Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

BAB II
LANDASAN TEORI

2.1 Komunikasi

Hakikat komunikasi adalah proses pernyataan antar manusia. Yang dinyatakan itu adalah pikiran atau perasaan seseorang kepada orang lain dengan menggunakan bahasa sebagai alat penyalurnya. Dalam "bahasa" komunikasi pernyataan dinamakan pesan (*message*), orang yang menyampaikan pesan disebut komunikator (*communicator*) sedangkan orang yang menerima pernyataan diberi nama komunikan (*communicatee*). Untuk tegasnya, komunikasi berarti proses penyampaian pesan oleh komunikator kepada komunikan, jika dianalisis pesan komunikasi terdiri dari dua aspek, pertama isi pesan (*the content of the message*), kedua lambang (*symbol*). Konkretnya isi pesan itu adalah pikiran atau perasaan, lambang adalah bahasa hidup manusia untuk berinteraksi dengan orang lain terkait dengan makhluk sosial (Effendy, 2000). Menurut Devito (dalam Widiantari dan Herdiyanto, 2013) sifat komunikasi terbagi dalam dua jenis, yaitu komunikasi secara langsung dan tidak langsung, komunikasi langsung merupakan komunikasi yang dilakukan dengan saling bertatap muka dalam suatu aktivitas komunikasi tanpa menggunakan perantara media. Sedangkan komunikasi secara tidak langsung merupakan komunikasi yang dilakukan dengan tidak bertemu secara langsung dalam suatu aktivitas komunikasi, dalam hal ini komunikasi tidak langsung dilakukan dengan perantara media seperti *e-mail* dan jejaring sosial.

2.2 Twitter

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh Twitter Inc., yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan *tweets* (Twitter, 2013). Mikroblog 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 pengirim dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Pengguna dapat melihat *tweets* pengguna lain yang dikenal dengan sebutan pengikut (*follower*).

Semua pengguna dapat mengirim dan menerima *tweets* melalui situs Twitter, aplikasi eksternal yang kompatibel (telepon seluler), atau dengan pesan singkat (SMS) yang tersedia di negara-negara tertentu (Twitter, 2013). 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 sampai 140 karakter disesuaikan dengan kompatibilitas dengan pesan SMS. Batas karakter 140 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 140 karakter (Twitter, 2013). Twitter menggunakan bit.ly untuk memperpendek otomatis semua *URL* yang dikirim tampil.

Terdapat 17 fitur yang digunakan untuk mengklasifikasikan akun Twitter dalam penelitian ini, yaitu :

1. *Retweets,* menghitung jumlah *retweets.*
2. *Likes,* menghitung jumlah *tweet* pengguna lain yang disukai.
3. *Plain status count,* menghitung jumlah *tweet* tanpa menggunakan *URL*, *hashtag* dan *mention*.
4. *Replies* , menghitung jumlah *replies* yang diterima dari pengguna lain.
5. *Mention count,* menghitung jumlah *mention* pada setiap *tweet.*
6. *URLs count,* menghitung jumlah keseluruhan *URLs* yang terdapat pada *tweet.*
7. *Hashtags count,* menghitung jumlah *Hashtags* yang terdapat pada *tweet*.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

8. *Collective influence,* menghitung jumlah *follower* ditambah *favorite*.
9. *Bio_length,* jumlah kata yang digunakan untuk membentuk biografi atau *description* pada Twitter.
10. *Numbers,* jumlah angka dalam setiap *tweet*.
11. *Long_words,* jumlah kata pada *tweet* yang lebih panjang dari 6 huruf.
12. *Emoticon,* jumlah *emoticon* dalam setiap *tweet*.
13. *Follower,* menghitung jumlah *follower*.
14. *Following,* menghitung jumlah *following*.
15. *All punctuation,* jumlah tanda baca pada setiap *tweet*.
16. *Words count,* jumlah kata pada setiap *tweet*.
17. *Media,* jumlah foto/video pada setiap *tweet*.

2.2.1 API Twitter

Application Programming Interface (API) merupakan fungsi-fungsi untuk menggantikan bahasa yang digunakan dalam *system calls* dengan bahasa yang lebih terstruktur dan mudah dimengerti oleh programmer. Fungsi yang dibuat dengan menggunakan API tersebut kemudian akan memanggil *system calls* sesuai dengan sistem operasinya. Tidak tertutup kemungkinan nama dari *system calls* sama dengan nama di API (Twitter, 2013).

