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.

BAB II
LANDASAN TEORI

2.1 *Text Mining*

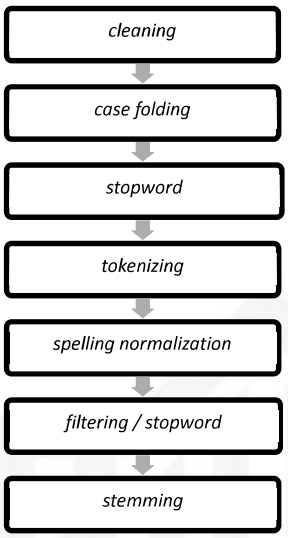
Text mining (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersebar, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman & Sanger, 2007). *Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction* dan *information retrival* (Berry & Kogan, 2010). Tahap-tahap *text mining* secara umum adalah *text preprocessing* dan *feature selection* (Feldman & Sanger, 2007), (Berry & Kogan, 2010).

2.1.1 *Text Preprocessing*

Menurut (Feldman & Sanger, 2007) *text preprocessing* merupakan tahapan proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah selanjutnya. Sekumpulan karakter yang bersambungan (teks) harus dipecah-pecah menjadi unsur yang lebih berarti, yang dapat dilakukan dalam tingkatan yang berbeda. Tahapan yang dilakukan dari *text preprocessing* dapat dilihat pada 2.1.

© 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.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU



Gambar 2.1 Proses *text preprocessing*

Tahapan *text preprocessing* terdiri dari proses *cleaning, case folding*, *tokenizing*, *filtering, spelling normalization, filtering dan stemming*.

a. *Cleaning* merupakan proses membersihkan review dari kata-kata yang tidak diperlukan untuk mengurangi proses noise pada proses klasifikasi. Kata-kata yang dihilangkan adalah karakter.

b. *Case folding* merupakan proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai z.

c. *Stopword* merupakan kumpulan kata-kata yang sering muncul dalam suatu dokumen. *Stopword* pada umumnya adalah sebuah kata penghubung yang tidak begitu penting, maka *stopword* dapat diabaikan dan tidak ikut dalam proses pengindeksan *stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopword* adalah"yang", "dan","di", dan sebagainya.

c. *Tokenizing* merupakan proses pemotongan kalimat menjadi sebuah kata dengan melakukan analisa terhadap dengan melalakukan analisa terhadap kumpulan data dengan memisahkan kata tersebut dan menentukan struktur dari setiap kata tersebut.

d. *Spelling normalization* (normalisasi) merupakan proses identifikasi kata silang dan penulisan kata berlebihan kemudian diganti dengan kata kamus KBBI (pusat bahasa departemen pendidikan nasional, 2008). Pada tahap ini setiap

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

dijumpai kata yang penggunaan huruf berlebihan dan kata yang tidak baku akan diubah. Adapun algoritma normalisasi yang dilakukan pada penelitian ini sebagai berikut:

1. Cari kata yang akan dinormalisasi dalam kamus. Jika ditemukan maka di asumsikan bahwa kata tersebut adalah *root word* (kata dasar), maka algoritma berhenti.
2. Jika tidak ditemukan hapus huruf berlebihan dimulai untuk setiap huruf pada kata, periksa huruf pertama kata tersebut, kemudian *recoding*. Periksa huruf selanjutnya jika huruf sama dengan huruf sebelumnya maka hapus huruf tersebut, jika tidak simpan huruf lakukan hal yang sama pada huruf selanjutnya.
3. Melakukan *recoding*
4. Jika telah diperiksa untuk setiap huruf periksa kata hasil proses sebelumnya pada kamus.
5. Jika ditemukan maka algoritma berhenti. Jika tidak ditemukan algoritma ini mengembalikan kata yang asli sebelum dilakukan penghapusan huruf berlebihan.
6. Selanjutnya periksa kata pada kamus silang.
7. Jika ditemukan lakukan perubahan kata menjadi kata baku. Jika tidak ditemukan maka kembalikan kata pada *root word* (kata dasar).

