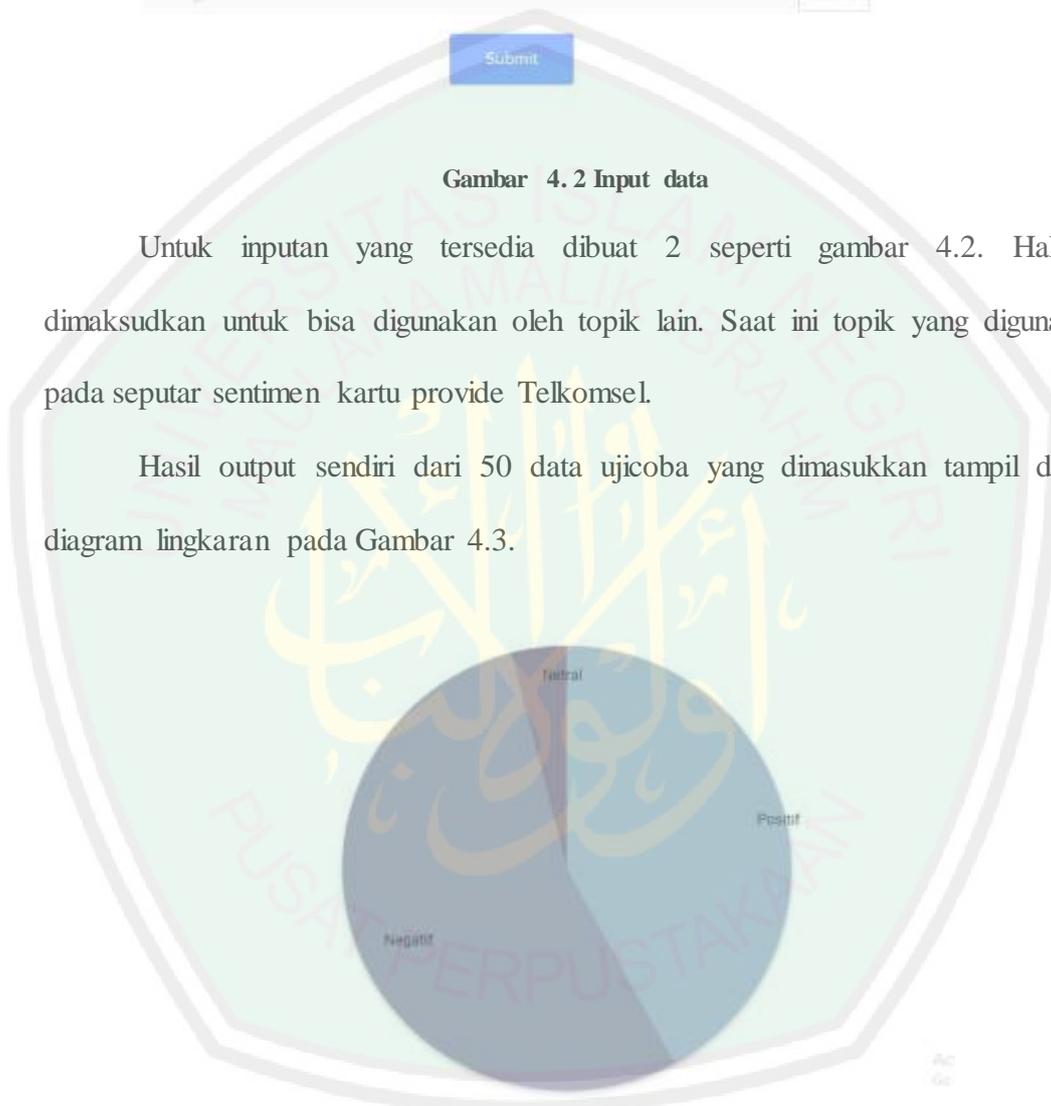
Upload your file here



Gambar 4.2 Input data

Untuk inputan yang tersedia dibuat 2 seperti gambar 4.2. Hal ini dimaksudkan untuk bisa digunakan oleh topik lain. Saat ini topik yang digunakan pada seputar sentimen kartu provide Telkomsel.

Hasil output sendiri dari 50 data ujicoba yang dimasukkan tampil dalam diagram lingkaran pada Gambar 4.3.



Gambar 4.3 Hasil Pie Chart

4.3 Pengujian

Pada penelitian kali ini jumlah data yang didapatkan dari hasil pencarian data yaitu 127 data tweet. Data tersebut mempunyai rentang waktu dari 01 Oktober-30 Oktober 2017.

4.3.1 Multinomial Naïve Bayes

Multinomial *Naïve Bayes* termasuk metode *supervised learning* yang digunakan untuk permodelan data dan prediksi. Hal ini karena selain mempunyai hasil yang akurat juga lebih mudah untuk implementasinya. Berikut adalah contoh dokumen *tweet* yang menjadi ujicoba.

Tabel 4.1 Kelas Positif

Tweet	Class	Probabilities Class	Positif	Negatif	Netral
njogeti admin telkom	Positif	4.31e-29	4.31e-29	5.37e-34	6.89e-34

Hasil diatas salah satu data *tweet* yang didapatkan mendapatkan hasil klasifikasi positif. Namun jika dilihat dari sisi manusia maka *tweet* tersebut mengandung sentimen negatif.