Pada awalnya perusahaan Summize yang menyediakan fasilitas mencari data di Twitter. Kemudian perusahaan Summize ini diakuisisi dan diganti merek menjadi *Twitter Search* sehingga *Search API* terpisah sebagai entitas sendiri. API Twitter terdiri dari 3 (tiga) bagian yaitu (Twitter, 2013) :

a. *Search API*

Search API dirancang untuk memudahkan user dalam mengelola *query search* di konten Twitter. User dapat menggunakannya untuk mencari *tweet* berdasarkan *keyword* khusus atau mencari *tweet* lebih spesifik berdasarkan username Twitter. *Search API* juga menyediakan akses pada data *Trending Topic*.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

b. ***Representational State Transfer (REST) API***

REST API memperbolehkan *developer* untuk mengakses inti dari Twitter seperti *timeline*, *status update* dan informasi user. *REST API* digunakan dalam membangun sebuah aplikasi Twitter yang kompleks yang memerlukan inti dari Twitter.

c. ***Streaming API***

Streaming API digunakan *developer* untuk kebutuhan yang lebih intensif seperti melakukan penelitian dan analisis data. *Streaming API* dapat menghasilkan aplikasi yang dapat mengetahui statistik *status update*, *follower* dan lain sebagainya.

2.3 Kepribadian

Kepribadian merupakan pola perilaku dan cara berpikir yang khas, yang menentukan penyesuaian diri individu terhadap lingkungan. Atkinson (dalam Retnowati dan Haryanthi, 2001) kepribadian mencakup kepribadian umum yang dapat diamati oleh orang lain dan kepribadian pribadi yang terdiri dari pikiran dan pengalaman yang jarang diungkapkan. Allport (Suryabrata, 2002) merumuskan kepribadian sebagai suatu organisasi yang dinamis dari sistem psikofisik individu yang menentukan tingkah laku dan pemikiran individu secara khas dalam menyesuaikan diri dengan lingkungannya. Eysenck (Suryabrata, 2002) memberikan definisi kepribadian sebagai keseluruhan pola perilaku, baik yang aktual maupun yang potensial dari organisme yang ditentukan oleh pembawaan dan lingkungan.

Hall dan Lindzey (dalam Retnowati dan Haryanthi, 2001) menganggap kepribadian sebagai suatu hal yang dapat memungkinkan prediksi tentang apa yang akan dilakukan individu dalam situasi tertentu, kepribadian berkenaan pada perilaku yang menyeluruh baik perilaku yang tampak maupun perilaku yang tidak tampak. Branca (dalam Retnowati dan Haryanthi, 2001) kepribadian

memperhatikan aspek-aspek yang tampak dari tingkah laku individu sebagai keseluruhan cara bertindak yang konsisten dari individu pada situasi tertentu. Hall dan Lindzey (dalam Retnowati dan Haryanthi, 2001) mendefinisikan kepribadian yang ditinjau dari sejarah organisme, fungsi kepribadian yang bersifat mengatur, memiliki ciri yang bersifat menetap dan berulang maupun baru dan unik, hakikat kepribadian yang abstrak dengan proses fisiologis yang mendasari proses-proses psikologis.

Berdasarkan uraian diatas, disimpulkan bahwa kepribadian merupakan suatu pola perilaku individu yang komprehensif, konsisten dan bersifat khas yang ditentukan oleh pembawaan serta lingkungan dalam ciri-ciri perilaku yang tampak maupun perilaku yang tidak tampak.

2.3.1 Karakteristik Kepribadian

Menurut Wulandari (dalam Retnowati dan Haryanthi, 2001) terdapat tiga karakteristik kepribadian yaitu pertama, kepribadian individu berkembang sepanjang kehidupan individu, yang ditandai dengan pengalaman hidup yang saling berintegrasi dan berakumulasi membentuk suatu kepribadian tertentu. Kedua, kepribadian individu bersifat unik dan khas, artinya bahwa kepribadian antara individu yang satu berbeda dengan kepribadian individu lain. Ketiga, perkembangan kepribadian sifatnya dinamis, tidak statis dengan cara-cara tertentu. Hal ini dipengaruhi oleh adanya faktor-faktor ekstrinsik berupa pola adaptasi dengan lingkungannya serta faktor intrinsik berupa pengalaman, motivasi dan faktor internal lainnya.

