Bulletin of Information System Research (BIOS)
Volume 3, No 1, December 2024 Page: 39-45
ISSN 2963-2455 (media online)
https://journal.grahamitra.id/index.php/bios

Komparasi Metode Decision Tree dan K-Nearest Neighbor (KNN) dalam Memprediksi Costumer Churn Pada Perusahaan Telekomunikasi

Khadisah Syah Riebhan Palluvi[1,*], Nadyari Syaada[2], Bunga Intan[3]

Universitas Bina Insan, Lubuklinggau, Indonesia
Email: [1,*]khadisah.syah21@gmail.com, [2]Nadyarisya31@gmail.com, [3]bungaintan@univbinainsan.ac.id
Email Penulis Korespondensi: khadisah.syah21@gmail.com

Abstrak−Prediksi customer churn bertujuan untuk mengklasifikasikan data pelanggan sebelumnya menjadi dua kategori: pelanggan yang akan berhenti berlangganan dan pelanggan yang akan terus berlangganan. Prediksi tersebut memanfaatkan ilmu data mining peran klasifikasi yang merupakan menempatkan variabel atau objek ke dalam beberapa kategori relevan yang telah ditetapkan sebelumnya. Dalam proses eksekusi data mining, diperlukan sebuah algoritma yang dapat mengklasifikasikan apakah customer churn atau tidak churn. Data yang digunakan terdiri dari 7043 rows dan 21 columns. Didalam data tersebut salah satu kolom akan dijadikan label yaitu kolom 'Churn'. Dalam proses prediksi churn, algoritma yang digunakan yaitu Decision Tree dan K-Nearest Neighbor. Dari hasil analisis yang dilakukan, pada algoritma KNN dihasilkan 76% dan Decision Tree 72%. Dengan hasil pemodelan akurasi 72% dan 76%, keduanya memenuhi kriteria kesuksesan >70%. Namun, model KNN dengan akurasi 76% lebih baik dan lebih diinginkan karena memberikan prediksi yang lebih akurat.

Kata Kunci: Decision Tree; KNN; Prediksi; Churn

Abstract−Customer churn prediction aims to classify previous customer data into two categories: customers who will unsubscribe and customers who will continue to subscribe. The prediction utilizes the science of data mining, the role of classification, which is placing variables or objects into several previously defined relevant categories. In the data mining execution process, an algorithm is needed that can classify whether a customer churns or not. The data used consists of 7043 rows and 21 columns. In the data, one of the columns will be used as a label, namely the 'Churn' column. In the churn prediction process, the algorithms used are Decision Tree and K-Nearest Neighbor. From the results of the analysis carried out, the KNN algorithm produced 76% and Decision Tree 72%. With the results of modeling accuracy of 72% and 76%, both meet the success criteria> 70%. However, the KNN model with an accuracy of 76% is better and more desirable because it provides more accurate predictions.

Keywords: Decision Tree; KNN; Prediction; Churn

1. PENDAHULUAN

Dalam era bisnis modern, mempertahankan pelanggan yang ada seringkali lebih menguntungkan dibanding menarik pelanggan baru dan juga membantu perusahaan menjual lebih banyak produk. Masalah yang dihadapi oleh perusahaan adalah bagaimana mencegah fenomena customer churn ini, yang terjadi ketika pelanggan berhenti berlangganan layanan atau produk perusahaan. Fenomena ini menjadi masalah kritis karena dapat menyebabkan penurunan pendapatan dan mengganggu stabilitas bisnis perusahaan. Untuk mengatasi masalah ini, diperlukan metode prediksi yang efektif [1][2].

Prediksi customer churn bertujuan untuk mengklasifikasikan data pelanggan sebelumnya menjadi dua kategori: pelanggan yang akan berhenti berlangganan dan pelanggan yang akan terus berlangganan. Prediksi tersebut memanfaatkan ilmu data mining peran klasifikasi yang merupakan menempatkan variabel atau objek ke dalam beberapa kategori relevan yang telah ditetapkan sebelumnya. Dalam proses eksekusi data mining, diperlukan sebuah algoritma yang dapat mengklasifikasikan apakah customer churn atau tidak churn[3][4].

