

BAB II

LANDASAN TEORI

2.1. *E-Comererce*

Pengertian *E-commerce* menurut (Laudon, C. K., & Laudon 2013) adalah suatu proses membeli dan menjual produk-produk secara elektronik oleh konsumen dan dari perusahaan ke perusahaan dengan komputer sebagai perantara transaksi bisnis, segala sesuatu berkaitan dengan transaksi jual beli yang dilakukan secara digital dengan menggunakan komputer yang terhubung dengan internet.

E-commerce adalah segala sesuatu yang berkaitan dengan kegiatan jual beli yang dilakukan secara elektronik dan menggunakan internet sebagai suatu jaringan penghubung antara penjual dan pembeli.

2.2. Persepsi

Nilai Pengertian Persepsi menurut (*Kotler_Keller 15th (1)*, n.d.) adalah proses yang digunakan oleh individu untuk memilih, mengorganisasi, dan menginterpretasi masukan informasi guna menciptakan gambaran dunia yang memiliki arti. Nilai pelanggan merupakan kombinasi kualitas, pelayanan, harga dari suatu penawaran produk. Nilai terhangat pada pelanggan adalah selisih antara jumlah nilai bagi pelanggan dan jumlah biaya dari pelanggan, dan jumlah nilai bagi pelanggan adalah sekelompok keuntungan yang diharapkan dari barang atau jasa tertentu. Nilai persepsi komentar mencerminkan pelanggan itu sendiri dalam saran keluhan dari baik maupun buruk dalam berbagai hal yang di dapat dalam ulasan.

2.3. Keputusan Pembelian

Pengertian keputusan pembelian menurut (Sumarwan dalam sari et al., 2017) keputusan pembelian merupakan suatu keputusan sebagai pemilihan suatu tindakan dari dua atau lebih pilihan alternatif. Sama halnya dengan yang dijelaskan oleh Setiadi dalam (Fitria, SE and Dwijananda 2016).

Keputusan pembelian adalah sesuatu yang berhubungan dengan rencana konsumen untuk membeli produk tertentu, serta berapa banyak unit produk yang dibutuhkan dalam periode tertentu. Menurut (Philip Kotler ; Kevin Lane Keller 2016) Keputusan pembelian konsumen adalah kegiatan perorangan dan rumah tangga yang membeli barang dan jasa untuk konsumsi pribadi.

Keputusan pembelian merupakan kegiatan individu yang secara langsung terlibat dalam pengambilan keputusan untuk melakukan pembelian terhadap produk yang ditawarkan oleh penjual. Perusahaan yaang cerdas memahami proses keputusan pembelian secara penuh, semua pengalaman mereka dalam pembelajaran, memilih, menggunakan dan bahkan menyingkirkan produk. Dari pengertian tersebut, dapat disimpulkan bahwa pengambilan keputusan merupakan suatu proses menentukan satu pilihan alternatif sebagai upaya untuk memecahkan masalah.

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 (Nurhuda, Sihwi, and Doewes 2013).

2.4. 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 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. 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 (Nurhuda, Sihwi, and Doewes 2013).

2.4.1 Normalisasi

Kalimat Bertujuan untuk menormalkan kalimat sehingga kalimat gaul menjadi normal (Adiasa 2013), sehingga bahasa gaul tersebut dapat dikenali sebagai bahasa yang sesuai dengan KBBI. Yang harus dilakukan untuk normalisasi kalimat adalah:

1. Meregangkan tanda baca (punctuation) dan symbol selain alphabet.
2. Meregangkan tanda baca adalah memberikan jarak terhadap tanda baca dari kata-kata sesudah atau sebelumnya, tujuannya agar tanda baca dan symbol selain alphabet tidak menjadi satu dengan kata-kata pada saat proses tokenisasi.
3. Mengubah menjadi huruf kecil semua.
4. Normalisasi kata.

2.4.2 Text Preprocessing

Text preprocessing merupakan tahapan proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah selanjutnya (Feldman, R., & Sanger 2017). 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

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*
- d. *Tokenizing* merupakan proses pemotongan kalimat menjadi sebuah kata dengan melakukan analisa terhadap dengan melakukan analisa terhadap kumpulan data dengan memisahkan kata tersebut dan menentukan struktur dari setiap kata tersebut.
- e. *Spelling normalization* (normalisasi) merupakan proses identifikasi kata slang dan penulisan kata berlebihan kemudian diganti dengan kata kamus KBBI (pusat bahasa departemen pendidikan nasional, 2008). Pada tahap ini setiap 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).
- f. *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*.
- g. *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.

. Kemudian dihitung *Inverse Document Frequency* (IDF) yaitu nilai bobot suatu *term* dihitung dari seringnya suatu *term* muncul di beberapa dokumen. Semakin sering suatu *term* muncul di banyak dokumen, maka

nilai IDF nya akan kecil. TF-IDF (*Term Frequency-Inverse Document Frequency*) (FELDMAN, Ronen, 2007).