2.3.2 Tipe Kepribadian Ekstrovert

Menurut Chaplin (2005) tipe kepribadian ekstrovert adalah satu kecenderungan untuk mengarahkan kepribadian lebih banyak ke luar daripada ke dalam diri sendiri. Seorang ekstrovert mempunyai ciri bersifat sosial, lebih banyak

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

berbuat daripada berkontemplasi (merenung, berpikir), dan seseorang dengan motif-motif yang dikondisikan oleh kejadian-kejadian eksternal.

Jung (dalam Alwisol, 2004) mengatakan bahwa ciri-ciri orang dengan tipe kepribadian ekstrovert adalah sosiabel, lincah, aktif, asertif, senang mencari sensasi, periang, dominan, bersemangat dan berani. Eysenck (dalam Suryabrata, 2002) mengemukakan ciri utama kepribadian ekstrovert adalah memilih berpartisipasi dalam kegiatan bersama, pesta hura-hura, olahraga beregu (sepak bola, arung jeram), minum alkohol dan menghisap mariyuana. Eysenck menghipotesiskan ekstroversi (dibandingkan dengan introversi) melakukan hubungan seksual lebih awal dan lebih sering, dengan lebih banyak pasangan, dan dengan perilaku seksual yang lebih bervariasi.

Selanjutnya Jung (dalam Sujanto, 2001) mengatakan bahwa orang yang ekstrovert terutama dipengaruhi oleh dunia obyektif, yaitu dunia diluar dirinya. Orientasinya terutama tertuju keluar, pikiran, perasaan, serta tindakan-tindakannya terutama ditentukan oleh lingkungannya, baik lingkungan sosial maupun lingkungan non sosial. Orang ekstrovert bersikap positif terhadap masyarakat, hatinya terbuka, mudah bergaul, dan hubungannya dengan orang lain lancar. Jadi dapat disimpulkan bahwa tipe kepribadian ekstrovert adalah satu kecenderungan untuk mengarahkan kepribadian lebih banyak ke luar daripada ke dalam diri sendiri dan tindakannya lebih ditentukanoleh faktor-faktor yang ada di luar dirinya.

2.3.3 Tipe Kepribadian Introvert

Menurut Chaplin (2005) tipe kepribadian introvert adalah satu kecenderungan untuk mengarahkan kepribadian untuk menarik diri dari kontak sosial dan minatnya lebih mengarah ke dalam pikiran-pikiran dan pengalamannya sendiri. Jung (dalam Alwisol, 2004) mengatakan bahwa ciri-ciri orang dengan tipe introvert adalah tidak sosial, pendiam, pasif, ragu, banyak fikiran, sedih, penurut, pesimis dan penakut. Eysenck (dalam Suryabrata, 2002) mengatakan

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

bahwa individu yang berkepribadian introvert memiliki kecenderungan menarik diri, menghindar dari keramaian situasi di sekelilingnya yang dapat membuatnya kelebihan rangsangan. Orang yang memiliki tipe kepribadian introvert memiliki aktifitas yang miskin rangsangan sosial seperti membaca, olahraga soliter (main ski, atletik), dan organisasi persaudaraan eksekutif.

Selanjutnya Jung mengatakan bahwa orang introvert terutama dipengaruhi oleh dunia subyektif, yaitu dunia di dalam dirinya sendiri. Orientasinya terutama tertuju ke dalam pikiran, perasaan, serta tindakan-tindakannya ditentukan oleh faktor-faktor subyektif. Penyesuaian dengan dunia luar kurang baik, jiwanya tertutup, sukar bergaul, sukar berhubungan dengan orang lain, dan kurang dapat menarik hati orang lain (Sujanto, 2001).

Jadi dapat disimpulkan bahwa tipe kepribadian introvert adalah satu kecenderungan untuk mengarahkan kepribadian untuk menarik diri dari kontak sosial dan minatnya lebih mengarah ke dalam pikiran-pikiran dan pengalamannya sendiri.

2.4 Tipe Data Kuantitatif

Variable data kuantitatif dibagi menjadi data diskrit dan data kontinyu. Berikut adalah penjelasan data diskrit dan data kontinyu (Sarwono, 2006).