Tabel 4.2 Kelas Positif Relevan

Tweet	Class	Probabilities Class	Positif	Negatif	Netral
@ telkomsel mohon info saya masih ada kuota gamemax 25gb misalnya kuota tsb saya gunakan utk flash reguler yg 24 jam apakah ini dimungkinkan dan bagaimana langkah2nya terima kasih	Positif	4.49e-173	4.492-173	5.26e-186	5. e-193

Pada *tweet* diatas mendapatkan hasil positif dan *tweet* tersebut juga positif. Hal ini merupakan pengaruh dari jumlah data yang mengandung kata yang mempunyai nilai probabilitas yang serupa. Seperti yang disebutkan *Naïve Bayes*

disebut demikian karena menganggap setiap nilai dari item tidak bergantung pada nilai yang lain atau independen. Sehingga menciptakan bias yang cukup tinggi.

Tabel 4.3 Kelas Negatif

Tweet	Class	Probabilities Class	Positif	Negatif	Netral
@telkomsel malam min untuk layanan data daerah jabar lagi gangguan kah ini udah 2 minggu lebih kira2 sampe kapan ya	Negatif	2.02e-175	9.43e-180	2.02e-175	6.55e-195

Pada dokumen yang negatif terdapat juga pada kelas positif. Sejauh ini tidak semua dokumen terklasifikasi sesuai dengan sentimen yang diinginkan. Namun sebagian besar sudah berhasil dengan sentimen yang sesuai dengan hasil.

4.3.2 Presisi dan Recall

Presisi adalah rasio dokumen relevan ditemukan dibagi total jumlah dokumen yang ditemukan. *Recall* adalah rasio dokumen relevan ditemukan kembali dibagi total jumlah dokumen dalam kumpulan dokumen yang dianggap relevan. Data dibagi menjadi 4: 1) Data ditemukan: relevan (tp), 2) Data ditemukan: tidak relevan (fp), 3) Data tidak ditemukan: relevan (fn), dan 4) Data tidak ditemukan: tidak relevan (mn)

$$\text{Presisi} = \frac{\text{data kelas yang sesuai}}{\text{total data kelas yang ditemukan}}$$

$$\text{Recall} = \frac{\text{data kelas yang sesuai}}{\text{total data kelas yang ditemukan}}$$

Dari total data 50 data uji coba ditemukan 27 kelas negatif, 21 kelas positif, dan 2 kelas netral. Total data tersebut dihasilkan dalam 1 kali percobaan dengan *dataset* yang sama dengan *dataset* percobaan lain.

Tabel 4.4 Presisi dan Recall Kelas

Kelas Positif	Relevan	Tidak Relevan
True	13	3
False	3	2

Pada tabel diatas pengukuran presisi dan *recall* yaitu:

$$\text{Presisi} = 13 / (13+3) = 0.8125$$

$$\text{Recall} = 13 / (13+5) = 0.8125$$

Tabel 4.5 Presisi dan Recall Kelas Negatif

Kelas Negatif	Relevan	Tidak Relevan
True	12	5
False	3	7

Pada tabel diatas pengukuran presisi dan *recall* yaitu:

$$\text{Presisi} = 12 / (12+5) = 0.8$$

$$\text{Recall} = 12 / (12+3) = 0.92$$

Tabel 4.6 Presisi dan Recall Kelas Netral

Kelas Positif	Relevan	Tidak Relevan
True	1	0
False	1	0

Pada tabel diatas pengukuran presisi dan *recall* yaitu:

$$\text{Presisi} = 1 / (1+0) = 1$$

$$\text{Recall} = 1 / (1+1) = 0.5$$

Akurasi yang didapatkan adalah 81.2 % untuk kelas positif. Nilai tersebut didapatkan dari preisis yang membagi antara jumlah yang ditemukan dengan hasil yang tweet yang relevan tapi salah. Untuk kelas negatif sebesar 80 % akurasi yang didapatkan untuk kelas netral adalah 50 %.

Untuk rata-rata akurasi yang dihasilkan oleh sistem adalah

$$Rata - rata = \frac{81,2 + 80 + 50}{3}$$

Maka didapatkan akurasinya adalah 70.4%.

4.3.3 Kecepatan

Pada aplikasi yang dibuat contoh pengukurannya. Pada penelitian ini pengukuran waktu dilakukan dengan menggunakan *library* dari python yakni *time*.

. Pada penelitian ini tersebut total waktu yang dibutuhkan untuk kedua data sejumlah 1.3904 detik. Jumlah data yang digunakan adalah 127 data latih dan 50 data ujicoba.

Proses paling banyak membutuhkan waktu tentu saja proses *Naïve Bayes*. Selain proses yang banyak juga komputasi yang lebih rumit. Selain itu penggunaan *resource* yang belum maksimal. Pada aplikasi yang dilakukan masih menggunakan satu *core cpu* saja sehingga proses dilakukan secara bertahap dan melakukan antrian pada 1 core cpu saja.

Proses pembobotan kata termasuk proses yang ringan. Selain karena algoritma yang digunakan tidak terlalu besar dalam hal komputasi. Skala sublinear juga masih wajar dalam proses yang dilakukan.

Pengukuran waktu pada proses pengambilan data tidak dilakukan karena proses tersebut bergantung pada banyaknya data Twitter yang dicari. Semakin

sedikit data yang didapat semakin cepat proses dilakukan. Variable data twitter tersebut tidak dapat diukur karena bergantung ketersediaan data di Twitter.

4.4 Pembahasan

Pemilihan pengambilan data pada *frontend twitter* membutuhkan teknik tersendiri daripada API yang disediakan oleh *Twitter* yang cenderung lebih mudah. Pada pengambilan data *Twitter* yang dilakukan peneliti lebih fleksibel karena *tweet* yang didapatkan bisa dalam rentang waktu yang lampau yang lama daripada API yang hanya menyediakan 7 hari sebelum pengambilan data dilakukan. Hal ini tentu mengurangi jumlah data yang didapat. Untuk mengatasi hal itu jika ingin tetap menggunakan API haruslah melakukan *Streaming* data selama rentang waktu tertentu. Namun hal ini juga membutuhkan sumberdaya yang berbeda.