e. *Filtering* adalah tahap mengambil kata - kata penting dari hasil tokenizing. Proses filtering dapat menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist / stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words.

f. *Stemming* merupakan proses mencari kata *root* / kata dasar dari setiap kata hasil dari proses normalisasi. Karena data komentar yang akan diklasifikasi menggunakan bahasa indonesia maka algoritma *stemming* untuk berbahasa indonesia yang mempunyai tingkat keakuratan yang lebih baik dibanding algoritma lainnya adalah algoritma ECS. Algoritma ini mengacu pada aturan

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

KBBI yang mengelompokkan imbuhan yang diperbolehkan atau imbuhan yang tidak diperbolehkan.

Berikut merupakan langkah-langkah algoritma ECS :

1. Kata yang belum di *stemming* dicari pada KBBI. Apabila kata langsung ditemukan, berarti kata tersebut adalah kata dasar, kata dikembalikan dan algoritma dihentikan.
2. Hilangkan *inflectional suffixes* terlebih dahulu, jika ini berhasil dan *suffix* adalah pertikel (“lah”atau“kah”), langkah ini dilakukan lagi untuk menghilangkan *inflectional possesive pronoun suffixes* (“ku”,“mu”atau “nya”).
3. Partikel *Derivational suffix* (“i”,”-an”,”-kan”) kemudian dihilangkan, langkah dilanjutkan lagi untuk mengecek apakah masih ada *derivational suffix* yang tersisa, jika ada maka akan dihilangkan. Apabila tidak ada lagi maka lakuan langkah selanjutnya.
4. *Derivational prefix* (“di-”, “ke-”, “se-”, “te-”, “me-”, “be-”, “pe-”) dihilangkan, kemudian langkah dilanjutkan lagi untuk mengecek apakah masih ada *derivational prefix* yang tersisa, jika ada maka akan dihilangkan. Apabila tidak ada lagi maka lakukan langkah selanjutnya.
5. Setelah tidak ada lagi imbuhan yang tersisa, maka algoritma ini dihentikan kemudian kata dasar tersebut di cari pada KBBI, jika kata ditemukan berarti algoritma ini berhasil tapi jika kata dasar tidak ditemukan maka dilakukan *recoding.*
6. Jika semua langkah telah dilakukan tetapi kata dasar tidak ditemukan pada kamus, maka algoritma ini mengembalikan kata yang asli sebelum dilakukan stemming.

2.1.2 *Text Transformation*

Tahap *analyzing* adalah tahap penentuan keterkaitan kata-kata antar dokumen. Pada tahapan ini pemprosesan teks dilanjutkan dengan proses transformasi teks menjadi data numerik sebagai reprentasi dari setiap dokumen.

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.

Pada text transformation ini kita hanya menentukan (TF) saja, karena hanya membobot jumlah kata yang akan digunakan untuk menghitung probilitas.

2.1.3 Penggalian Informasi Pada *Text Mining*

Tahap akhir penggalian informasi pada *text mining* yaitu ekstraksi ilmu pengetahuan (*knowledge discovery*), dimana terdapat beberapa jenis kategori utama yang bisa dilakukan sebagai berikut (Miner, dkk, 2012).

2.1.3.1 Klasifikasi

Klasifikasi adalah bentuk analisis data yang mengekstrak model untuk menggambarkan kelas data. Model yang dibangun meliputi pengklasifikasian dan prediksi kategori label kelas. Klasifikasi data mempunyai dua tahapan proses, yaitu tahap pembelajaran (*learning step*) dimana model klasifikasi dibangun berdasarkan label yang sudah diketahui sebelumnya dan tahapan klasifikasi (*classification step*) dimana model digunakan untuk memprediksi label kelas dari data yang diberikan. Klasifikasi memiliki berbagai aplikasi, termasuk deteksi penipuan, penargetan marketing, prediksi kinerja, manufaktur, diagnosis medis, dan banyak lainnya. Sebagai contoh, kita dapat membangun sebuah model klasifikasi untuk mengkategorikan apakah suatu aplikasi pinjaman bank termasuk aman atau beresiko.karena pada awal pembangunan model label kelas dari data telah diketahui, klasifikasi juga disebut sebagai metode *supervised learning.*

