Jurnal Mantik, 6 (3) (2022), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Application of the K-NN Algorithm for Sales Effectiveness of Pastries at Mami's Kitchen Bakery

Leliana Harahap[1], Sartika Dewi Purba[2], Kamson Sirait[3], Jonas Franky R Panggabean[4]
[1,2,3,4] Akademi Informatika dan Komputer Medicom Medan

A R T I C L E I N F O

Article history:

Received Nov 02, 2022
Revised Nov 16, 2022
Accepted Nov 30, 2022

Keywords:

Algorithm;
Effectiveness;
K-NN;
Sales.

ABSTRACT

This research is very helpful for companies in generating sales effectiveness so as to generate large profits. To get a big profit, the company must be able to achieve maximum sales targets. This bakery sells various types of pastries that are in demand by the public. Mami's kitchen bakery must really make policies and decisions that really attract people to buy pastries. Therefore this research can help mama's kitchen bakery in determining sales effectiveness so that it can find out the number of product sales so that it continues to increase. The K-NN algorithm is very suitable for solving sales effectiveness problems

Corresponding Author:

Leliana Harahap,
Akademi Informatika dan Komputer Medicom Medan
Jl. Darat No.74, Petisah Hulu, Kec. Medan Baru, Kota Medan, Sumatera Utara 20152
Email: leliharahap05@gmail.com

1. INTRODUCTION

In running a business, sales are the determining factor for a business to develop or not. Sales is one of the business goals, where sales are the determinant of income from the business. Good company marketing activities will generate large profits. To achieve profits, strong marketing is needed which is always developing and can collaborate with the changing times. Research conducted**,** this algorithm is used to predict furniture sales data on CV. Great Octo Jepara. The results showed that the proposed method was successfully implemented to solve sales prediction cases with an error rate of 6 percent and an accuracy of 94 percent (R. A. Pangestu, 2018)

Research conducted by Nobertus Krisandi. et al with the title "K-Nearest Neighbor Algorithm in the Classification of Palm Oil Production Data at PT. Minamas District of Parindu". The data used are data on palm oil production (tonnage) from 50 farmer groups in the period July-December 2011 at PT. Minasa, Sanggau Regency. The k values used are k=1, k=3, k=5 and k=7. The results showed that the dominant production yield was with a value of k = 7 which had an accuracy value of 34%. This also indicates that the K-Nearest Neighbor (KNN) is affected by the amount of data clustering (B. Sawit, 2019)

Bakeries are one of the phenomenal businesses nowadays because all people like them. Mami's kitchen bakery sells various kinds of pastries, where pastries have been one of the foods or snacks that have always been of interest to the public from ancient times. However, bakery kitchens need to increase their sales effectiveness because their competitiveness is quite high. So that it can disrupt sales that generate company profits. To be able to increase effectiveness, it is necessary to have sales data for a year and to classify the number of types that are most purchased by consumers.

With the classification of data on mama's kitchen bakery, it can be seen whether the number of sales has increased or decreased. Therefore it is necessary to carry out data mining processes in processing data such as using artificial intelligence techniques, mathematics and statistics to extract and identify useful data from various databases and linked to the K-NN (K-Nearest Neighbor) method.

Research conducted by Ricky Imanuel Ndaumanu, Kusrini, M. Rudyanto Arief entitled Analysis of Prediction of Student Resignation Rates Using the K-Nearest Neighbor Method. In this study, it was stated that from the number of new student registrations, many students also withdrew each year due to various problems. Because there are students who resign, the author analyzes student resignation using the K-Nearest Neighbor algorithm. This study aims to predict the level of accuracy of STIKOM UYELINDO Kupang student resignation (B. Suhartini1, 2019).

The K-NN method is very suitable for use because it is able to provide quite significant results so that you can find out the development of sales of the Mami's Kitchen Bakery, whether sales are increasing or losing. If the Kitchen Mami Bakery knows the percentage of sales, it can create a good marketing strategy. Therefore, it is necessary to conduct a study, namely the application of the K-NN Algorithm for the Effectiveness of Sales of Pastries at Mami's Kitchen Bakery.

2. METHOD

1. Definition of Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. Data Mining is a multidisciplinary scientific field, describing work areas that include database technology, machine learning, statistics, pattern recognition, information retrieval, artificial neural networks, knowledge-based systems, artificial intelligence, high performance computing, and data visualization. Data Mining is defined as data mining or attempts to extract valuable and useful information on very large databases (Senubekti, 2022) .

Data mining Is the process of searching and extracting information from piles of large amount of data pile is the main process of data mining, the main purpose of processing the data is to produce new information (Muliono, 2019).