Setelah data didapatkan data dibagi menjadi 2 *dataset* data ujilatih dan data ujicoba. Data ujilatih digunakan sebagai data latihan yang sebelumnya harus dilakukan penandaan pada data. Penandaan ini berfungsi untuk klasifikasi data yang selanjutnya digunakan sebagai data latih.

Pengkategorian atau pengklasifikasian *tweet* ini dilakukan oleh peneliti dengan bantuan media online maupun dari rekan peneliti yang lain. Tentu saja dalam hal ini peneliti dibantu rekan dari disiplin ilmu berbeda yang lebih memahami tentang pemberian emosi atau sentimen pada kalimat. Namun karena keterbatasan peneliti hanya mampu melakukan 127 data *tweet* yang bisa dikategorikan.

Berikut contoh tweet pengkategorian oleh peneliti:

1	username	date	retweets	favorites	Tweet	geo	mentions	hashtags	permalink	Class
49	@alibasartech	5/9/2018 14:35	0	0	Baik... Alhamdulillah sudah berhasil!				https://twitter.com/alibasartech/status/994118643150548993	1
50	@chundyeol	5/9/2018 22:19	0	0	Telkomsel sih lya kayanya				https://twitter.com/chundyeol/status/994235453025449600	0
51	@ucupsaklek	5/9/2018 22:19	0	0	Ada yang ngerasa jaringan Telkom (ntah Indihome atau Telkomsel) lagi lemot banget?				https://twitter.com/ucupsaklek/status/994235457339703296	-1

Gambar 4.4 Contoh Data Latih

Proses ekstraksi fitur yang digunakan adalah pembersihan *tweet* tanda baca, hastag, *url*, tag html, dan hal-hal yang dianggap gangguan. Hal ini dilakukan agar saat proses klasifikasi berjalan hal-hal tersebut tidak menjadikan proses kalkulasi tidak berjalan dengan baik.

Setiap rincian dari ekstraksi fitur terdapat pada *pseudocode* pada subbab Ekstraksi fitur. Metode yang digunakan adalah melakukan Regex pada *tweet* yang menjadi data pada penelitian ini. Setiap *tweet* harus melalui proses tersebut.

Stopword Removal termasuk dalam Ekstraksi fitur. Namun dalam hal ini tidak semua yang seharusnya dianggap sebagai *stopword* dihapus. Dalam hal ini peneliti menganggap kata tersebut penting dalam suatu kalimat. Sehingga tidak harus dihilangkan.

Pada pembahasan sebelumnya disebutkan metode *PortStemmer* sebagai metode pengakaran kata. *PortStemmer* hanya salah satu metode untuk melakukan pengakaran kata. Salah satu yang sering digunakan adalah dari penelitian dari nazrief tentang algoritma untuk mengakaran kata.

Dalam kasus tertentu pengakaran kata dibutuhkan untuk mengetahui sifat dan kedudukan dari kata tersebut. Misalnya saja untuk mengetahui frasa dari kata tersebut. Namun kendala utama dari *stemmer* adalah penggunaan Bahasa Indonesia yang tidak baku dan terlebih lagi pengguna Bahasa Indonesia sendiri yang majemuk sehingga menyebabkan *stemmer* dapat memiliki arti yang berbeda dari arti sebenarnya. Dalam hal majemuk ini seringkali penggunaan Bahasa Indonesia tercampur dengan Bahasa Daerah sehingga terkadang berbeda arti maupun maknanya.

Pada penelitian ini *PortStemmer* tidak digunakan karena pada pengujian pertama yang dilakukan oleh peneliti hasil *tweet* tidak sesuai dengan apa yang dimaksud oleh penulis *tweet* tersebut. Hal ini diakibatkan acuan pengakaran kata merujuk pada Bahasa Indonesia terlebih khusus pada Kamus Besar Bahasa Indonesia (KBBI) yang mungkin sekali memiliki kemiripan dengan Bahasa daerah yang sudah diserap kedalam Bahasa Indonesia dan mengalami pergeseran arti ataupun makna.

Pembobotan kata menggunakan *tfidf* dihitung menggunakan waktu yang dibutuhkan dalam uji latih maupun uji coba. Dalam uji latih data sample yang digunakan 127 tweet sedangkan data ujicoba menggunakan 50 data. Pada 127 tweet telah dikategorikan menjadi 3 yaitu: positif, negatif, netral. Pengkategorian ini masih berkaitan dengan sentimen tweet terhadap telkomsel. Tidak hanya produk, namun juga pelayanan serta brand telkomsel.

Pada hasil data latih pembobotan kata terhadap data latih disajikan pada gambar 4.6 kumpulan matrix tersebut dibentuk dari seluruh kosat kata yang terdapat pada seluruh tweet kemudian dihitung. Seperti yang disebutkan diatas

proses pembobotan kata menggunakan sublinear. Pada subliner data dihasilkan lebih cepat. Namun dengan bias yang lebih banyak. Tapi hal ini berkurang jika data latih yang diberikan semakin banyak dan bervariasi.

