JIM: Jurnal Ilmu Multidisiplin
E-ISSN: 2829-4580
P-ISSN: 2829-4599
GREENATION PUBLISHER
+62 812 1046 7572 http://greenpub.org greenation.info@gmail.com
DOI: https://doi.org/10.38035/jim.v4i4

Klasifikasi Sentimen untuk Prediksi Churn Pengguna Aplikasi Mamikos Menggunakan Support Vector Machine (SVM) dan Naïve Bayes

Fitri Nur Utami[1], Hani Dewi Ariessanti[2], Riya Widayanti[3], Arief ichwani[4]
[1]Universitas Esa Unggul, Jakarta, Indonesia, fu75375@student.esaunggul.ac.id
[2]Universitas Esa Unggul, Jakarta, Indonesia, haniariessanti@esaunggul.ac.id
[3]Universitas Esa Unggul, Jakarta, Indonesia, riya.widayanti@esaunggul.ac.id
[4]Universitas Esa Unggul, Jakarta, Indonesia

Corresponding Author: fu75375@student.esaunggul.ac.id

Abstract: *The high rate of user churn or cessation from using the Mamikos application is a serious challenge for the company. This study aims to analyze sentiment and predict the likelihood of user churn using the Support Vector Machine (SVM) and Naïve Bayes algorithms. This method was chosen because SVM excels in handling high-dimensional data, while Naïve Bayes is effective in probability-based classification. The dataset was obtained from the Mamikos company through an independent study program, then went through the stages of preprocessing, feature extraction, labeling, and modeling. Evaluation was carried out using a confusion matrix to measure accuracy, precision, recall, and f1-score. The results show that the combination of the two algorithms provides fairly good accuracy in predicting churn based on user reviews. This research contributes to the development of machine learning-based customer retention strategies.*

Keyword: *Churn, sentiment analysis, Mamikos, Support Vector Machine, naïve Bayes, machine learning*

Abstrak: Tingginya tingkat churn atau berhentinya pengguna dalam menggunakan aplikasi Mamikos menjadi tantangan serius bagi perusahaan. Penelitian ini bertujuan untuk menganalisis sentimen dan memprediksi kemungkinan churn pengguna menggunakan algoritma Support Vector Machine (SVM) dan Naïve Bayes. Metode ini dipilih karena SVM unggul dalam menangani data berdimensi tinggi, sedangkan Naïve Bayes efektif dalam klasifikasi berbasis probabilitas. Dataset diperoleh dari perusahaan Mamikos melalui program studi independen, kemudian melalui tahapan preprocessing, ekstraksi fitur, pelabelan, dan pemodelan. Evaluasi dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan f1-score. Hasil menunjukkan bahwa kombinasi kedua algoritma memberikan akurasi yang cukup baik dalam memprediksi churn berdasarkan ulasan pengguna. Penelitian ini memberikan kontribusi dalam pengembangan strategi retensi pelanggan berbasis machine learning.

Kata Kunci: Churn, analisis sentiment, Mamikos, Support Vector Machine, naïve Bayes, machine learning

PENDAHULUAN

Aplikasi penyedia hunian Seperti Mamikos menjadi andalan masyarakat urban dalam mencari informasi tempat tinggal. Namun, mempertahankan pengguna aktif menjadi tantangan serius, terutama Ketika pengguna beralih ke platform lain. Fenomena ini disebut churn, yaitu keadaan dimana pengguna berhenti menggunakan layanan. Prediksi churn sangat penting untuk mempertahankan loyalitas pengguna dan meningkatkan strategi pemasaran berbasis data.

Industri aplikasi mobile di indonesia berkembang pesat, salah satunya ditunjukan oleh keberhasilan Mamikos sebagai platform pencarian kost terkemuka dengan lebih dari lima juta unduh(Pokhrel, 2024). Namun, Mamikos menghadapi tantangan churn, yaitu fenomena berhentinya pengguna dari penggunaan layananan, yang dapat berdampak pada penurunan pendapatan dan reputasi(Fikri et al., 2020).

