



Sentiment Analysis of PLN Mobile Application Review Using Naïve Bayes Classifier and K-Nearest Neighbor Algorithm

Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor

Syafrizal^{1*}, M. Afdal², Rice Novita³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indoneisa

E-Mail: ¹11950315040@students.uin-suska.ac.id,
²m.afdal@uin-suska.ac.id, ³rice.novita@uin-suska.ac.id,

Received Sep 6th 2023; Revised Nov 10th 2023; Accepted Dec 10th 2023
Corresponding Author: Syafrizal

Abstract

The real proof that PLN continues to improve its services is by launching an application, namely PLN Mobile. Many customers feel the convenience of the application. But now, some customers are encountering problems such as failing to load locations when making complaints and purchasing tokens with virtual accounts. The balance has been deducted, but the token code does not appear. This research conducts sentiment analysis of PLN Mobile application user reviews using a text mining approach. This approach can perform sentiment classification on user reviews quickly. Data was collected using scrapping techniques on the Google Play Store, and obtained 3000 data lines. The data was then labelled by an expert, resulting in 2099 positive sentiments (69.97%), 368 neutral (12.27%) and 533 negatives (17.77%). Furthermore, modelling uses the NBC and KNN algorithms with K-Fold Cross Validation as a validation technique. The results show that the NBC model is better than KNN with an accuracy of 77.69%, recall of 53.14%, precision of 59.84% and F1-Score of 54.09%. Furthermore, the analysis process is carried out using word cloud data visualization. The result is that the PLN Mobile application provides convenience to customers in using PLN services such as token purchases, complaints, and other features. However, the PLN Mobile application still has several problems often reviewed by users, one of which is when making token payments.

Keyword: Classification, K-Nearest Neighbor, Naïve Bayes Classifier, PLN Mobile, Sentiment Analysis.

Abstrak

Bukti nyata PLN terus meningkatkan pelayanannya adalah dengan meluncurkan sebuah aplikasi yaitu PLN Mobile. Banyak pelanggan yang merasakan kemudahan dengan adanya aplikasi tersebut. Namun kini beberapa pelanggan mulai menjumpai permasalahan seperti gagal memuat lokasi saat melakukan pengaduan dan saat pembelian token dengan virtual account, saldo telah terpotong namun kode token tidak muncul. Penelitian ini melakukan analisis sentimen terhadap ulasan pengguna aplikasi PLN Mobile menggunakan pendekatan text mining. Pendekatan ini dapat melakukan klasifikasi sentimen pada ulasan pengguna dengan cepat. Data dikumpulkan menggunakan teknik scrapping pada Google Play Store dan mendapatkan 3000 baris data. Data tersebut kemudian diberi label oleh seorang pakar sehingga menghasilkan 2099 sentimen positif (69,97%), 368 netral (12,27%) dan 533 negatif (17,77%). Selanjutnya dilakukan pemodelan menggunakan algoritma NBC dan KNN dengan K-Fold Cross Validation sebagai teknik validasi. Hasilnya menunjukkan model NBC lebih baik dibandingkan KNN dengan akurasi sebesar 77,69%, recall 53,14%, precision 59,84% dan F1-Score 54,09%. Selanjutnya proses analisis dilakukan dengan visualisasi data menggunakan word cloud. Hasilnya yaitu dengan adanya aplikasi PLN Mobile memberikan kemudahan kepada pelanggan dalam menggunakan layanan PLN seperti pembelian token, pengaduan, dan berbagai fitur lainnya. Namun aplikasi PLN Mobile masih memiliki beberapa permasalahan yang sering menjadi ulasan penggunaannya salah satunya adalah saat melakukan pembayaran token.

Kata Kunci: Analisis Sentimen, Klasifikasi, K-Nearest Neighbor, Naïve Bayes Classifier, PLN Mobile

1. PENDAHULUAN

Perusahaan Listrik Negara atau dikenal dengan PT. PLN (Persero) merupakan perusahaan yang bergerak dibidang jasa penyedia listrik di Indonesia dan termasuk Badan Usaha Milik Negara (BUMN) [1]. Dalam pelayanan dan pendistribusian listrik, PLN membagi fungsi unit induknya menjadi tiga, yaitu pembangkitan, transmisi, dan distribusi [2]. Karena luasnya cakupan wilayah kerja PLN, maka dari itu PLN memiliki berbagai unit yang tersebar Indonesia dan bertanggung jawab pada setiap fungsi tertentu sesuai dengan unit induknya [3]. Pelayanan informasi yang diberikan PLN sangat penting karena listrik yang padam dapat menimbulkan kerugian di berbagai bidang, seperti ekonomi, kesehatan, pendidikan, dan transportasi [4].