Industri penyedia jasa telekomunikasi merupakan industri yang terus berkembang dan selalu dibutuhkan masyarakat. Dengan semakin banyaknya jumlah perusahaan telekomunikasi baik penyedia layanan GSM (*Global System Mobile*) maupun CDMA (*Code Division Multiple Access*), masing-masing akan saling menerapkan strategi untuk memperebutkan perhatian pelanggan dan menguasai pasar. Berbagai cara dilakukan dalam mendukung strategi tersebut, seperti: penerapan tarif murah, penyediaan layanan/ fitur khusus kepada pelanggan, undian berhadiah, bonus pulsa, jaminan minimalisasi call drop, ataupun lainnya. Semua hal tersebut bertujuan untuk mempertahankan atau menambah revenue yang didapat oleh perusahaan, serta landasan bahwa biaya untuk mempertahankan pelanggan akan lebih murah dibandingkan biaya untuk menarik pelanggan baru[5]. Untuk itu dalam proses prediksi churn algoritma yang digunakan adalah decision tree dan K-Nearest Neighbor (K-NN).

Algoritma K-nearest neighbor (K-NN) merupakan penelitian menggunakan metode dengan mencari kedekatan antara kriteria kasus baru dengan beberapa kriteria kasus lama berdasarkan criteria kasus yang paling mendekati. Algoitma K-nearest neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data latih yang jaraknya paling dekat dengan objek tersebut. Ketepatan algoritma K-NN ini sangat di pengaruhi oleh ada atau tidaknya kriteria-kriteria yang tidak relevan, atau jika bobot kriteria tersebut tidak setara dengan relevansinya terhadap klasifikasi[6][7][8].

Metode Decision tree merupakan metode yang ada pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon pohon keputusan yang mempresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara jumlah calon variable input dengan sebuah variabel target[9][10][11].

Bulletin of Information System Research (BIOS)
Volume 3, No 1, December 2024 Page: 39-45
ISSN 2963-2455 (media online)
https://journal.grahamitra.id/index.php/bios

Penelitian ini akan melakukan perbandingan antara algoritma K-nearest neighbor dan Metode Decision tree, penelitian ini diharapkan dapat memberikan beberapa manfaat, antara lain: peningkatan akurasi prediksi churn, pengembangan strategi retensi pelanggan yang lebih efektif, pengembangan model prediktif yang handal, dan kontribusi terhadap literasi data mining.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan fokus pada analisis data yang dikumpulkan dari Github. Data ini akan diolah melalui klasifikasi menggunakan algoritma Decision Tree dan K-Nearest Neighbor (K-NN).

2.1 Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan guna mendapatkan suatu informasi dalam menyelesaikan permasalahan. Berikut merupakan metode pengumpulan data yang digunakan penulis dalam melakukan penelitian:

1. Studi Pustaka
   Studi Pustaka merupakan suatu proses pengumpulan data kepustakaan dilakukan dengan cara mengumpulkan materi melalui jurnal, literatur, buku ataupun situs internet sebagai sumber Pustaka yang berkaitan dengan penelitian terutama tentang memprediksi costumer churn dengan metode decision tree dan K-Nearest Neighbor (K-NN).
2. Data Primer
   Data primer merupakan data yang diperoleh secara langsung dari objek penelitian. Data dalam penelitian ini diperoleh dari Github.

2.2 Metode Analisa

Penelitian ini menggunakan metode analisis dengan menerapkan algoritma decision tree dan K- Nearest Neigbor (K-NN). Adapun penjelasannya sebagai berikut:

1. *Decision Tree*
   Decision Tree merupakan salah satu algoritma pada pengolahan data metode klasifikasi. Algoritma machine learning yang satu ini mempresentasikan strukturnya seperti struktur pohon untuk menemukan hasil keputusan[12][13]. Konsep dari Decision Tree ini adalah dengan menyajikan pernyataanpernyataan bersyarat pada setiap langkah yang bercabang untuk pengambilan keputusan berdasarkan perhitungan dari data set itu sendiri. Dalam metode Klasifikasi, algoritma Decision Tree merupakan yang paling banyak digunakan[14][15][16].
2. *K-Nearest Neighbor* (KNN)
   Berbeda dengan model algoritma Decision Tree, K-Nearest Neighbor (KNN) merupakan algoritma yang melakukan klasifikasi data berdasarkan jarak paling dekat dengan objek target[17]. Konsep dari K-Nearest Neighbor (KNN) sangat sederhana, algoritma ini berjalan dengan mengklasifikasikan berdasarkan kesamaan ciri-ciri terhadap kelompok tertentu dari titik data terdekat atau tetangganya[18]. Dengan konsep seperti itu, K-Nearest Neighbor (KNN) akan memberikan hasil akhir yang kompetitif[19][20].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan terdiri dari 7043 rows dan 21 columns. Didalam data tersebut salah satu kolom akan dijadikan label yaitu kolom ‘Churn’. Data tersebut dapat dilihat pada gambar 1 berikut:

```
[246] df = pd.read_csv('Telco-Customer-Churn1.csv')

[247] df
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtectio
0	7590-VHVEG	Female	0	Yes	No	1.0	No	No phone service	DSL	No	...	N
1	5575-GNVDE	Male	0	No	No	34.0	Yes	No	DSL	Yes	...	Ye
2	3668-QPYBK	Male	0	No	No	2.0	Yes	No	DSL	Yes	...	N
3	7795-CFOCW	Male	0	No	No	45.0	No	No phone service	DSL	Yes	...	Ye
4	9237-HQITU	Female	0	No	No	2.0	Yes	No	Fiber optic	No	...	N
...	...	...	...	...	...	...	...	...	...	...	...	.
7038	6840-RESVB	Male	0	Yes	Yes	24.0	Yes	Yes	DSL	Yes	...	Ye

Gambar 1. Data Mentah

Tipe data yang terkumpul yaitu object, int dan float. Data yang bukan numerik akan diubah menjadi numerik agar dapat melakukan pemodelan KNN dan Decision Tree. Dapat dilihat pada gambar 2 berikut:

Bulletin of Information System Research (BIOS)
Volume 3, No 1, December 2024 Page: 39-45
ISSN 2963-2455 (media online)
https://journal.grahamitra.id/index.php/bios

```
df.dtypes

customerID          object
gender              object
SeniorCitizen        int64
Partner             object
Dependents          object
tenure             float64
PhoneService        object
MultipleLines       object
InternetService     object
OnlineSecurity      object
OnlineBackup        object
DeviceProtection    object
TechSupport         object
StreamingTV         object
StreamingMovies     object
Contract            object
PaperlessBilling    object
PaymentMethod       object
MonthlyCharges     float64
TotalCharges        object
Churn               object
dtype: object
```

Gambar 2. Tipe Data

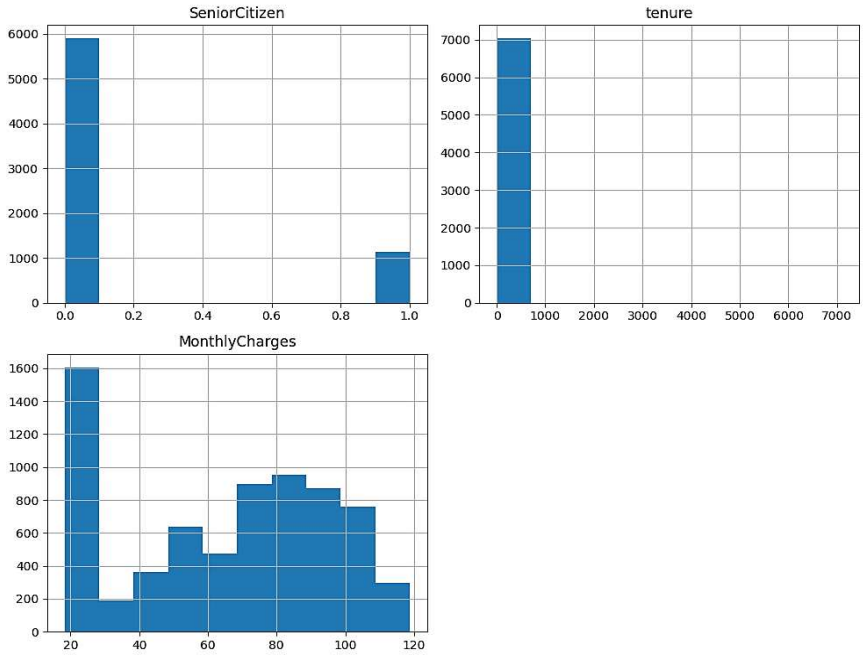
Adapun deskripsi statistic dasar yang menampilkan kolom dengan tipe data numerik, yaitu SeniorCitizen, tenure, MonthlyCharges.