Analisis churn yang dikombinasikan dengan analisis sentiment memberikan memberikan nilai strategis, karena ulasan dan umpan balik pengguna mencermikan kepuasan mereka terhadap layanan yang di terima(Damanik & Jambak, 2023). Penelitian sebelumnya telah menunjukan bahwa algoritma Support Vector Machine (SVM) berhasil mencapai akurasi 78,10% dalam mengklasifikasikan sentiment ulasan pengguna Mamikos. Berdasarkan hal tersebut, penelitiam ini mengusulkan kombinasi dua algoritma popular, Yaitu SVM dan Naïve Bayes, yang masing masing memiliki keunggulan dalam klasifikasi tekls dan pemodelan probabilistik(Lubis & Setyawan, 2024).

Integrasi dua pendekatan tersebut diharapkan dapat pendekatan tersebut diharapkan dapat meningkatkan akurasi dalam mendeteksi churn pengguna berdasarkan sentiment. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan kontribusi nyata terhadap pengembangan strategi bisnis Mamikos berbasis data. Dengan demikian, pendekatan analitik berbasis machine learning ini dapat dimanfaatkan untuk meningkatkan retensi pelanggan dan kualitas layanan aplikasi secara keseluruhan(Maulana et al., 2024).

Landasan Teori

Aplikasi Mamikos

Mamikos merupakan platform pencarian kost yang menghubungkan pemilik dan pencari property sewa secara langsung. Layanan ini menyediakan informasi lengkap mengenai kost, apartemen, dan kategori lainnya, serta dilengkapi dengan ulasan pengguna dan verifikasi lokasi melalui tim Mamichecker(Fikri et al., 2020). Dengan cakupan lebih dari dua juta kamar kos di lebih dari 140 kota, Mamikos menjadi sumber data yang potensial dalam penelitian ini untuk menganalisis pengalaman dan kepuasan pengguna(Jin et al., 2020).

Konsep Churn

Churn merupakan indikator penting dalam menilai keberhasilan layanan, di mana pengguna memutuskan berhenti menggunakan aplikasi dalam jangka waktu tertentu(Kulsum et al., 2022). Faktor penyebab churn dapat mencakup rendahnya kualitas layanan, tingginya persaingan pasar, kurangnya inovasi, serta pengalaman pengguna yang buruk(Muthmainnah & Voutama, 2023). Memahami penyebab churn memungkinkan pengembangan strategi retensi yang lebih efektif.

Analisis Sentimen

Analisis sentimen digunakan untuk mengevaluasi opini pengguna melalui teks ulasan, dengan tujuan mengukur polaritas sentimen apakah positif, negatif, atau netral(Pokhrel,

2024). enelitian ini menggunakan dataset dari Mamikos dan melakukan pelabelan sentimen melalui proses labeling dan koreksi oleh ahli bahasa untuk mendapatkan akurasi yang optimal dalam pengklasifikasian(Guswandri et al., 2022).

Algoritma Support Vector Machine (SVM)

SVM merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane terbaik yang memisahkan dua kelas dengan margin maksimum. SVM juga mampu menangani data non-linear melalui penggunaan fungsi kernel yang memproyeksikan data ke ruang berdimensi lebih tinggi(Ipmawati et al., 2024). Evaluasi model dilakukan dengan menggunakan confusion matrix dan metrik akurasi, precision, recall, dan f1-score(Maulana et al., 2024).

Algoritma Naïve Bayes

Naïve Bayes merupakan metode klasifikasi berdasarkan Teorema Bayes yang menghitung probabilitas dari fitur terhadap kelas target secara independen(Syafrianto, 2022). Algoritma ini cocok untuk klasifikasi teks Churn dan Tidak Churn dan digunakan dalam penelitian ini untuk menganalisis churn berdasarkan sentimen ulasan pengguna. Evaluasi model dilakukan melalui confusion matrix untuk mengukur performa klasifikasi(Guswandri et al., 2022)

METODE

Objek Penelitian

Objek utama penelitian adalah data historis pengguna aplikasi Mamikos. Fokus penelitian tertuju pada data interaksi pengguna, khususnya umur akun dan status churn, untuk memahami pola perilaku yang mengarah pada penghentian penggunaan aplikasi.

Populasi dan Sampel

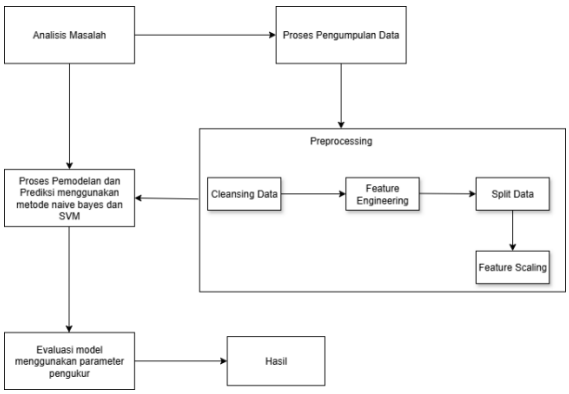
Populasi dalam penelitian ini mencakup seluruh pengguna aplikasi Mamikos dengan riwayat interaksi dalam periode tertentu. Sampel terdiri dari 113.248 data pengguna dengan status churn, yang digunakan sebagai data utama untuk klasifikasi.

Analisis Permasalahan

Masalah utama dalam penelitian yang diangkat adalah tingginya churn pengguna Mamikos. Untuk mengatasi hal tersebut, analisis dilakukan dengan pendekatan machine learning dan teknik analisis sentimen, guna mengidentifikasi dan memprediksi pola churn berdasarkan umpan balik pengguna. Dua algoritma digunakan, yakni Support Vector Machine (SVM) dan Naïve Bayes, untuk mengevaluasi efektivitas klasifikasi berdasarkan sentimen.

Diagram Proses Penelitian

Tahapan-tahapan dalam penelitian yang dilakukan untuk menganalisis churn pengguna menggunakan algoritma Naïve Bayes dan SVM. Penelitian dimulai dengan analisis masalah untuk memahami latar belakang dan tujuan dari studi yang dilakukan. Selanjutnya, dilakukan pengumpulan data dari sumber yang relevan, yaitu data pengguna aplikasi Mamikos.



Gambar alur analisis

Tahap pra-pemrosesan data mencakup beberapa proses utama, yaitu:

1. Cleansing data untuk menghapus data yang tidak konsisten atau tidak lengkap,
2. Feature engineering untuk membuat fitur yang merepresentasikan perilaku pengguna,
3. Split data untuk membagi dataset menjadi data latih dan data uji, serta
4. Feature scaling agar skala nilai antar fitur seragam dan mendukung kinerja model.

Teknik Pengumpulan Data

Pada Penelitian ini Pengumpulan data dilakukan dengan dua cara:

1. Studi Literature untuk memperoleh referensi ilmiah terkait churn dan analisis sentiment
2. Pengambilan data internal perusahaan Mamikos selama program studi independent, disertai dengan perjanjian Non-Disclosure Agreement (NDA)

Preprocessing Data

Preprocessing Data dilakukan untuk menyiapkan data sebelum digunakan pada model klasifikasi. Tahapan terdiri dari:

1. Cleansing Data
    a. Menghapus nilai kosong (Missing Value) dan data duplikasi untuk menjaga kualitas data. Feature Engineering
2. Feature Engineering
    a. Feature engineering merupakan langkah krusial untuk meningkatkan kemampuan model dalam memahami pola-pola churn. Proses ini mencakup pemilihan fitur penting berdasarkan korelasi dengan target, konversi umur akun ke dalam kategori usia, serta transformasi waktu ke dalam representasi numerik. Selain itu, dilakukan encoding terhadap fitur kategorikal agar dapat dibaca oleh algoritma. Analisis statistik deskriptif digunakan untuk memastikan distribusi data sesuai. Teknik seleksi fitur juga diterapkan untuk mengurangi kompleksitas dan meningkatkan efisiensi model. Dengan pendekatan ini, representasi data menjadi lebih informatif dan mendukung peningkatan akurasi model.
3. Feature Scaling
    a. Normalisasi data numerik dengan Min-Max Scaling agar seluruh fitur berada dalam skala [0,1], menghindari bias dalam algoritma SVM.
4. Split Dataset
5. Membagi data menjadi 80% untuk pelatihan dan 20% untuk pengujian.

Pemodelan Naïve Bayes dan Support Vector Machine (SVM)

1. Naïve Bayes

a. Model Algoritma ini Digunakan varian Multinomial, sesuai untuk data hasil transformasi teks. Model ini cepat, efisien, dan cocok untuk data berlabel dengan dimensi tinggi.

2. **Support Vector Machine (SVM)**
   a. Algoritma Support Vector Machine (svm) yaitu Algoritma yang Menggunakan LinearSVC yang dikalibrasi dengan CalibratedClassifierCV agar dapat menghasilkan estimasi probabilitas. Parameter dioptimasi menggunakan GridSearchCV dengan evaluasi F1-score.
3. **Evaluasi Model**
   a. Evaluasi model dilakukan dengan menggunakan metrik akurasi, precision, recall, dan f1-score. Selain itu, confusion matrix digunakan untuk menilai sejauh mana model berhasil membedakan pengguna churn dan tidak churn. Perbandingan hasil antara dua algoritma menjadi dasar untuk menentukan pendekatan paling optimal dalam memprediksi churn pengguna Mamikos.

HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan ini menjelaskan secara terperinci setiap tahapan yang telah dilakukan dalam penelitian, mulai dari proses pengumpulan data, tahap pra-pemrosesan, pelatihan model, hingga evaluasi akhir. Seluruh tahapan tersebut dipaparkan secara sistematis pada bagian berikut:

Data Penelitian

Penelitian ini menggunakan data internal dari PT Git Grow Ayo, yang merupakan perusahaan pengelola platform penyewaan kamar kos, Mamikos. Dataset mencakup riwayat aktivitas pengguna dari tanggal 28 Juni 2018 hingga 30 Juni 2023. Fokus utama adalah perilaku *churn* atau perpindahan pengguna yang dianalisis berdasarkan aktivitas harian dan diolah menjadi agregasi bulanan.

Dari hasil visualisasi, terlihat tren peningkatan signifikan dalam interaksi pengguna sejak awal tahun 2020 hingga pertengahan 2023, mencapai puncaknya pada hampir 10.000 kunjungan. Penurunan interaksi tercatat pada awal 2021, yang diduga berkaitan dengan pembatasan mobilitas saat pandemi COVID-19.

Pra-pemrosesan Data

Library Pemrograman

Analisis dilakukan menggunakan bahasa pemrograman Python dengan bantuan beberapa pustaka, seperti:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import LinearSVC
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import classification_report, accuracy_score
from sklearn.impute import SimpleImputer
from imblearn.over_sampling import SMOTE
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder
from google.colab import files
from sklearn.model_selection import GridSearchCV
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
import matplotlib.pyplot as plt
import seaborn as sns
```

Gambar library

1. Pandas dan NumPy untuk manipulasi data.

2. Scikit-learn untuk proses pemodelan, transformasi fitur, dan evaluasi.
3. Imbalanced-learn (SMOTE) untuk penanganan distribusi data yang tidak seimbang.
4. Matplotlib untuk visualisasi hasil.
5. Google Colab files untuk integrasi data input/output di lingkungan cloud

Penanganan Data Kosong

Proses awal mencakup identifikasi dan penanganan nilai kosong (missing values). Kolom seperti signup_type, submit_booking, dan total_favorite memiliki tingkat ketidakterisian data lebih dari 50%. Sementara kolom first_device_type memiliki lebih dari 99% nilai kosong dan diputuskan untuk dihapus. Seperti pada gambar :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113249 entries, 0 to 113248
Data columns (total 15 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   tenant_id             113249 non-null  int64
 1   date_account_created  111305 non-null  object
 2   gender                113249 non-null  object
 3   signup_method         111305 non-null  object
 4   signup_type           31560 non-null   object
 5   first_device_type     884 non-null     object
 6   submit_booking        46913 non-null   object
 7   domisili_code         113249 non-null  object
 8   total favorite        19941 non-null   float64
 9   total_chat            113249 non-null  int64
 10  total_submit booking  113249 non-null  int64
 11  total visit           113249 non-null  int64
 12  total visit listing   113249 non-null  int64
 13  total click booking   113249 non-null  int64
 14  churn                 113249 non-null  object
dtypes: float64(1), int64(6), object(8)
memory usage: 13.0+ MB
```

Gambar Data missing value

Penanganan data kosong dilakukan dengan teknik imputasi, seperti mengisi dengan nilai rata-rata (mean), nilai terbanyak (mode), atau label khusus seperti ‘unknown’. Maka akan dihasilkan seperti gambar berikut:

```
RangeIndex: 113249 entries, 0 to 113248
Data columns (total 21 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   tenant_id             113249 non-null  int64
 1   date_account_created  113249 non-null  datetime64[ns]
 2   gender                113249 non-null  object
 3   signup_method         113249 non-null  object
 4   signup_type           113249 non-null  object
 5   first_device_type     113249 non-null  object
 6   submit_booking        113249 non-null  object
 7   domisili_code         113249 non-null  object
 8   total favorite        113249 non-null  int64
 9   total_chat            113249 non-null  int64
 10  total_submit booking  113249 non-null  int64
 11  total visit           113249 non-null  int64
 12  total visit listing   113249 non-null  int64
 13  total click booking   113249 non-null  int64
 14  churn                 113249 non-null  object
 15  account_age_days      113249 non-null  int64
 16  lama_akun             113249 non-null  int64
 17  remarks_lama_akun     113249 non-null  object
 18  remarks_visit         113249 non-null  object
 19  remarks_list_visit    113249 non-null  object
 20  remarks_click_booking 113249 non-null  object
dtypes: datetime64[ns](1), int64(9), object(11)
```

Gambar data clean

Rekayasa Fitur

Kemudian setelah data bersih peneliti membuat rekaya fitur (Feature Engineering) Untuk meningkatkan kemampuan model dalam mengenali pola churn, beberapa fitur baru dikembangkan, antara lain:

account_age_days	lama_akun	remarks_lama_akun	remarks_visit	remarks_list_visit	remarks_click_booking
2245	2245	akun lama	ramai	rendah	sering
2498	2498	akun lama	tidak ramai	tinggi	sering
2752	2752	akun lama	tidak ramai	tinggi	sering
2751	2751	akun lama	ramai	sedang	sedang
2750	2750	akun lama	ramai	sedang	sering
...	...	...	...	...	...
748	748	akun lama	sedang	tinggi	jarang
748	748	akun lama	ramai	tinggi	sedang

Gambar Hasil rekayasa fitur

1. account_age_days: menghitung jumlah hari sejak pengguna mendaftar.
2. remark_lama_akun: klasifikasi akun baru/lama dengan ambang 180 hari.
3. remarks_visit, remarks_visit_listing, dan remarks_click_booking: klasifikasi berdasarkan kuartil distribusi untuk memahami intensitas aktivitas pengguna

Langkah ini bertujuan menyederhanakan fitur numerik menjadi bentuk kategorikal yang lebih bermakna bagi algoritma klasifikasi.

Transformasi Data

Fitur numerik distandarisasi menggunakan StandardScaler, sementara fitur kategorikal diubah menjadi vektor biner menggunakan OneHotEncoder. Seperti gambar berikut:

```
# Feature Scaling
preprocessor = ColumnTransformer(transformers=[
    ('num', StandardScaler(), numeric_cols),
    ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_cols)
])
```

Gambar Penerapan ColumnTransformer

Kedua jenis transformasi digabung dalam satu pipeline menggunakan ColumnTransformer dan hanya diterapkan pada data pelatihan guna menghindari kebocoran informasi ke data uji.

Pembagian Data dan Penyeimbangan

Dataset dibagi menjadi data pelatihan (80%) dan pengujian (20%) menggunakan fungsi train_test_split. Karena data churn bersifat tidak seimbang, diterapkan SMOTE (Synthetic Minority Over-sampling Technique) pada data pelatihan untuk menyeimbangkan distribusi antara kelas churn dan tidak churn.

Tabel split data

Komponen	**Keterangan**
Total Data	113.249 sampel
Training Data	90.599 data (80%)
Testing Data	22.650 data (20%)
Metode	Train_test_split (scikit-learn)
Parameter	Test_size=0,2, random_state=42
Fitur (X)	Total_submit_booking, remark_lama_akun, remark_visit, remark_click_booking, account_age_days

Target (y)	churn (1 = churn, 0 = tidak churn)

Implementasi Algoritma Gaussian Naïve Bayes

Model pertama yang digunakan adalah Gaussian Naive Bayes, yang sesuai untuk fitur numerik yang terdistribusi normal. Model ini dilatih menggunakan data yang telah diseimbangkan dengan SMOTE. Setelah proses prediksi, hasil berupa angka 1 dan 0 dikonversi menjadi label "positif" (tetap menggunakan aplikasi) dan "negatif" (berhenti menggunakan aplikasi) menggunakan fungsi np.where().

```
# ====================
# 📌 NAIVE BAYES
# ====================
nb = GaussianNB()
nb.fit(X_train_sm.toarray() if hasattr(X_train_sm, "toarray") else X_train_sm, y_train_sm)
y_pred_nb = nb.predict(X_test_pre.toarray() if hasattr(X_test_pre, "toarray") else X_test_pre)
```

Gambar Code Naive Bayes

Algoritma Gaussian Naïve Bayes digunakan sebagai salah satu metode pembanding dalam klasifikasi churn pengguna. Proses pelatihan model dilakukan dengan memanfaatkan data hasil oversampling (X_train_sm, y_train_sm). Untuk memastikan kompatibilitas format data, digunakan fungsi .toarray() apabila data berbentuk sparse matrix. Setelah model selesai dilatih, dilakukan prediksi terhadap data uji (X_test_pre) dengan pendekatan yang sama dalam penyesuaian format data. Hasil prediksi kemudian dianalisis untuk membandingkan performa algoritma ini dengan model lain seperti SVM.

Implementasi Algoritma Support Vector Machine (SVM)

Algoritma kedua yang digunakan adalah Support Vector Machine dengan kernel linier. Untuk memungkinkan prediksi probabilistik, model dikalibrasi menggunakan CalibratedClassifierCV. SVM efektif dalam menangani data berdimensi tinggi dan kasus klasifikasi biner seperti churn.

```
# 📌 SVM (dengan Calibrated Classifier untuk probabilitas)
# ====================
svm = CalibratedClassifierCV(LinearSVC(random_state=42))
svm.fit(X_train_sm, y_train_sm)
y_pred_svm = svm.predict(X_test_pre)
```

Gambar Code Support Vector Machine (SVM

Algoritma Support Vector Machine (SVM) digunakan untuk mengklasifikasikan potensi churn pengguna. Untuk meningkatkan akurasi dan memungkinkan prediksi probabilistik, model dikalibrasi menggunakan CalibratedClassifierCV dengan dasar algoritma LinearSVC. Proses pelatihan dilakukan menggunakan data yang telah dioversampling dengan teknik SMOTE, yaitu X_train_sm dan y_train_sm. Setelah proses pelatihan selesai, model digunakan untuk memprediksi data uji (X_test_pre), menghasilkan label prediksi (y_pred_svm) yang kemudian dievaluasi untuk menilai kinerja model.

Evaluasi Model

Kinerja Model Gaussian Naïve Bayes

Model Naive Bayes menunjukkan tingkat akurasi sebesar 95,83%. Hasil evaluasi berdasarkan metrik klasifikasi:

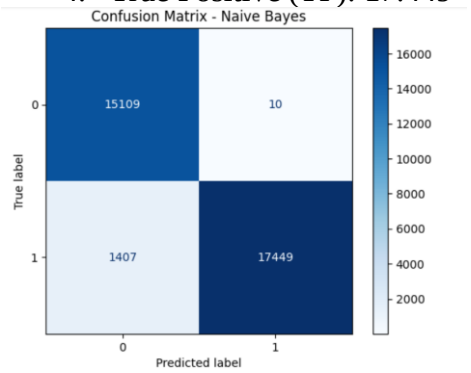
1. Precision untuk kelas positif: 1.00
2. Recall untuk kelas positif: 0.93
3. F1-Score untuk kelas positif: 0.96

```
=== NAIVE BAYES ===
              precision    recall  f1-score   support

           0       0.91      1.00      0.96     15119
           1       1.00      0.93      0.96     18856

    accuracy                           0.96     33975
   macro avg       0.96      0.96      0.96     33975
weighted avg       0.96      0.96      0.96     33975

Akurasi: 0.9582928623988226
```

Gambar Hasil Prediksi Naive Bayes

Confusion matrix model menunjukkan:

1. True Negative (TN): 15.109
2. False Positive (FP): 10
3. False Negative (FN): 1.407
4. True Positive (TP): 17.449



Gambar Confusion Matrix Naive Bayes

Model ini sangat presisi dalam mengidentifikasi pengguna yang tidak churn, namun cenderung melewatkan sebagian pengguna yang benar-benar churn (tinggi FN).

Kinerja Model Support Vector Machine (SVM)

SVM mencatat akurasi sebesar 95,72% dengan rincian metrik:

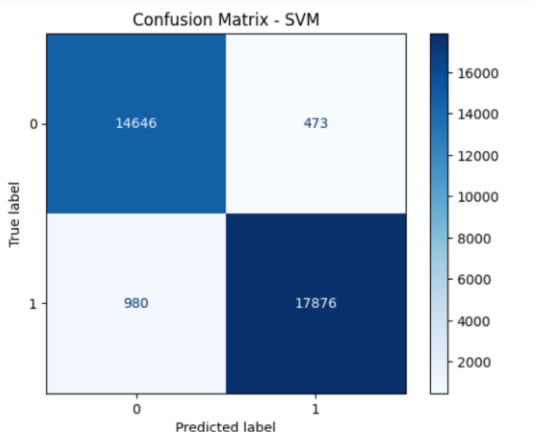
1. Precision kelas positif: 0.97
2. Recall kelas positif: 0.95
3. F1-Score kelas positif: 0.96

```
=== SVM ===
              precision    recall  f1-score   support

           0       0.94      0.97      0.95     15119
           1       0.97      0.95      0.96     18856

    accuracy                           0.96     33975
   macro avg       0.96      0.96      0.96     33975
weighted avg       0.96      0.96      0.96     33975

Akurasi: 0.957233259749816
```

Gambar Hasil Prediksi Support Vector Machine (SVM

Confusion matrix model SVM:

1. True Negative (TN): 14.646
2. False Positive (FP): 473
3. False Negative (FN): 980
4. True Positive (TP): 17.876



Gambar Confusion Matrix SVM

Meskipun sedikit lebih rendah dalam presisi, SVM unggul dalam hal recall, yang berarti model ini lebih responsif dalam mendeteksi pengguna yang berpotensi churn.

KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi potensi churn pengguna aplikasi Mamikos melalui pendekatan machine learning berbasis analisis sentimen. Dua algoritma yang digunakan dalam proses klasifikasi adalah Support Vector Machine (SVM) dan Naïve Bayes. Berdasarkan hasil evaluasi model dan analisis data, dapat disimpulkan bahwa algoritma SVM memiliki performa yang lebih unggul dibandingkan Naïve Bayes. Keunggulan tersebut terlihat dari capaian metrik evaluasi seperti akurasi, precision, recall, dan f1-score yang lebih tinggi, serta kemampuan model dalam menghasilkan prediksi yang lebih seimbang terhadap data churn dan non-churn. Analisis terhadap data sentimen menunjukkan bahwa sebagian besar pengguna memberikan ulasan yang bernada positif terhadap aplikasi. Meskipun demikian, keberadaan churn masih teridentifikasi pada kelompok pengguna dengan sentimen baik. Hal ini mengindikasikan bahwa perilaku pengguna yang berujung pada churn tidak hanya dipengaruhi oleh ekspresi opini secara eksplisit, tetapi juga oleh faktor-faktor lain yang lebih kompleks dan tidak langsung tergambar dari sentimen semata.

Proses pra-pemrosesan dan rekayasa fitur juga memberikan kontribusi besar terhadap keberhasilan model. Penerapan metode seperti SMOTE untuk penyeimbangan data, normalisasi nilai numerik, serta pengelompokan fitur berdasarkan kuartil, berhasil meningkatkan kualitas data yang digunakan untuk pelatihan model. Hal ini secara langsung berdampak pada peningkatan akurasi model dalam mengenali pola churn yang tersembunyi di dalam data. Penggunaan dua algoritma berbeda dalam penelitian ini memberikan sudut pandang komparatif yang penting. Algoritma SVM terbukti efektif untuk data yang kompleks dan berdimensi tinggi, sedangkan Naïve Bayes menunjukkan kecepatan dan efisiensi dalam proses klasifikasi berbasis probabilitas. Melalui pendekatan ini, penelitian mampu memberikan gambaran yang lebih menyeluruh dalam memahami potensi churn dan karakteristik perilaku pengguna.

REFERENSI

Damanik, S. D., & Jambak, M. I. (2023). Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma C4.5. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *3*(6), 1303–1309. https://doi.org/10.30865/klik.v3i6.829

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, *10*(02), 71–76. https://doi.org/10.32664/smatika.v10i02.455

Guswandri, A., Cahyono, R. P., Akutansi, S. I., & Komputer, T. (2022). Penerapan Sentimen Analis Menggunakan Metode Naïve Bayes Dan Svm. *Ilmudata.Org*, *2*(12), 1–16.

Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 247–256. https://doi.org/10.57152/malcom.v4i1.1066

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12343 LNCS*, 503–515. https://doi.org/10.1007/978-3-030-62008-0_35

Kulsum, U., Jajuli, M., & Sulistiyowati, N. (2022). Analisis Sentimen Aplikasi WETV di Google Play Store Menggunakan Algoritma Support Vector Machine. *Journal of Applied Informatics and Computing*, *6*(2), 205–212. https://doi.org/10.30871/jaic.v6i2.4802

Lubis, A. Y., & Setyawan, M. Y. H. (2024). Analisis Sentimen Terhadap Aplikasi Pospay Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *6*(3), 514–521. https://doi.org/10.47233/jteksis.v6i3.1310

Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 375–384. https://doi.org/10.57152/malcom.v4i2.1206

Muthmainnah, T. N., & Voutama, A. (2023). Volume 6 ; Nomor 2. *Juli*, *6*, 463–471. https://ojs.trigunadharma.ac.id/index.php/jsk/index

Pokhrel, S. (2024). No TitleΕΛΕΝΗ. *Αγαη*, *15*(1), 37–48.

Syafrianto, A. (2022). Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Sentimen Analisis. *The Indonesian Journal of Computer Science Research*, *1*(2), 1–15. https://doi.org/10.59095/ijcsr.v1i2.11