Bukti nyata upaya PLN terus meningkatkan pelayanan adalah PLN meluncurkan sebuah aplikasi yaitu PLN mobile yang tersedia bagi pelanggan melalui smartphone yang berbasis Android dan iOS. Aplikasi PLN Mobile pertama kali diluncurkan oleh Direksi PLN pada tahun 2016 dan diluncurkan kembali pada tahun 2020 dengan fitur-fitur baru. Sedangkan untuk Unit PLN ULP Panam (Pekanbaru) Aplikasi PLN Mobile mulai di terapkan pada tahun 2018 memiliki fitur-fitur yang masih sederhana kemudian melakukan inovasi dengan menambahkan berbagai fitur. Jumlah pelanggan PLN ULP Panam per juni 2023 sebanyak 125.375 dengan downloader 41.328. Aplikasi ini merupakan suatu produk hasil inovasi PT. PLN yang bekerjasama dengan PT Indonesia Comnet Plus, anak perusahaan PT. PLN [5].

Aplikasi PLN Mobile adalah aplikasi mobile yang terintegrasi dengan dua aplikasi lain, yaitu Aplikasi Pengaduan Terpadu dan Aplikasi Pelayanan Pelanggan Terpusat. Aplikasi ini digunakan untuk menyampaikan pengaduan secara terintegrasi. Pelanggan dapat dengan mudah mendapatkan informasi tentang fitur-fitur aplikasi, seperti pembayaran tagihan listrik dan pembelian token, pencatatan angka meter mandiri, perubahan atau penambahan daya, pengaduan gangguan dan keluhan, pemantauan pembelian token, pemantauan pemakaian listrik pascabayar, notifikasi tagihan, notifikasi pemadaman listrik, informasi perkembangan penyelesaian gangguan, hingga pemeliharaan jaringan listrik. Aplikasi ini mudah dan gratis untuk diunduh dari Google Play Store untuk smartphone berbasis Android [6].

Seiring berjalannya waktu banyak pelanggan yang melaporkan keluhan dan gangguan via PLN Mobile. Diantara para pelanggan tersebut masih tidak puas terhadap pelayanan PLN di aplikasi PLN Mobile seperti, terjadi kesalahan saat mengajukan pengaduan “gagal memuat lokasi”, tidak adanya notifikasi pemadaman listrik, kilowatt hour (kWh) sering muncul code “Periksa”, lamanya proses pengaduan dan yang paling sering terjadi adalah pembelian token dengan virtual account saldo terpotong namun kode token tidak muncul sehingga pelanggan memberikan komentar pengaduannya ke pihak PLN melalui Aplikasi PLN mobile pada platform Google Play Store. Peringkat ulasan di Google Play Store tidak selalu menggambarkan kualitas aplikasi dengan akurat, karena pengguna sering memberikan peringkat yang tidak sesuai dengan ulasan mereka. Selain itu, membaca semua ulasan aplikasi PLN Mobile akan memakan waktu dan tenaga. Oleh karena itu, sistem klasifikasi digunakan untuk mengukur sentimen publik terhadap aplikasi PLN Mobile.

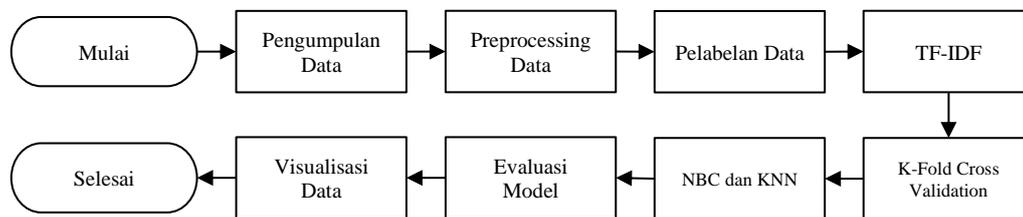
NBC merupakan algoritma klasifikasi sederhana berdasarkan pada penerapan teorema bayes dengan asumsi independensi [7]. Algoritma ini banyak diterapkan pada beberapa kasus seperti analisis sentimen, klasifikasi teks, filter spam, dan lainnya [8]. Sedangkan KNN merupakan salah satu algoritma yang mengklasifikasikan objek berdasarkan data pembelajaran yang paling dekat dengan objek tersebut [8]. Algoritma KNN sangat sederhana namun cukup efektif dalam mengkategorikan teks [9]. Kedua algoritma tersebut telah banyak diimplementasikan dan tentunya memiliki kelebihan maupun kekurangannya masing-masing, sehingga perlu dilakukan pengujian pada dataset yang akan dianalisis dan dipelajari untuk melihat algoritma mana yang tepat pada dataset tersebut.

Terdapat beberapa penelitian sebelumnya yang telah menggunakan algoritma NBC dan KNN pada kasus sentimen ulasan aplikasi pada Google Play Store. Penelitian [10] melakukan analisis sentimen ulasan aplikasi PLN Mobile menggunakan algoritma NBC dan pelabelan otomatis dengan *Lexicon Vader* menghasilkan akurasi model sebesar 70%. Selanjutnya penelitian [11] melakukan klasifikasi ulasan aplikasi Mobile JKN menggunakan KNN dan NBC menghasilkan model KNN sebagai yang terbaik dengan akurasi sebesar 87,59%. Kemudian penelitian [12] menggunakan algoritma NBC dan KNN untuk analisis sentimen ulasan aplikasi Tokopedia menghasilkan model KNN sebagai yang terbaik dengan akurasi sebesar 86,09%. Penelitian lainnya [13] melakukan analisis sentimen pada aplikasi Soptify menggunakan algoritma SVM dan NBC menghasilkan model NBC sebagai yang terbaik dengan akurasi sebesar 86,4%.

Penelitian ini bertujuan untuk melakukan analisis sentimen pada ulasan aplikasi PLN Mobile menggunakan algoritma NBC dan KNN. Melalui penelitian ini PT. PLN akan dengan mudah mengetahui tingkat kepuasan pengguna terhadap aplikasi PLN Mobile. Perbedaan penelitian ini dengan penelitian sebelumnya adalah pada penelitian ini melakukan perbandingan algoritma antara NBC dan KNN. Selain itu pelabelan pada penelitian ini dilakukan oleh seorang pakar bahasa sehingga hasil yang didapatkan lebih akurat dan sesuai dengan konteksnya. Hasil penelitian ini diharapkan dapat membantu PT. PLN dalam mengevaluasi dan mengembangkan layanan aplikasi mereka lebih jauh dengan strategi bisnis profesional yang berfokus pada kepuasan pengguna.

2. METODOLOGI PENELITIAN

Penelitian ini melakukan analisis sentiment pada aplikasi PLN Mobile dengan algoritma NBC dan KNN. Setiap tahapan pada penelitian ini disajikan melalui diagram alir agar penelitian dapat berjalan sesuai tujuan yang akan dicapai. Tahapan penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data yang diambil adalah data ulasan pengguna dari Google Play Store pada aplikasi PLN Mobile. Pengambilan data dilakukan dengan menggunakan teknik *scrapping* menggunakan bahasa pemrograman *Python*. Rentang waktu pengambilan data adalah selama satu tahun terakhir, mulai dari Juli 2022 hingga Juli 2023. Hasilnya, jumlah data yang didapatkan adalah sebanyak 3000 baris data.

2.2 Preprocessing Data

Tahapan *preprocessing* dilakukan untuk mempersiapkan kumpulan data mentah yaitu data teks agar dapat diolah lebih lanjut [9]. Terdapat beberapa tahapan dari *preprocessing* data teks, yaitu [14]:

1. *Cleaning*
Pada tahap ini dilakukan pembersihan data teks dari kesalahan, inkonsistensi, dan tidak relevan meliputi menghapus karakter khusus seperti tanda baca, simbol, dan angka serta mengubah format penulisan menjadi huruf kecil.
2. *Tokenizing*
Di tahap ini setiap data teks akan dipecah menjadi unit-unit yang lebih kecil, yang disebut token. Token tersebut dapat berupa kata, frasa, atau kalimat.
3. *Filtering*
Tahapan ini akan menghilangkan kata yang tidak memiliki arti penting seperti kata hubung, kata ganti orang, dan lainnya.
4. *Stemming*
Stemming merupakan tahapan terakhir dari *preprocessing*. Tahapan ini akan mengekstrak kata menjadi bentuk dasar atau kata dasar dengan cara menghilangkan imbuhan atau akhiran dari kata.

2.3 Pelabelan

Dataset yang telah didapatkan melalui proses *scrapping* masih bertipe *unsupervised* yang berarti data tersebut belum memiliki label atau kategori. Maka dari itu data tersebut perlu dilakukan pelabelan terlebih dahulu sebelum melakukan pemodelan. Pada penelitian ini, proses pelabelan data dilakukan secara manual oleh seorang pakar bidang Bahasa Indonesia untuk memastikan bahwa setiap data memiliki label yang tepat sesuai dengan kontennya.

2.4 TF-IDF

Term Frequency dan *Inverse Document Frequency* (TF-IDF) adalah ukuran statistik yang menggambarkan pentingnya suatu istilah terhadap sebuah dokumen dalam sebuah korpus. *Term Frequency* (TF) adalah frekuensi kemunculan suatu istilah dalam suatu dokumen. Semakin sering suatu istilah muncul, maka semakin besar nilai TF-nya. Sementara *Inverse Document Frequency* (IDF) adalah ukuran kelangkaan suatu istilah dalam sebuah kumpulan dokumen. Semakin jarang suatu istilah muncul, maka akan semakin besar nilai IDF-nya [11].

2.5 K-Fold Cross Validation

K-Fold Cross Validation merupakan teknik validasi yang membagi data menjadi K bagian yang sama, kemudian model akan dilatih dan dievaluasi sebanyak K kali [7]. Teknik ini digunakan karena dinilai mampu memperoleh hasil yang maksimal saat melakukan pemodelan [15]. Pada penelitian ini nilai K-Fold yang ditetapkan adalah 10 yang artinya data akan dipisahkan menjadi 10 bagian dimana 1 bagian akan menjadi data uji dan 9 lainnya menjadi data latih. Data *training* digunakan untuk melatih algoritma mempelajari pola dari data sedangkan data *testing* digunakan untuk menguji seberapa baik algoritma yang telah dilatih [16].

2.6 Naïve Bayes Classifier

NBC merupakan algoritma klasifikasi yang sederhana dengan konsep penerapan teorema bayes dengan asumsi independensi [7]. Algoritma ini dapat digunakan pada data numerik maupun kategorikal [17]. Selain itu NBC dapat menangani dataset yang berukuran besar dan data noise [18]. NBC banyak diterapkan pada beberapa kasus seperti analisis sentimen, klasifikasi teks, filter spam, dan lainnya [8]. Berikut adalah persamaan dari algoritma NBC.

$$P(X|H) = P(H|X) P(H) / P(X) \tag{1}$$

2.7 K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan salah satu algoritma yang mengklasifikasikan suatu objek berdasarkan kumpulan data pembelajaran yang terdekat dengan objek tersebut [8]. Algoritma KNN sangat sederhana namun cukup efektif dalam mengkategorikan teks [9]. Namun KNN memiliki kekurangan yaitu sangat bergantung pada parameter K dan memerlukan komputasi yang tinggi [19]. Berikut adalah persamaan dari algoritma KNN.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n ((a_r(x_i) - a_r(x_j)))^2} \tag{2}$$

2.8 Evaluasi Model

Untuk mengetahui kemampuan model melakukan klasifikasi maka dilakukan evaluasi. Proses evaluasi model dalam kasus klasifikasi melibatkan penggunaan *confusion matrix* untuk dapat mengukur akurasi, *precision*, dan *recall* pada model [16]. Berikut adalah persamaan yang dapat digunakan.

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \tag{3}$$

$$Precision = (TP) / (TP + FP) \tag{4}$$

$$Recall = (TP) / (TP + FN) \tag{5}$$

$$F1\ Score = 2 * (Recall * Precision) / (Recall + Precision) \tag{6}$$

2.9 Visualisasi Data

Visualisasi data merupakan tahapan terakhir dari penelitian ini. Tujuan melakukan visualisasi data adalah untuk menampilkan beberapa kata yang sering muncul berdasarkan sentimennya. Hal ini tentu akan memudahkan saat melakukan analisis terhadap opini yang disampaikan oleh pengguna aplikasi PLN Mobile tanpa perlu membaca semua ulasan pengguna satu-persatu.

3. HASIL DAN ANALISIS

Pengambilan data dilakukan dengan menggunakan teknik scrapping menggunakan bahasa pemrograman Python dengan rentang waktu selama satu tahun terakhir, mulai dari Juli 2022 hingga Juli 2023. Hasilnya, jumlah data yang didapatkan adalah sebanyak 3000 baris data. Data yang didapatkan melalui *scrapping* dapat dilihat pada Tabel 1.

Tabel 1. Pengumpulan Data

No	Komentar
1	Aplikasi nya sangat membantu seperti kalau mau tambah daya jadi mudah, dan kalau ada pemeliharaan ada notifikasi pemberitahuan selain itu ada notifikasi pembayaran tagihan dan pembelian pulsa jadi ga lupa. Mantapppp PLN
2	Maaf ya masih 4 rating, menurut pengalaman saya memakai aplikasi ini sudah cukup bagus hanya saja kurang Adanya fitur memasukkan id pelanggan/no meteran menggunakan scan barcode. Padahal di setiap meteran listrik sudah tersedia barcode. Justru pengalaman Saya memakai fitur scan ini sudah ada di aplikasi Sh*p*e. Jadi sudah gaperlu repot masukkan angka lagi. Mana tulisannya kecil2, jadi harapan saya supaya kedepan fitur tsb segera terealisasi
...	...
2999	sdh 2 hari,waktu pengajuan tambah daya & pasang baru ,pada kolom " lanjutkan" setelah verifikasi NIK ktp muncul tulisan "data error",tidak bisa lanjut ke menu selanjutnya,mohon segera diperbaiki,terima kasih
3000	Aplikasinya mantap sukses trus.

Data yang telah terkumpul kemudian dilakukan *preprocessing*. Tahapan awal dari *preprocessing* adalah melakukan *cleaning* data dengan mengubah format teks menjadi *lowercase* (huruf kecil), menghapus

karakter selain huruf seperti simbol, hyperlink, tanda baca, angka dan lainnya. Data yang telah melalui tahap cleaning dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Cleaning* Data

No	Komentar
1	aplikasi nya sangat membantu seperti kalau mau tambah daya jadi mudah dan kalau ada pemeliharaan ada notifikasi pemberitahuan selain itu ada notifikasi pembayaran tagihan dan pembelian pulsa jadi ga lupa mantapppp pln
2	maaf ya masih rating menurut pengalaman saya memakai aplikasi ini sudah cukup bagus hanya saja kurang adanya fitur memasukkan id pelanggan no meteran menggunakan scan barcode padahal di setiap meteran listrik sudah tersedia barcode justru pengalaman saya memakai fitur scan ini sudah ada di aplikasi sh p e jadi sudah gaperlu repot masukkan angka lagi mana tulisannya kecil jadi harapan saya supaya kedepan fitur tsb segera terealisasi
...	...
2999	sdh hari waktu pengajuan tambah daya pasang baru pada kolom lanjutkan setelah verifikasi nik ktp muncul tulisan data error tidak bisa lanjut ke menu selanjutnya mohon segera diperbaiki terima kasih
3000	aplikasinya mantap sukses trus

Proses berikutnya adalah *tokenizing*, yaitu memecah kalimat menjadi penggalan kata atau disebut dengan token. Hasil dari tahapan *tokenizing* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Tokenizing*

No	Komentar
1	['aplikasi', 'nya', 'sangat', 'membantu', 'seperti', 'kalau', 'mau', 'tambah', 'daya', 'jadi', 'mudah', 'dan', 'kalau', 'ada', 'pemeliharaan', 'ada', 'notifikasi', 'pemberitahuan', 'selain', 'itu', 'ada', 'notifikasi', 'pembayaran', 'tagihan', 'dan', 'pembelian', 'pulsa', 'jadi', 'ga', 'lupa', 'mantapppp', 'pln']
2	['maaf', 'ya', 'masih', 'rating', 'menurut', 'pengalaman', 'saya', 'memakai', 'aplikasi', 'ini', 'sudah', 'cukup', 'bagus', 'hanya', 'saja', 'kurang', 'adanya', 'fitur', 'memasukkan', 'id', 'pelanggan', 'no', 'meteran', 'menggunakan', 'scan', 'barcode', 'padahal', 'di', 'setiap', 'meteran', 'listrik', 'sudah', 'tersedia', 'barcode', 'justru', 'pengalaman', 'saya', 'memakai', 'fitur', 'scan', 'ini', 'sudah', 'ada', 'di', 'aplikasi', 'sh', 'p', 'e', 'jadi', 'sudah', 'gaperlu', 'repot', 'masukkan', 'angka', 'lagi', 'mana', 'tulisannya', 'kecil', 'jadi', 'harapan', 'saya', 'supaya', 'kedepan', 'fitur', 'tsb', 'segera', 'teralisasi']
...	...
2999	['sdh', 'hari', 'waktu', 'pengajuan', 'tambah', 'daya', 'pasang', 'baru', 'pada', 'kolom', 'lanjutkan', 'setelah', 'verifikasi', 'nik', 'ktp', 'muncul', 'tulisan', 'data', 'error', 'tidak', 'bisa', 'lanjut', 'ke', 'menu', 'selanjutnya', 'mohon', 'segera', 'diperbaiki', 'terima', 'kasih']
3000	['aplikasinya', 'mantap', 'sukses', 'trus']

Setelah melakukan *tokenizing*, tahapan selanjutnya adalah *filtering* yaitu membuang kata yang tidak memiliki makna dan informasi dari teks seperti kata hubung atau lainnya. Pada implementasinya, tahapan *filtering* dilakukan dengan menggunakan bantuan pustaka *Natural Language Toolkit* (NLTK). Hasil dari tahapan *filtering* dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Filtering*

No	Komentar
1	['aplikasi', 'nya', 'membantu', 'daya', 'mudah', 'pemeliharaan', 'notifikasi', 'pemberitahuan', 'notifikasi', 'pembayaran', 'tagihan', 'pembelian', 'pulsa', 'ga', 'lupa', 'mantapppp', 'pln']
2	['maaf', 'ya', 'rating', 'pengalaman', 'memakai', 'aplikasi', 'bagus', 'fitur', 'memasukkan', 'id', 'pelanggan', 'no', 'meteran', 'scan', 'barcode', 'meteran', 'listrik', 'tersedia', 'barcode', 'pengalaman', 'memakai', 'fitur', 'scan', 'aplikasi', 'sh', 'p', 'e', 'gaperlu', 'repot', 'masukkan', 'angka', 'tulisannya', 'harapan', 'kedepan', 'fitur', 'tsb', 'teralisasi']
...	...
2999	['sdh', 'pengajuan', 'daya', 'pasang', 'kolom', 'lanjutkan', 'verifikasi', 'nik', 'ktp', 'muncul', 'tulisan', 'data', 'error', 'menu', 'mohon', 'diperbaiki', 'terima', 'kasih']
3000	['aplikasinya', 'mantap', 'sukses', 'trus']

Stemming merupakan proses terakhir dari *preprocessing*. Tahap ini akan mengubah suatu kata menjadi kata dasar dengan cara menghilangkan imbuhan atau akhiran dari kata. Pada implementasinya, proses *stemming* dilakukan menggunakan bantuan dari pustaka *Sastrawi*. Hasil dari tahapan *preprocessing* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Stemming*

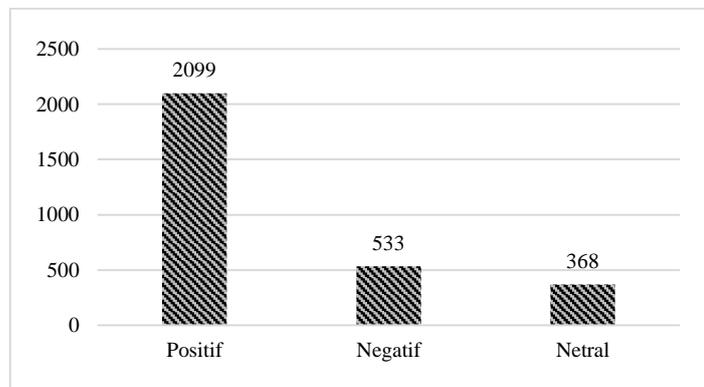
No	Komentar
1	['aplikasi', 'nya', 'bantu', 'daya', 'mudah', 'pelihara', 'notifikasi', 'pemberitahuan', 'notifikasi', 'bayar', 'tagih', 'beli', 'pulsa', 'ga', 'lupa', 'mantapppp', 'pln']
2	['maaf', 'ya', 'rating', 'alam', 'pakai', 'aplikasi', 'bagus', 'fitur', 'masuk', 'id', 'langgan', 'no', 'meter', 'scan', 'barcode', 'meter', 'listrik', 'sedia', 'barcode', 'alam', 'pakai', 'fitur', 'scan', 'aplikasi', 'sh', 'p', 'e', 'gaperlu', 'repot', 'masuk', 'angka', 'tulis', 'harap', 'depan', 'fitur', 'tsb', 'realisasi']
...	...
2999	['sdh', 'aju', 'daya', 'pasang', 'kolom', 'lanjut', 'verifikasi', 'nik', 'ktp', 'muncul', 'tulis', 'data', 'error', 'menu', 'mohon', 'baik', 'terima', 'kasih']
3000	['aplikasi', 'mantap', 'sukses', 'trus']

Data yang telah dilakukan *preprocessing* kemudian akan diberi label. Untuk mendapatkan label yang akurat dan sesuai dengan konteks maka pada penelitian ini proses pelabelan dilakukan oleh seorang pakar yang memiliki pengalaman di bidang bahasa Indonesia. Label yang diberikan pada penelitian ini terdiri dari tiga kelompok sentimen yaitu positif, netral dan negatif. Hasil pelabelan dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pelabelan

No	Komentar	Sentimen
1	aplikasi nya bantu daya mudah pelihara notifikasi pemberitahuan notifikasi bayar tagih beli pulsa ga lupa mantapppp pln	Positif
2	maaf ya rating alam pakai aplikasi bagus fitur masuk id langgan no meter scan barcode meter listrik sedia barcode alam pakai fitur scan aplikasi sh p e gaperlu repot masuk angka tulis harap depan fitur tsb realisasi	Netral
...
2999	sdh aju daya pasang kolom lanjut verifikasi nik ktp muncul tulis data error menu mohon baik terima kasih	Netral
3000	aplikasi mantap sukses trus	Positif

Hasil pelabelan yang dilakukan oleh seorang pakar mendapatkan total 2099 sentimen positif (69,97%), 368 sentimen netral (12,27%) dan 533 sentimen negatif (17,77%) dari total 3000 baris data. Ini mengindikasikan bahwa aplikasi PLN Mobile memiliki layanan aplikasi yang memuaskan pengguna. Visualisasi dari hasil pelabelan data dapat dilihat pada Gambar 2.



Gambar 2. Sentimen Aplikasi PLN Mobile Berdasarkan Hasil Pelabelan

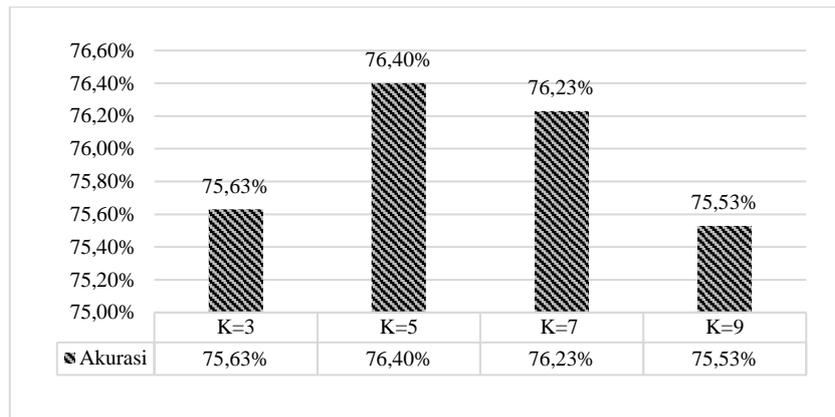
Tahapan selanjutnya adalah melakukan TF-IDF untuk mendapatkan nilai atau bobot dari suatu kata. Semakin sering suatu term muncul, maka akan semakin memengaruhi nilai bobot. Perhitungan TF-IDF dilakukan menggunakan Python dengan bantuan pustaka dari *Scikit-Learn*. Hasil dari tahapan TF-IDF dapat dilihat pada Tabel 7.

Tabel 7. Sampel Hasil Pembobotan dengan TF-IDF

No	aplikasi	listrik	mobile	pln	periksa	notifikasi	meter	pengaduan
1	0.0879	0	0	0,9491	0.2186	0	0,5527	0
2	0.8971	0,6364	0	0	0	0.7251	...	0
...
2999	0	0	0	0	0	0	0	0

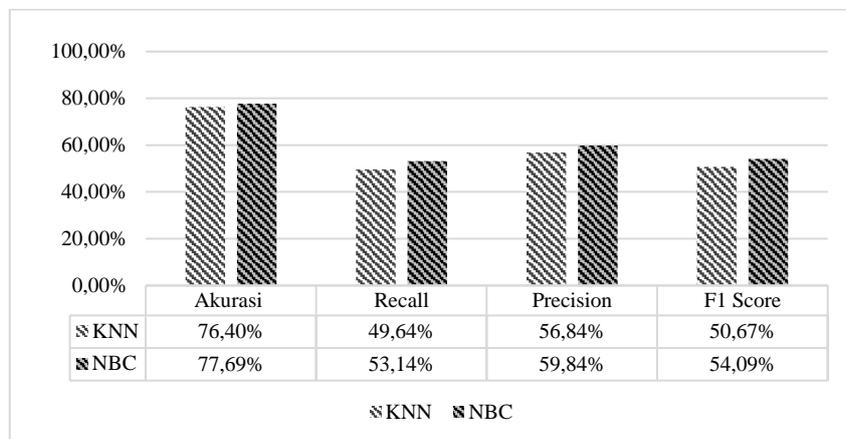
No	aplikasi	listrik	mobile	pln	periksa	notifikasi	meter	pengaduan
3000	0	0	0	0	0	0	0	0

Pemodelan sentimen aplikasi PLN mobile dilakukan menggunakan dua algoritma yaitu NBC dan KNN. Selain itu proses validasi dilakukan menggunakan teknik *K-Fold Cross Validation* dengan 10-Fold. Tahap pemodelan pertama kali dilakukan menggunakan algoritma NBC. Pemodelan dengan algoritma NBC menghasilkan nilai akurasi sebesar 77,69%. Selanjutnya algoritma KNN digunakan pada pemodelan dengan melakukan pengujian parameter K yaitu 3, 5, 7 dan 9. Pemodelan algoritma KNN dengan pengujian parameter K menghasilkan K = 5 sebagai yang terbaik dengan nilai akurasi sebesar 76,40%, sedangkan akurasi terendah terdapat pada parameter k = 9 dengan nilai akurasi sebesar 75,53%. Hasil pemodelan KNN dengan pengujian parameter K dapat dilihat pada Gambar 3.



Gambar 3. Hasil Pemodelan dengan Algoritma KNN

Hasil pemodelan dari kedua algoritma tersebut kemudian akan dibandingkan untuk mendapatkan model terbaik berdasarkan nilai akurasi, *precision*, *recall*, dan *F1-Score*. Hasil perbandingan model NBC dan KNN dapat dilihat pada Gambar 4.



Gambar 4. Perbandingan Algoritma NBC dan KNN

Berdasarkan hasil perbandingan model yang dilakukan didapatkan bahwa algoritma NBC lebih baik dibandingkan KNN, dengan tingkat akurasi sebesar 77,69%, *recall* 53,14%, *precision* 59,84% dan *F1-Score* 54,09%. Sedangkan pada model KNN hanya memiliki nilai akurasi sebesar 76,4%, *recall* 49,64%, *precision* 56,84% dan *F1-Score* 50,67%.

Melalui hasil pemodelan tersebut, algoritma NBC dapat dengan baik melakukan klasifikasi sentimen dibandingkan KNN. Meskipun algoritma ini sederhana namun memberikan kinerja yang lebih efisien terutama dalam hal komputasi. Kemudian algoritma KNN dapat dengan cukup baik menangani data teks meskipun tidak sebaik algoritma NBC. Namun hasil yang didapatkan dari algoritma KNN sangat bergantung pada nilai K sehingga perlu dilakukan serangkaian pengujian yang tentunya membutuhkan waktu komputasi yang lebih lama dibandingkan NBC untuk mendapatkan hasil yang optimal.

Selanjutnya, untuk mengetahui opini yang disampaikan oleh pengguna aplikasi PLN Mobile, maka perlu dilakukan tahapan visualisasi data. Tahap visualisasi ini dilakukan menggunakan *word cloud* yang akan

menampilkan kata yang sering muncul pada sentimen positif, netral, dan negatif. Pada sentiment positif terdapat tiga kata dengan frekuensi tertinggi yaitu “aplikasi”, “mudah”, dan “bantu”. Ini mengindikasikan bahwa pelanggan berpendapat dengan adanya PLN Mobile sangat membantu mereka dengan fitur-fitur yang ada di aplikasi PLN Mobile. Hasil visualisasi sentimen positif pada aplikasi PLN Mobile dapat dilihat pada Gambar 5.



Gambar 5. Visualisasi Kata Positif

Selanjutnya pada sentimen netral terdapat tiga kata dengan frekuensi tertinggi yaitu “aplikasi”, “bayar”, “token”. Ini mengindikasikan bahwa pelanggan sedikit mengalami kendala dalam melakukan pembayaran token melalui aplikasi dan juga memberi saran dengan adanya kata “mohon” dan beberapa kata positif lainnya. Hasil visualisasi sentimen netral pada aplikasi PLN Mobile dapat dilihat pada Gambar 6.



Gambar 6. Visualisasi Kata Netral

Kemudian pada sentimen negatif terdapat tiga kata dengan frekuensi tertinggi yaitu “aplikasi”, “bayar”, dan “token”. Kata tersebut mirip dengan yang ada pada sentimen netral yaitu pelanggan mengalami kendala dalam melakukan pembayaran token melalui aplikasi. Namun bedanya dengan sentimen netral, di sentimen negatif diiringi dengan kata “susah”, “gagal”, “lambat”, “kecewa”, dan “ganggu” yang mengindikasikan pelanggan mulai kurang puas dan meluapkan kekecewaannya terhadap layanan dari aplikasi PLN Mobile. Hasil visualisasi sentiment negatif pada aplikasi PLN Mobile dapat dilihat pada Gambar 7.

Melalui hasil analisis dari visualisasi tersebut dapat diketahui adanya aplikasi PLN Mobile memberikan kemudahan kepada pelanggan dalam menggunakan layanan PLN seperti pembelian token, pengaduan, dan berbagai fitur lainnya. Namun aplikasi PLN Mobile masih memiliki beberapa permasalahan dalam melakukan pembayaran token. Untuk itu pihak PLN diharapkan dapat melakukan evaluasi terkait fitur yang masih belum maksimal dan mengembangkan layanan yang baik lagi dan berfokus pada kepuasan pengguna.

- [8] H. Wisnu, M. Afif, and Y. Ruldevyani, "Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes," *Journal of Physics: Conference Series*, vol. 1444, no. 1, p. 12034, 2020, doi: 10.1088/1742-6596/1444/1/012034.
- [9] S. G. Setyorini and Mustakim, "Application of the nearest neighbor algorithm for classification of online taxibike sentiments in indonesia in the google playstore application," *Journal of Physics: Conference Series*, vol. 2049, no. 1, p. 12026, 2021, doi: 10.1088/1742-6596/2049/1/012026.
- [10] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [11] C. Annisa, M. Afdal, and T. K. Ahsyar, "Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest Neighbor Pada Sentimen Review Aplikasi Mobile Jkn," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1033–1040, 2023, doi: 10.30865/mib.v7i3.6242.
- [12] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p. 1329, 2022, doi: 10.30865/jurikom.v9i5.4774.
- [13] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [14] A. R. Atmadja, W. Uriawan, F. Pritisen, D. S. Maylawati, and A. Arbain, "Comparison of Naive Bayes and K-nearest neighbours for online transportation using sentiment analysis in social media," *Journal of Physics: Conference Series*, vol. 1402, no. 7, 2019, doi: 10.1088/1742-6596/1402/7/077029.
- [15] P. F. Pratama, D. Rahmadani, R. S. Nahampun, D. Harmutika, A. Rahmadeyan, and M. F. Evizal, "Random Forest Optimization Using Particle Swarm Optimization for Diabetes Classification," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 1, no. 1, pp. 41–46, 2023.
- [16] A. Rahmadeyan, Mustakim, I. Ahmad, A. D. Alexander, and A. Rahman, "Phishing Website Detection with Ensemble Learning Approach Using Artificial Neural Network and AdaBoost," in *2023 International Conference on Information Technology Research and Innovation (ICITRI)*, 2023, pp. 162–166, doi: 10.1109/ICITRI59340.2023.10249799.
- [17] P. Kamal and S. Ahuja, "An ensemble-based model for prediction of academic performance of students in undergrad professional course," *Journal of Engineering, Design and Technology*, vol. 17, no. 4, pp. 769–781, 2019, doi: 10.1108/JEDT-11-2018-0204.
- [18] A. W. Syaputri, E. Irwandi, and M. Mustakim, "Naïve Bayes Algorithm for Classification of Student Major's Specialization," *Journal of Intelligent Computing & Health Informatics*, vol. 1, no. 1, p. 17, 2020, doi: 10.26714/jjichi.v1i1.5570.
- [19] A. Putri, C. Syaficha Hardiana, E. Novfuja, F. Try Puspa Siregar, Y. Fatma, and R. Wahyuni, "Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.