2.4.1 Variable Diskrit

Variabel Diskrit adalah variabel yang berupa data pengkategorian atau membedakan atau mengelompokkan jenis tertentu. Data jenis ini disebut Data Nominal atau Data Dikotomik. Misalnya: Data dikotomik 1 untuk kategori benar dan 0 untuk kategori salah. Data dikotomi ini sering digunakan untuk mengkatogorikan jawaban benar dan salah dari tes-tes jenis objektif semisal pilihan ganda (Multiple Choice Question (MCQ). Contoh lainnya adalah simbol 1 untuk pria dan 2 untuk wanita. Data tersebut bersifat tetap (setara) dan tentu tidak dapat digunakan dalam operasi hitung. Selain itu, data yang termasuk dalam Variabel

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak Cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Diskrit (Data Diskrit) adalah data bilangan bulat. Bilangan bulat adalah bilang yang tidak dalam bentuk pecahan/desimal.

2.4.2 Variable Kontinyu

Berbeda dengan variabel diskrit yang mana data hanya berfungsi sebagai label, maka variabel kontinyu merupakan data yang dapat digunakan untuk operasi hitung. Data kontinyu adalah data yang diperoleh dari hasil penghitungan atau pengukuran, sehingga data tidak hanya berupa bilangan bulat, tetapi juga bisa dalam bentuk desimal, misalnya 2,5. Data kontinyu juga bisa dalam bentuk bilangan bulat, namun kelompok data tersebut memungkinkan variasinya ke dalam bentuk pecahan. Contoh dari data untuk variabel ini adalah jumlah benar atau salah dalam suatu tes, skor nilai, ranking, tinggi badan, berat badan, panjang, jarak, dan lain-lain. Data tersebut dapat berubah-ubah atau bervariasi.

Berikut ini jenis-jenis data dalam Variabel Kontinyu:

a) Data Ordinal

Data ordinal merupakan data peringkat. Misalnya juara 1, juara 2 dan juara 3. Angka tersebut mempunyai makna, lebih dari sekedar label seperti pada data nominal di atas. Juara 1 tentu lebih pintar dari juara 2 dan seterusnya.

b) Data Interval (Rentangan)

Data interval merupakan data jenis rentangan yang sudah dapat digunakan dalam operasi hitung. Selain itu, data interval mempunyai jarak yang jelas di antara masing-masing data.

c) Data Rasio

Data rasio merupakan data pengukuran yang paling kompleks dan tentu dapat digunakan dalam opreasi hitung. Angka dalam data rasio merupakan angka yang sesungguhnya, bukan hanya sebagai simbol. Apabila ada angka 0 berarti memang angka 0 yang sebenarnya atau mutlak.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak Cipta Milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

2.5 Naïve Bayes Classifier

Naïve Bayes adalah salah suatu metode *machine learning* yang menggunakan perhitungan probalitas. Menurut Kusumadewi (dalam Citra, 2015) *Naïve Bayes* digunakan untuk melakukan klasifikasi data pada kelas tertentu. Klasifikasi-klasifikasi *bayes* adalah klasifikasi statistik yang dapat memprediksi suatu kelas probabilistas. Kelebihan *NBC* adalah sederhana tetapi memiliki akurasi yang tinggi (Rodiyansyah dan Winarko Edi, 2013).

Dalam metode *IR* konsep seperti ini biasa di tandai dengan adanya satu set data yang dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Sekelompok data yang akan diproses dan dicari kelasnya disebut dengan data *testing*, sedangkan data *training* merupakan data yang telah dihitung sebelumnya yang kemudian dibandingkan nilai dengan sejumlah fitur yang ada didalam data testing.

2.5.1 Teorema Bayes

Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan pengaruh yang didapat dari hasil observasi. Teorema Bayes menerangkan hubungan antara probabilitas terjadinya peristiwa A dengan syarat peristiwa B telah terjadi dan probabilitas terjadinya peristiwa B dengan syarat peristiwa A telah terjadi. Teorema ini didasarkan pada prinsip bahwa tambahan informasi dapat memperbaiki probabilitas. Teorema Bayes ini bermanfaat untuk mengubah probabilitas yang dihitung dengan tersedianya data dan informasi tambahan.