```
df.describe()
```

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7040.000000	7043.000000
mean	0.162147	35.043892	64.761692
std	0.368612	115.282871	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	7100.000000	118.750000

Gambar 3. Deskripsi Statistic

Berdasarkan Gambar 3 diatas, Karakteristik data yang terkumpul dengan visualisasi grafik dapat dilihat sebagai berikut:



Gambar 4. Histogram Distribusi Data

Bulletin of Information System Research (BIOS)
Volume 3, No 1, December 2024 Page: 39-45
ISSN 2963-2455 (media online)
https://journal.grahamitra.id/index.php/bios

Gambaran 4 distribusi data. Jika distribusi data simetris dan berbentuk, maka data cenderung normal. Jika distribusi memiliki ekor panjang di salah satu sisi, maka data mungkin memiliki skewness. Skewness merupakan ukuran statistik yang menggambarkan kemiringan atau asimetri dari distribusi data. Ini mengindikasikan seberapa simetris atau tidak simetris distribusi data tersebut. Kesimpulan dari visualisasi histogram diantaranya: 1) SeniorCitizen, Mayoritas pelanggan bukan warga senior. 2) Tenure, Sebagian besar pelanggan memiliki masa keanggotaan yang sangat pendek, dengan beberapa outlier yang mungkin perlu diteliti lebih lanjut. 3) MonthlyCharges, Pelanggan membayar biaya bulanan yang bervariasi, dengan sebagian besar membayar di kisaran rendah hingga menengah.

3.2 Pre-processing

Tahap selanjutnya setelah data didapatkan adalah melakukan preprocessing data. Preprocessing data adalah salah satu tahap yang dilakukan setelah dataset berhasil didapatkan, kemudian dibersihkan dari berbagai elemen yang tidak relevan untuk penelitian dan akhirnya mendapatkan dataset yang berkualitas(Rizka Yudana et al., 2023). Dalam penelitian ini, dataset akan melalui preprocessing data dengan beberapa. Pertama ada Feature Selection, dimana terdapat satu atribut bernama “No_rows” yang tidak akan dimasukkan ke dalam pemrosesan data karena tidak relevan dan hanya berfungsi sebagai penomoran data di luar pemrosesan. Kemudian terdapat Missing Data Treatment yang dilakukan kepada beberapa atribut yang memiliki missing value. Adapun atribut yang memiliki missing value dapat dilihat sebagai berikut:

```
# Mengecek missing values
print(df.isnull().sum())
```

```
customerID          0
gender              5
SeniorCitizen       0
Partner             0
Dependents          0
tenure              3
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges        0
Churn               0
dtype: int64
```

Gambar 5. Missing Values

Dengan hasil Analisa Gambar 4 di atas ditemukan nilai missing values pada kolom gender dan tenure, hal tersebut diperlukan untuk membersihkan data atau menggunakan Teknik imputasi untuk mengisi nilai yang hilang. Setelah berhasil membersihkan dataset dari missing value, selanjutnya adalah memberikan label kepada salah satu atribut. Label sendiri sangat diperlukan pada metode klasifikasi karena role atribut ini digunakan sebagai identifier dan mengklasifikasi bagi objek atau data yang sedang olah(Yudiana et al., 2023). Pada penelitian data set ini, yang berperan sebagai label adalah atribut “Churn”.

3.3 Pengujian Akurasi Algoritma

Pengimplementasian kepada algoritma Decision Tree, K-Nearest Neighbor (KNN) pertama-tama dilakukan untuk testing dan menentukan algoritma mana yang memiliki kelayakan dan akurasi paling tinggi pada dataset target. Testing dilakukan menggunakan hasil split data yang telah dihasilkan sebelumnya. 80% data yang merupakan data training akan dihubungkan/dimasukkan ke dalam algoritma untuk menjadi model atau sumber perhitungan. Kemudian algoritma yang akan diukur keakuratannya dihubungkan dengan operator Apply Model bersamaan dengan 20% data tersisa sebagai data testing. Pengujian pertama dilakukan kepada Algoritma Decision Tree, Pada hasil pengujian ini didapatkan akurasi dari Decision Tree dengan akurasi sebesar 72%. Berikut Gambar 6 rincian hasil testing Decision Tree:

Bulletin of Information System Research (BIOS)
Volume 3, No 1, December 2024 Page: 39-45
ISSN 2963-2455 (media online)
https://journal.grahamitra.id/index.php/bios

```
# Membuat dan melatih model DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)

# Memprediksi label untuk data uji
dt_pred = dt.predict(X_test)

# Menghitung akurasi
print('Decision Tree Accuracy:', accuracy_score(y_test, dt_pred))
print('Decision Tree Classification Report:\n', classification_report(y_test, dt_pred))
```

```
Decision Tree Accuracy: 0.7281760113555713
Decision Tree Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.81      0.81      1036
           1       0.49      0.50      0.49       373

    accuracy                           0.73      1409
   macro avg       0.65      0.65      0.65      1409
weighted avg       0.73      0.73      0.73      1409
```

Gambar 6. Pemodelan Decision Tree

Feature importance yang dihasilkan pada model ini yang pertama yaitu MonthlyCharges, dilanjutkan TotalCharges, tenure dan gender.

```
Decision Tree Feature Importance:
                  Importance
MonthlyCharges      0.377592
TotalCharges        0.365444
tenure              0.219856
gender              0.037108
```

Gambar 7. Decision Tree Feature Importance

Selanjutnya, pengujian dilakukan kepada Algoritma K-Nearest Neighbor (KNN). Pengujian dilakukan dengan split data yang masih sama. Setelah dilakukan pengujian didapatkan hasil akurasi sebesar 76%.

```
# Membuat dan melatih model KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train)

# Memprediksi label untuk data uji
knn_pred = knn.predict(X_test)

# Menghitung akurasi
print('KNN Accuracy:', accuracy_score(y_test, knn_pred))
print('KNN Classification Report:\n', classification_report(y_test, knn_pred))
```

```
KNN Accuracy: 0.7686302342086586
KNN Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.87      0.85      1036
           1       0.57      0.49      0.53       373

    accuracy                           0.77      1409
   macro avg       0.70      0.68      0.69      1409
weighted avg       0.76      0.77      0.76      1409
```

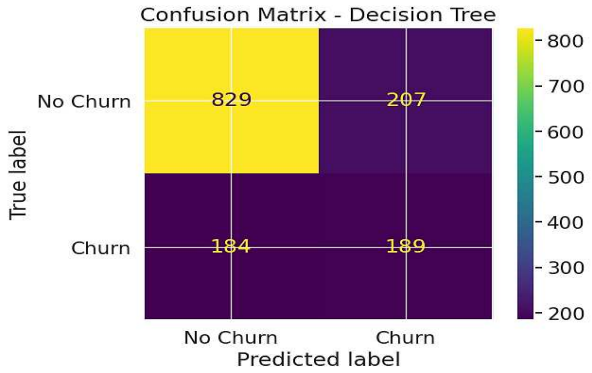
Gambar 8. Pemodelan K-Nearest Neighbor

Untuk model K-Nearest Neighbors (KNN) tidak memiliki atribut feature_importances. Atribut feature_importances_ adalah khusus untuk beberapa model, seperti Decision Tree, Random Forest, dan model tree-based lainnya, yang bisa mengukur seberapa penting setiap fitur dalam membuat prediksi teks tebal.

Bulletin of Information System Research (BIOS)
Volume 3, No 1, December 2024 Page: 39-45
ISSN 2963-2455 (media online)
https://journal.grahamitra.id/index.php/bios

3.4 Confusion Matrix

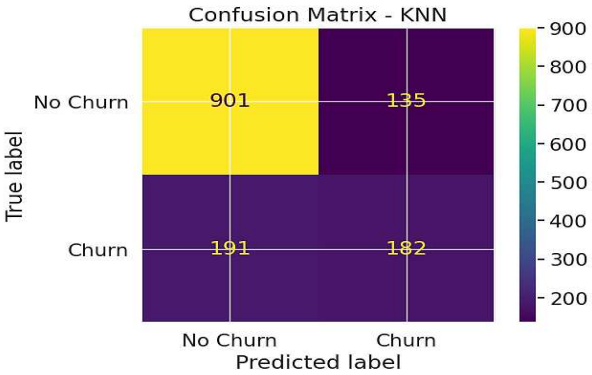
1. Decision Tree



Gambar 9. Confusion Matrix Decision Tree

Berdasarkan gambar 9 diatas, dapat dijelaskan bahwa True Label "No Churn" (Baris pertama) terdiri atas Predicted "No Churn": 829 pelanggan diprediksi tidak churn dan mereka benar-benar tidak churn dan Predicted "Churn": 207 pelanggan diprediksi churn, tetapi mereka sebenarnya tidak churn. Sedangkan True Label "Churn" (Baris kedua) terdiri atas Predicted "No Churn": 184 pelanggan diprediksi tidak churn, tetapi mereka sebenarnya churn dan Predicted "Churn": 189 pelanggan diprediksi churn dan mereka benar-benar churn.

2. K-Nearest Neighbor (KNN)



Gambar 10. Confusion Matrix K-Nearest Neighbor

Berdasarkan gambar 10 diatas, dapat dijelaskan bahwa True Label "No Churn" (Baris pertama) terdiri atas Predicted "No Churn": 901 pelanggan diprediksi tidak churn dan mereka benar-benar tidak churn. Dan Predicted "Churn": 135 pelanggan diprediksi churn, tetapi mereka sebenarnya tidak churn. Sedangkan True Label "Churn" (Baris kedua) terdiri atas Predicted "No Churn": 191 pelanggan diprediksi tidak churn, tetapi mereka sebenarnya churn. Dan Predicted Churn": 182 pelanggan diprediksi churn dan mereka benar-benar churn.

4. KESIMPULAN

Dari hasil analisis yang dilakukan, pada algoritma KNN dihasilkan 76% dan Decision Tree 72%. Dengan hasil pemodelan akurasi 72% dan 76%, keduanya memenuhi kriteria kesuksesan >70%. Namun, model KNN dengan akurasi 76% lebih baik dan lebih diinginkan karena memberikan prediksi yang lebih akurat. Kedua model mungkin belum mencapai performa optimal jika tidak dilakukan tuning hyperparameter yang maksimal. Untuk penelitian selanjutnya, dapat digunakan teknik seperti *grid search* atau *random search* untuk menemukan konfigurasi terbaik bagi KNN dan Decision Tree. Ini bisa memengaruhi hasil akhir secara signifikan.

REFERENCES

[1] D. Putriani, A. P. A. Prayogi, A. I. Shofyana, A. Ristyawan, and E. Daniati, "Prediksi Customer Churn Menggunakan Algoritma Decision Tree," *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 8, no. 1, pp. 85–94, 2024, doi: 10.29407/inotek.v8i1.4914.

[2] S. D. Damanik and M. I. Jambak, "Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma C4. 5," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1303–1309, 2023, doi: 10.30865/klik.v3i6.829.

[3] D. Fatmawati, W. Trisnawati, Y. Jumaryadi, and G. Triyono, "Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1056–1062, 2023, doi: 10.30865/klik.v3i6.803.
[4] C. Hardjono and S. M. Isa, "Implementation of Data Mining for Churn Prediction in Music Streaming Company Using 2020 Dataset," *Journal on Education*, vol. 5, no. 1, pp. 1189–1197, 2022, doi: 10.31004/joe.v5i1.740.
[5] M. Yunus and N. K. A. Pratiwi, "Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskemas Cakranegara," *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, vol. 4, no. 4, pp. 221–231, 2023, doi: 10.35746/jtim.v4i4.328.
[6] R. Rismala, I. Ali, and A. R. Rinaldi, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 585–590, 2023, doi: 10.36040/jati.v7i1.6419.
[7] A. P. Silalahi, H. G. Simanullang, and M. I. Hutapea, "Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita," *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 7, no. 1, pp. 144–149, 2023, doi: 10.46880/jmika.Vol7No1.pp144-149.
[8] A. A. Karim, M. A. Prasetyo, and M. R. Saputro, "Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia," *Seminar Nasional Teknologi & Sains*, vol. 2, no. 1, pp. 377–382, 2023, doi: 10.29407/stains.v2i1.2877.
[9] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4. 5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," *Jurnal Krisnadana*, vol. 1, no. 2, pp. 21–32, 2022, doi: 10.58982/krisnadana.v1i2.108.
[10] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and R. Rahmaddeni, "Implementasi Algoritma Decision Tree C4. 5 dan Support Vector Regression untuk Prediksi Penyakit Stroke: Implementation of Decision Tree Algorithm C4. 5 and Support Vector Regression for Stroke Disease Prediction," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 61–67, 2022, doi: 10.57152/malcom.v2i2.426.
[11] A. Fatkhudin, M. Y. Febrianto, F. A. Artanto, M. W. N. Hadinata, and R. Fahlevi, "Algoritma Decision Tree C. 45 dalam analisa kelulusan mahasiswa Program Studi Manajemen Informatika UMPP," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 8, no. 2, pp. 83–86, 2022, doi: 10.35329/jiik.v8i2.240.
[12] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru: Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.
[13] B. N. Azmi, A. Hermawan, and D. Avianto, "Analisis Pengaruh komposisi data training dan data testing Pada penggunaan PCA Dan Algoritma decision tree untuk KLASIFIKASI Penderita Penyakit liver," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 281–290, 2023, doi: 10.35746/jtim.v4i4.298.
[14] A. H. Nasrullah, "Implementasi algoritma Decision Tree untuk klasifikasi produk laris," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.
[15] Y. A. Singgalen, "Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 323–332, 2023, doi: 10.30865/klik.v4i1.1153.
[16] I. L. Kharisma, D. A. Septiani, A. Fergina, and K. Kamdan, "Penerapan Algoritma Decision Tree untuk Ulasan Aplikasi Vidio di Google Play," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 218–226, 2023, doi: 10.25077/TEKNOSI.v9i2.2023.218-226.
[17] A. Homaidi and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
[18] A. R. D. Nugraha, K. Auliasari, and Y. A. Pranoto, "Implementasi Metode K-Nearest Neighbor (KNN) Untuk Seleksi Calon Karyawan Baru (Studi Kasus: BFI Finance Surabaya)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 2, pp. 14–20, 2020, doi: 10.36040/jati.v4i2.2656.
[19] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
[20] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.