

BAB 2

LANDASAN TEORI

2.1 Sistem Pemerintahan Berbasis Elektronik

E-government, juga dikenal sebagai pemerintah berbasis elektronik adalah penggunaan dari internet dan banyaknya teknologi digital yang sudah berkembang untuk melakukan transaksi, menyediakan sebuah layanan publik, memudahkan dalam komunikasi, dan mengatur kordinasi manajemen organisasi dalam pemerintahan. Sistem ini yang mencakup, interaksi antara pemrintah ke pemrintah, hubungan antara pemerintah dan bisnis, dan layanan pemerintah kepada masyarakat. E-Government ditujukan untuk penerepan teknologi oleh pemerintah, terutama aplikasi berbasis web yang dapat memperluas akses dan meningkatkan kualitas layanan pemerintah kepada berbagai pihak, termasuk masyarakat umum, sektor bisnis, pegawai internal, dan institusi pemerintah lainnya. E-government juga merupakan proses pembaruan dalam metode kerja pemerintahan, sistem berbagi informasi, dan penyediaan layanan kepada pihak internal maupun eksternal yang memberikan manfaat bagi semua pemangku kepentingan termasuk pemerintah, masyarakat, dan pelaku usaha. Penggunaan beragaam teknologi informasi seperti internet, website, dan komputer yang diberikan oleh lembaga pemerintah kepada masyarakat, sektor bisnis dan cabang pemerintahan lain berusaha untuk meningkatkan pelayanan publik, memperkuat hubungan dengan dunia usaha dan industri, memberdayakan masyarakat dengan memberikan akses yang lebih mudah kepengetahuan dan data, dan menciptakan sistem pemerintahan yang lebih efektif dan efisien[9].

2.2 Google Play Store

Google menyediakan sebuah platform yang dikenal sebagai Play Store, yang menampung banyaknya produk digital seperti permainan, aplikasi, film, lagu, buku, komik, dan lain-lain dalam berbagai jenis. Sistem penilaian dan ulasan di Play Store adalah fitur yang memungkinkan pengguna memberikan pendapat mereka tentang produk atau barang yang telah mereka coba[10]. Google Play Store, yang merupakan App Store utama, berfungsi sebagai platform resmi untuk aplikasi Android yang berjalan pada sistem operasi Android. Lebih dari satu juta aplikasi dan lebih dari setengah juta penerbit dapat ditemukan di Google Play Store[11].

2.3 Andal by Taspen

Salah satu inovasi layanan digital PT Taspen (Persero) yaitu aplikasi Andal by Taspen, membantu orang lebih mudah dalam mengelola layanan pensiun mereka[12]. Penggunaan digitalisasi melalui aplikasi Andal by Taspen dimaksudkan untuk meningkatkan keamanan transaksi, mempercepat pencairan dana, dan mempersingkat tugas manajemen. Sebelumnya proses pembayaran gaji pensiun sering menghadapi masalah. Salah satu masalah utama yang paling umum adalah keterlambatan dan masalah administrasi, yang mempengaruhi kepuasan penerima manfaat[13].

Selain itu, pengguna dapat mengetahui hak dan kewajiban peserta, memperbarui E-SPTB, menyampaikan keluhan di Taspen care, dan menemukan lokasi layanan dengan menggunakan aplikasi ini. Selain itu, Anda dapat melaporkan JKK (Jaminan Kecelakaan Kerja) dengan menggunakan aplikasi ini untuk menghitung manfaat THT dan pensiun. Selain itu, Anda akan menemukan informasi tentang produk Taspen Life, prosedur autentikasi, formulir dan syarat untuk mengajukan klaim, dan cara mengatur jadwal kehadiran Anda di Mal Pelayanan Publik (MPP)[14]. PT Taspen (Persero) adalah badan usaha milik negara yang dipilih oleh pemerintah untuk mengelola program asuransi sosial untuk Pegawai Negeri Sipil, seperti Dana Pensiun dan Tabungan Hari Tua[15].