Sesuai dengan probabilitas subyektif, bila seseorang mengamati kejadian B dan mempunyai keyakinan bahwa ada kemungkinan B akan muncul, maka probabilitas B disebut probabilitas prior. Setelah ada informasi tambahan bahwa misalnya kejadian A telah muncul, mungkin akan terjadi perubahan terhadap perkiraan semula mengenai kemungkinan B untuk muncul. Probabilitas untuk B sekarang adalah probabilitas bersyarat akibat A dan disebut sebagai probabilitas

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

posterior. Teorema Bayes merupakan mekanisme untuk merubah probabilitas dari prior menjadi probabilitas posterior.

Prediksi *Bayes* didasarkan pada teorema *Bayes* dengan formula umum sebagai berikut (Han dkk, 2006) :

$$P(H|E) = \frac{P(E|H) x\, P(H)}{P(E)} \qquad (2.1)$$

Keterangan :

P (H|E) : Probabilitas akhir bersyarat suatu hipotesis H terjadi jika bukti E terjadi.

P(E|H) : Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

P(H) : Probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun.

P(E) : Probabilitas awal bukti E terjadi tanpa memandang hipotesis atau bukti yang lain.

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan bukti (E) yang diamati. Ada beberapa hal penting dari aturan *Bayes* tersebut, yaitu (Han dkk, 2006) :

1. Sebuah probabilitas awal H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau P(H|E) adalah probabilitas dari suatu hipotesis setelah bukti diamati

2.5.2 *Naive Bayes* Untuk Klasifikasi dan Prediksi

Menurut Prasetyo (2012) *Bayes* merupakan teknik prediksi berbasis probabilitas sederhana dengan asumsi independen (ketergantungan) yang kuat. Model yang digunakan dalam *naïve bayes* adalah model fitur independen, yang dimaksud dengan independen yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. *Naïve Bayes Classifier* menggunakan pendekatan teorema *Bayes* untuk

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

mengitung probabilitas kategori berdasarkan dokumen yang telah diketahui (Prasetyo, 2012).

Penghitungan probabilitas tersebut menggunakan persamaan :

$$\boldsymbol{P(c) = \frac{N_c}{N}} \tag{2.2}$$

Keterangan :

P(c) = Nilai prior setiap kelas.

Nc = Banyak dokumen dalam suatu kelas (n)

N = Jumlah keseluruhan dokumen.

Untuk klasifikasi dengan data bertipe kontinyu digunakan rumus *densitas Gauss* (George, dkk (2012). Fungsi densitas mengekpresikan probabilitas relatif. Rumus ini dikarakterisasi dengan dua parameter yaitu mean (μ) dan varian (σ^2) sebagai berikut :

$$\boldsymbol{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \tag{2.3}$$

Keterangan :

x : Nilai atribut

μ : Nilai *mean*

σ : Nilai *standar deviasi*

Dimana nilai *mean* dan *standar deviasi* dapat diestimasi dari data untuk setiap kelas data. Nilai *mean* dihitung dengan :

$$\boldsymbol{\mu = \frac{\sum_{i=1}^{n} xi}{n}} \tag{2.4}$$

Keterangan :

n : Jumlah data dalam satu kelas

x_i : nilai atribut ke-i

Sedangkan *standar deviasi* dihitung dengan :

$$\boldsymbol{\sigma^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (xi-\mu)^2} \tag{2.5}$$

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Sultan Syarif Kasim Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Keterangan :
μ : mean
n : jumlah data dalam satu kelas
x_i : nilai atribut ke-i

Setelah didapatkan nilai *conditional probabilities densitas gauss*. Maka selanjutnya penentuan kelas menggunakan persamaan :

$$\mathbf{P}\left(\mathbf{c} \mid \mathbf{dn}\right) = \mathbf{P}(\boldsymbol{c}) * \prod \boldsymbol{p}(\boldsymbol{w}|\boldsymbol{c}) \qquad (2.6)$$

Keterangan :
P(c|dn) : *Choosing a class*
Pc : *Priors*
$\prod p(w|c)$: Total *Conditional Probabilities Densitas Gauss*

Sebuah keuntungan dari NBC adalah bahwa ia memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan varian dari variabel) yang diperlukan untuk klasifikasi. Karena variabel diasumsikan independen, hanya varian dari variabel-variabel untuk setiap kelas yang perlu ditentukan dan bukan keseluruhan *covariance matrix*.