Pada kolom dibawah hasil sebagian hasil dari pembobotan kata.

```
([[0.0, 3.8142592685777856, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, , 0.0,
2.252762968495368, 2.252762968495368, 0.0, 0.0, 0.0,
2.9459101490553135, 0.0, 0.0, 0.02.252762968495368, 0.0]], ['a', 'abe',
'abe,', 'telkomsel', 'ngawur', 'sinyal', 'banyak', 'provider', 'indosat', 'luar',
'andai', 'gabut', 'at', 'confronting', 'malang', 'daerah', 'negara', 'jawa", 'last,',
'mahal,', 'presiden', 'telkom', 'pecat', 'bot', 'penak.', 'miskin.', 'lemot', 'ganti',
'problem:', 'puas', 'rasa', 'entah', 'to', 'kacau', 'tower', 'simcard', 'harus',
```

Dataset yang digunakan pada penelitian ini merupakan *tagged dataset*. *Tagged dataset* digunakan sebagai proses belajar algoritma *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* sebagai algoritma untuk mengklasifikasi *dataset* ujjicoba. Penelitian ini dilakukan untuk menguji metode *Multinomial Naïve Bayes* dalam hal ketepatan dalam melakukan klasifikasi dan kecepatan dalam kalkulasi.

Dari hasil yang didapatkan dari proses pengujian presisi dan *recall* setiap tabel kelas mempunyai nilai yang berbeda. Nilai presisi kelas positif berbeda dari kelas negatif maupun kelas netral. Nilai tersebut bergantung pada hasil klasifikasi oleh *Multinomial Naïve Bayes*. Selain itu *dataset* yang digunakan juga mempengaruhi hasil. Meskipun begitu *dataset* yang sudah dilakukan penandaan oleh peneliti juga berpengaruh dalam hasil klasifikasi.

Sebagai contoh pada tabel hasil kelas positif menunjukkan relevansi yang berbeda meski sama-sama menghasilkan probabilitas positif. Hal ini terkait juga

dengan susunan kata yang pada *tweet* tersebut. Dari kedua *tweet* tersebut mempunyai susunan kata yang berbeda.

Pada hasil presisi dan *recall* hasil perhitungan sangat bervariasi. Meskipun juga bergantung pada jumlah data acuan namun hasil tersebut merupakan kalkulasi dari hasil penelitian.

Pada kelas positif hasilnya 0.8125 dan 0.8125 dari hasil tersebut terlihat banyak *tweet* yang dihasilkan mampu mendeteksi *tweet* yang berkategori positif namun untuk presisi sendiri memiliki angka yang lebih besar.

Pada kelas negatif hasilnya 0.8 dan 0.9 hasil ini berbeda dengan kelas positif. Berbanding terbalik dengan kelas positif. Selain dari jumlah *dataset* yang berbeda juga berpengaruh dari variasi kata yang menjadi penyusun kalimat. Dengan model *Naïve Bayes* yang menganggap setiap kata independen sehingga susunan kata tidak mempengaruhi dari pengkategorian. Namun nilai dari bobot kata yang menjadi acuan utama untuk proses klasifikasi.

Pengaruh dari besaran *dataset* yang digunakan juga berpengaruh dalam pengkategorian. Selain *dataset* utama yang masih dalam skala kecil dan tidak terlalu bervariasi dalam kelas sehingga persebaran kelas juga tidak terlalu merata.

Dari sisi kecepatan kalkulasi aplikasi yang dibuat cukup cepat dengan data yang terbilang sedikit apabila dibandingkan dengan *library* untuk metode yang sama yakni *Multinomial Naïve Bayes*. Hal ini terlihat dari proses pembobotan kata maupun proses *Multinomial Naïve Bayes* sendiri yang cenderung lebih cepat sepersekian detik dari *Library Scipy*.

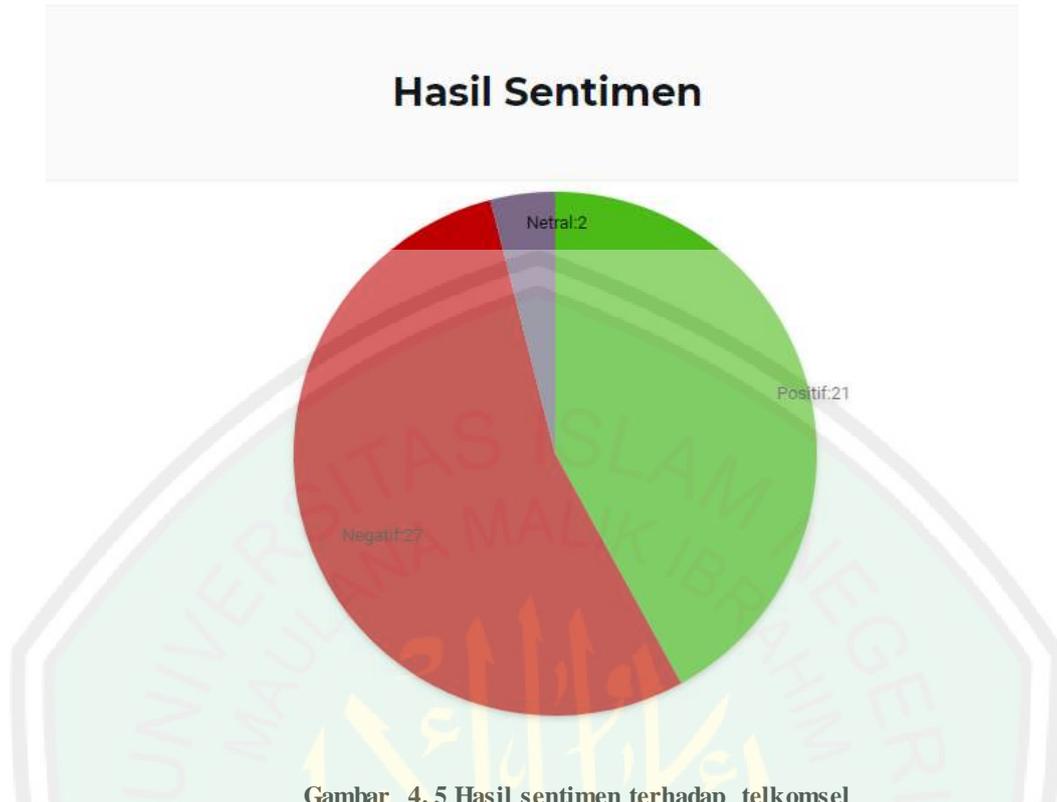
Pengaruh kecepatan ini terdapat pada proses perhitungan yang digunakan. Pada proses yang dilakukan peneliti proses kalkulasi menggunakan hitungan yang

