



Prediksi Kerajinan Siswa Berdasarkan GPA Menggunakan *K-Nearest Neighbors* (KNN)

Maulana Fansyuri^{1*}, Deni Setiawan², Amalia Azzahra³, Dhaifina Nabila⁴, Fijriani Silviana⁵, Lusiyanti⁶

^{1,2,3,4,5,6}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ^{1*}dosen02359@unpam.ac.id, ⁵fijrianis@gmail.com

(* : coressponding author)

Abstrak – Penelitian ini bertujuan untuk memprediksi kesibukan siswa berdasarkan Nilai IPK (Indeks Prestasi Kumulatif) menggunakan algoritma *K-Nearest Neighbors* (KNN). Prestasi akademik, yang tercermin dalam IPK, merupakan indikator penting yang mempengaruhi kesibukan siswa dalam aktivitas belajar serta keterlibatan mereka dalam kegiatan ekstrakurikuler. Penelitian ini menggunakan dataset yang mencakup 500 rekaman siswa, dengan atribut seperti jam belajar per minggu, tingkat kehadiran, partisipasi dalam kegiatan ekstrakurikuler, jenis kelamin, dan informasi demografis lainnya. Algoritma KNN diterapkan untuk menganalisis data dan membangun model prediktif yang dapat digunakan untuk mengidentifikasi tingkat kesibukan siswa. Akurasi model diuji melalui teknik validasi silang untuk memastikan keandalannya. Hasil awal menunjukkan bahwa faktor-faktor seperti jam belajar dan tingkat kehadiran memiliki hubungan signifikan dengan tingkat kesibukan siswa. Penelitian ini menyoroti potensi algoritma KNN dalam konteks pendidikan, memberikan wawasan berharga bagi pendidik untuk mendukung siswa yang mungkin memerlukan perhatian lebih. Selain itu, penelitian ini membuka peluang untuk penelitian lebih lanjut guna meningkatkan akurasi prediksi dan merancang intervensi yang lebih efektif.

Kata Kunci: Kesibukan Siswa; Indeks Prestasi Kumulatif (IPK); *K-Nearest Neighbors* (KNN); Prediksi Akademik.

Abstract – This study aims to predict student diligence based on their Grade Point Average (GPA) using the *K-Nearest Neighbors* (KNN) algorithm. Academic performance, as indicated by GPA, is a crucial factor influencing students' engagement in learning activities and their overall diligence. The research utilizes a dataset comprising 500 student records, which includes attributes such as GPA, study hours per week, attendance rates, participation in extracurricular activities, and demographic information. The KNN algorithm is employed to analyze the data and build a predictive model. The model's accuracy is validated through cross-validation techniques to ensure reliability. Preliminary results indicate that factors such as study hours and attendance significantly correlate with student diligence levels. This research highlights the potential of machine learning techniques like KNN in educational settings, providing valuable insights for educators to identify students who may require additional support. Furthermore, it opens avenues for future research aimed at enhancing prediction accuracy and developing targeted interventions.

Keywords: Student Diligence; Grade Point Average (GPA); *K-Nearest Neighbors* (KNN); Educational Data Analysis.

1. PENDAHULUAN

Kerajinan siswa dalam belajar merupakan salah satu faktor penting yang menentukan keberhasilan akademik, namun pengukuran tingkat kerajinan sering kali sulit dilakukan secara langsung. Salah satu indikator yang sering digunakan untuk menilai hasil akademik siswa adalah nilai Grade Point Average (GPA). GPA mencerminkan kemampuan dan upaya siswa dalam menyelesaikan tugas akademik, namun tidak selalu menggambarkan tingkat kerajinan mereka dalam belajar. Oleh karena itu, untuk memprediksi tingkat kerajinan siswa secara lebih akurat, diperlukan pendekatan yang lebih terstruktur dengan memanfaatkan data yang tersedia, seperti GPA (Gharbi & Jaziri, 2021). Penerapan algoritma *K-Nearest Neighbors* (KNN) sebagai metode pembelajaran mesin menawarkan potensi untuk memetakan hubungan antara GPA dan kerajinan siswa secara lebih efektif (Singh & Mishra, 2020).

Dalam beberapa tahun terakhir, berbagai penelitian telah mencoba mengembangkan model prediksi kinerja akademik siswa menggunakan teknik pembelajaran mesin, termasuk algoritma KNN (Gharbi & Jaziri, 2021) mengidentifikasi bahwa data GPA merupakan prediktor yang baik dalam memodelkan kinerja akademik siswa, sementara (Al-Zubaidi & El-Sayed, 2020)



menunjukkan keberhasilan KNN dalam memprediksi performa akademik siswa berdasarkan berbagai parameter pendidikan. Penelitian-penelitian tersebut menunjukkan bahwa dengan menggunakan algoritma seperti KNN, institusi pendidikan dapat melakukan prediksi yang lebih akurat tentang potensi kerajinan siswa, yang berguna dalam merancang intervensi atau dukungan lebih lanjut.

Penelitian ini mengusulkan penggunaan algoritma *K-Nearest Neighbors* (KNN) untuk memprediksi tingkat kerajinan siswa berdasarkan nilai GPA mereka. Algoritma KNN dipilih karena kemampuannya untuk mengenali pola dalam data dan membuat prediksi berdasarkan kedekatan data. Dengan membagi data siswa ke dalam kategori kerajinan tertentu dan menghubungkannya dengan nilai GPA mereka, model ini dapat memberikan gambaran yang lebih jelas mengenai hubungan antara GPA dan kerajinan siswa, serta memprediksi tingkat kerajinan siswa di masa depan. Dalam penelitian ini, data yang digunakan akan mencakup informasi GPA siswa yang dibagi dalam kategori tertentu untuk analisis lebih lanjut.

Penelitian ini membawa nilai keterbaruan dengan menggabungkan konsep prediksi akademik menggunakan KNN dengan fokus khusus pada tingkat kerajinan siswa. Meskipun sudah banyak penelitian yang menggunakan KNN untuk memprediksi performa akademik, sangat sedikit yang mengeksplorasi prediksi tingkat kerajinan secara langsung berdasarkan data GPA. Pendekatan ini memungkinkan untuk memberikan wawasan yang lebih dalam mengenai pola belajar siswa, yang dapat digunakan oleh lembaga pendidikan untuk merancang strategi pembelajaran yang lebih personal dan berbasis data. Dengan mengembangkan model prediktif ini, diharapkan dapat membantu meningkatkan upaya intervensi akademik yang lebih efektif, terutama dalam mendukung siswa yang membutuhkan perhatian lebih.

2. METODE

Penelitian ini dirancang untuk menganalisis hubungan antara kebiasaan belajar siswa dan performa akademik mereka dengan menggunakan algoritma *K-Nearest Neighbors* (KNN). Tujuan utamanya adalah untuk memprediksi nilai rata-rata siswa (GPA) berdasarkan berbagai variabel, seperti kebiasaan belajar (jam belajar per minggu dan tingkat kehadiran), faktor demografis (usia, jenis kelamin, dan jurusan), serta kegiatan tambahan seperti partisipasi dalam ekstrakurikuler dan pekerjaan paruh waktu. Dataset yang digunakan diperoleh dari situs Kaggle.com, yang merupakan salah satu basis data dalam kompetisi di bidang penglihatan komputer dan dapat diakses oleh publik dengan judul *Student Performance Prediction* (Farokhah, 2020). Dataset ini relevan dengan tujuan penelitian karena informasi yang disediakan mendukung analisis pola belajar siswa secara komprehensif. Data tersebut kemudian diproses secara pre-processing untuk memastikan bahwa data siap digunakan sebagai data target. Total data target yang terkumpul adalah 150 record. Proses klasifikasi melibatkan pencarian model atau fungsi yang dapat menggambarkan serta membedakan antara konsep atau kelas-kelas data. Tujuannya adalah untuk memprediksi kelas dari objek yang tidak memiliki label, sehingga memberikan perkiraan mengenai kelasnya (Krisandi, Helmi, & Prihandono, 2013). Data latih selanjutnya dimasukkan ke dalam algoritma K-Nearest Neighbor untuk mengembangkan model klasifikasi. Setelah model klasifikasi terbentuk, dilakukan uji akurasi menggunakan data uji. Alur dari model klasifikasi yang dibangun menggunakan K-Nearest Neighbor digunakan untuk mencari tetangga terdekat dari data uji dan melakukan klasifikasi berdasarkan mayoritas label kelas dari tetangga terdekat tersebut (Munandar & Munir, 2022).

2.1 Data Mining

Data mining adalah proses penemuan pola atau pengetahuan yang tersembunyi dalam sejumlah besar data. Proses ini melibatkan teknik analisis statistik, machine learning, dan algoritma untuk mengekstraksi informasi berharga dari dataset yang besar dan kompleks. Data mining digunakan untuk menganalisis data dengan cara yang memungkinkan untuk membuat prediksi atau mendapatkan wawasan yang tidak terduga sebelumnya. Tujuannya adalah untuk mengidentifikasi pola yang relevan, asosiasi, dan tren yang dapat membantu dalam pengambilan keputusan bisnis, prediksi, atau pemahaman lebih dalam mengenai fenomena tertentu.



Data Mining adalah proses menemukan pola dan pengetahuan menarik dari sejumlah besar data. Proses ini mencakup beberapa langkah, mulai dari pengumpulan data, pembersihan data, pemodelan, hingga evaluasi hasil. Beberapa teknik yang digunakan dalam data mining antara lain klasifikasi, regresi, clustering, dan asosiasi. Dengan memanfaatkan teknik-teknik ini, data mining dapat digunakan dalam berbagai aplikasi seperti analisis pasar, prediksi performa akademik, deteksi penipuan, hingga analisis perilaku pelanggan.

Dalam konteks pendidikan, data mining dapat diterapkan untuk memprediksi performa siswa, menganalisis kebiasaan belajar, atau mengidentifikasi faktor-faktor yang mempengaruhi hasil belajar. Misalnya, dengan menggunakan algoritma klasifikasi, kita dapat memprediksi siswa yang berpotensi mengalami kesulitan akademik, sehingga tindakan preventif dapat diambil untuk meningkatkan hasil belajar mereka.

2.2. Metode KNN (*K-Nearest Neighbors*)

K-Nearest Neighbors (KNN) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi (Han, Kamber, & Pei, 2012). Dalam konteks prediksi kerajinan siswa berdasarkan GPA, KNN digunakan untuk memprediksi tingkat kerajinan siswa dengan membandingkan data siswa yang akan diprediksi dengan siswa lain yang sudah ada dalam dataset pelatihan, berdasarkan kedekatannya. KNN mengklasifikasikan atau memprediksi dengan mencari k tetangga terdekat yang paling mirip dengan data yang ingin diprediksi, dan hasil prediksi didasarkan pada mayoritas nilai dari tetangga tersebut.

Langkah-langkah untuk menghitung metode Algoritma KNN :

a. Menentukan Nilai k

Nilai k merujuk pada jumlah tetangga terdekat yang dipertimbangkan dalam proses klasifikasi atau prediksi. Sebagai contoh, jika $k=3$, maka prediksi untuk siswa yang bersangkutan akan didasarkan pada tiga siswa yang memiliki kedekatan paling tinggi dengan data siswa yang akan diprediksi. Pemilihan nilai k yang optimal sangat penting untuk menghindari overfitting atau underfitting dalam model.

b. Menghitung Jarak

Untuk menentukan kedekatan antara data siswa yang akan diprediksi dan data pelatihan, kita harus menghitung jarak antar data. Dalam kasus ini, salah satu metrik yang sering digunakan adalah Euclidean Distance, yang dirumuskan sebagai berikut:

c. Menentukan Kelas atau Nilai Prediksi

Setelah menghitung jarak antara data yang akan diprediksi dengan semua data pelatihan, KNN kemudian memilih k tetangga terdekat berdasarkan jarak terkecil. Hasil prediksi diperoleh dengan dua cara, tergantung pada jenis masalah yang ditangani:

1. **Klasifikasi:** Jika masalahnya adalah klasifikasi (misalnya, mengklasifikasikan siswa sebagai "Rajin" atau "Tidak Rajin"), maka kelas yang paling sering muncul di antara k tetangga terdekat dipilih sebagai prediksi.
2. **Regresi:** Jika masalahnya adalah regresi (misalnya, memprediksi skor atau kerajinan siswa dalam skala tertentu), maka nilai rata-rata dari k tetangga terdekat dihitung dan digunakan sebagai nilai prediksi.

Dalam konteks penelitian ini, untuk memprediksi kerajinan siswa, kita akan menggunakan teknik regresi KNN, dengan menghitung rata-rata kerajinan siswa berdasarkan k tetangga terdekat.

3. ANALISA DAN PEMBAHASAN

3.1 Dataset

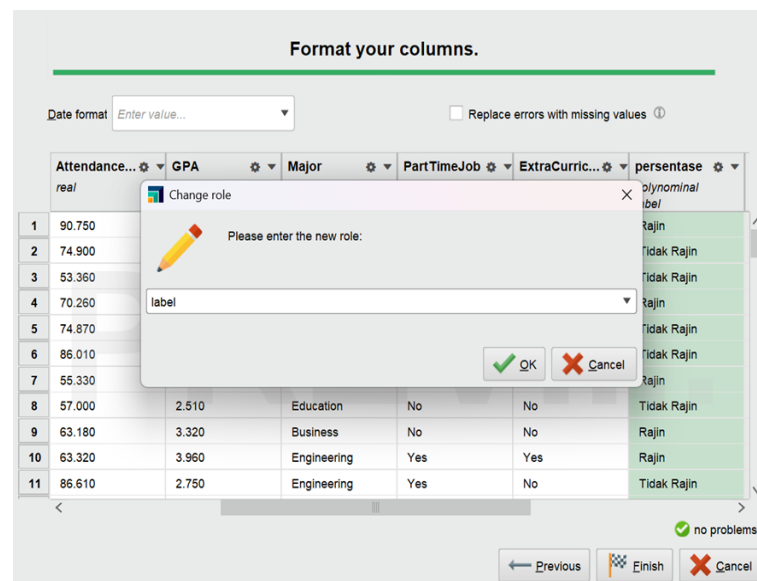
Tahapan pertama pada analisa ini yakni dengan memilih data yang bersumber dari website www.kaggle.com dengan judul "Student Performance Prediction". Pada penelitian ini menggunakan

satu dataset yang berisi 500 data terdiri dari 9 atribut, 4 atribut bertipe numerik, 5 atribut bertipe kategorik. Untuk lebih jelasnya dapat dilihat pada tabel 2 berikut ini.

Table 1. Atribut dan tipe data

Atribut	Tipe	Keterangan
StudentID	Numerik	ID unik untuk setiap siswa
Gender	Kategorik	Jenis kelamin siswa (laki-laki atau perempuan)
StudyHoursPerWeek	Numerik	Jumlah jam yang dihabiskan siswa untuk belajar per minggu
AttendanceRate	Numerik	Tingkat kehadiran siswa dalam persentase
GPA	Numerik	Nilai rata-rata siswa
Major	Kategorik	Jurusan atau program studi yang diambil siswa
PartTimeJob	Kategorik	Menunjukkan apakah siswa memiliki pekerjaan paruh waktu (Ya atau Tidak)
ExtraCurricularActivities	Kategorik	Menunjukkan apakah siswa berpartisipasi dalam kegiatan ekstrakurikuler (Ya atau Tidak)
persentase	Kategorik	Kategori kehadiran siswa (Rajin atau Tidak Rajin)

3.2 Transformation (Transformasi Data)



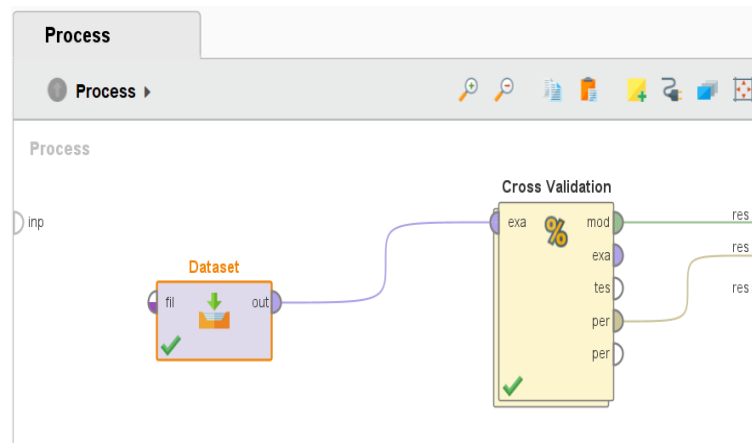
Gambar 1. Mengubah role atribut

Pada gambar 1 merupakan proses transformasi data yang akan diberikan label yang dibutuhkan algoritma K-Nearest Neighbor.

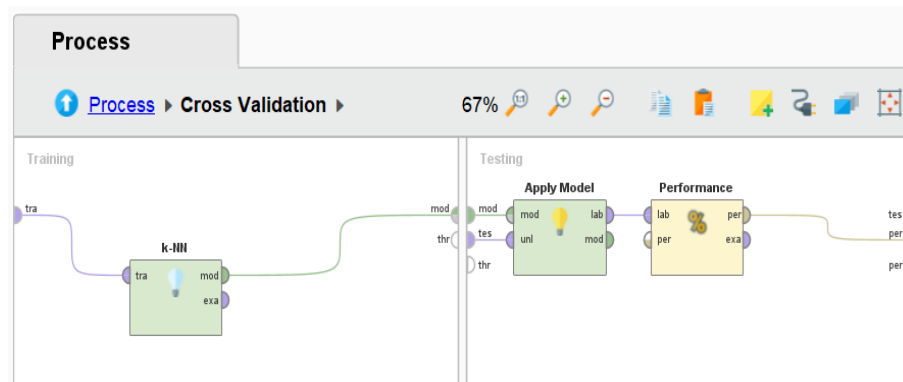
3.3 Data Mining

Pada bagian ini, kita akan menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan tool RapidMiner untuk menganalisis bagaimana faktor-faktor seperti jam belajar (*StudyHoursPerWeek*), tingkat kehadiran (*AttendanceRate*), pekerjaan paruh waktu (*PartTimeJob*), dan kegiatan

ekstrakurikuler (*ExtraCurricularActivities*) dapat memprediksi Indeks Prestasi Kumulatif (IPK) atau GPA mahasiswa.



Gambar 2. Process



Gambar 3. Cross Validation

Berdasarkan gambar diatas adalah model dari implementasi algoritma K-Nearest Neighbor (KNN) dengan tools RapidMiner yaitu dengan menjalankan beberapa operator diantaranya:

a. *Read Excel*

Operator ini digunakan untuk membaca data dari file Excel yang berisi dataset siswa. Di dalamnya terdapat informasi tentang StudentID, Gender, StudyHoursPerWeek, AttendanceRate, GPA, Major, PartTimeJob, ExtraCurricularActivities, dan persentase.

b. *Cross Validation*

Operator ini digunakan untuk mengevaluasi kinerja model KNN dengan membagi dataset menjadi beberapa bagian (fold). Setiap fold akan digunakan secara bergantian sebagai data testing, sedangkan fold lainnya sebagai data training. Hal ini dilakukan untuk mendapatkan estimasi kinerja model yang lebih akurat dan mencegah overfitting (Witten, Frank, & Hall, 2016).

c. *K-NN*

Operator ini merupakan algoritma K-Nearest Neighbor yang digunakan untuk memprediksi kelas mahasiswa. Algoritma ini akan mencari "tetangga" terdekat dari setiap data dalam data testing berdasarkan data training, dan memprediksi nilai IPK berdasarkan nilai IPK dari tetangga-tetangga tersebut. Parameter k menentukan jumlah tetangga terdekat yang akan dipertimbangkan (Altman, 1992).

d. *Apply Model*

Operator ini digunakan untuk menerapkan model KNN yang telah dilatih pada data testing. Hasilnya adalah prediksi IPK untuk setiap mahasiswa dalam data testing.

e. *Performance*

Operator ini digunakan untuk mengevaluasi kinerja model KNN. Metrik evaluasi yang umum digunakan adalah root mean squared error (RMSE), mean absolute error (MAE), dan R-squared. Metrik-metrik ini mengukur seberapa baik model dalam memprediksi IPK dengan benar (Powers, 2011).

3.4 Evaluasi

Bagian ini menjelaskan evaluasi kinerja model K-Nearest Neighbor (KNN) yang telah dibangun untuk memprediksi kategori kehadiran siswa ("Rajin" atau "Tidak Rajin"). Evaluasi dilakukan menggunakan confusion matrix seperti yang ditunjukkan pada gambar.

accuracy: 55.60% +/- 3.50% (micro average: 55.60%)

	true Rajin	true Tidak Rajin	class precision
pred. Rajin	139	112	55.38%
pred. Tidak Rajin	110	139	55.82%
class recall	55.82%	55.38%	

Gambar 4. Hasil akurasi

Interpretasi Confusion Matrix:

- True Positive (TP): 139 siswa diprediksi "Rajin" dan memang benar "Rajin".
- True Negative (TN): 139 siswa diprediksi "Tidak Rajin" dan memang benar "Tidak Rajin".
- False Positive (FP): 112 siswa diprediksi "Rajin" padahal sebenarnya "Tidak Rajin".
- False Negative (FN): 110 siswa diprediksi "Tidak Rajin" padahal sebenarnya "Rajin".

Accuracy, 55.68%. Ini menunjukkan bahwa model KNN secara keseluruhan benar memprediksi kategori kehadiran siswa sekitar 55.68% dari total data.

Class Precision, "Rajin": 55.38%. Dari semua siswa yang diprediksi "Rajin", 55.38% memang benar "Rajin", "Tidak Rajin": 55.82%. Dari semua siswa yang diprediksi "Tidak Rajin", 55.82% memang benar "Tidak Rajin".

Class Recall, "Rajin": 55.82%. Dari semua siswa yang sebenarnya "Rajin", model berhasil memprediksi 55.82% dengan benar. "Tidak Rajin": 55.38%. Dari semua siswa yang sebenarnya "Tidak Rajin", model berhasil memprediksi 55.38% dengan benar.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi tingkat kerajinan siswa berdasarkan nilai GPA menggunakan algoritma *K-Nearest Neighbors* (KNN). Dari hasil analisis, ditemukan bahwa faktor seperti jam belajar per minggu dan tingkat kehadiran memiliki hubungan signifikan terhadap kerajinan siswa. Model yang dibangun memiliki tingkat akurasi 55.68%, meskipun masih terdapat ruang untuk peningkatan performa melalui pengoptimalan parameter dan penggunaan dataset yang lebih beragam.

Pendekatan ini menunjukkan potensi penerapan algoritma KNN dalam konteks pendidikan, khususnya untuk membantu institusi dalam mengidentifikasi siswa yang membutuhkan perhatian lebih. Selain itu, penelitian ini memberikan dasar yang kuat untuk eksplorasi lebih lanjut dalam meningkatkan akurasi prediksi dengan melibatkan metode pembelajaran mesin lainnya atau dengan menambahkan fitur baru ke dalam dataset.



Untuk pengembangan di masa depan, penelitian ini dapat ditingkatkan dengan mengevaluasi algoritma lain, memperluas cakupan atribut dataset, dan mengimplementasikan model pada populasi siswa yang lebih luas. Dengan demikian, model ini dapat memberikan kontribusi yang lebih signifikan dalam mendukung perencanaan intervensi pendidikan berbasis data.

REFERENCES

- Altman, N. S. (1992). An Introduce to Kernel and Nearest Neighbor Nonparametric Regression. *The American Statistician*, 175-185.
- Al-Zubaidi, M., & El-Sayed, A. (2020). A novel prediction model for student performance using KNN algorithm. *Journal of Educational Computing Research*, 420-438.
- Farokhah, L. (2020). IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BUNGA DENGAN EKSTRAKSI FITUR WARNA RNB. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 1129-1136.
- Gharbi, A., & Jaziri, A. (2021). Predicting student performance using data mining algorithms: A survey. *International Journal of Advanced Computer Science and Applications*, 99-105.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques (3rd ed.)*. Amsterdam: Elsevier.
- Krisandi, N., Helmi, & Prihandono, B. (2013). ALGORITMA K-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL PRODUKSI SAWIT PADA PT.MINAMAS KECAMATAN PARINDU. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, 33-39.
- Munandar, T. A., & Munir, A. Q. (2022). Implementasi K-Nearest Neighbor Untuk Prototype Sistem Pakar Identifikasi Penyakit Jantung. *Jurnal Teknologi Informasi*, 44-50.
- Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 37-63.
- Singh, D., & Mishra, S. (2020). Predictive modeling of student performance using *K-Nearest Neighbors* and other machine learning algorithms. *International Journal of Educational Technology in Higher Education*, 22-34.
- Witten, I. H., Frank, A., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan kaufman.