PT TASPEN Persero telah menunjukkan komitmennya terhadap transformasi digital melalui peluncuran superapps "Andal by Taspen". Aplikasi ini merupakan manifestasi dari Customer Digital Services yang mengintegrasikan seluruh layanan digital TASPEN dalam satu platform. Berdasarkan tinjauan literatur Andal by Taspen, aplikasi ini mewujudkan prinsip-prinsip e-government dalam konteks BUMN dengan menawarkan berbagai fitur komprehensif, termasuk pendaftaran, pengajuan klaim, informasi pensiun bulanan, dan sistem autentikasi yang disederhanakan. Corporate Secretary TASPEN menekankan bahwa pengembangan aplikasi ini tidak hanya ditujukan untuk kemudahan akses, tetapi juga sebagai solusi terhadap tantangan layanan digital masa depan. Fitur autentikasi berbasis swafoto mencerminkan upaya inovasi yang berpusat pada kebutuhan pengguna, terutama pensiunan lanjut usia. Sistem ini memungkinkan verifikasi klaim dan akses layanan dapat dilakukan tanpa batasan waktu dan tempat[3].

2.4 Analisis Sentimen

Analisis sentimen merupakan usaha untuk mengenali perasaan dan membagi polaritas dalam sebuah teks atau frasa. Dengan cara ini, setiap bagian teks dapat diwakili sebagai perasaan positif, negatif, atau netral. Jejaring sosial banyak digunakan untuk analisis sentimen untuk mengklasifikasikan persepsi[16]. Analisis sentimen berhubungan dengan emosi, opini, sikap, dan perasaan. Ini disebabkan oleh kebutuhan untuk menganalisis ide yang menyatakan sentimen dan mengumpulkan pengetahuan untuk mengeksplorasi perusahaan yang membutuhkan data teks. Analisis sentimen adalah teknik untuk menentukan apakah sebuah ulasan bersifat positif, negatif, atau netral. Analisis ini fokus pada ungkapan daripada topik yang dibahas dalam teks[17]. Dalam penelitian, algoritma klasifikasi naive bayes akan digunakan. Metode klasifikasi yang sederhana, algoritma Naive Bayes menghitung probabilitas dengan mengumpulkan dan mengintegrasikan nilai dari dataset[18].

2.5 Naive Bayes

Dalam data mining, algoritma Naive Bayes digunakan untuk mengklasifikasikan objek menggunakan teknik yang berbasis statistik dan probabilitas. Metode ini mampu melakukan prediksi kemungkinan suatu kejadian dengan memanfaatkan pola-pola yang telah teridentifikasi dari data yang telah dipelajari sebelumnya [19]. Metode statistik yang penting untuk pengenalan pola adalah teorema Bayes. Memanfaatkan algoritma naive bayes, penelitian ini memproses dan mengkategorikan ulasan aplikasi Andal by Taspen dari Google Play Store menjadi tiga kategori sentimen, positif, negatif dan netral. Rumus dasar persamaan teorema Bayes[20], rumus perhitungan Multinomial Naive Bayes [21].

Rumus 2.1 menunjukkan cara perhitungan *Teorema Bayes*

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

Ket :

X: Informasi data yang masih belum dikenal.

H: Hipotesis mengenai data X adalah kategori spesifik.

$P(H | X)$: Kemungkinan dari H berdasarkan dari kondisi X.

P(H): Kemungkinan dari hipotesis terkait H.

$P(X | H)$: Kemungkinan yang diperoleh dari X berlandaskan dari Hipotesis H.

$P(X)$: Kemungkinan X yang merupakan data contoh.

Rumus 2.2 merupakan bentuk penyederhanaan dari Teorema Bayes yang digunakan dalam proses klasifikasi, termasuk pada analisis sentimen berbasis algoritma Multinomial Naive Bayes. Rumus ini digunakan untuk menghitung probabilitas suatu hipotesis atau kelas H (positif, netral, atau negatif) berdasarkan data fitur X yang diamati, dalam hal ini berupa kata-kata dalam sebuah ulasan.

$$P(H|X) = P(X|H) \cdot P(H) \quad (2.2)$$

Ket:

$P(H | X)$: Probabilitas kelas H berdasarkan fitur X.

$P(X | H)$: Probabilitas data X muncul dalam kelas H

$P(H)$: Probabilitas awal dari kelas H, berdasarkan data latih.

2.6 Multinomial Naive Bayes

Dalam klasifikasi teks, model Multinomial Naive Bayes, yang sering digunakan, memperhitungkan berapa kali setiap kata muncul dalam setiap ulasan[22]. Metode klasifikasi yang dikenal sebagai Algoritma Multinomial Naive Bayes menggunakan probabilitas untuk membantu memutuskan apakah dokumen yang diuji mengandung sentimen yang baik atau buruk. Hasil klasifikasi ditunjukkan dengan jumlah total dokumen positif dan negatif[23].