sederhana namun juga tidak mengabaikan proses ketepatan dari algoritma. Namun melihat hasil yang memiliki bias yang cukup banyak dari kedua algoritma dibutuhkan metode untuk menurunkan nilai bias tersebut. Nilai bias ini terlihat dari hasil yang tidak sesuai dengan harapan. Terlihat dari *tweet* yang relevan maupun relevan tapi tidak dianggap benar.

Pada penelitian ini *provider* terbatas pada telkomsel sehingga kata untuk melakukan pencarian data terbatas pada *provider* telkomsel. Pencarian kata untuk mendapatkan data di *Twitter* juga dilakukan dengan proses *trial and error*. Hal ini dilakukan agar data yang didapatkan sesuai dengan yang diharapkan.

Hasil sentimen umum sebagai hasil klasifikasi pada layanan ini juga sekaligus sebagai jawaban dari rumusan masalah yang sebelumnya didapatkan dari latar belakang masalah. Hal ini untuk mengetahui seberapa besar sentimen masyarakat pada *provider*. Meskipun bisa diketahui dari data ujilatih namun data tersebut hanya sebagian kecil dari seluruh populasi data yang didapatkan.

Sentimen terhadap provider telkomsel dari hasil penelitian yang diperlihatkan pada gambar 4.5. Pengujian ini dilakukan dengan 127 data latih dengan rincian 50 data positif, 2 data netral, 75 data negatif. Diujicobakan pada 50 data. Dengan hasil sesuai gambar 4.5.



Gambar 4.5 Hasil sentimen terhadap telkomsel

Meskipun dalam percobaan random akurasi yang didapatkan dari aplikasi cenderung mendekati dengan perhitungan. Namun harus dilakukan koreksi ulang. Melihat hasil dari yang didapatkan sentimen terhadap telkomsel cenderung negatif. Hal ini juga terlihat dari data yang didapatkan dari data latih yang cenderung lebih banyak negatif.

Namun klasifikasi positif juga menunjukkan data yang cukup banyak. Tidak hanya keluhan yang disampaikan namun juga pertanyaan, apresiasi dari pengguna juga ikut tersampaikan yang terlihat dari kelas positif data ujilatih maupun data ujicoba.

Namun mengingat variasi data yang belum begitu banyak bias yang dihasilkan juga hampir berbanding lurus dengan hasil yang didapat. Hal ini bisa diatasi dengan cara memperbanyak data latih. Untuk mendapatkan data latih diperlukan waktu *crawling* yang lebih banyak.

Dalam hal ini *Multinomial Naïve Bayes* berhasil sebagai salah satu klasifikasi yang ditandai dengan nilai yang cukup untuk semua kelas. Terutama mampu menyelesaikan masalah analisis sentimen pada penelitian ini.

4.5 Kajian Islam

Aplikasi yang dibangun pada dasarnya ialah tentang menganalisis sentiment suatu produk dimana digunakan sebagai salah satu tolak ukur dalam melakukan pelayanan jasa kepada konsumen ataupun pengguna. Sebagaimana diketahui tingkat persaingan dalam industri perdagangan sangatlah ketat, sehingga membutuhkan tolak ukur dalam melakukan pelayanan dan respon yang cepat sehingga pelanggan dapat memberikan hasil yang positif juga. Inilah yang menjadi acuan dalam jasa tersebut dalam melakukan inspeksi baik ataupun tidak.

Hal ini berkaitan dengan hadist Rasulullah SAW, sebagai berikut:

“ فَلَمْ يَزَلْ عُمَرُ يُرَاجِعُنِي فِيهِ حَتَّى شَرَحَ اللَّهُ لِي ذَلِكَ صَدْرِي، وَرَأَيْتُ الَّذِي رَأَى عُمَرُ

Artinya “Umar senantiasa membujukku untuk mengevaluasi pendapatku dalam permasalahan itu hingga Allah melapangkan hatiku dan akupun berpendapat sebagaimana pendapat Umar” [HR. Bukhari]. Seseorang dapat terbantu untuk mengevaluasi diri dengan bermusyawarah bersama dengan niat untuk mencari kebenaran. Imam Bukhari mengeluarkan suatu riwayat yang menceritakan usul Umar kepada Abu Bakr *radhiallahu anhu* untuk mengumpulkan al-Quran. Tatkala saat itu Abu Bakr menolak usul tersebut, akan tetapi Umar terus mendesak Beliau dan mengatakan bahwa hal itu merupakan kebaikan. Abu Bakr tidak bersikukuh dengan pendapatnya dan ketika terdapat usulan yang baik dan kedudukan Beliau lebih tinggi tidak menghalangi untuk menerima kebenaran pihak yang memiliki

pendapat yang berbeda. Hadist tersebut mengajarkan kita bahwa tidak ada manusia yang sempurna yang tak luput dari salah dan lupa sehingga kita sebagai sesama umat manusia wajib untuk saling mengingatkan.