Pembobotan TF-IDF adalah jenis pembobotan yang sering digunakan dalam information retrieval dan *Text Mining*. Pembobotan ini adalah suatu pengukuran statistik untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen. Tingkat kepentingan meningkat ketika sebuah kata muncul beberapa kali dalam sebuah dokumen tetapi diimbangi dengan frekuensi kemunculan kata tersebut dalam kumpulan dokumen. TF IDF dapat dirumuskan sebagai berikut,

$$TF\ IDF(t_k, d_j) = TF(t_k, d_j) \cdot IDF(t_k)$$

- h. Dimana sebelumnya dihitung terlebih dahulu *Term Frequency* (TF) yaitu frekuensi kemunculan suatu *term* di tiap dokumen. Kemudian dihitung *Inverse* Algoritma ini mengacu pada aturan

2.4.3 *Term Frequency–Inverse Document Frequency* (TF-IDF)

TF-IDF (*Term Frequency-Inverse Document Frequency*) (FELDMAN, Ronen; SANGER 2007) Pembobotan TF-IDF adalah jenis pembobotan yang sering digunakan dalam information retrieval dan *Text Mining*. Pembobotan ini adalah suatu pengukuran statistik untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen.

Tingkat kepentingan meningkat ketika sebuah kata muncul beberapa kali dalam sebuah dokumen tetapi diimbangi dengan frekuensi kemunculan kata tersebut dalam kumpulan dokumen. TF IDF dapat dirumuskan sebagai berikut,

$$TF\ IDF(t_k, d_j) = TF(t_k, d_j) \cdot IDF(t_k)$$

Dimana sebelumnya dihitung terlebih dahulu *Term Frequency* (TF) yaitu frekuensi kemunculan suatu *term* di tiap dokumen *Document*

Frequency (IDF) yaitu nilai bobot suatu *term* dihitung dari seringnya suatu *term* muncul di beberapa dokumen. Semakin sering suatu *term* muncul di banyak dokumen, maka nilai IDF nya akan kecil. Berikut rumus-rumus TF dan IDF.

$$TF(t_k, d_j) = f(t_k, d_j)$$
$$IDF(t_k) = \log \frac{N}{df(t)}$$

Pembobotan untuk menghitung perkata paling banyak keluar dari persamaan kata yang mewakili saran kritik dari komentar pelanggan

2.4.4 *Naïve Bayes Classifier*

Algoritma *Naïve Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman, Ronen, 2007). Model ditemukan berdasarkan analisis data training (objek data yang kelasnya diketahui)(Kim, Sang-Bum and Han, Kyoung-Soo and Rim, Hae-Chang and Myaeng 2006).

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah learning (fase training), dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk aturan klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari aturan klasifikasi Proses klasifikasi didasarkan pada empat komponen (Gorunescu 2011):

1. Kelas

Variabel dependen berupa kategori yang merepresentasikan “label” yang terdapat pada objek. Contohnya: rating.

2. *Predictor*

Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: rating buruk atau tidak, rating baik atau tidak

3. *Training dataset*

Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.

4. *Testing dataset*

5. Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.4.5 *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 and Affandes 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 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 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. Metode ini untuk mengurangi ketidakseimbangan pembagian data train and test yang akan di ukur dalam proses *naïve bayes*.

2.4.6 Naïve Bayes

Naïve Bayes Menurut Oslan dan Delen dalam (Kurniawan 2013) Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu dengan memprediksi masa depan berdasarkan data dari masa lalu.

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada *teorema Bayes* . Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes* . Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi atau kejadian.

Naïve Bayes untuk setiap kelas keputusan menghitung probabilitas dengan syarat bahwa kelas keputusan adaah benar,mengingat vector informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. *Probabilitas* yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

Naïve bayes adalah algoritma model machine learning dan memiliki ke unikan tersendiri ada beberapa tipe algoritma seperti:

1. Bernoulli Naive Bayes

Pada algoritma ini, prediktor adalah variabel boolean. Dikarenakan prediktornya variabel boolean, maka satu-satunya nilai yang ada adalah benar atau salah. Algoritma ini digunakan ketika data sesuai dengan distribusi *bernoulli multivariat*.

2. *Naive Bayes Multinomial*

Algoritma ini digunakan untuk memecahkan masalah klasifikasi dokumen. Contohnya, apabila ingin menentukan apakah suatu dokumen termasuk dalam suatu kategori, algoritma ini bisa digunakan untuk memilahnya. *Naive bayes* menggunakan frekuensi kata-kata sekarang sebagai *fitur*.

3. *Gaussian Naive Bayes*

Algoritma ini digunakan apabila prediktor tidak diskrit namun memiliki nilai kontinu. Prediktor tersebut diasumsikan sebagai sampel dari distribusi gaussian.