Rumus 2.3 menunjukkan cara perhitungan *Multinomial Naive Bayes*

$$P(X|H) = \frac{\text{count}(X, H) + 1}{(\sum_{w \in V} \text{count}(X, H)) + |V|} \quad (2.3)$$

Keterangan:

$\text{count}(X|H)$ = Jumlah kali kata kunci muncul dalam kelas atau kategori.

$\sum_{w \in V} \text{count}(X|H)$ = Berasal dari jumlah semua kata terdapat pada kategori atau kelas.

$|V|$ = Jumlah dari semua kata unik yang terdapat pada setiap kategori.

2.7 TF-IDF

TF-IDF, juga dikenal sebagai Frekuensi Kata Terbalik Frekuensi Dokumen, memberikan nilai penting pada kata atau teks dalam dokumen atau kalimat berdasarkan dua faktor, jumlah kali kata muncul dalam dokumen atau kalimat dan jumlah dokumen yang mengandung kata-kata tersebut. Nilai frekuensi kata diberikan untuk menunjukkan betapa pentingnya kata tersebut.[24].

2.7.1 Rumus TF-IDF

Term Frequency (TF) merupakan pembobotan kata dalam dokumen. Inversed document frequency (IDF) frekuensi dari dokumen yang mengandung kata. Rumus dalam melakukan tahap TF-IDF dapat dilihat pada rumus 2.4, 2.5, 2.6

Rumus 2.4 menunjukkan cara perhitungan *Term Frequency (TF)*

$$tf(t, d) = \frac{f_{t,d}}{\max(tf)} \quad (2.4)$$

Keterangan:

$tf(t, d)$ = Frekuensi yang terdapat pada term (TF).

$\max(tf)$ = Total dari keseluruhan data yang terdapat dalam dokumen.

tf = Jumlah kemunculan term terbanyak didokumen yang sama.

$f_{t,d}$: Frekuensi term t dalam dokumen d

Rumus 2.5 menunjukkan cara perhitungan *Inversed Document Frequency (IDF)*

$$idf_t = \log \left(\frac{D}{df_t} \right) \quad (2.5)$$

Keterangan:

$idf(t)$ = Setiap dokumen harus memiliki term t .

df_t = jumlah dari dokumen yang mengandung term t

D = Total dari keseluruhan dokumen.

Rumus 2.6 menunjukkan cara perhitungan *TF-IDF*

$$W_{t,d} = tf(t, d) \times idf_t \quad (2.6)$$

Keterangan:

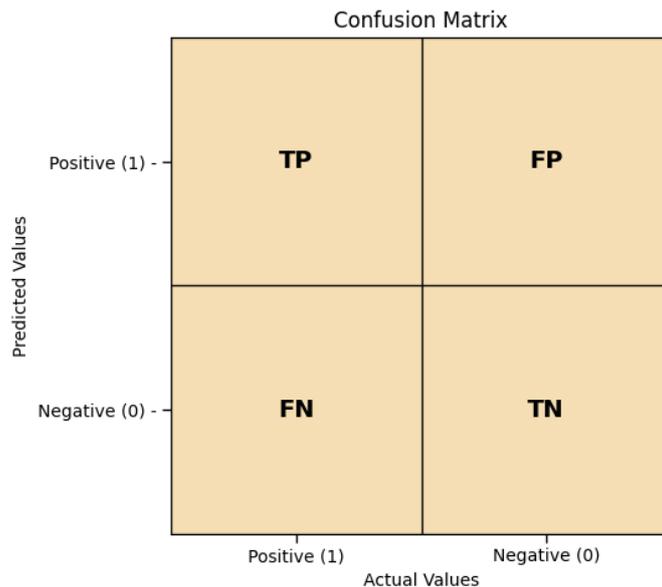
$W_{t,d}$ = bobot term dalam suatu dokumen

$idf(t)$ = bobot dari kemunculan term t pada setiap dokumen.

$tf(t, d)$ = Frekuensi term (TF).

2.8 Confusion Matrix

Confusion matrix merupakan teknik yang biasa dipakai untuk mengevaluasi tingkat ketepatan dalam aktivitas data mining[25]. Tabel confusion matrix menampilkan empat kategori kombinasi antara nilai aktual dan prediksi. Dalam confusion matrix, frase "true positif atau kata yang benar positif, true negatif adalah kata negatif asli, false positif atau kata positif palsu, dan false negatif atau kata negatif yang salah" menggambarkan hasil proses klasifikasi[26].



Gambar 2.1. Confusion Matrix

Keterangan:

TP: Kata prediksi yang benar benar positif, dan itu benar disebut True Positive.