2.1.3.2 Pengelompokan

Tidak seperti klasifikasi, pada model *clustering* pengelompokan data dilakukan dengan menggunakan algoritma yang sudah ditentukan dan selanjutnya data akan diproses oleh algoritma untuk dikelompokkan menurut karakteristik alaminya. Tidak ada unsur pembimbingan (dengan pemberian label kelas), melainkan algoritma akan berjalan dengan sendirinya untuk mengelompokkan data tersebut. Data yang lebih dekat (mirip) dengan data lain akan berkelompok dalam satu *cluster*, sedangkan data yang lebih jauh (berbeda) dari data yang lain akan berpisah dalam kelompok yang berbeda. Untuk masalah pengelompokkan data berdasarkan kemiripan/ketidakmiripan antar data tanpa ada label kelas yang

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.

diketahui sebelumnya disebut dengan pembelajaran tidak terbimbing atau *unsupervised learning.* Dalam konteks yang lain, pembelajaran tidak terbimbing disebut juga pengelompokan atau *clustering*. Menurut struktur, *clustering* terbagi menjadi dua, yaitu *hierarki* dan *partisi*. Dalam pengelompokan berbasis hierarki, satu data tunggal bisa dianggap sebuah *cluster*, dua atau lebih *cluster* kecil dapat bergabung menjadi sebuah *cluster* besar, begitu seterusnya hingga semua data dapat bergabung menjadi sebuah *cluster*. Di sisi lain, pengelompokan berbasis partisi membagi set data ke dalam sejumlah *cluster* yang tidak bertumpang-tindih antara satu *cluster* dengan *cluster* yang lain, artinya setiap data hanya menjadi anggota satu *cluster* saja.

2.1.3.3 Asosiasi

Asosiasi merupakan proses pencarian hubungan antar elamen data. Dalam dunia industri retail, analisis asosiasi biasanya disebut *market basket analysis* Asosiasi tersebut dihitung berdasarkan ukuran *Support* (presentase dokumen yang memuat seluruh konsep suatu produk A dan B) dan *confidence* (presentase dokumen yang memuat seluruh konsep produk B yang berada dalam subset yang sama dengan dokumen yang memuat seluruh konsep produk A).

2.1.3.4 Analisis Tren

Tujuan dari analisis tren yaitu untuk mencari perubahan suatu objek atau kejadian oleh waktu. Salah satu aplikasi analisis tren yaitu kegiatan identifikasi evolusi topik pada penelitian artikel.

2.2 BUKALAPAK

Bukalapak merupakan salah satu online marketplace terkemuka di Indonesia yang menyediakan sarana jual–beli dari konsumen ke konsumen. Semua orang dapat membuka toko online di Bukalapak dan melayani pembeli dari seluruh Indonesia untuk transaksi satuan maupun banyak. Bukalapak merupakan bagian dari PT Kreatif Media Karya group (https://www.bukalapak.com/about).

Adapun fitur fitur yang terdapat pada bukalapak adalah:

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

1. Push

   Fitur push merupakan cara untuk mempromosikan barang sehingga berada di posisi pertama di halaman daftar barang. Fitur push hanya dapat dibayar dengan saldo buka dompet. Menu paket push bisa dilihat di bawah menu buka dompet dan juga menyediakan beli paket push di halaman barang yang dijual.

2. Cetak Alamat Pengiriman

   Fitur ini merupakan dari halaman transaksi, dimana penjual dapat mencetak alamat dari si pembeli dengan lengkap dari transaksi yang sedang berlangsung,

3. Notifikasi SMS setiap ada pesanan

   Penjual akan mendapatkan SMS setiap ada pesanan.

4. Feedback positif setiap transaksi sukses

   Penjual akan mendapatkan feedback positif secara otomatis dari sistem walaupun pembeli tidak memberikan feedback saat transaksi.