Dari aplikasi bisa diketahui respon masyarakat/konsumen terhadap pelayanan apakah masyarakat menilai positif atau negatif. Metode yang digunakan dalam penelitian ini adalah Multinomial Naive Bayes yang mana dalam pengimplementasian dilakukan pemilihan kata/crawling yang akan menghasilkan hasil yang akurasi dalam menentukan nilai positif dan negatif yang sudah diakurasi terhadap data training. Penulis sangat berhati-hati dalam pengerjaan aplikasi mengingat hadist Rasulullah yang berbunyi:

عَنْ جُبَيْرِ بْنِ سَعِيدٍ عَنِ الْأَعْلَى عَبْدِ عَنْ سُفْيَانَ حَدَّثَنَا السَّرِيُّ بْنُ بَشْرٍ حَدَّثَنَا غَيْلَانُ بْنُ مَحْمُودٍ حَدَّثَنَا
بِعَنْبَرِ الْقُرْآنِ فِي قَالَ مَنْ وَسَلَّمَ عَلَيْهِ اللَّهُ صَلَّى اللَّهُ رَسُولُ قَالَ قَالَ عَنْهُمَا اللَّهُمَّ رَضِيَ عَبَّاسُ ابْنِ
صَحِيحٌ حَسَنٌ حَدِيثٌ هَذَا عَيْسَى أَبُو قَالَ النَّارِ مِنْ مَفْعَدَهُ فَلْيَتَّبِعُوا عِلْمِ

Artinya : (Al-Turmuzi berkata): Mahmud bin gailan telah menceritakan kepada kami, (Mahmud berkata): Bisyr bin al-Syariyy menceritakan kepada kami, (Bisyr berkata): Sufyan menceritakan kepada kami dari ‘Abd al-A-‘la dari Sa’id bin Jubair dari Ibn “Abbas Ra, ia berkata : Rasulullah SAW bersabda : siapa yang mengatakan tentang (isi) al-Qur’an dengan tanpa landasan pengetahuan, maka hendaklah ia menempati tempat duduknya dari api neraka” Abu “isa (al-Turmuzi) berkata: hadist ini hasan sahih.

Hadist ini mengingatkan kepada kita bahwa penafsiran sesuatu apapun tidak bisa dilakukan semena mena termasuk aplikasi ini tidak bisa dan tidak boleh mengklasifikasikan tweet positif dan negatif. Oleh kerennya diperlukan data training sebagai acuan pengaplikasian Multinomial Naive Bayes.

BAB V PENUTUP

5.1 Kesimpulan

Pembahasan pada sebelumnya menunjukkan pegujian metode *Multinomial Naïve Bayes*. Metode *Multinomial Naïve Bayes* terbukti berfungsi baik pada penelitian ini. Hal ini terbukti hasil presisi dan *recall* yang bernilai 0.8125 dan 0.8125 pada kelas positif dan 0.8 dan 0.9 pada kelas negatif.

Pengaruh yang paling besar terlihat pada *dataset* yang digunakan. Pada *dataset* yang digunakan masih pada jumlah yang terlampaui sedikit hanya $\frac{1}{2}$ dari data ujicoba. Sehingga variasi data tidak terlalu banyak. Terbukti dari kelas positif yang memiliki data sedikit nilai presisi dan *recall* tidak mencapai 0.9.

Dari proses presisi dan *recall* dapat akurasi sebesar 70.4 %. Hasil ini didapatkan dari kedua proses presisi. Sehingga dapat dikatakan berhasil untuk sentimen analisis yang dibuat pada penelitian ini.

Pada sentimen layanan provider klasifikasi dapat digambarkan dari peneliti penunjukkan sentimen negatif yang cukup banyak. Hal ini ditunjukkan pada grafik pada bab sebelumnya. Dimana sentimen terhadap provider telkomsel dari total data 50 data uji ditemukan 27 kelas negatif, 21 kelas positif dan 2 kelas netral. Total data tersebut dihasilkan dalam 1 kali percobaan dengan dataset yang sama dataset percobaan lain. Meskipun tidak berbeda terlalu jauh dengan kelas positif tentunya ini tidak terlalu baik untuk bisnis tersebut. Sehingga dapat menjadi perhatian yang lebih untuk pihak *provider* untuk melakukan analisis sentimen lanjutan.

5.2 Saran

Setelah mengembangkan aplikasi ini, dan beberapa saran yang harus diterapkan guna pengembangan lebih lanjut:

1. Peneliti menyarankan untuk penelitian selanjutnya aplikasi ini bisa berkembang menjadi lebih baik dengan menambahkan fitur-fitur yang dibutuhkan
2. Kiranya pengembangan program aplikasi ini dapat digunakan bukan hanya untuk menganalisa sentiment brand Telkomsel saja namun dapat digunakan untuk public figure serta perusahaan lain yang ingin mengetahui respon pasar terhadap produk dan eksistensi yang dihasilkannya
3. Aplikasi ini diharapkan dapat dikembangkan sehingga bisa dikatakan aplikasi *User Friendly*

DAFTAR PUSTAKA

- Calvin, S. J. (2014). *Using Text Mining to Analyze Mobile Phone Provider Service Quality (Case Study : Sosial Media Twitter)*. *International Journal of machine Learning and Computing*, 106-109. doi:<http://doi.org/10.7763/IJMLC.2014.V4.395>
- Christopher D. Manning, P. R. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Coletta LFS, D. S. (2014). *Combining classification and clustering for tweet sentiment analysis*. *Proceedings-2014 Brazilian Conference on Intelligent Systems* (pp. 210-215). Brazil: BRACIS. doi:<https://doi.org/10.1371/journal.pone.0079449>
- Gamallo P, G. M. (2014). *Citius: A Naïve_Bayes Strategy for Sentimen Analysis on English Tweets*. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEva 2014)*.
- Haibing Wu, X. G. (2014). *Balancing Between Over-Weighting and Under-Weighting in*. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1604/1604.04007.pdf>
- Imam Fahrur Rozi, S. H. (2012). *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi*. *Jurnal EECCIS*, 6.
- Kamaruzaman, S. C. (2014). *Text Categorization using Association Rule and Naive Bayes Classifier*. *Asian Journal of Information Technology*, 3(9), 657-665.
- Kibriya Ashraf M., F. E. (2014). *Multinomial Naïve Bayes for Text Categorization Revisited*. *Australian joint conference on artificial(17)*.
- Kohlhase, M. (2017, August 24). *Provisional Lecture Notes : Artificial Intelligence (Künstliche Intelligenz 1)*. FAU Erlangen-Nürnberg.
- Ng, A. (2017, September 22). *Generative Learning algorithms*. *CS229 Lecture notes*. Retrieved from <http://cs229.stanford.edu/notes/cs229-notes2.pdf>
- Pak, A. & Paroubek, P. 2010. *Twitter as a Corpus for Sentimen Analysis and Opinion Mining*. Dalam *Proceedings of the Seventh International*

Conference on Language Resources and Evaluations (LREC'10.Valletta: Malta)

Riszki Wijayatun Pratiwi, Y. S. (2016). *Prediksi Rating Film Menggunakan Metode Naïve Bayes*. *Jurnal Teknik Elektro*, 8.

Shimodaira, H. (2015, February 10). *Text Clasification using Naive Bayes. Learning and Data note 7*. Retrieved from <http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn-note07-2up.pdf>

Wijaya H, E. A. (2013). *Twitter Sentimen Analisis and Insight for Indonesian Mobile Operators*. In *Information System International Conference (ISICO)*, 367.

Wilson TA, W. J. (2009). *Recognizing Contextual Polarity: an exploration of features for phrase-level sentimen analysis*. *Computational Linguistics*, 399-433. doi:<http://doi.otg/10.1162/coli.08-012-R1-06-90>

LAMPIRAN 1

Tweet	Class	Probabilities Class	Positif	Negatif	Netral
njogeti admin telkom	Positif	4.32E-29	4.32E-29	5.38E-34	6.89E-32
telkomsel lemot lagi yah	Negatif	8.50E-22	1.27E-28	8.50E-22	2.09E-29
@ telkomsel ini responnya lama kah atau emang gak bisa min ane isi pulsa jam 16 penawaran baru muncul jam 18 pic twitter com/lyujspbe9	Negatif	1.10E-159	2.96E-169	1.10E-159	1.35E-179
@ telkomsel no penghipnotis	Negatif	2.31E-31	1.38E-31	2.31E-31	1.30E-42
duh ful	Negatif	0.595238095	0.3968254	0.5952381	0.007936508
@ myxl xl knp ya tiap sore jam 7 sampe malam gini lemot bgt coba tlp 817 aja ngga bisa nyambung kudu ganti ke @ telkomsel ni kayanya	Negatif	1.12E-209	1.26E-224	1.12E-209	4.88E-218
@ telkomsel tolong	Negatif	1.97E-90	5.43E-	1.97E-	4.00E-

nomer saya dicek 085228252121 kenapa tidak bisa kirim pesan			104	90	105
karyawannya liburan mau long weekend~	Negatif	3.94E-08	2.39E-08	3.94E-08	3.51E-11
dri tdi ditunggu kak	Positif	3.17E-13	3.17E-13	1.77E-13	2.68E-15
cek dm ya abang	Negatif	4.61E-26	1.73E-29	4.61E-26	7.34E-33
oke baik kak endra saya tunggu informasi dari pihak @ telkomsel terima kasih kembali	Positif	8.57E-102	8.57E-102	3.40E-121	4.10E-119
min anjir lama banget balesnyaa saya butuh kepastian bisa atau tidakkk	Negatif	7.32E-41	7.57E-46	7.32E-41	3.19E-51
@ telkomsel mohon info saya masih ada kuota gamemax 25gb misalnya kuota tsb saya gunakan utk flash reguler yg 24 jam apakah ini dimungkinkan dan	Positif	4.49E-173	4.49E-173	5.26E-186	5.00E-193

bagaimana langkah2nya terima kasih					
oke min sudah saya dm ya tolong ya min sinyal yg di dapat cuma h+ padahal baru beli kuota	Negatif	5.23E-158	3.26E-164	5.23E-158	3.13E-181
jam sgini telkomsel jelek bangeet	Negatif	5.37E-23	1.46E-26	5.37E-23	2.92E-30
how is it that my phone is showing a 4g signal but it is takin 10 minutes to download a 12mb app telkomsel you dirty sob	Netral	4.50E-110	1.31E-112	2.83E-110	4.50E-110
@ dominos_id mau tanya 1 berapa minimal pembayaran pakai bca dan @ tcash_id by @ telkomsel 2 apakah bisa pakai flash bca / mandiri / bri minimal berapa thxyou	Positif	1.79E-73	1.79E-73	2.92E-77	2.74E-87
iya oper aja sana sini kyak bola aneh dah	Netral	5.64E-53	1.90E-59	2.17E-55	5.64E-53
@ telkomsel tolong	Positif	2.56E-206	2.56E-	3.05E-	4.32E-

bantuannya saya udh regis no tapi koo bonus 10gb belum dpt dan pulsanya saya masih 8k tapi kok gak bisa dipakai dan selalu dibilang kurang pulsanya bahkan hanya untuk sms tolong penjelasannya			206	227	234
@ telkomsel haloo mau tanya app langit musik itu cuma untuk lagu indo	Positif	2.48E-69	2.48E-69	4.88E-79	1.58E-77
tolong kamu camkan ituyayuk	Negatif	7.39E-23	4.15E-23	7.39E-23	4.16E-27
dalam seri red dead redemption 2 sepertinya bakal menjadi akhir era para koboi lewat trailer barunya di sini pic twitter com/m2pbyxzsp7	Positif	9.32E-77	9.32E-77	3.22E-77	4.64E-87
@ telkomsel min sya suda registrasi dgn no ktp & kk yg benar tpi jawabnya tetap sama pic	Positif	1.20E-127	1.20E-127	1.16E-133	1.28E-141