2.5 *Supervised learning*

Supervised learning merupakan suatu pembelajaran yang terawasi dimana jika output yang diharapkan telah diketahui sebelumnya. Biasanya pembelajaran ini dilakukan dengan menggunakan data yang telah ada (Pustejovsky, J. and Stubbs 2012). *Supervised learning* biasanya di pakai dalam klasifikasi *machine learning*.

2.6 *Machine learning*

Menurut (Mohri, Rostamizadeh, and Talwalkar 2012) *machine learning* dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa dijadikan data pembelajar

1. Dalam pembelajaran machine learning, terdapat skenario-skenario seperti :
Supervised Learning Penggunaan skenario supervised learning, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.
2. *Unsupervised Learning* Penggunaan skenario *unsupervised learning*, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi

label. Setelah itu mencoba untuk mengelompokan data berdasarkan karakteristik-karakteristik yang ditemui.

3. Reinforcement learning pada skenario reinforcement learning fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

2.7 Visualisasi data

Visualisasi Menurut (Andrews and Hirsch 2016) “Information visualisation (InfoVis) is the visual presentation of abstract information spaces and structures, together with accompanying interactions, so as to facilitate their rapid assimilation and understanding.” Definisi di atas menyatakan bahwa, visualisasi informasi adalah sebuah teknik penyajian informasi dari data abstrak dan terstruktur dengan cara representasi visual yang efisien untuk mempermudah pengguna dalam memahami informasi serta memiliki kemampuan interaktif terhadap pengguna agar mempermudah dalam melakukan analisa informasi yang disajikan.

Menurut saya visual adalah salah satu alat bantu dalam proses machine learning yaitu untuk menentukan probabilitas dan juga menentukan hasil yang akan di pilih berdasar kan akurasi atau rekomendasi dalam sebuah bentuk gambar maupun statistika.

2.8 Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python bisa dibilang bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat source code mudah dibaca. Python juga memiliki library yang lengkap sehingga memungkinkan programmer untuk membuat aplikasi yang mutakhir dengan menggunakan source code yang

tampak sederhana (Perkovic 2012). Python merupakan bahasa pemrograman terstruktur, pemrograman yang dapat digunakan untuk membuat aplikasi, perintah komputer, dan melakukan analisis data.

2.9 Wordcloud

Word cloud adalah salah satu hasil dari metode text mining yang menampilkan kata-kata populer terkait dengan kata kunci internet dan data teks. Menurut PBC (Arkham agustinah 2015) word cloud sering digunakan untuk menyoroti istilah populer atau trend berdasarkan frekuensi pengguna. Word cloud merupakan pendekatan yang dapat menjelaskan pertanyaan penelitian dengan sangat cepat dan mudah, dapat menjelajahi word cloud secara singkat dan dapat melakukan analisis yang komprehensif. Kata yang paling sering muncul di dalam data teks akan memiliki bentuk yang paling besar, begitu pula sebaliknya.

Menurut saya berdasarkan penjelasan diatas word clouds menampilkan kata merdasarkan nilai banyak yang paling sering keluar semakin banyak kata yang keluar akan semakin dominan atau besar kata yang akan terlihat pada visual tersebut.

2.10 Confussion Matrix

Pada bidang pengenalan pola (*pattern recognititon*) dan temu kembali informasi (*information retrieval*) terdapat proses perhitungan yang umum digunakan sebagai alat ukur untuk menghitung kinerja sebuah sistem atau metode yang digunakan yaitu *confusion matrix*. *Confusion matrix* merupakan suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Sehingga hasil evaluasi dengan *confusion matrix* berupa nilai akurasi, presisi dan *recall*. Presisi dan *recall* merupakan istilah yang akan muncul apabila sistem yang sudah dibuat dapat menampilkan hasil (*retrieve*) suatu hasil baik berupa klasifikasi, prediksi dan pencarian.

Sedangkan didalam buku “Pengolahan Citra Digital” yang ditulis oleh Pulung Nurtation Andono *et al* (2017), *confusion matrix* adalah matrix yang

berisikan mengenai hasil yang diperoleh dari prediksi klasifikasi dan data actual yang dilakukan oleh sistem klasifikasi. Jadi, *confusion matrix* berisikan suatu informasi dari sistem yang sudah melakukan perbandingan antara hasil pengelompokkan yang diharapkan atau seharusnya keluar. Umumnya kinerja sistem klasifikasi dihitung dengan memerlukan beberapa data yang ada dalam tabel 2.1 *confusion* matriks menggunakan 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP) dan *False Negatif* (FN) sebagai berikut:

Tabel 2.1 *Confusion Matrix*

		Nilai Sebenarnya	
		True	False
Nilai Prediksi	True	TP (True Positif) <i>Corect Result</i>	FP (False Positif) <i>Unexpected Result</i>
	False	FN (False Negatif) <i>Missing Resukt</i>	TN (True Negatif) <i>Corect Obsence of Result</i>

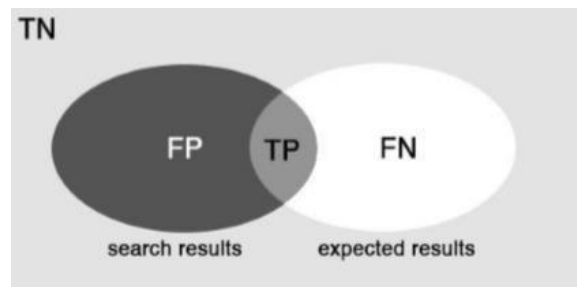
Jadi akurasi, Presisi dan *Recall* dapat di rumuskan sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+TN+TP+FN} \times 100 \% \quad (1)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \% \quad (3)$$

Ilustrasi dari tabel *confusion matrix* terdapat dalam gambar 2.1 seperti berikut :



Gambar 2.1 Diagram Ilustrasi Tabel *Confusion Matrix*

TP : Hasil pencarian relevan dan hasil muncul sesuai harapan.

FP : Hasil pencarian tidak relevan tapi hasil muncul secara tidak terduga.

FN : Hasil pencarian relevan tapi hasil tidak muncul sesuai harapan, hilang.

TN : Hasil pencarian tidak relevan dan tidak adanya hasil revisi yang muncul.

Akurasi (*accuracy*) adalah derajat kedekatan antara nilai prediksi dengan nilai actual (sesungguhnya). Nilai akurasi dapat diperoleh dengan Persamaan 1. Presisi (*precision*) atau *confidence* yaitu derajat dalam ketepatan atau kesesuaian antara informasi yang diminta oleh *user* dengan jawaban yang diberikan oleh sistem (*recommendation system*). Dapat diartikan juga bahwa *precision* merupakan alat ukur kualitas untuk melihat seberapa manfaatnya sistem pencarian (*Precision* → *How useful* → *Quality*). Nilai presisi dapat diperoleh dengan Persamaan 2. Sedangkan *Recall* atau *sensitivity* didefinisikan sebagai derajat keberhasilan sistem dalam menemukan kembali sebuah informasi pencarian (*Recall* → *How Complete* → *Quality*). Nilai *recall* dapat diperoleh dengan Persamaan 3. Error adalah kasus yang diidentifikasi salah dalam sejumlah data, sehingga dapat dilihat seberapa besar tingkat kesalahan pada sistem yang digunakan. Nilai error dapat diperoleh dengan Persamaan 4

$$Error = \frac{FP}{TP} \times 100 \% \quad (4)$$

Nilai yang biasanya diberikan untuk akurasi, presisi dan *recall* biasanya berbentuk presentase 1 - 100%, jika nilai yang didapatkan diatas 50% maka dapat dikatakan efektif dan begitu juga sebaliknya jika nilai yang didapatkan dibawah 50% maka

dapat dikatakan tidak efektif sistem tersebut. Di sarankan untuk menghitung dan mengevaluasi hasil sebuah sistem pencarian informasi, minimal menggunakan dua parameter yaitu presisi dan *recall* karena jika nilai persentase *recall* ataupun presisinya tinggi., maka sistem informasi tersebut dianggap baik. Contohnya, jika seseorang sedang mencari sebuah informasi mengenai “resep memasak sayur bayam” pada sebuah sistem pencarian dan didalam sistem tersebut terdapat 100 macam-macam resep memasak sayur bayam. Maka sistem tersebut dianggap baik apabila mampu menemukan 100 macam-macam resep memasak sayur bayam.

2.11 TextBlob

Textblob merupakan salah satu library dalam Python untuk pemrosesan data bentuk teks dan menggunakan NLTK untuk Natural Language Processing (NLP) yang bersifat unlabelled (Kaur & Sharma 2020). Textblob secara optimal digunakan untuk menganalisis sentimen pada data teks dengan bahasa formal. Dalam pengolahan analisis sentimen Textblob menggunakan dua nilai numerik untuk menentukan jenis sentimen, yakni subjectivity dan polarity. Subjectivity merupakan nilai dari opini bersifat subjektif yang ada di dalam sebuah data teks. Skor untuk subjectivity ada pada kisaran 0 sampai dengan 1. Skor 0 menunjukkan tingkat opini yang rendah di dalam suatu data teks dan sebaliknya. Polarity adalah variabel yang akan menentukan sentimen suatu data teks meliputi sentimen positif, negatif, dan netral. Skor untuk polarity berkisar antara -1 sampai dengan 1. Skor diidentifikasi sebagai data teks dengan sentimen negatif yang tinggi dan sebaliknya.