5. Kepastian menerima uang pembayaran

   Penjual dapat langsung menerima uang pembayaran setelah pembeli mengkonfirmasi penerimaan barang atau 1×24 jam setelah barang dikirim oleh kurir.

6. Prioritas di mesin pencarian

   Setiap halaman barang penjual telah melalui proses SEO (Search Engine Optimization) agar tampil lebih unggul.

7. Perhitungan ongkos kirim otomatis

   Penjual tidak perlu lagi repot menghitung ongkos kirim karena sudah di sediakan sistem yang otomatis akan menghitung ongkos kirim yang harus dibayar oleh pembeli.

2.3 *Analisis Sentimen*

Analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen bersifat positif , negatif atau netral. Penelitian mengenai analisis sentimen telah

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.

berkembang sejak tahun 2003 dan merupakan bagian dari *text mining* yang merupakan penelitian komputasi berdasarkan sentimen, *emoticon*, pendapat, komentar dan setiap ekspresi yang diungkapkan oleh teks. Analisis sentimen difokuskan untuk *review* klasifikasi berdasarkan polaritas. Berdasarkan klasifikasi, analisis sentimen dibagi menjadi dua kelompok utama. Yaitu dokumen klasifikasi ke pendapat atau fakta, atau dikenal sebagai klasifikasi subjektivitas (*subjectivity classification*) dan dokumen klasifikasi ke dalam positif atau negatif, atau dikenal sebagai analisis sentimen. Hal ini adalah proses yang penting untuk menentukan dokumen yang memiliki opini dan dokumen yang menyimpulkan opini bernilai positif, negatif maupun netral (Faishol Nurhuda, Sari Widya Sihwi, dan Afrizal Doewes, 2013).

2.4 *Naïve Bayes Classifier*

Algoritma *Naïve Bayes Classifier* merupakan algortima yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger, 2007).

Dalam penelitian ini yang menjadi data uji adalah komentar konsumen bukalapak. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah di kategorinya, sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya. Sebuah keuntungan dari *Naïve Bayes Classifier* adalah bahwa ia memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan varian dari variable) yang diperlukan untuk klasifikasi. Karena variable diasumsikan independen yang mana hanya varian dari variable-variabel untuk setiap kelas yang perlu ditentukan dan bukan keseluruhan *covariance matrix.* Penghitungan nilai probabilitas tersebut menggunakan persamaan (Jurafsky, 2011).

$$\boldsymbol{P(c)} = \frac{\boldsymbol{N_C}}{\boldsymbol{N}} \tag{2.1}$$

Nc = Banyak dokumen dalam suatu kelas (n)
N = Jumlah keseluruhan dokumen data latih dan data uji

$$\mathbf{P}\left(\mathbf{c} \,\middle|\, \mathbf{dn}\right) = \mathbf{P}(\boldsymbol{c}) * \boldsymbol{\Pi p(w|c)} \tag{2.2}$$

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.

P(c|dn) = *Choosing a class*

Pc = *Priors*

$\Pi p(w|c)$ = Total *Conditional Probabilities*

Selanjutnya,

$$\mathbf{P(p/n/net) = \frac{N(pos/neg/net)}{N}} \quad (2.3)$$

Keterangan :

P(p/n/net) = *Priors* positif atau negatif.

N (pos/neg) = Dokumen mengandung komentar positif atau negatif

N = Jumlah dokumen

$$\mathbf{P(w|c) = \frac{count\ (w,c+1)}{count\ (c)+|V|}} \quad (2.4)$$

Count (w,c) = Frekuensi kata **w** pada kelas **c**

Count (c) = Total frekuensi kata pada masing-masing kelas **c**

I V I = Total kata unik pada keseluruhan kelas c

$$\mathbf{P\left(p/n/net \mid d7\right) = P\ (p/n/net) * \Pi p(w|p/n/net)} \quad (2.5)$$

Keterangan :

$\mathbf{P\left(p/n/net \mid d7\right)}$ = *Choosing a class*

$\mathbf{P\ (p/n/net)}$ = Probabilitas kelas positif atau negatif

$\mathbf{\Pi p(w|p/n/net)}$ = Total *Conditional Probabilities* kata pada kelas positif atau Negatif

2.5 *Blackbox*

Blackbox testing merupakan pengujian yang dilakukan hanya mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari perangkat lunak. Jadi digambarkan seperti melihat kotak hitam yang hanya bisa melihat penampilan luarnya saja, tanpa tahu ada apa dibalik bungkus hitamnya. Sama seperti pengujian *blackbox*, mengevaluasi hanya dari tampilan luarnya (interface), fungsionalitasnya tanpa mengetahui apa sesungguhnya yang terjadi didalam proses detailnya (hanya mengetahui input dan output) (Mustaqbal, 2015).

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.

Metode uji dapat diterapkan pada semua tingkat pengujian perangkat lunak, unit, integrase, fungsional, sistem dan penerimaan. Biasanya terdiri dari kebanyakan jika tidak semua pengujian pada tingkat yang lebih tinggi, tetapi juga bisa mendominasi unit *testing* juga. *Blackbox Testing* cenderung untuk menemukan hal-hal berikut :

1. Fungsi yang tidak benar atau tidak ada.
2. Kesalahan antarmuka (*interface errors*).
3. Kesalahan pada struktur data dan akses basis data.
4. Kesalahan performansi (*performance errors*).
5. Kesalahan inisialisasi dan terminasi.

Pengujian didesain untuk menjawab pertanyaan-pertanyaan berikut:

1. Bagaimana fungsi-fungsi diuji agar dapat dinyatakan valid?
2. Input seperti apa yang dapat menjadi bahan kasus uji yang baik?
3. Apakah sistem sensitif pada input-input tertentu?
4. Bagaimana sekumpulan data dapat diisolasi?
5. Berapa banyak rata-rata data dan jumlah data yang dapat ditangani sistem?
6. Efek apa yang dapat membuat kombinasi data ditangani spesifik pada operasi sistem?

Kelebihan *Blackbox Testing* :

1. Spesifikasi program dapat ditentukan di awal.
2. Dapat digunakan untuk menilai konsistensi program.
3. *Testing* dilakukan berdasarkan spesifikasi.
4. Tidak perlu melihat kode program secara detail.
5. Dapat memilih *subset test* secara efektif dan efisien.
6. Dapat menemukan cacat.
7. Memaksimalkan *testing* investmen.

Kekurangan *Blackbox Testing* :

1. Bila spesifikasi program yang dibuat kurang jelas dan ringkas, maka akan sulit membuat dokumentasi setepat mungkin.
2. Tester tidak pernah yakin apakah PL tersebut benar – benar lulus uji.

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.

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* adalah sebagai berikut (Mustika, 2015) :

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.
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 berlansung, 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 data set. Pendekatan ini memiliki keuntungan dalam penggunaan data sebanyak mungkin untuk training (pengujian). *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.

2.7 *Confusion Matrix*

Metode ini menggunakan tabel matriks seperti yang terlihat pada 2.1 berikut ini jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negative (Kohavi and Provost, 1998).

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.

Tabel 2.1 ***Confusion matrix***

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	True positives	False positives
-	False negatives	True negatives

True positives 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 positif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity* (*recall*), *specificity*, *precision* dan *accuracy*.*Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah record yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan di bawah ini:

$$Sensivity = \frac{TP}{(TP+FN)} \quad (2.6)$$

$$Specifity = \frac{TN}{(FP+TN)} \quad (2.7)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2.8)$$

$$Accuracy = \frac{TP+TN}{(P+N)} \quad (2.9)$$

Keterangan:

TP = jumlah *true positives*

TN = jumlah *true negatives*

FP = jumlah *false positives*

FN = jumlah *false negative*

2.8 Penelitian Terkait

Pada tabel 2.2 berikut dapat dilihat beberapa penelitian sebelumnya mengenai analisis sentiment.