2.6 *K-fold Cross Validation*

K-fold cross validation adalah teknik yang dapat digunakan jika memiliki jumlah data yang terbatas. Cara kerja *k-fold cross validation* sebagai berikut :

1. Seluruh data dibagi menjadi K bagian.
2. *Fold* ke -1 adalah ketika bagian ke-1 menjadi data uji *(testing* data) dan sisanya menjadi data latih (*training* data). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.
3. Fold ke-2 adalah ketika bagian ke-2 menjadi data uji (testing data) dan sisanya menjadi data latih (*training* data). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.
4. Demikian seterusnya hingga mencapai *fold* ke-K.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

5. Hitung rata-rata akurasi dari N buah akurasi di atas. Rata-rata akurasi ini menjadi akurasi final.

Metode *k-fold cross validation* melakukan generalisasi dengan membagi data kedalam **k** bagian berukuran sama. Selama proses berlangsung, salah satu dari partisi dipilih untuk data uji, dan sisanya digunakan untuk data latih. Langkah ini di ulangi **k** kali sehingga setiap partisi digunakan untuk data uji tepat satu kali. Metode *k-fold cross validation* menetapkan **k = N** (ukuran dari dataset). Pendekatan ini memiliki keuntungan dalam penggunaan data sebanyak mungkin untuk *training*. *Test set* secara efektif mencakup keseluruhan data set. Kekurangan data pendekatan ini adalah banyaknya komputasi untuk mengulangi prosedur sebanyak **N** kali. *K-fold cross validation* adalah salah satu teknik untuk mengevaluasi keakuratan model klasifikasi (Citra, 2015).

2.7 Confusion Matrix

Salah satu cara untuk menghitung akurasi adalah dengan menggunakan metode *confusion matrix*. *Confusion matrix* merupakan sebuah cara yang berguna untuk menganalisis seberapa baik *classifier* mengenali data. Metode ini menggunakan tabel matriks seperti terlihat pada Tabel 2.1 berikut ini, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif.

Tabel 2.1 Model *Confusion Matrix*

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	True positive	False positives
-	False negative	True Negatives

True positive adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *false negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

positif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif, kemudian masukkan data uji (Citra, 2015). Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-niai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity* (*recall*), *specifity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specifity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif.

Untuk menghitung digunakan persamaan dibawah ini :

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \tag{2.7}$$

$$\text{Specifity} = \frac{TN}{(FP+TN)} \tag{2.8}$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{2.9}$$

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)} \tag{2.10}$$

Keterangan :

TP = jumlah *true positives*

TN = jumlah *true negatives*

FP = jumlah *false positives*

FN = jumlah *false negatives*

2.8 Penelitian Terkait

Penelitian terkait mengenai *text mining* menggunakan jejaring sosial Twitter. Penelitian terkait dapat dilihat pada Tabel 2.2 berikut :

Tabel. 2.2 Penelitian Terkait

Peneliti	Judul	Bahasa	Metode Klasifikasi	Ekstraksi Fitur	Domain	Akurasi
Sarwani & Mahmudy (2015)	Analisa Twitter Untuk	Indonesia	*NBC*		*Tweet*	Dengan mengguna kan 10

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Peneliti	Judul	Bahasa	Metode Klasifikasi	Ekstraksi Fitur	Domain	Akurasi
	Mengetahui Karakter Seseorang Menggunakan Algoritma Naïve Bayess Classifier					data latih dan 10 data uji maka didapatkan akurasi 100%
Damanik & Khodra (2015)	Prediksi Kepribadian *Big 5* Pengguna Twitter Dengan *Support Vector Regression*	Indonesia	*Support Vector Regression*	Bigram	*Tweet*	Nilai *Person Correlation* 85%
Pamungkas, dkk (2015)	Analisis sentiment pada sosial media Twitter menggunakan Naïve Bayes Classifier terhadap kata kunci	Indonesia	*NBC*		*Tweet*	91 %

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Peneliti	Judul	Bahasa	Metode Klasifikasi	Ekstraksi Fitur	Domain	Akurasi
	"kurikulum 2013"					
Febrianto, Prasetia, & Wijaya (2015)	Pembuatan Sistem Prediksi Kepribadian " *The Big Five Traits* " dari Media Sosial Twitter	Indonesia	*Neural Network, SVM, Naïve Bayes*		*Tweet*	71 %
Citra (2015)	Implementasi Naïve Bayes Classification untuk menentukan sentimen terhadap operator seluler pada *tweet*	Indonesia	*NBC*		*Tweet*	Tanpa *stopword* (86%), menggunakan *stopword* (90,5%)

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU