JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i1.5811

PREDIKSI CHURN PELANGGAN PADA LAYANAN DESAIN GRAFIS *HOME* DESAIN MENGGUNAKAN ALGORITMA *NAÏVE BAYES*

Iip Imron Daipah[1], Rini Astuti[2], Willy Prihartono[3]

[1,3] Program Studi Teknik Informatika STMIK IKMI Cirebon; Jl. Perjuangan No. 10B, Majasem, Cirebon, Jawa Barat.
[2] Program Studi Sistem Informasi STMIK LIKMI Bandung ; Jl. Ir. H. Juanda No. 96, Lebakgede, Coblong, Bandung, Jawa Barat.

Received: 25 Desember 2024
Accepted: 14 Januari 2025
Published: 20 Januari 2025

Keywords:
Model prediksi *churn*, *Naïve Bayes*, strategi retensi, analisis data, sektor jasa kreatif.

CorespondentEmail:
iipimrondaipah@gmail.com

Abstrak. Penelitian ini bertujuan untuk mengembangkan model prediksi *churn* pelanggan di Home Desain menggunakan algoritma *Naïve Bayes*. Dataset yang digunakan mencakup data transaksi dari tahun 2021 hingga 2024, dengan fitur-fitur seperti jenis layanan, jumlah revisi, harga, status pembayaran, kategori revisi, dan keterlambatan pembayaran. Proses analisis mengikuti tahapan dalam metode *Knowledge Discovery in Databases (KDD)*, yang mencakup seleksi, praproses, transformasi data, pembuatan model, dan evaluasi. Pada tahap seleksi data, fitur yang relevan dengan *churn* pelanggan dipilih, sementara praproses dilakukan untuk memastikan konsistensi data. Transformasi data digunakan untuk mempersiapkan data sebelum dimasukkan ke dalam model prediksi. Model dikembangkan menggunakan algoritma *Naïve Bayes*, yang terbukti efektif dalam menangani data besar dan masalah klasifikasi. Hasil evaluasi menunjukkan bahwa model yang dikembangkan memiliki performa sangat baik, dengan akurasi 99%, presisi 1.00, recall 0.99, dan F1-score 0.99, yang menandakan bahwa model ini sangat handal dalam mengidentifikasi pelanggan yang berisiko *churn*. Temuan ini membuktikan bahwa *Naïve Bayes* adalah alat yang efektif dalam merancang strategi retensi yang lebih efisien, serta berguna bagi perusahaan di sektor jasa kreatif untuk meningkatkan pengambilan keputusan berbasis data. Penelitian ini memberikan kontribusi signifikan terhadap literatur mengenai penggunaan algoritma prediktif dalam industri kreatif, membuktikan bahwa *Naïve Bayes* dapat diandalkan dalam menganalisis data besar dan memprediksi churn pelanggan dengan akurasi tinggi. Temuan ini juga membantu perusahaan merancang strategi yang lebih efektif untuk mempertahankan pelanggan dan meningkatkan daya saing pasar.

Abstract. *This study aims to develop a customer churn prediction model for Home Design using the Naïve Bayes algorithm. The dataset includes transaction data from 2021 to 2024, with features such as service type, revision count, price, payment status, revision category, and payment delays. The analysis follows the Knowledge Discovery in Databases (KDD) process, including data selection, preprocessing, transformation, model development, and evaluation. Relevant features related to churn are selected, and data preprocessing ensures consistency. The Naïve Bayes algorithm, effective for handling large datasets and classification problems, is used to build the model. The evaluation results show outstanding performance with 99% accuracy, 1.00 precision, 0.99 recall, and 0.99 F1-score, indicating the model's reliability in identifying churn-risk customers. The study demonstrates that Naïve Bayes is an effective tool for designing efficient retention strategies and supports data-driven decision-making in the creative services sector, contributing to competitive advantage in the market.*

1. PENDAHULUAN

Perkembangan teknologi yang terus mengalami kemajuan seiring berjalannya waktu. telah mendorong meningkatnya kebutuhan masyarakat terhadap informasi dan komunikasi. Berbagai perangkat teknologi digunakan secara luas untuk mengatasi tantangan dalam memenuhi kebutuhan[1]. Perkembangan teknologi komputasi, khususnya di bidang analisis data dan kecerdasan buatan, telah membuka peluang baru bagi bisnis untuk memanfaatkan data dalam mendukung pengambilan keputusan berskala besar. Salah satu dampak utamanya adalah peningkatan efisiensi dalam menganalisis perilaku pelanggan, termasuk kemampuan memprediksi kemungkinan pelanggan berhenti menggunakan layanan *churn*[2] .

Dengan berkembangnya bisnis, khususnya dalam sektor jasa desain grafis, salah satu tantangan yang semakin menonjol adalah meningkatnya persaingan di pasar. Dalam kondisi pasar yang sangat kompetitif, penyedia jasa sering menghadapi masalah utama berupa tingginya tingkat pelanggan yang berhenti menggunakan layanan, dikenal sebagai *churn***.** Pelanggan dianggap mengalami *churn* ketika mereka berhenti melakukan transaksi atau menghentikan langganan terhadap layanan yang ditawarkan oleh perusahaan [3].

Salah satu dampak utamanya adalah peningkatan efisiensi dalam menganalisis perilaku pelanggan, termasuk kemampuan memprediksi kemungkinan pelanggan churn atau berhenti berlangganan[4].Di tengah persaingan bisnis yang semakin kompetitif, terutama di industri jasa desain grafis dan percetakan, keberhasilan perusahaan sangat bergantung pada kemampuan mereka untuk menjaga loyalitas pelanggan [5]. Prediksi *churn* pelanggan dapat dilimplementasikan menggunakan teknik data mining.

Data mining merupakan teknik untuk mengumpulkan serta menganalisis data dengan tujuan mengungkap pola-pola tersembunyi yang terdapat dalam kumpulan data. [6].Untuk menjalankan proses data mining, diperlukan algoritma yang dapat mengklasifikasikan pelanggan ke dalam kategori *churn* atau *non-churn* secara akurat [7]. Sebagai solusi, penggunaan algoritma *Naïve Bayes* sangat disarankan.

Naive Bayes menerapkan pendekatan berbasis probabilitas dalam perhitungannya. Secara keseluruhan, metode ini dikenal karena tingkat akurasinya yang tinggi dan sering dimanfaatkan untuk mengatasi berbagai permasalahan prediksi dalam klasifikasi [8].

2. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

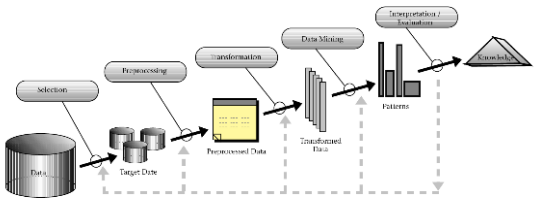
Penelitian terkait prediksi *churn* pelanggan telah banyak dilakukan di berbagai sektor yang menekankan pentingnya loyalitas pelanggan. Salah satu penelitian oleh [7] Mengkaji penggunaan algoritma *Naïve Bayes* untuk klasifikasi *churn* pelanggan dalam kerangka *CRISP-DM*, mencakup tahap pemahaman bisnis, persiapan data, pemodelan, dan evaluasi. Perangkat yang digunakan dalam penelitian tersebut meliputi *RapidMiner*, *NetBeans*, dan *Microsoft Excel*. Hasilnya menunjukkan bahwa model Naïve Bayes mencapai akurasi 84,95% dan *AUC (Area Under the Curve)* sebesar 83,20%, yang menunjukkan kinerja yang baik. Penelitian ini memberikan kontribusi yang signifikan bagi UMKM dengan menyediakan model efektif untuk menganalisis perilaku pelanggan dan meningkatkan retensi melalui prediksi churn. Implementasi model ini membuktikan bahwa *Naïve Bayes* adalah alat yang efisien untuk pengelolaan data pelanggan pada bisnis kecil dan menengah.

Dalam sektor bisnis ritel, penelitian yang dilakukan oleh [9] Menerapkan algoritma *Naïve Bayes* untuk memprediksi *customer churn* dengan tingkat akurasi mencapai 80% dan precision sebesar 100%. Penelitian ini menggunakan beberapa variabel data seperti jenis kelamin, grade point, kepemilikan kartu kredit, rentang usia, nilai transaksi rata-rata, serta durasi keanggotaan. Evaluasi kinerja model dilakukan melalui *confusion matrix* dan grafik *ROC*, yang menunjukkan bahwa Naïve

Bayes efektif dalam memprediksi pelanggan yang berpotensi berhenti.

3. METODE PENELITIAN

Penelitian ini menggunakan tahapan *Knowledge Discovery in Databases (KDD)* untuk mengidentifikasi pola yang relevan dan menghasilkan wawasan akurat dalam mengukur Churn pelanggan[10] .Gambar 3.1:



Gambar 3. 1 Tahap Proses KDD

3.1. Data Selection

Tahap ini, dilakukan pemilihan dan ekstraksi data yang relevan dari basis data pelanggan di *Home* Desain. Data yang dipilih mencakup variabel-variabel yang dapat mempengaruhi churn, seperti jenis layanan, pelanggan, jumlah revisi, harga, status pembayaran, total revisi, kategori revisi, dan keterlambatan pembayaran. Langkah awal yang krusial dalam memastikan relevansi dan dukungan data terhadap tujuan penelitian adalah pemilihan data yang tepat untuk memprediksi potensi churn pelanggan.

3.2. Prapemrosesan

Tahap ini meliputi penghapusan nilai yang hilang, perbaikan inkonsistensi, dan penanganan data duplikat. Merubah variabel kategorikal. Data numerik kemudian dinormalisasi atau distandarisasi menggunakan *StandardScaler* untuk menghindari dominasi fitur dengan skala berbeda.

3.3. Transformasi

Tahap transformasi dilakukan untuk mengatasi variabel kategorikal yang kompleks dan data numerik dengan distribusi tidak normal. Fitur yang tidak relevan akan dihapus, sementara fitur penting bisa digabung atau diubah, seperti menerapkan transformasi logaritmik pada data yang distribusinya miring.Tahap ini bertujuan agar data lebih optimal, sehingga model prediksi dapat bekerja lebih efisien dan menghasilkan akurasi yang lebih baik.

3.4. Data Mining

Tahap ini, membangun model prediksi *churn* menggunakan *Naïve Bayes.* Data yang telah diproses dan di transformasi akan dibagi menjadi dua bagian: data pelatihan dan data uji. Model *Naïve Bayes* kemudian dilatih dengan data pelatihan untuk memahami hubungan probabilistik antara fitur dan variabel target *churn.* Setelah dilatih, model diuji menggunakan data uji untuk mengukur akurasi, presisi, recall, dan F1-score, yang akan menunjukkan seberapa akurat model memprediksi pelanggan yang berisiko *churn*. probabilitas suatu kelas diberikan oleh rumus *Bayes*:

$$\mathrm{P\ (C|X)} = \frac{P(C|X)P(C)}{P(X)}$$

Penjelasan Rumus Algoritma *Naïve Bayes* :
a.P(C|X) : probabilitas terjadinya C (*churn* atau tidak *churn*) diberikan fitur X
b.P(X|C) : *likelihood*, probabilitas mengamati fitur X diberikan kelas C
c.P(C) : probabilitas prior kelas C
d.(X) : probabilitas *evidence* atau fitur X

3.5. Evaluasi

Evaluasi dilakukan untuk mengukur kinerjanya dalam memprediksi *churn.* Proses ini menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk memastikan bahwa model memiliki tingkat keakuratan yang memadai serta mampu menghasilkan prediksi yang andal. Hasil evaluasi memberikan gambaran mengenai keunggulan dan kelemahan model, yang sekaligus dapat digunakan untuk mengidentifikasi area yang memerlukan perbaikan lebih lanjut.

4. HASIL DAN PEMBAHASAN

Menerapkan Algoritma *Naïve Bayes* dalam analisis prediksi churn pada data transaksi home desain.Metodologi ini mengikuti langkah-langkah *Knowledge Discovery in Databases (KDD).*

4.1 Data *Selection*

Data ini diperoleh dari pelaku bisnis di bidang jasa desain grafis percetakan yang berlokasi di Ciamis Jawa Barat dengan jumlah

datanya adalah 464 pelanggan selama 3 tahun dari 2022-2024 dengan format *excel*. data pelanggan ini mempunyai 9 atribut yaitu ; Tanggal,Nama Klien,Jenis Layanan,Jumlah Revisi,Harga , Status Pembayaran ,Total Revisi, Kategori Revisi dan Keterlambatan Pembayaran. Berikut, hasil seleksi kolom yang relevan . Gambar 4.1 :

Data setelah seleksi:

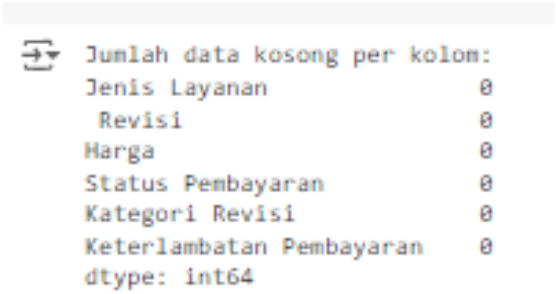
	Jenis Layanan	Revisi	Harga	Status Pembayaran	Kategori Revisi	Keterlambatan Pembayaran
0	Desain Undangan Khitan	3	30000	lunas	Tinggi	0
1	Desain Logo	4	20000	Belum Lunas	Sedang	1
2	Desain Kemasan	2	20000	lunas	Rendah	0
3	Desain Vektor	6	20000	lunas	Tinggi	0
4	Desain Poster	4	25000	lunas	Tinggi	0
...	...	...	...	...	...	...
395	Desain Pamflet	3	20000	lunas	Sedang	0
396	Desain Poster	3	20000	lunas	Sedang	0
397	Desain PDH	4	20000	lunas	Sedang	0
398	Desain Jersey Voli	4	20000	lunas	Sedang	0
399	Desain Poster	2	20000	lunas	Rendah	0

400 rows × 8 columns

Gambar 4. 1 Data Hasil Seleksi

4.2 Prapemrosesan

Tahapan pra-pemrosesan dimulai dengan mengidentifikasi jumlah nilai kosong pada setiap kolom dataset . Berikut hasil verifikasi menunjukkan tidak adanya lagi *missing values*. Gambar 4.2 :



```
Jumlah data kosong per kolom:
Jenis Layanan                0
 Revisi                      0
Harga                        0
Status Pembayaran            0
Kategori Revisi              0
Keterlambatan Pembayaran     0
dtype: int64
```

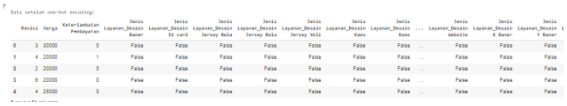
Gambar 4. 2 Data Nilai Kosong

4.3 Tranformation

tahap ini, dilakukan perubahan tye data dan melakukan teknik *one hot encoding* dan *normalisasi* data.

```
Menggunakan one-hot encoding pada kolom kategori
final_df_encoded = pd.get_dummies(final_df, columns=['Jenis Layanan', 'Status Pembayaran', 'Kategori Revisi'], drop_first=True)

Menampilkan data setelah encoding
print("\nData setelah one-hot encoding:")
final_df_encoded.head()
```

Dalam penelitian ini, hasil *one hot encoding* merubah kolom *kategorikal* menjadi *numeric biner* .Gambar 4.3 :

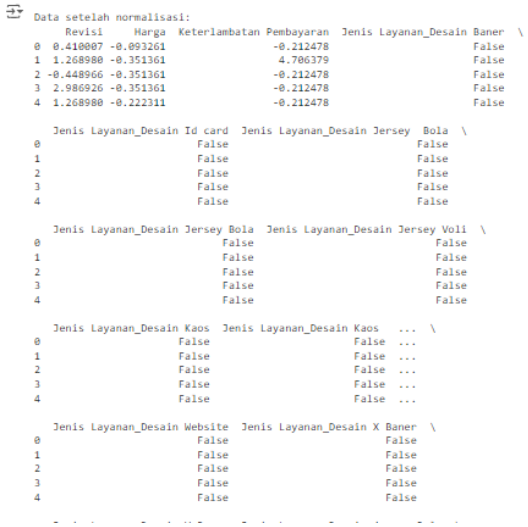


Data setelah one-hot encoding:

	Revisi	Harga	Keterlambatan Pembayaran	Jenis Layanan_Desain Baner	Jenis Layanan_Desain Id card	Jenis Layanan_Desain Jersey Bola	Jenis Layanan_Desain Jersey Voli	Jenis Layanan_Desain Kaos	Jenis Layanan_Desain Kaos	...	Jenis Layanan_Desain Website	Jenis Layanan_Desain X Baner	Jenis Layanan_Desain Y Baner	L
0	3	30000	0	False	False	False	False	False	False	...	False	False	False	
1	4	20000	1	False	False	False	False	False	False	...	False	False	False	
2	2	20000	0	False	False	False	False	False	False	...	False	False	False	
3	6	20000	0	False	False	False	False	False	False	...	False	False	False	
4	4	25000	0	False	False	False	False	False	False	...	False	False	False	

Gambar 4. 3 Hasil *One Hot Encoding*

Langkah kedua melakukan normalisasi data numerik menggunakan *StandardScaler*.Hasil normalisasi data. Gambar 4.4 :



Data setelah normalisasi:

	Revisi	Harga	Keterlambatan Pembayaran	Jenis Layanan_Desain Baner	Jenis Layanan_Desain Id card	Jenis Layanan_Desain Jersey Bola	Jenis Layanan_Desain Jersey Voli	Jenis Layanan_Desain Kaos	Jenis Layanan_Desain Kaos	...	Jenis Layanan_Desain Website	Jenis Layanan_Desain X Baner	Jenis Layanan_Desain Y Baner	L
0	0.410007	-0.093261	-0.212478	False	False	False	False	False	False	...	False	False	False	
1	1.268980	-0.351361	4.706379	False	False	False	False	False	False	...	False	False	False	
2	-0.448966	-0.351361	-0.212478	False	False	False	False	False	False	...	False	False	False	
3	2.986926	-0.351361	-0.212478	False	False	False	False	False	False	...	False	False	False	
4	1.268980	-0.222311	-0.212478	False	False	False	False	False	False	...	False	False	False	

Gambar 4. 4 Hasil *Normalisasi*

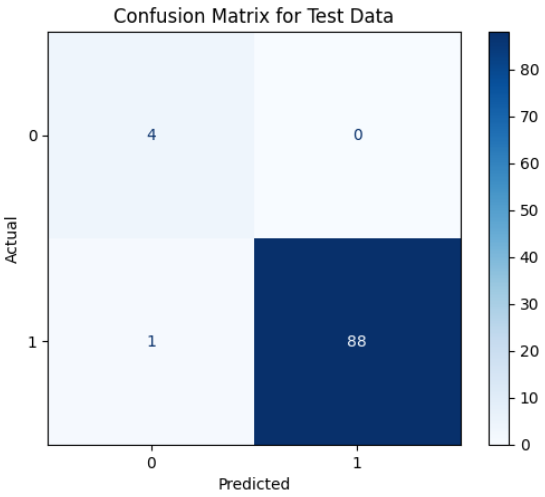
4.4 Data Mining

Pada tahapan ini algoritma *Naïve Bayes (GaussianNB)* digunakan untuk memprediksi *churn* pelanggan. Dataset dibagi menjadi data pelatihan dan pengujian menggunakan fungsi train_test_split dilatih menggunakan data pelatihan dengan metode *model.fit*() dan digunakan untuk memprediksi *churn* pada data pengujian menggunakan *model.predict().*

```
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
import matplotlib.pyplot as plt

# Prediksi pada data latih
y_train_pred = model.predict(X_train_scaled)

# Hitung confusion matrix
cm = confusion_matrix(y_train, y_train_pred)

# Tampilkan confusion matrix
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=model.classes_)
disp.plot(cmap=plt.cm.Blues)

# Menambahkan judul dan label
plt.title("Confusion Matrix for Train Data")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# Menampilkan nilai confusion matrix
print("Confusion Matrix:")
print(cm)
```

Gambar 4. 5 *Script Code Confusion Matrix*

Confusion Matrix divisualisasikan untuk menunjukkan jumlah prediksi yang benar dan salah. Gambar 4.5 :



Gambar 4. 6 Hasil *Konfusion Matrix*

Confusion Matrix menggambarkan bahwa model memiliki kinerja yang sangat baik dalam mendeteksi pelanggan churn. Terdapat 88 True Positives (TP) dan hanya 1 False Negative (FN), yang menandakan bahwa model sangat andal dalam mengidentifikasi pelanggan yang benar-benar berisiko churn.

Pada tahapan berikutnya menggunakan fungsi classification_report untuk mengevaluasi performa model dengan memberikan laporan terperinci mengenai metrik evaluasi, seperti presisi, recall, dan F1-score untuk setiap kelas. Dalam hal ini, laporan dibagi menjadi dua kelas: *Churn* dan tidak *Churn*. Gambar 4.6 :

```
Classification Report:
              precision    recall  f1-score   support

 Tidak Churn       0.80      1.00      0.89         4
       Churn       1.00      0.99      0.99        89

    accuracy                           0.99        93
   macro avg       0.90      0.99      0.94        93
weighted avg       0.99      0.99      0.99        93
```

Gambar 4. Hasil *Classification Report*

Hasil klasifikasi ini memberikan evaluasi kinerja model untuk kedua kelas: *Churn* dan Tidak *Churn*. Untuk kelas Tidak *Churn*, presisi adalah 0.80, recall 1.00, dan F1-score 0.89, yang menunjukkan jumlah prediksi benar[11].Untuk kelas *Churn*, presisi mencapai 1.00, recall 0.99, dan F1-score 0.99, menunjukkan bahwa model sangat akurat dalam memprediksi kasus positif dengan sangat sedikit kesalahan. Secara keseluruhan, akurasi model pada data uji adalah 0.99, dengan nilai rata-rata macro F1-score sebesar 0.94 dan weighted F1-score 0.99, yang menandakan performa yang sangat baik dalam mengklasifikasikan kedua kelas.

Pada tahapan berikutnya , ROC-AUC score dihitung untuk menilai kemampuan model dalam membedakan antara kelas positif dan negatif. Skor ini merepresentasikan area di bawah kurva ROC, yang menggambarkan performa keseluruhan model.

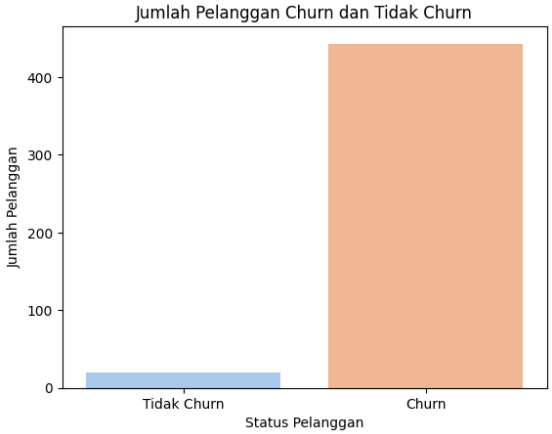
```
# Menghitung ROC-AUC Score
roc_auc = roc_auc_score(y_test, model.predict_proba(X_test_scaled)[:, 1])
```

ROC-AUC score sebesar 0.99 menunjukkan bahwa model memiliki kemampuan yang luar biasa dalam membedakan antara kelas positif dan negative [12]. Skor ini sangat mendekati nilai maksimum 1, yang menunjukkan bahwa model hampir sempurna dalam memprediksi kedua kelas, dengan hanya sedikit kesalahan dalam klasifikasi. Secara keseluruhan, nilai ini mencerminkan kinerja model yang sangat baik dalam mengklasifikasikan data uji. Hasil ROC gambar 4.7 :

```
ROC-AUC Score: 0.99
```

Gambar 4. 7 Hasil *ROC-AUC*

Tahapan berikutnya memvisualisasikan data pelanggan *churn* dan tidak *churn* menggunakan diagram batang. Gambar 4.8 :

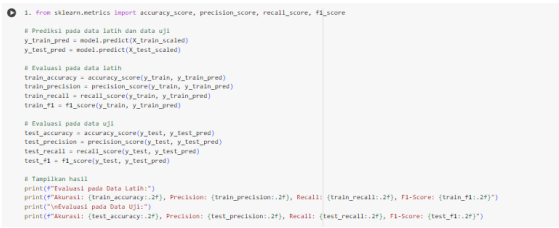


Gambar 4. 8 *Churn* VS Tidak *Churn*

Berdasarkan grafik distribusi status pelanggan di atas, terlihat bahwa jumlah pelanggan yang mengalami *churn* jauh lebih banyak dibandingkan pelanggan yang tidak churn.

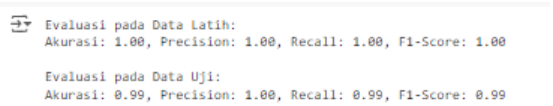
4.5 Evaluasi

Pada tahap terakhir , dilakukan evaluasi performa model menggunakan metrik utama, yaitu accuracy, precision, recall, dan F1-score, baik pada data latih maupun data uji.



```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Prediksi pada data latih dan data uji
y_train_pred = model.predict(X_train_scaled)
y_test_pred = model.predict(X_test_scaled)

# Evaluasi pada data latih
train_accuracy = accuracy_score(y_train, y_train_pred)
train_precision = precision_score(y_train, y_train_pred)
train_recall = recall_score(y_train, y_train_pred)
train_f1 = f1_score(y_train, y_train_pred)

# Evaluasi pada data uji
test_accuracy = accuracy_score(y_test, y_test_pred)
test_precision = precision_score(y_test, y_test_pred)
test_recall = recall_score(y_test, y_test_pred)
test_f1 = f1_score(y_test, y_test_pred)

# Tampilkan hasil
print(f"Evaluasi pada Data Latih:")
print(f"Akurasi: {train_accuracy:.2f}, Precision: {train_precision:.2f}, Recall: {train_recall:.2f}, F1-Score: {train_f1:.2f}")
print("\nEvaluasi pada Data Uji:")
print(f"Akurasi: {test_accuracy:.2f}, Precision: {test_precision:.2f}, Recall: {test_recall:.2f}, F1-Score: {test_f1:.2f}")
```

Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik pada data latih dan data uji. Gambar 4.9 :



```
Evaluasi pada Data Latih:
Akurasi: 1.00, Precision: 1.00, Recall: 1.00, F1-Score: 1.00

Evaluasi pada Data Uji:
Akurasi: 0.99, Precision: 1.00, Recall: 0.99, F1-Score: 0.99
```

Gambar 4. 9 Hasil Evaluasi Data Latih dan Data Uji

Pada data latih, model mencapai skor sempurna untuk semua metrik (accuracy, precision, recall, dan F1-score), yaitu 1.00, yang mengindikasikan bahwa model memprediksi seluruh data pelatihan dengan benar. Pada data uji, model mencapai akurasi 0.99, yang berarti 99% prediksi benar. Precision sebesar 1.00 menunjukkan bahwa semua prediksi churn benar, tanpa ada prediksi positif palsu. Recall sebesar 0.99 menunjukkan bahwa hampir semua kasus churn terdeteksi, dengan sedikit data churn yang tidak terprediksi. F1-score sebesar 0.99, sebagai rata-rata harmonik antara precision dan recall, mengonfirmasi keseimbangan yang baik dalam prediksi. Secara keseluruhan, model menunjukkan kemampuan generalisasi yang sangat baik.

5. KESIMPULAN

Berdasarkan hasil penelitian, yang berjudul "PREDIKSI *CHURN* PELANGGAN PADA LAYANAN DESAIN GRAFIS MENGGUNAKAN ALGORITMA *NAÏVE BAYES*" mempunyai 2 kesimpulan utama yang diperoleh adalah sebagai berikut :

1. hasil analisis menunjukkan bahwa terdapat enam atribut utama yang secara signifikan memengaruhi *churn* pelanggan pada sektor jasa desain grafis dan percetakan, yaitu jenis layanan, jumlah revisi, harga layanan, status pembayaran, kategori revisi, dan keterlambatan pembayaran. Atribut-atribut ini memberikan wawasan penting terkait perilaku pelanggan yang berpotensi berhenti menggunakan layanan. Pelanggan yang memilih jenis layanan dengan tingkat kompleksitas tinggi, ditambah dengan jumlah revisi yang banyak, cenderung lebih sering mengalami *churn*. Ketidak puasan pelanggan terhadap kualitas layanan yang tercermin dari revisi yang sering, serta harga layanan yang dianggap tidak kompetitif, meningkatkan kemungkinan *churn*. Selain itu, keterlambatan pembayaran dan status pembayaran yang tertunda juga menjadi faktor utama yang menunjukkan risiko pelanggan berhenti berlangganan. Pemahaman yang mendalam mengenai faktor-faktor ini sangat penting bagi perusahaan dalam merancang strategi retensi yang lebih efektif dan terarah.
2. performa model prediksi *churn*, algoritma *Naïve Bayes* terbukti sangat efektif dengan menunjukkan hasil evaluasi yang luar biasa, yakni akurasi sebesar 99%, presisi 1.00, recall 0.99, F1-score 0.99, dan ROC-AUC 0.99. Keunggulan utama Naïve Bayes terletak pada kemampuannya dalam mengklasifikasikan pelanggan *churn* dan *non-churn* dengan sangat baik, sekaligus menawarkan efisiensi dalam pemrosesan data besar. Dibandingkan dengan metode lain seperti XGBoost yang memiliki akurasi 97% dan AUC 0.995, Naïve Bayes unggul dalam hal efisiensi waktu

komputasi dan kemudahan implementasi. Dengan demikian, model prediksi *churn* menggunakan Naïve Bayes memberikan kontribusi yang signifikan bagi perusahaan dalam pengambilan keputusan strategis berbasis data. Model ini memungkinkan identifikasi pelanggan berisiko secara lebih cepat dan akurat, serta memfasilitasi perusahaan dalam mengarahkan sumber daya secara lebih efisien untuk mencegah *churn*, menjadikannya pilihan yang sangat relevan untuk diterapkan dalam sektor jasa desain grafis dan percetakan.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang mendalam kepada semua pihak yang telah memberikan dukungan terhadap penelitian ini, khususnya kepada ayah dan ibu tercinta yang telah memberikan tenaga, materi, ilmu, serta do'a yang tiada henti. Ucapan terima kasih juga ditujukan kepada adik tercinta yang menjadi sumber motivasi untuk terus belajar dan menggali ilmu baru. Selain itu, penulis juga menghaturkan apresiasi kepada kakek, nenek, dan seluruh keluarga besar atas segala bentuk dukungan yang diberikan.

DAFTAR PUSTAKA

[1] A. P. Kusandi and S. Informasi, "Fakultas sains dan teknologi universitas buddhi dharma tangerang 2022," 2022.

[2] Miryam Clementine and Arum, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi NaÃ¯ve Bayes dan ID3," *J. Process.*, vol. 17, no. 1, pp. 9–18, 2022, doi: 10.33998/processor.2022.17.1.1170.

[3] D. Rika Widianita, "No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title," *AT-TAWASSUTH J. Ekon. Islam*, vol. VIII, no. I, pp. 1–19, 2023.

[4] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," *Appl. Soft Comput.*, 2023, doi: 10.1016/j.asoc.2023.110103.

[5] S. Leonova and I. Gotskivska, "Building customer loyalty using smm tools on the example of beauty studio lima beauty studio," 2021, doi: 10.32840/2522-4263/2021-1-16.

[6] J. O. Ogunleye, *The Concept of Data Mining*. 2021. doi: 10.5772/INTECHOPEN.99417.

[7] R. Novendri, R. Andreswari, and O. N. Pratiwi, "Implementasi Data Mining Untuk Memprediksi Customer Churn Menggunakan Algoritma Naive Bayes," *e-Proceeding Eng.*, vol. 8, no. 2, pp. 2762–2773, 2021.

[8] C. A. Rahayu, "Prediksi Penderita Diabetes Menggunakan Metode Naive Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, 2023, doi: 10.23960/jitet.v11i3.3055.

[9] A. Yulianto and F. Firmansyah, "Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes," *Remik*, vol. 6, no. 1, pp. 41–47, 2021, doi: 10.33395/remik.v6i1.11196.

[10] R. R. Adhitya, Wina Witanti, and Rezki Yuniarti, "Perbandingan Metode Cart Dan Naïve Bayes Untuk Klasifikasi Customer Churn," *INFOTECH J.*, vol. 9, no. 2, pp. 307–318, 2023, doi: 10.31949/infotech.v9i2.5641.

[11] "Churn Prediction Models Using Gradient Boosted Tree and Random Forest Classifiers," International Conferences on Computing Advancements, 2023. doi: 10.1109/ICCA51723.2023.10181933.

[12] R. C. Admanegara and W. Handayani, "Customer Churn Analysis Using Machine Learning to Improve Customer Retention on Vissie Net," vol. 12, no. 09, pp. 7379–7387, 2024, doi: 10.18535/ijsrm/v12i09.em05.