Data mining is a process of discovering meaningful new correlations, patterns, and trends with sifting through large amounts of data stored in in the repository, using recognition technology patterns and statistical and mathematical techniques (Zulfa Nabila, 2021)

Data mining is an iterative and interactive process to discover new patterns or models which is authentic (perfect), useful and understandable in a very large database (massive databases). Data mining contains the search for the desired trend or pattern in the database Great for helping decision making in the future (Syahdan, 2018)

2. Definition of Application

According to the Big Indonesian Dictionary (KBBI), the notion of application is an act of applying, whereas according to some experts, application is an act of practicing a theory, method, and other matters to achieve certain goals and for an interest desired by a group or groups that have planned and arranged beforehand.

An application is a program on a computer or cellphone that is used to run a program that has been created (Dewi K N, 2021).

3. Stages of Data Mining

There are several stages of data mining, which include:

a. Data Selection Selection (selection)
b. Cleaning Data
c. Data Transformation

d. Datamining
e. Interpretation/Evaluation

4. Classification

The result of the data description model is an analysis of the classification of the data. Which aims to get a model that can separate / distinguish between class data to be able to calculate objects that are not found (B. Sawit, 2019).

Classification algorithms that are widely used, namely Decision/Classification Trees, Bayesian Classifiers/Naïve Bayes Classifiers, Neural Networks, Statistical Analysis, Genetic Algorithms, Rough Sets, K-Nearest Neighbor, Rule Based Methods, Memory Based Reasoning, and Support Vector Machines (SVMs). The classification process is based on four components (Yahya, 2018).

a. Variable Class.
b. Variable Predictors.
c. Training Datasets Training.
d. Test datasets

5. K-NN Algorithm (K-Nearest Neighbor)

K-Nearest Neighbor (K-NN) is a method which uses a supervised algorithm where results of the newly classified test sample based on the majority of the categories on K-NN. The purpose of this algorithm is classifying new objects based on attributes and a training sample. Classification does not use a model anything to match and only based on memory (Winda, 2020)

The K-Nearest Neighbor (KNN) algorithm is an algorithm that is used to classify an object, based on the k pieces of training data that are closest to the object. The condition for the value of k is that it cannot be greater than the number of training data, and the value of k must be odd and more than one (Rivki.M, 2017).

KNN (K-Nearest Neighbor) is a classification method based on the closest object to the object or feature, the most commonly used data in learning data (Salsabila, 2021). To calculate the K-NN algorithm, the following steps are needed:

a. Determine parameter K (Sum of closest data).
b. Calculate each squared Euclidean distance (query instance) with the sample.
c. Collect category Y (Nearest Neighbor Classification)

There are many ways to measure the shortest distance between new data and old data (training data), such as Euclidean distance and Manhattan distance (city block distance), Euclidean distance is the way to be used, namely (Bachtiar, 2017):

$$\sqrt{(a1-b1)^2+(a2-b2)^2+.+(an-bn)^2}$$

Where a = a1, a2,..., an, and b = b1, b2, ..., bn are the nth attribute values of the two records. On category values, measurements with euclidean distance are not unsuitable.

Can be replaced with the following function:

Different (ai, bi) = {0 if ai=bi1.

Where ai and bi are equal to category values. If the attribute values between the 2 records are compared with the same then the distance value is 0, meaning similar, otherwise, if it is different then the closeness value is one, meaning there is no similarity.

To calculate similarity:

$$Similarity\ (p,q) = \frac{n\sum_{t}^{n} = 1\ \int(pi,qi)X\ wi}{wi}$$

Information:

P = New case
q = Cases in storage
n = Number of attributes in each case
i = Individual attribute between 1 to n
f = the function of the similarity attribute i between cases p and case q
w = the weight given to the i-the attribute

6. Datasets

The data set is a data set which consists of one or a table in the database, the contents of each table column are representative of the variables used and the contents of the rows are records of the data set that is filled. The data sets have values on all the variables used with datum terms which consist of various files.

Dataset Preprocessing is a process that is executed before analyzing the dataset where the dataset will be processed into a normalized format (Haryo Bagas Assyafah, 2021).

7. Rapid Miner

Rapid Miner is an open source software application that can be programmed in the Java language, this application does not have a special programming language because all the tools or facilities are ready to use, such as making various models, all you have to do is adjust the method to be used.

RapidMiner is an application or software that functions as a learning tool in data mining science. The platform is developed by a company dedicated to all steps involving large amounts of data in commercial business, research, education, training and learning. RapidMiner has around 100 learning solutions for clustering, classification and regression analysis (Vincentius Riandaru Prasetyo, 2021). RapidMiner also supports about 22 file formats, such as .xls, .csv and so on

3. RESULTS AND DISCUSSION

1. Data collection technique

In collecting data using several methods that will be used, including:

a. Technical Observation, data collection directly to Mami's Kitchen Bakery.
b. Interview (Interview), conduct a question and answer session directly to the owner or person in charge of the shop.
c. Literature study, adjusting to previous related research.

2. Data Processing Engineering

1. K-NN Algorithm Experiment

In the following table is an example of a sales dataset at Mami's Kitchen Bakery with 7 records for training data number 1 to 7 and for testing data number 8

Table 1. Example of sales dataset at Mami's Kitchen Bakery

No	Type	Profit	Income	Amount	Category
1	Kastengel	500.000	3.000.000	150	Achieve the target
2	Snow Princess	300.000	2.500.000	300	Achieve the target
3	Cheese stick	1.000.000	6.000.000	230	Achieve the target
4	Peanuts	1.000.000	6.000.000	300	Achieve the target
5	Choco chips	800.000	5.000.000	250	Achieve the target
6	Nastar	5.000.000	10.000.000	1000	Achieve the target

7	Baking pan	4.000.000	8.000.000	1200	Achieve the target
8	Syringe	500.000	1.500.000	200	?

The steps of the K-NN algorithm:

a. Determination of parameter k=4 (number of nearest neighbours). Here in determining the parameter k=4
b. Calculates Euclidean distance
c. Sort the codes with the smallest distance.

Table 2 Sequence of the smallest Euclidian distance.

No	Type	Rank
1	Kastengel	6
2	Snow Princess	7
3	Cheese stick	3
4	Peanuts	4
5	Choco chips	5
6	Nastar	1
7	Baking pan	2
8	Syringe	8

d. Collect category Y (Nearest Neighbor Classification). The number of K specified K=5.

Table 3. The number of K=5

No	Type	Rank
1	Kastengel	1
2	Snow Princess	2
3	Cheese stick	3
4	Peanuts	4
5	Choco chips	5

e. Use the majority category, then the classification and processing results are:

Table 4. The classification and processing results

No	Type	Rank	Category
1	Nastar	1	Achieve the target
2	Baking pan	2	Achieve the target
3	Cheese stick	3	Achieve the target
4	Peanuts	4	Achieve the target
5	Choco chips	5	Achieve the target

2. Achieve the target

Several steps are needed to process the data, namely:

a. To analyze, training data is needed which will be processed and entered into Rapidminer and using various formats such as csv, xls, mdb, and others. The data format used by the author is xls.
b. Open the rapidminer application, then the start page will appear, then click File and select New Process.

There are several stages used in processing data, including:

1. Data analysis, training data such as excel format is needed.
2. Open the rapidminer application, select new process

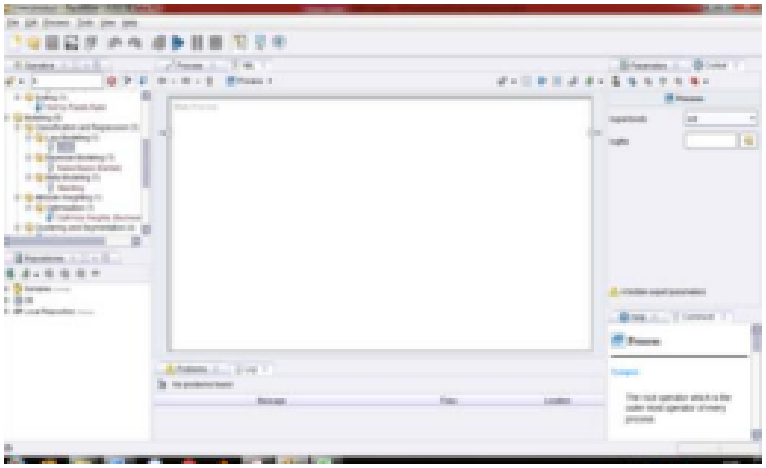

Figure 1. View of the Rapidminer Process

3. Enter the data that we have prepared from excel format, select the Import Wizard sub.
4. The next step displays the data to be imported in the wizard which contains the training data file.
5. In the data view, specify the storage location.
6. Then specify the data label, if the attribute is not needed, remove the tick at the top and adjust the order of the attributes that are not used. Then select the finish section to proceed to the next stage.
7. Next, add the algorithm model that we are testing, select the Operator- Modeling-Classification and Regression-Lazy Modeling- k-NN tab. Then slide K-NN towards the Main Process section and so on.
8. Then connect to be able to see the results of processed data using the algorithm we created, select the Process-RUN menu.

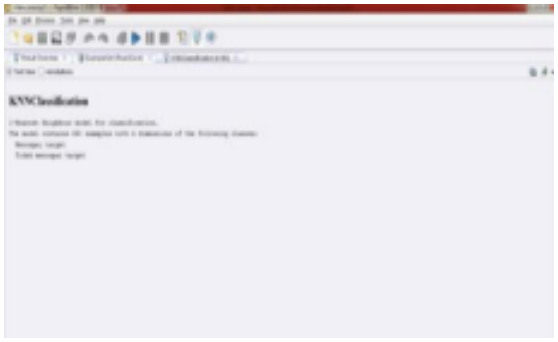

Figure 2. Results of KNN algorithm data processing

3. Cross validation

In testing using 273 test data, it is then carried out in order to find out the value of k with the level of truth/accuracy of the data, the higher the amount of k data used, the better the test results. The best k value depends on the type of data used. The way to get a good k value is by parameter optimization, for example by cross validation. In this case, the K-NN classification is predicted with the closest training data, with 8 cross validations. The K-NN algorithm conducts data training with data separated by cross validation into 2 parts, namely one part for training data and another for data testing.

The training section consists of the K-NN method and the testing section consists of the apply model and performance sections.

4. Confusion Matrix

The number of TP (true positive) data is 130 records which are classified as target classes and FN (False Negative) there are 30 records, but in reality there are several classes whose targets have not been reached. 132 records are false positives which are classified as the class not reaching the target and 19 false positive records as the class not being reached but are classified as the class reaching the target. By using the K-NN algorithm there are 4 validations of 87.6% and are calculated by.

5. CONCLUSION

From the results of the research that has been done, it is concluded that sales data is processed by applying data mining techniques with the K-NN algorithm. Sales data that is known to produce an average number of representatives per type of sales data of 5 pieces can be said to be a category capable of producing targets and if sales are below 5 units, then they are not able to produce targets. This category can produce a number of items that are the most sold or in demand by the public at Mami's Kitchen Bakery. The research was tested by calculating according to the rules and steps of the K-NN algorithm so as to produce a model. The experimental results have been carried out with a K-Fold Validation value of 4 with a flow rate of 87.6%. Therefore this algorithm is very suitable in classifying data on the number of goods that are best sold. The best-selling or most sold items are nastar, baking pan flowers, cheese sticks, peanuts, choco chips. So it can be concluded that data analysis with this method is very accurate in sales data analysis.

REFERENCES

B. Sawit, S. B. (2019). Metode Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada Pt Bumi Sawit Sukses (Bss) Menggunakan Metode K-Nearest Neighbor. 198–207.

B. Suhartini1, H. (2019). Klasifikasi Algoritma KNearest Neighbor Berbasis Particle Swarm Optimization Untuk Kelayakan Bantuan Rehabilitasi Rumah Tidak Layak Huni Pada Desa Lenek Duren Kecamatan Aikmel Kabupaten Lombok Timur . 79– 85.

Bachtiar, M. R. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *J. Sist. Inf*, 31.

Dewi K N, I. H. (2021). Konsep Aplikasi E-Dakwah Untuk Generasi Milenial Jakarta. *Ikraith Informatika*.

Haryo Bagas Assyafah, D. T. (2021). Analisis Datasetmenggunakan Sentiment Analysis(Studi KasusPada Tripadvisor). *Jurnal Strategi*.

Muliono, R. a. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *Journal of Computer Engineering, System and Science*, 2502–2714.

R. A. Pangestu, S. R. (2018). Aplikasi Web Berbasis Algoritma K-Nearest Neighbour Untuk Menentukan Klasifikasi Barang Studi Kasus : Perum Peruri.

Rivki.M, B. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 31-37.

Salsabila, A. Y. (2021).). Identifikasi Citra Jenis Bunga menggunakan Warna HSV dan Tekstur GLCM. *Technomedia Journal*.

Senubekti, D. (2022). Prinsip Klasifikasi Dan Data Mining Dengan Algoritma C4.5. *Jurnal Nuansa Informatika*, 87-93.

Syahdan, S. (2018). Data Mining Penjualan Produk Dengan Metode Apriori. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 56-63.

Vincentius Riandaru Prasetyo, H. L. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier. *Jurnal Nasional Teknologi dan Sistem Informasi* , 008-017.

Winda, Y. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape. *Jurnal Informatika dan Teknologi*, 104-114.

Yahya. (2018). Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decition Tree(C4.5). 56–63.

Zulfa Nabila, A. R. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-means. *Jurnal Teknologi dan Sistem Informasi*, 100-108.