twitter com/8xv6u5yez					
sudah bisa makasih	Negatif	2.32E-28	3.19E-30	2.32E-28	9.70E-33
@ telkomsel	Negatif	5.51E-07	1.29E-07	5.51E-07	9.71E-12
kmren telkomsel parah	Negatif	0.001357952	0.0005238	0.00135795	1.60E-06
maauf aku telkomsel	Positif	1.28E-13	1.28E-13	9.37E-14	8.65E-18
oh ya min kalo biaya sms penerimaan bukti sukses atauctransaksi melalui m-bca berapa telkomsel dan indosat	Positif	2.49E-54	2.49E-54	6.13E-55	1.68E-61
mending dadi sadel becak	Negatif	0.595238095	0.3968254	0.5952381	0.007936508
ok sudah saya dm	Negatif	1.46E-38	1.09E-38	1.46E-38	6.28E-45
@ telkomsel malem ka mau tanya ka no ku ko g bisa pake 4g ya	Negatif	7.94E-98	1.34E-106	7.94E-98	5.72E-111
bukan itu nomornya yang benar udah saya dm	Positif	2.88E-58	2.88E-58	3.90E-60	3.99E-68
satu jam yg lalu saya	Positif	1.63E-225	1.63E-	1.03E-	4.61E-

juga dention seperti ini mas sudah saya lakukan apa yg diperintah tetapi sampai sekarang belum ada respon juga dan masih dengan masalah yg sama			225	236	251
@ telkomsel malam min untuk layanan data daerah jabar lagi gangguan kah ini udah 2 minggu lebih kira2 sampe kapan ya	Negatif	2.02E-175	9.44E-180	2.02E-175	6.56E-195
@ telkomsel apakah paket seperti di gambar kuota 15gb full untuk flash atau di bagi dengan kuota videomax pic twitter com/z6brmkr1ti	Negatif	1.20E-129	2.19E-138	1.20E-129	1.65E-147
tumben bgt nih telkomsel ngadat gini udah complain dari minggu kemaren ditungguin tindak lanjutnya gak ada	Negatif	1.54E-127	8.90E-148	1.54E-127	1.06E-150

juga sampe gue complain lagi for the second time					
@ telkomsel admin tolong kode puk untuk nomor 082195661986 621002956266198601	Positif	7.70E-67	7.70E- 67	4.02E- 68	8.98E-72
hihihi mau telpon yang seru gaess pakek popcall aja dijamin deh aktifin disini yak *567*1*2#	Negatif	2.93E-102	8.30E- 111	2.93E- 102	1.63E- 105
@ tcash_id tcash belum bisa dipakai untuk operator selain telkomsel ya pic twitter com/a2mcbzgs7y	Positif	8.82E-81	8.82E- 81	3.67E- 81	4.23E-93
ya sudahlah signal @ telkomsel seminggu ini tiap habis maghrib kayak ga tau jalan pulang ke rumah mungkin aku disuruh kelarin novel aja biar pinter # tutupgadget	Positif	2.56E-174	2.56E- 174	5.71E- 182	2.17E- 186
@ telkomsel min ini jaringan internet kpn	Positif	5.91E-89	5.91E- 89	3.86E- 101	1.58E- 104

normal kembali mohon tindakan perbaikan segera diselesaikan					
ini untuk wilayah sidoarjo 4g emang lagi gangguan atau gimana dari kemaren kok lelet bener paket internet naik tapi kok ga dibarengi kualitas yg oke sih @ telkomsel	Negatif	1.18E-237	7.36E-281	1.18E-237	2.97E-270
@ telkomsel malam min mohon info cara mengaktifkan tcash wallet saya sudah download aplikasinya dan mendaftar tapi pada proses masukkan pin tcash terhenti karena pin yang saya masukkan salah hingga pin diblokir mohon bantuannya di nomor 081241344349	Positif	2.33E-171	2.33E-171	1.56E-183	1.78E-196
@ telkomsel mohon	Negatif	3.56E-128	9.77E-	3.56E-	5.29E-

konfirmasi sinyal 4g di daerah kami kabupaten kep yapen - serui kenapa hilang pdhl bru 2 hr mncul makasih			136	128	144
begini amat yaa jaringan @ telkomcare sama @ telkomsel padahal bumh	Negatif	1.60E-84	3.65E- 89	1.60E- 84	5.92E-94
admin telkomsel nih hahahaha	Negatif	4.11E-28	3.35E- 31	4.11E- 28	1.87E-39
@ telkomsel mau tanya saya isi pulsa dri tadi kok gak masuk2 trs cek pulsa jg gak bisa	Positif	7.18E-140	7.18E- 140	5.23E- 149	1.39E- 158
aku di purwokerto aku mau aktifasi tcash wallet haruskah ke grapariaku baru isi pulsaini dapat cashback 100rb	Positif	2.27E-83	2.27E- 83	3.06E- 91	1.50E-96
woo tak pikir cuma nomer ku tok jebule liane enek durung enek 1 jam iki mau sinyal ilang 2x di komplek yo muk	Positif	1.87E-206	1.87E- 206	3.75E- 209	3.49E- 217

kon restart hp mindah jaringan tapi yo gak ngaruh di cek nomer e jare admin e telkomsel ra ono gangguan wkwkwk					
--	--	--	--	--	--