TN: Kata negatif asli, yang memprediksi kata negatif dan itu benar disebut True Negative.

FP: Kata positif palsu, yang menduga kata positif dan salah disebut False Positive.

FN: Kata negatif yang salah, memprediksi kata negatif dan salah disebut False Negative.

Setelah jumlah dari keempat elemen confusion matrix diperoleh, metrik akurasi, presisi, dan recall dapat dihitung untuk mengevaluasi model. Ini dapat dilakukan dengan menggunakan rumus[7].

A Accuracy

Accuracy adalah penguruan dari semua data yang telah dikategorikan, baik yang positif maupun negatif. Dapat menghitungnya dengan menggunakan rumus berikut ini.

Rumus 2.7 menunjukkan cara perhitungan *Accuracy Confusion Matrix*

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.7)$$

B Precision

Perbandingan rasio benar benar positif oleh jumlah data yang menghasilkan hasil prediksi positif disebut precision[27]. Dapat menghitungnya dengan menggunakan rumus berikut ini.

Rumus 2.8 menunjukkan cara perhitungan *Precision*

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

C Recall

Recall adalah perkiraan rasio benar positif yang dibandingkan dengan perhitungan semua data benar positif[27]. Dapat menghitungnya dengan menggunakan rumus berikut ini.

Rumus 2.9 menunjukkan cara perhitungan *Recall*

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.9)$$

D F1-Score

Ketika nilai dari hasil precision dan recall digabungkan, maka akan tercipta sebuah metrik yang disebut F1-score, Metrik ini merupakan rata-rata harmonik

yang diberikan bobot berdasarkan nilai precision dan recall. Untuk menghitungnya, menggunakan rumus berikut[7].

Rumus 2.10 menunjukkan cara perhitungan *F1-Score*

$$F1-Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (2.10)$$

2.9 Text-preprocessing

Proses awal dalam mengolah data mentah agar dapat digunakan oleh model atau analisis lanjutan dikenal sebagai preprocessing[28]. Preprocessing merupakan tahap krusial dalam text mining, pemrosesan bahasa alami, dan pengambilan informasi. Data preprocessing digunakan oleh text mining untuk memperoleh wawasan yang menarik dan signifikan dari teks yang tidak terstruktur[29]. Text preprocessing terdiri dari beberapa proses, yang dikelompokkan sebagai berikut[30].

1. Cleaning Data

Dalam proses ini, username, hashtag, simbol, emotikon, tautan, karakter tunggal, angka, ruang kosong, dan tanda baca yang tidak mengandung informasi emosional dikeluarkan dari data teks.

2. Case Folding

Tahapan casefolding ini yaitu mengubah dari huruf besar atau kapital menjadi huruf kecil.

3. Tokenizing

Tahap tokenizing ini proses memisahkan teks menjadi beberapa kata.

4. Normalisasi

Pada tahap ini dilakukan dengan membuat sebuah file csv yang isinya dari kata yang salah dalam penulisan serta dengan perbaikannya.

5. Stopwords Removal

Dalam tahap stopwords removal, kata-kata yang tidak mempunyai makna khusus, seperti penghubung, dihapuskan.

6. Stemming

Dalam proses stemming, pustaka sastrawi digunakan untuk menghapus atau

mengambil kata-kata yang memiliki afiks hingga bentuk dasarnya dapat dihasilkan.

2.10 TextBlob

TextBlob merupakan pustaka Python yang bisa dimanfaatkan untuk mengolah data teks. ini adalah kerangka kerja sumber terbuka dengan API yang konsisten yang dapat digunakan untuk melakukan berbagai tugas proses pengolahan bahasa natural, termasuk ekstraksi frasa kata benda, penandaan bagian ucapan, analisis sentimen, klasifikasi, dan terjemahan. Ini menilai subjektivitas dan polaritas perasaan. Polaritas memilih apakah sentimen dapat positif, negatif, atau netral, angka floating point antara $[-1, 1]$, dimana $+1$ menunjukkan sentimen yang positif, dan -1 menunjukkan sentimen yang negatif. Subjektivitas menentukan apakah pernyataan tersebut bergantung pada keyakinan, opini, atau asumsi serta apakah emosi dan perasaan pribadi memengaruhinya. Angka ini adalah titik yang bergerak dalam kisaran $[0, 1]$ [31]. TextBlob akan mempertimbangkan kata-kata dan ekspresi yang dapat digunakan untuk menerapkan titik ekstrem dan titik tengah, mengabaikan istilah yang tidak dikenal[32].

