Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Prediction of Criminal Theft Locations at the Binjai Police Station using Historical Data and the KNN Algorithm

Rizky Pambudi [1*], Yusfrizal [2], Hermansyah Sembiring [3]

[1,2,3]STMIK Kaputama

ekiaja002@gmail.com[1*], yusfrizal80@gmail.com[2], hermansyah.sembiring@gmail.com[3]

Abstract

This research aims to predict the locations of theft crimes in the Binjai Police Department by utilizing historical data obtained from Binjai Police. The implementation of the *K-Nearest Neighbor* (KNN) algorithm was carried out to analyze crime patterns based on variables such as district location, time of occurrence, day of occurrence, and type of crime. The historical data were processed through normalization, splitting into training and testing datasets, and model evaluation using RapidMiner software. The results show that the KNN algorithm is able to classify with a fairly good level of accuracy, making it a useful basis for providing predictive information on theft-prone locations. These findings are expected to assist the police and the Binjai city government in formulating crime prevention strategies and increasing public awareness.

Keywords: *Historical Data*, *Prediction*, *Theft*, *K-Nearest Neighbor*, *RapidMiner*

1. Introduction

Public safety and order are essential for the development of a region. As one of the growing cities in Indonesia, Binjai faces many challenges in terms of security stability. The increasing crime rate, particularly theft, can negatively impact the economic development and quality of life of the community in the area. Therefore, a method that can be used to analyze and predict theft crime locations is needed to assist authorities in making better prevention plans and to provide the public with warnings about areas prone to theft crimes.

In today's digital era, the utilization of technology and data analysis has become an innovative solution in various fields, including security. One method that can be applied is the implementation of machine learning algorithms to predict the locations of theft crimes based on historical data. One of the most popular machine learning methods for prediction is K-Nearest Neighbors (KNN). The KNN algorithm is a method for classifying new objects based on training data that have the nearest neighbors to the given object [1]. This algorithm falls into the category of supervised learning, where new data can be classified based on the majority class of the nearest neighbors. The grouping category is the most commonly used. This method enables researchers to analyze theft data, predict or discover patterns, and classify data using KNN as the underlying algorithm [2].

Using the K-Nearest Neighbors (KNN) algorithm, the prediction of theft crime locations at the Binjai Police Station can be carried out based on various factors, such as the location of the incident, the day of the incident, the time of the incident, and the type of theft. The results of this prediction can be utilized by the police, local government, and the community in designing more effective strategies for crime prevention and control.

2. Literatur Riview

2.1. Data Mining

Data mining is the process of extracting useful information and patterns from a large amount of data, which includes data collection, data extraction, data analysis, and data statistics. Data mining is also known as knowledge discovery, knowledge extraction, data or pattern analysis, information harvesting, and so on [3].

2.1.1. Data Mining Process

Data mining has a series of processes for extracting data that have not been previously discovered. These processes or stages start from raw data and end with processed knowledge or information [4]. These stages can be seen in the image below.

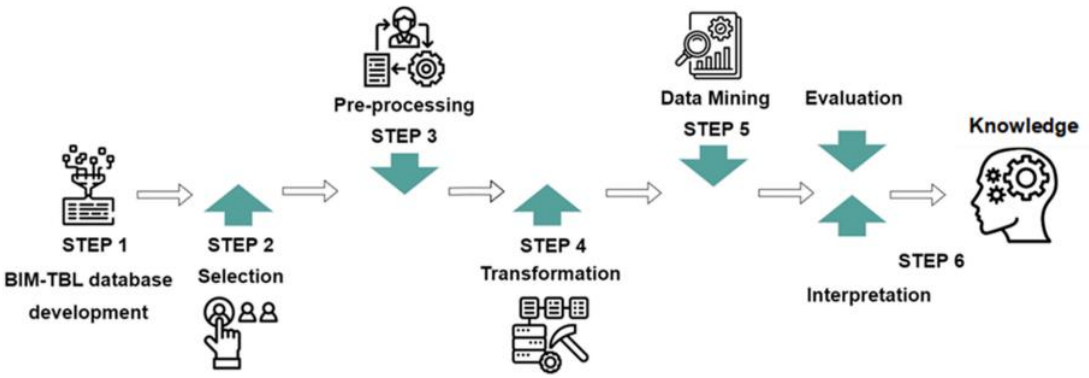

Fig. 1 : Data Mining Steps

2.1.2. Data Mining Algorithms

In the data mining process, an algorithm is a procedure or method used to analyze and discover patterns in data. Some of the most common types of algorithms used in data mining include [5]:

a. Classification: Algorithms used to categorize data into specific classes. Examples of classification algorithms include Decision Trees, Random Forest, and Support Vector Machine.
b. Regression: Used to predict a continuous value based on independent variables. Linear and polynomial regression are two popular regression algorithms.
c. Clustering: This is a technique for grouping data into clusters based on similar characteristics. K-Means, Hierarchical Clustering, and DBSCAN are frequently used clustering algorithms.
d. Association: Used to identify relationships between variables in a dataset. Apriori and FP-Growth association algorithms are often used for market basket analysis.
e. Anomaly Detection: These are algorithms used to identify data that is unusual or deviates from the common pattern. Some methods frequently used in this technique include Isolation Forest and Local Outlier Factor.

2.2. K-Nearest Neighbour

The K-Nearest Neighbors (KNN) classification algorithm, which is based on the idea of "closeness" between data points, predicts the class of a new data sample based on the majority class of its k-nearest neighbors in the training data. KNN is a distance-based supervised learning algorithm. KNN can be used for both regression and classification prediction [6].

2.3. Location

According to the KBBI (Great Dictionary of the Indonesian Language), "lokasi" (location) refers to a let's (position or whereabouts). The word itself doesn't have a direct definition, but its synonym, "letak," means the position or place where something is located.

2.4. Prediction

Prediction is the forecasting or estimation of a future state by using information from the past and present to reduce errors. Based on this understanding, prediction is a structured activity of forecasting what might happen in the future by using information from the past and present [7].

2.5. Crime

Crime is a form of deviation from societal norms that is seen as a serious threat to social values and order. This behavior becomes a humanitarian and social problem because it can disrupt both individuals and society at large. Social norms serve as the foundation of order, and if violations are not addressed, it could endanger social stability [8].

2.6. Theft

In criminal law, theft is a criminal act that involves the unlawful or unauthorized taking or possession of someone else's property. This act is recognized as a serious offense in various legal systems around the world. Theft generally involves the intention to take, is done without right, and causes a transfer of ownership from the rightful owner to the perpetrator [8].

2.7. Historical Data

In a descriptive qualitative study, historical data is defined as secondary data or past research findings that are relevant to the current study, such as documents or data obtained from related institutions. This historical data is used to provide context without having to be obtained directly by the current researcher [9]. However, in this research, historical data is used as primary data due to its role as core data for predicting new data to be sought.

2.8. RapidMiner

RapidMiner is an open-source software designed to support various data analysis needs. As a popular solution in the field of data science, RapidMiner allows users to effectively perform data mining, text mining, and predictive analysis. The software combines a wide range of descriptive and predictive analytical techniques to help users gain deeper insights from their data. With these comprehensive features, RapidMiner facilitates more accurate and strategic decision-making based on the analysis results. RapidMiner is a standalone program for data analysis and data mining that can be integrated into its own products. Written in Java, it can run on all operating systems (Hidayati et al. 2023).

3. Analysis And Design

3.1. Research Methodology

The research methodology section explains the stages used in research using the K-Nearest Neighbor (KNN) algorithm. These stages include preparation, theoretical study, data collection, data analysis, system testing and implementation, and a final stage of conclusions and suggestions.

3.2. Model Implementation

In this study, a flowchart was designed to illustrate the process of predicting criminal theft data at the Binjai Police Station. The proposed flowchart can be seen in the image below.

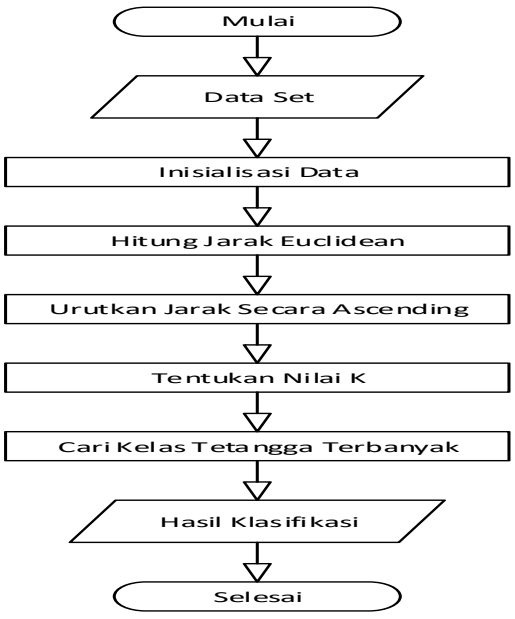

Fig. 2 : Flowchart

1. Start: The process begins to run the KNN algorithm.
2. Dataset: Provide historical data containing features (attributes) and labels (classes) that will be used for training and testing.
3. Data Initialization: The data is converted into a numerical format so that it can be calculated using the Euclidean Distance formula.
4. Calculate Euclidean Distance: Measure the distance between the test data and each piece of training data using the Euclidean distance formula.
5. Sort Distances in Ascending Order: All calculated distances are sorted from the smallest (closest) to the largest (farthest).
6. Determine the value of K: Choose the number of nearest neighbors (e.g., k=5) that will be used in the classification.
7. Find the Majority Class of Neighbors: From the 'k' closest data points, identify the class that appears most frequently.
8. Classification Result: The test data is classified into the class with the most neighbors.
9. End: The prediction or classification process is complete.

3.2.1. Data on criminal theft in Binjai

For data prediction, historical data is needed to serve as the basis for analysis and method calculations, allowing for the best possible alternatives to be obtained. In this study, the data used is criminal theft data from the Binjai Police Station, which was acquired from the police station itself. The data below is ready to be used for implementing the K-Nearest Neighbor (KNN) algorithm.

Table. 1 : Research data sample

No	Kecamatan	Waktu Kejadian	Hari Kejadian	Jenis Pidana	Tipe Lokasi TKP
1	Binjai Timur	Siang	Sabtu	Curanmor	Tempat Tinggal & Hunian
2	Binjai Utara	Pagi	Senin	Curat	Perdagangan, Jasa & Komersial
3	Binjai Kota	Pagi	Selasa	Curat	Perdagangan, Jasa & Komersial
4	Binjai Utara	Malam	Rabu	Curat	Tempat Tinggal & Hunian
5	Binjai Timur	Malam	Minggu	Cubis	Infrastruktur & Transportasi
6	Binjai Utara	Dini Hari	Jum'at	Curat	Tempat Tinggal & Hunian
7	Binjai Barat	Dini Hari	Sabtu	Curat	Tempat Tinggal & Hunian
8	Binjai Kota	Sore	Kamis	Curanmor	Tempat Tinggal & Hunian
9	Binjai Kota	Malam	Rabu	Curanmor	Tempat Tinggal & Hunian
10	Binjai Utara	Malam	Senin	Curas	Tempat Tinggal & Hunian
11	Binjai Selatan	Malam	Senin	Curas	Infrastruktur & Transportasi
12	Binjai Utara	Sore	Minggu	Cubis	Tempat Tinggal & Hunian

To perform data normalization, you first initialize the features in the research data. Then, you initialize the data into numerical form. This data can then be expressed in independent variables: District (X1), Time (X2), Day (X3), and Type of Crime (X4).

Table. 1 : Normalized data sample

No	Kecamatan	Waktu	Hari	Jenis Pidana	Tipe Lokasi
	X1	X2	X3	X4	
1	4	3	6	3	Tempat Tinggal & Hunian
2	2	2	1	2	Perdagangan, Jasa & Komersial
3	1	2	2	2	Perdagangan, Jasa & Komersial
4	2	5	3	2	Tempat Tinggal & Hunian
5	4	5	7	4	Infrastruktur & Transportasi
6	2	1	5	2	Tempat Tinggal & Hunian
7	5	1	6	2	Tempat Tinggal & Hunian
8	1	4	4	3	Tempat Tinggal & Hunian
9	1	5	3	3	Tempat Tinggal & Hunian
10	2	5	1	1	Tempat Tinggal & Hunian
11	3	5	1	1	Infrastruktur & Transportasi
12	2	4	7	4	Tempat Tinggal & Hunian

4. Discussion And Implementation

In this study, historical crime data from the Binjai Police Station was used. This data was processed using the K-Nearest Neighbor (KNN) method through several stages in the RapidMiner application.The initial stage begins with the **data import** process, where an Excel file containing the raw data is loaded into RapidMiner. Next, a **set role** operation is performed to define which attribute will serve as the label or prediction target.

After the preparation is complete, the KNN algorithm is applied. This method classifies test data based on its proximity to the training data. The system calculates the closest distance between the test data and the training data, then determines the most dominant class from a specified number of nearest neighbors.

4.1. Impementation

Based on the analysis results, most of the predictions align with the actual labels. This indicates that the K-NN model is quite capable of recognizing patterns in both the training and testing data. However, there are still some cases where the prediction results do not match the original labels, such as on row 123, which was predicted as "perkantoran & industri" (offices & industry). From this explanation, the display of the analysis results can be seen as follows.

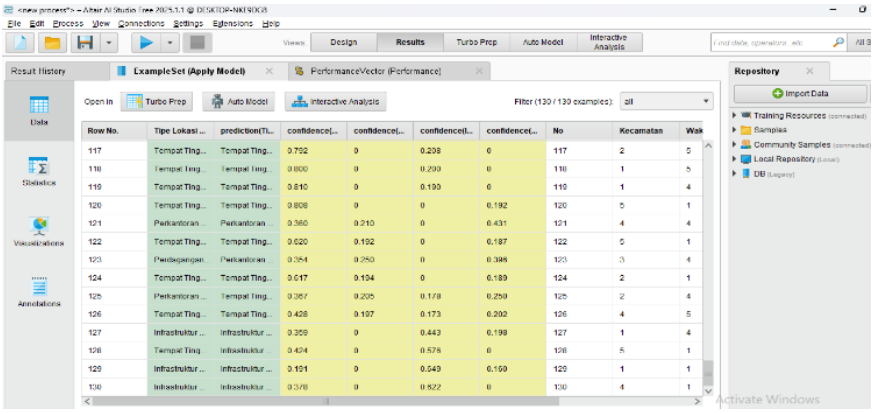

Fig. 3 : Prediction Results

Testing the prediction of criminal theft locations using the K-Nearest Neighbor (KNN) method in the **RapidMiner** application was carried out through several stages, from data preparation to evaluating the prediction results. The following shows a fairly good accuracy, with a stronger performance in predicting theft locations, achieving an accuracy of 73.08%, which can be seen in the image below.

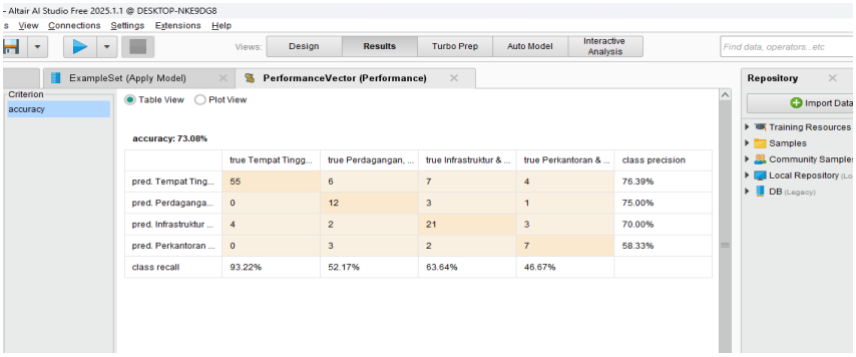

Fig. 4 : Accuracy Results

5. Conclusion and Recommendations

5.1. Conclusion

Some conclusions that can be drawn from this research are as follows:

1. There are several important steps in applying the KNN algorithm to predict the location of criminal theft. First, you prepare the dataset and then clean the data. Next, you calculate the Euclidean distance between the training data and the testing data. After the distances are obtained, the KNN model is applied with the specified conditions. In this study, the value of *k* is 5. Therefore, each piece of test data will be assigned the class of its 5 nearest neighbors based on the majority class among them.
2. The testing conducted using RapidMiner resulted in a model accuracy of 73.08% with a 75% data split. This indicates that the system is quite good at predicting the location of criminal theft.
3. The implementation of KNN can help police and other related agencies understand the distribution and patterns of criminal activity based on historical data. This information can then be used as a basis for developing crime prevention and response strategies in the city of Binjai.

5.2. Recommendations

As a follow-up and further development of this research, the following are recommended:

1. Future research is advised to add other variables, such as the victim's age, occupation, or environmental factors, to achieve more accurate prediction results.
2. It is recommended that a larger dataset be used, for example, 500-1000 rows, so that the model can recognize crime patterns more accurately.

References

[1] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” Build. Informatics, Technol. Sci., vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.

[2] F. Hermawan and J. Prianggono, “Crime of theft prediction using Machine Learning K-Nearest Neighbour Algorithm at Polresta Bandar Lampung,” Sinkron, vol. 8, no. 3, pp. 1515–1527, 2023, doi: 10.33395/sinkron.v8i3.12422.

[3] S. S. M. K. Muhammad Arhami and S. T. M. T. Muhammad Nasir, Data Mining - Algoritma dan Implementasi. Andi Offset, 2020. [Online]. Available: https://books.google.co.id/books?id=AtcCEAAAQBAJ

[4] Y. Ardilla *et al.*, DATA MINING DAN APLIKASINYA. Penerbit Widina, 2021. [Online]. Available:

https://books.google.co.id/books?id=53FXEAAAQBAJ
[5] S. Syam *et al.*, Data Mining : Teori dan Penerapannya dalam Berbagai Bidang. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=hTAxEQAAQBAJ
[6] B. A. S. Fairuz, Panduan Praktis Machine Learning Klasifikasi Menggunakan Python: Diandra Kreatif. Diandra Kreatif, 2024. [Online]. Available: https://books.google.co.id/books?id=Wr5AEQAAQBAJ
[7] L. Suryadi, N. E. Pratiwi, F. Ardhy, and P. Riswanto, "Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor (K-Nn)(Studi Kasus: Toko Zerita Meubel)," JUSIM (Jurnal Sist. Inf. Musirawas), vol. 7, no. 2, pp. 174–184, 2022.
[8] H. Hamdiyah, "Analisis Unsur-Unsur Tindak Pidana Pencurian: Tinjauan Hukum," J. Tahqiqa J. Ilm. Pemikir. Huk. Islam, vol. 18, no. 1, pp. 98–108, 2024, doi: 10.61393/tahqiqa.v18i1.216.
[9] A. D. Cahya and T. Anggraini, "Analisis Problematika Bank Syariah Indonesia Setelah Merger Studi Kasus Bank Syariah Indonesia (BSI)," Syntax Lit. J. Ilm. Indonesia., vol. 7, no. 12, pp. 17230–17241, 2022.