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

Tabel 2.2 Penelitian terkait

No	Peneliti	Judul	Metode	Kesimpulan
1	Faishal Nuruz Zuhri, Andry Alamsyah, S.Si., M.Sc (2017)	Analisis Sentimen Masyarakat Terhadap *Brand* Smartfren Menggunakan *Naive Bayes Classifier* di Forum Kaskus	NBC	Analisis sentimen terhadap data Kaskus mengenai *brand* Smartfren dilakukan dengan metode *Naïve Bayes Classifier*, dengan menggunakan 1000 data latih, didapatkan *precision* sebesar 98.42%, *recall* sebesar 98.40%, dan tingkat akurasi sebesar 98.40%. Hal ini menunjukkan bahwa sistem dapat memisahkan data bersentimen positif dan negatif dengan sangat baik. Dan juga memperoleh nilai koefisien kappa sebesar 0.948.
2	Juen Ling, I Putu Eka N. Kencana, Tjokorda Bagus Oka (2014)	Analisis Sentimen Menggunakan Metode *Naïve Bayes Classifier* dengan Seleksi Fitur *Chi Square*	NBC	Berdasarkan hasil yang diperoleh dapat disimpulan bahwa kemunculan frekuensi fitur pada kategori yang diharapkan dan kategori yang tidak diharapkan memiliki peranan penting dalam seleksi fitur *Chi Square,* oleh karena itu seleksi fitur *Chi Square* baik digunakan dalam penyeleksian fitur dibandingkan dengan metode *frequency-based*. Serta pembangunan sistem analisis sentimen menggunakan metode NBC dengan Bahasa pemrograman Java memperoleh akurasi sebesar 83% dan rata-rata harmonik sebesar 90,713%. Terdapat kesalahan klasifikasi karena pada data uji terdapat fitur yang muncul pada bukan kategorinya.
3	Ghulam Asrofi Buntoro (2016)	Analisis Sentimen Hatespeech Pada Twitter dengan Metode *Naïve Bayes Classifier* dan *Support Vector Machine*	NBC dan SVM	Nilai akurasi Analisis sentimen twit Bahasa Indonesia dengan tagar Hatespeech tertinggi didapat dengan metode klasifikasi *Support Vector Machine (SVM)* dengan tokenisasi *unigram, stopword list* Bahasa Indonesia dan *emoticons,* dengan nilai rata-rata akurasi mencapai 66,6%, nilai presisi 67,1%, nilai *recall* 66,7% nilai *TP rate* 66,7% dan nilai *TN rate* 75,8%. Dapat diketahui metode klasifikasi *Support Vector Machine (SVM)* lebih tinggi akurasinya dibandingkan metode klasifikasi *Naïve Bayes Classifier (NBC).*
4	Menurut Syahmia Gusriani, Kartina Diah Kusuma Wardhani dan Muhammad Ihsan Zul	Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi	NBC	Kesimpulan yang diperoleh dari penelitian ini ialah *Naïve bayes* dapat dijadikan metode klasifikasi untuk analisis sentimen dengan keakuratan 93.7%.

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.

No	Peneliti	Judul	Metode	Kesimpulan
	(2016)	Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)		
5	Akhmad Pandhu Wijaya, Heru Agus Santoso (2016)	Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government	NBC	Teknik klasifikasi dokumen dengan NBC dan pembobotan fitur metode *tf-idf* menghasilkan nilai yang pasti dan akurasi yang baik karena bobot memperkecil kemungkinan kesalahan pada pengklasifikasian, fitur yang mempunyai frekuensi tertentu dapat mempengaruhi keakuratan dalam klasifikasi bergantung pada frekuensi fitur dan dokumen yang mengandung fitur tersebut.Hasil dari klasifikasi dokumen menggunakan NBC. Pada penelitian ini dengan data *training* sebanyak 260 dokumen politik dan 222 dokumen ekonomi menggunakan 40 data *testing* menunjukkan nilai akurasi yang baik pada keseluruhan klasifikasi, dengan akurasi keseluruhan klasifikasi sebesar 85%.

© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU