International Journal Software Engineering and Computer Science (IJSECS) 3 (3), **2023, 457-466**
Published Online December 2023 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v3i3.1878.

Classification of Production Machine Spare Part Stock Data Request Needs Using The K-Nearest Neighbor Method

Hamdi Yansyah *
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa Cikarang, Bekasi Regency, West Java Province, Indonesia.
Email: hamdiyns@mhs.pelitabangsa.ac.id

Sifa Fauziah
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa Cikarang, Bekasi Regency, West Java Province, Indonesia.
E-mail: sifa_fauziah@pelitabangsa.ac.id

Donny Maulana
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa Cikarang, Bekasi Regency, West Java Province, Indonesia.
E-mail: donny.maulana@pelitabangsa.ac.id

Received: 30 November 2023; Accepted: 15 December 2023; Published: 30 December 2023.

Abstract: Spare parts encompass various items that are offered, owned, utilized, or consumed to fulfill consumer desires and requirements. This research implements the K-Nearest Neighbor algorithm on a test dataset consisting of 100 data objects, resulting in a novel classification perspective. The study includes a comprehensive model evaluation process involving Cross Validation on both training and testing datasets, comprising 1000 records with 36 critical and 64 non-critical outcomes. Performance assessment and testing utilizing the RapidMiner Studio application yield optimal results under various modeled scenarios. The accuracy of this algorithm model stands at 98.00%, with a standard deviation of +/- 4.00%.

Keywords: Spare Parts; Machinery; Products; Data Mining; K-NN; Classification.

1. Introduction

One of the businesses in the manufacturing industry is experiencing a very rapid increase and the process for managing the availability of spare parts that will be processed for machines means that stock requirements must increase, so it is necessary to collect spare part stock data. The need for spare parts to be processed every day is always increasing, so there is a need to increase the need for appropriate stock and grouping of goods according to needs. Inventory of goods is very important for a company because it will affect production and sales levels. Inventory is needed to create production results with the aim of producing products. Therefore, inventory control is very necessary to control the stock of spare parts needed for production. In its implementation, companies can experience inventory control problems such as excess inventory and inventory shortages.

Apart from that, process management to regulate the availability of inventory of needed goods is also very necessary to maximize certain stocks that are most needed by machines. There are several analytical approaches that can be used to carry out classification, one of which is data mining. Data mining is a process of finding new relationships that have patterns, meanings and habits based on existing data using mathematical, statistical, artificial intelligence and machine learning techniques [1]. One of the methods contained in data mining is the K-Nearest Neighbor method. The aim of this research is to provide an alternative inventory management policy by predicting needs using this method, in determining stock security for one of the spare parts.

Previous research by Izzan, Ratna, & Zainal (2022) applied the KNN method to predict product sales at Maju Jaya Fish Processing MSMEs [2]. This study aimed to assess the prediction accuracy and validation errors of processed fish production levels using the KNN method in comparison to other forecasting methods (SMA, DMA, SES, DES, and Naive) until December 2022. The research involved the development of six models to explore the impact of various analytical factors. The findings indicated that KNN time series forecasting outperformed other methods, displaying lower Root Mean Square Error (RSME) values. Furthermore, the proposed recommendation, which achieved a net Benefit-Cost Ratio of 2.82, proved to be financially viable. The investment in purchasing a laptop for this purpose was deemed feasible.

Research conducted by Bahrun Said Renhoran, Nova Nurhandayani (2018), Application of the C4.5 Algorithm to Determine Stock Data and Material Demand Targets Most Needed by the Logistics Warehouse at PT PLN (Persero) Kebon Jeruk Area. The aim of this research is to classify stock data and material demand targets for users/work supervisors, which will then create a data set classification with the number of cases most needed [3]. In the research process that has been carried out from this stage, there is a need for data summarization or a process of converting raw data into data that is easy to manage. The raw data will be entered into their respective categories and then the data will also be selected, so that the data is suitable for the data mining process which has been carried out in the Implementation of the C4.5 Algorithm Classification by processing the data using the Decision Tree method and making it easier to run in the RapidMiner application system. The Decision Tree method is very useful in this research, namely, to determine material/goods stock data from target warehouse material requests that are most frequently requested and needed by users. So, it can be concluded that the problem in determining material/goods stock data can be solved using data mining techniques, namely with the C4.5 Algorithm and getting the level of accuracy produced by the system using the Decision Tree method in the RapidMiner application is 100%. So, it can be concluded that this research can help warehouse users/supervisors in selecting and preparing the required stock of materials according to standards quickly and adequately.

2. Research Method

In this research the author used a Data Mining approach to carry out data analysis. The data that will be used as a dataset in this research is product stock data taken over a period of one year, which is obtained in the form of a spreadsheet file in Excel format and the data is still undergoing a selection and preprocessing process following the stages of the data processing process. In carrying out analysis and looking for results on basic needs data, these steps are made to research in research, namely, to make it easier and in accordance with the desired goals.

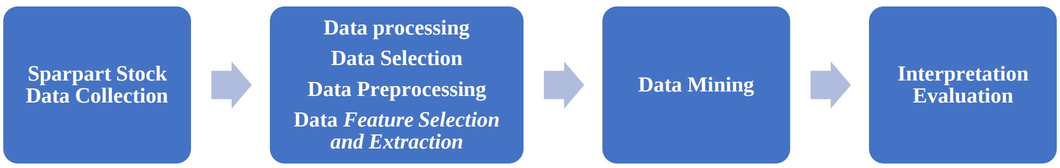
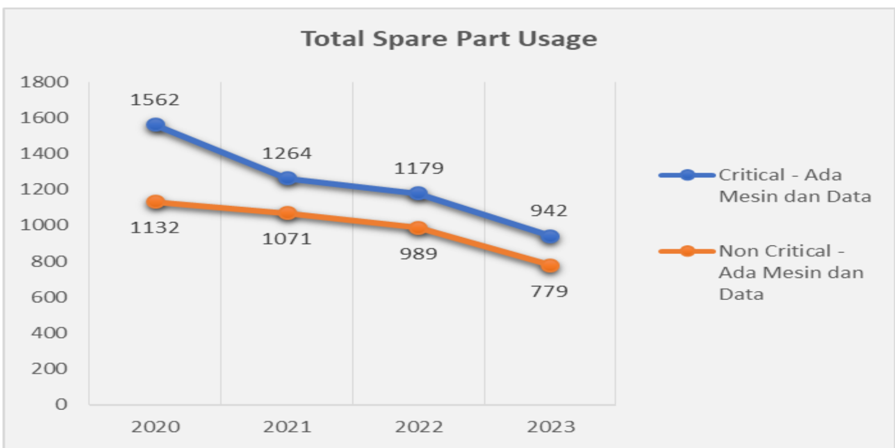


Figure 1. Research Stage

2.1 Data collection
At this stage, the data required for the research is collected. The data used is as follows:



Graph 1. Dataset

2.2 Data Selection
Data selection involves the process of reducing the dataset's size while still representing the original data accurately. This stage encompasses data cleaning, which includes removing missing values, duplicate data, checking for data

inconsistencies, and rectifying errors within the data. This cleaning process was carried out manually using spreadsheet software. Data cleaning addressed issues such as empty entries, incorrect data input, and spare part stock data lacking categorization [20].

2.3 Data Preprocessing
This research primarily focuses on data cleaning, dealing with noisy and incomplete data. The selection of attributes that will be utilized for the dataset and analysis is performed, with certain irrelevant attributes being excluded. This exclusion is because these attributes do not contribute to the calculation of the K-NN algorithm to be employed.

2.4 Feature Selection and Extraction
Feature selection is a preprocessing activity aimed at selecting influential features while excluding irrelevant ones in modeling or data analysis tasks.

2.5 Data Mining
Data mining involves the extraction of hidden insights from a large database, which is subsequently transformed into a testable dataset. The results of confusion tables are employed to assess accuracy, recall, and precision in the classification algorithm. Accuracy represents the percentage of correctly predicted values compared to actual values. Recall measures the algorithm's success rate, while precision indicates the accuracy of the predicted class [17].

2.6 Interpretation Evaluation
Evaluation is conducted by closely observing and analyzing the outcomes of the algorithm to verify their alignment with the research's objectives. This evaluation serves the purpose of understanding the results of the analyzed calculations and assessing the functionality of the employed methods and algorithms. Testing utilizes the RapidMiner tool to verify if the data aligns with the tool's outcomes. Furthermore, method and algorithm validation are performed by measuring the analysis results to determine how new objects are classified based on attributes and training samples. Given a query point, multiple K objects or training points that are closest to the query point are identified, and the query's predicted value is subsequently determined.

3. Result and Discussion

3.1 Results
Model testing is carried out with the aim of knowing the results of the calculations being analyzed and measuring whether the methods and algorithms used are functioning well or not. The application of the model with the K-NN algorithm in this research uses the K-NN algorithm which can be utilized using the RapidMiner Studio application. The author uses the K-Fold Cross Validation evaluation method because this function not only creates several test data samples repeatedly but divides the dataset into separate parts of the same size. The model is trained by a subset of training data and validated by a validation subset (test data) of 3 K. With *K-Fold Cross Validation* you can reduce computing time while maintaining the accuracy of model estimates. The model performance testing using the RapidMiner Studio application is done with the following steps:

1) Import the data needed for the process in the Rapid Miner tool. In the Rapid Miner application, select and click Import Data, then select the *SP File Name* data that will be used and then determine the attributes and labels that will be used.

	Name on SP File *polynominal*	Min *integer*	Max *integer*	Status SP *polynominal*
1	Bearing 6002 ZZ SKF	2	9	Critical
2	Bearing 6004 NR ZZ SKF	3	3	Non Critical
3	Bearing 6004 ZZ SKF	1	11	Critical
4	Bearing 6005 ZZ SKF	3	13	Critical
5	Bearing 6006 ZZ SKF	2	2	Non Critical
6	Bearing 6008 ZZ SKF	2	4	Non Critical
7	Bearing 6009 ZZ SKF	2	4	Non Critical
8	Bearing 6010 ZZ SKF	2	2	Non Critical
9	Bearing 6013 ZZ SKF	1	2	Non Critical
10	Bearing 606 ZZ SKF	2	2	Non Critical
11	Bearing 608 ZZ SKF	1	19	Critical
12	Bearing 614 2125 T2 SHI	1	1	Non Critical

Figure 2. Import Data

2) Click the Design menu, in the process view, add a dataset on the folder to the process display screen.
3) Next on the menu Validation, select the *Cross Validation* function operator, to apply performance evaluation to the K-Nearest Neighbor algorithm model that will be carried out. The *Cross Validation* function operator has parameters that can be used. Here *the number of folds parameter* is used to give the value *k* (number of iterations). Then the *sampling type parameter* is used to select a *sampling technique* that divides the dataset. In this test, *the number of folds parameter is determined* to be 10 (default) and *the sampling type* value is *automatic*.
4) Double click on the *Cross Validation function operator*, then on the Training section, select the Modeling menu, in the Predictive submenu, select *the k-NN function operator*.
5) Next, in the Testing section of the *Cross Validation function, add the* Scoring menu, select the *Apply Model operator* and drag it to the process display screen, through this function you can set the application of the model from the dataset used in this process to the prediction of the label that will be applied.
6) Add the *Performance operator* obtained from the Validation menu and Performance submenu and drag it to the process screen and place it in the Testing section of the *Cross Validation function*.

Connect all these commands to form a process flow diagram showing the flow of stages as in Figure 3 below.

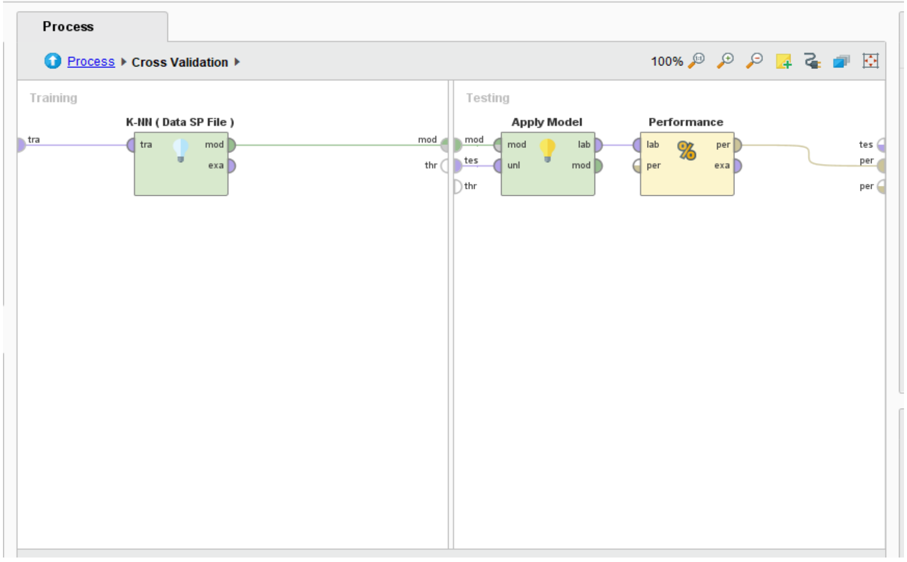


Figure 3. Evaluation of KNN Model with Cross Validation

Testing the performance of the classification model with the KNN algorithm in the RapidMiner Studio application obtained results as in Figure 4 below .

PerformanceVector

```
PerformanceVector:
accuracy: 98.00% +/- 4.00% (micro average: 98.00%)
ConfusionMatrix:
True:   Critical        Non Critical
Critical:       34      0
Non Critical:   2       64
precision: 97.14% +/- 5.71% (micro average: 96.97%) (positive class: Non Critical)
ConfusionMatrix:
True:   Critical        Non Critical
Critical:       34      0
Non Critical:   2       64
recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Non Critical)
ConfusionMatrix:
True:   Critical        Non Critical
Critical:       34      0
Non Critical:   2       64
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Non Critical)
AUC: 0.800 +/- 0.245 (micro average: 0.800) (positive class: Non Critical)
AUC (pessimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Non Critical)
```

Figure 4 Performance Results

And here is a model *Confusion Matrix* produced in evaluating the AUC classification model in the RapidMiner Studio application

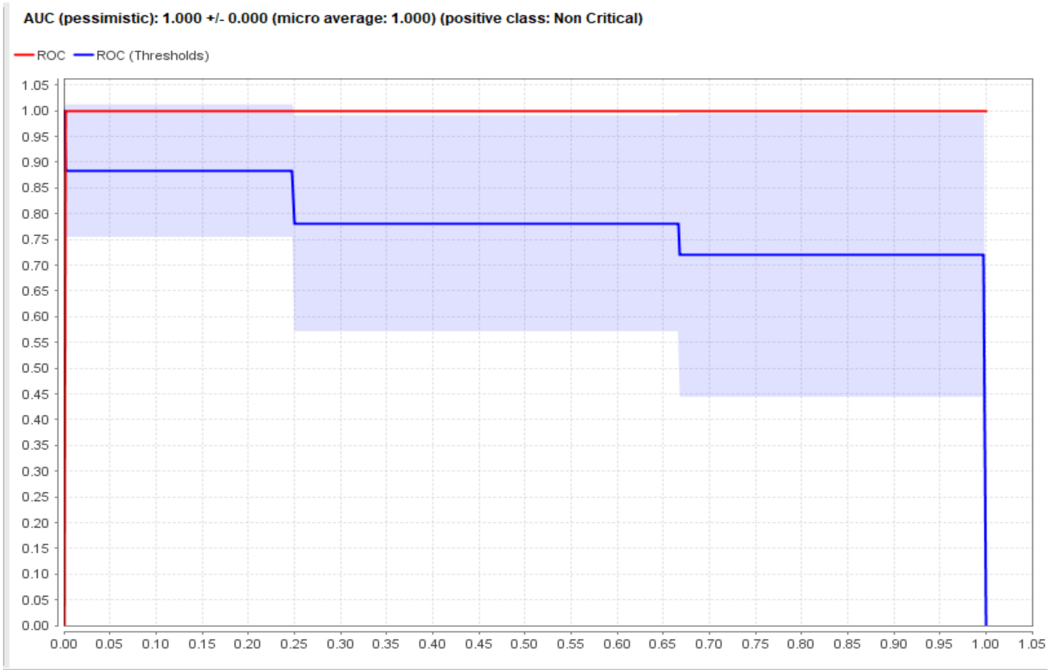


Figure 5. *AUC* results

3.2. Discussion

By using a 90 : 1 0 comparison scenario for *training data* and data *testing* from the dataset that is owned, then the model stage in applying the *K-Nearest Neighbor algorithm* produces predictions for objects to group into labels critical , non-critical labels . The total dataset used is 1000 data, so with this composition there are 900 data used as *training data* and 1000 data as *testing* data . In the simulation example, the calculation of the implementation of the *K-Nearest Neighbor algorithm* is carried out using 100 data taken randomly from the test dataset *(testing data)* . The results show that from the 100 datasets there are 2 objects included in the critical and non-critical categories , 36 objects included in the non-critical category , and 64 objects included in the critical category .

Table 1. Distribution of test data classifications *(testing data)* used in calculation simulations .

Name on SP File	Min	Max	Status SP
Bearing 6002 ZZ SKF	2	9	Critical
Bearing 6004 NR ZZ SKF	3	3	Non Critical
Bearing 6004 ZZ SKF	1	11	Critical
Bearing 6005 ZZ SKF	3	13	Critical
Bearing 6006 ZZ SKF	2	2	Non Critical
Bearing 6008 ZZ SKF	2	4	Non Critical
Bearing 6009 ZZ SKF	2	4	Non Critical
Bearing 6010 ZZ SKF	2	2	Non Critical
Bearing 6013 ZZ SKF	1	2	Non Critical

Next, in the ratio used in the modeling scenario , 1000 pieces of data are used as data *testing* and processed via the RapidMiner Studio application . The results of this process show that there are 36 objects entered in category classification critical, next there are 64 objects that are classified as non-critical at. The following is a graph of the results of the classification in the RapidMiner Studio application.

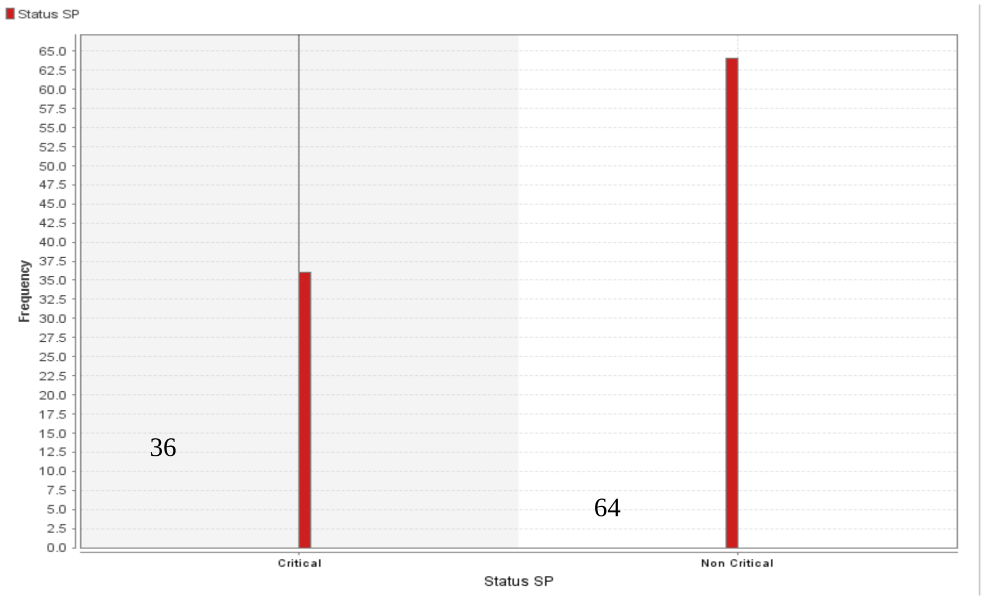


Figure 6. Classification distribution graph in the RapidMiner Studio application

Comparison The dataset used as test data produces relevant classifications in the RapidMiner Studio application as can be seen in the following comparison table.

Table 2. Comparison of the distribution of *testing* data classifications used in calculation simulations with the RapidMiner Studio application

ExampleSet (100 examples, 1 special attribute, 3 regular attributes)

Row No.	Status SP	Name on SP...	Min	Max
1	Critical	Bearing 6002...	2	9
2	Non Critical	Bearing 6004...	3	3
3	Critical	Bearing 6004...	1	11
4	Critical	Bearing 6005...	3	13
5	Non Critical	Bearing 6006...	2	2
6	Non Critical	Bearing 6008...	2	4
7	Non Critical	Bearing 6009...	2	4
8	Non Critical	Bearing 6010...	2	2
9	Non Critical	Bearing 6013...	1	2
10	Non Critical	Bearing 606 ...	2	2
11	Critical	Bearing 608 ...	1	19
12	Non Critical	Bearing 614 ...	1	1
13	Critical	Bearing 6201...	1	8
14	Critical	Bearing 6202...	1	6
15	Critical	Bearing 6204...	19	77
16	Critical	Bearing 6205...	1	13
17	Critical	Bearing 6206...	12	33

The author also discussed the testing of the implementation of the K-NN algorithm model using the RapidMiner Studio application by utilizing model evaluation using the *Cross Validation function* because this function not only creates several test data samples repeatedly, but divides the dataset into separate parts of the same size. Meanwhile, by using the stages of the model evaluation process with *Cross Validation* on training data *(training)* and test data *(testing)* , namely 1000 *record* dataset , *a Confusion Matrix* is formed which can describe *the Accuracy, Precision and Recall values* of the model applied, namely in evaluating the application of the *KN earest Neighbor algorithm* on product stock . Then there is an *Accuracy* value of 98.00 % with standard deviation +/- 4.00 % .

Copyright © 2023 IJSECS

Table View Plot View

accuracy: 98.00% +/- 4.00% (micro average: 98.00%)

	true Critical	true Non Critical	class precision
pred. Critical	34	0	100.00%
pred. Non Critical	2	64	96.97%
class recall	94.44%	100.00%	

Figure 7. *Accuracy* value of the classification model in the RapidMiner Studio application

Then there is a *Precision* value of 97.14 % with standard deviation +/- 6, 71 % .

Table View Plot View

precision: 97.14% +/- 5.71% (micro average: 96.97%) (positive class: Non Critical)

	true Critical	true Non Critical	class precision
pred. Critical	34	0	100.00%
pred. Non Critical	2	64	96.97%
class recall	94.44%	100.00%	

Figure 8. *Precision results* for the classification model Studio

Then there is a *Recall* value of 100.00 % with standard deviation +/- 0.00 % .

Table View Plot View

recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Non Critical)

	true Critical	true Non Critical	class precision
pred. Critical	34	0	100.00%
pred. Non Critical	2	64	96.97%
class recall	94.44%	100.00%	

Figure 9. Results *Recall* of Classification Models

By measuring model performance with *Cross Validation*, the resulting accuracy has a standard deviation or standard deviation value, which aims to see the distance between the average accuracy and the accuracy of each trial (iteration).

4. Related Work

In the field of predictive analytics and data-driven decision making, many methodologies have been explored and applied to address various business challenges. This section discusses the main studies and approaches relevant to the research. Niar *et al.* (2020) presents the implementation of the Naïve Bayes algorithm to predict rattan supplies. Their work focused on ensuring sustainable production and raw material availability during the Covid-19 pandemic, demonstrating the importance of predictive modeling in inventory management [1]. Arimi *et al.* (2022) proposed the use of the K-Nearest Neighbor (KNN) method to predict product sales in small and medium enterprises (MSMEs) engaged in fish processing. This approach highlights the applicability of KNN in forecasting sales trends of perishable products [2]. Darmi and Setiawan (2016) explored the application of the K-Means clustering method for product sales classification, especially in the context of minimarkets. Their research highlights the potential of K-Means in categorizing products based on their sales performance, thereby aiding inventory management [4]. Adiana *et al.* (2018) addresses customer segmentation using a combination of RFM models and clustering techniques. This study shows the effectiveness of data mining approaches in identifying customer behavior patterns for targeted marketing strategies [5].

Suyanto *et al.* (2022) introduced a new framework for diabetes detection based on nearest neighbor algorithm. By combining KNN with clustering and dimensionality reduction techniques, they achieved competitive levels of accuracy, highlighting an innovative approach to healthcare analytics [6]. Susena *et al.* (2018) optimized Support Vector Machine (SVM) parameters with Particle Swarm Optimization (PSO) for blood donor classification. This study provides an example of the importance of parameter tuning in machine learning models for correct predictions [7]. Ulya *et al.* (2021) focuses on optimizing the K-Nearest Neighbor algorithm to prioritize village development assistance. Their work shows the importance of parameter selection in KNN for accurate classification [10]. Liklikwatil and Noersasongko (2018) presented K-Nearest Neighbor optimization using Particle Swarm Optimization to predict rubber commodity prices. This study emphasizes the potential of hybrid algorithms in improving prediction accuracy [11]. Miftahuddin *et al.* (2020) explored font type identification using the Genetic Modified K-Nearest Neighbor method. Their research shows the application of machine learning in font recognition, which has an impact on the design and graphics industry [12]. Fahlevi (2020) applied Genetic Modified K-Nearest Neighbor to classify recipients of welfare rice distribution. This research shows the versatility of GMKNN in solving classification problems [15]. Angriani (2017) discusses inventory management for retailers using the Single Exponential Smoothing method. This study emphasizes the importance of time series forecasting techniques in inventory control for retailers [19].

The differences between this research and previous research lie in the methodology, research focus, and application context. This research, in predictive analytics and data-driven decision making, describes various methods such as Naïve Bayes, K-Nearest Neighbor, K-Means clustering, RFM models, and others, applied in various business situations. Past research, on the other hand, covers a wide range of topics such as rattan stock prediction, product sales, customer segmentation, diabetes detection, blood donor classification, machine learning parameter optimization, and typeface recognition. Meanwhile, previous research may be more focused on certain aspects of data analysis such as classification, parameter optimization, or the use of certain algorithms in more specific contexts. In addition, this research and previous research may have differences in the businesses they focus on. This research covers various aspects such as inventory management, sales forecasting, customer segmentation, and health analysis, whereas previous research may have focused more on one or two of these aspects. In terms of methodology, this research and previous research may have different approaches and data analysis techniques, depending on the problem to be solved. Previous research may have used more general or specific methods, while this research introduces a variety of different methods. However, they both have one thing in common, namely that they both contribute to further understanding of how data analysis and machine learning methods can be applied in business and produce valuable insights for practitioners and researchers in this field.

5. Conclusion

Based on the results of research conducted by the author, the conclusions obtained are as follows: 1) The K-Nearest Neighbor algorithm model on testing data of 100 data objects, obtaining results that show new insights in the form of classification. Meanwhile, using the stages of the model evaluation process with Cross Validation on training and testing data, namely 1000 dataset record that has 36 critical and 64 non-critical results, and 2) Performance evaluation and testing using the RapidMiner Sstudio application can provide optimal results with the modeled scenarios. This algorithm model has an accuracy value of 98.00 % with standard deviation +/- 4.00 %.

Acknowledgments

This paper and the research behind it would not have been possible without the permission of Allah SWT. Mrs. Sifa Fauziah and Mr. Donny Maulana, for their enthusiasm, knowledge and exacting attention to detail have been an inspiration and kept my work on track of this paper. My colleagues at Pelita Bangsa University have also supported me during my study. I am also very thankful for the insights contained in various references that provide materials to keep me knowledged. The generosity and expertise of one and all have improved this study in innumerable ways and saved me from many errors; those that inevitably remain are entirely my own responsibility.

References

[1] Niar, Y., Komariah, K., Surip, A., Saputra, R. and Ali, I., 2020. Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Rotan. *KOPERTIP: Scientific Journal of Informatics Management and Computer*, *4*(1), pp.28-34. DOI: https://doi.org/10.32485/kopertip.v4i1.112.

[2] Arimi, I., Purwaningsih, R. and Rosyada, Z.F., 2022. METODE K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI PENJUALAN PRODUK PADA UMKM PENGOLAHAN IKAN MAJU JAYA. *Industrial Engineering Online Journal*, *12*(1).

[3] Renhoran, B.S., Nurhandayani, N. and Septiana, L., 2018. Penerapan Algoritma C4. 5 Untuk Menentukan Data Stok Dan Target Permintaan Material Yang Paling Dibutuhkan Gudang Logistik Pada PT. PLN (Persero) Area Kebon Jeruk. *INTI Nusa Mandiri*, *12*(2), pp.13-20.

[4] Darmi, Y.D. and Setiawan, A., 2016. Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama*, *12*(2). DOI: https://doi.org/10.37676/jmi.v12i2.418.

[5] Adiana, B.E., Soesanti, I. and Permanasari, A.E., 2018. Analisis segmentasi pelanggan menggunakan kombinasi RFM model dan teknik clustering. *Jurnal Terapan Teknologi Informasi*, *2*(1), pp.23-32. DOI: https://doi.org/10.21460/jutei.2018.21.76.

[6] Suyanto, S., Meliana, S., Wahyuningrum, T. and Khomsah, S., 2022. A new nearest neighbor-based framework for diabetes detection. *Expert Systems with Applications*, *199*, p.116857. DOI: https://doi.org/10.1016/j.eswa.2022.116857.

[7] Susena, I.G.N.E., Furqon, M.T. and Wihandika, R.C., 2018. Optimasi Parameter Support Vector Machine (SVM) dengan Particle Swarm Optimization (PSO) Untuk Klasifikasi Pendonor Darah Dengan Dataset RFMTC. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *2*(12), pp.7278-7284.

[8] Retno Tri Vulandari, 2017. *Data Mining* . Yogyakarta: Gava Media.

[9] Saepudin, A., Aryanti, R., Fitriani, E. and Dahlia, D., 2020. Optimasi Algoritma SVM Dan k-NN Berbasis Particle Swarm Optimization Pada Analisis Sentimen Fenomena Tagar# 2019GantiPresiden. *Jurnal Khatulistiwa Informatika*, *6*(1), pp.95-102.

[10] Ulya, S., Soeleman, M.A. and Budiman, F., 2021. Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa. *Techno. Com*, *20*(1), pp.83-96. DOI: https://doi.org/10.33633/tc.v20i1.4215.

[11] Liklikwatil, R.D., Noersasongko, E. and Supriyanto, C., 2018. Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet. *E-JURNAL JUSITI: Jurnal Sistem Informasi dan Teknologi Informasi*, *7*(2), pp.172-182. DOI https://doi.org/10.36774/jusiti.v7i2.252.

[12] Miftahuddin, Y., Umaroh, S. and Dwiutama, A.A., 2020. Identifikasi Jenis Font Menggunakan Metode Genetic Modified K-Nearest Neighbor. *Rekayasa Hijau: Jurnal Teknologi Ramah Lingkungan*, *4*(3), pp.157-166. DOI: https://doi.org/10.26760/jrh.v4i3.157-166.

[13] Hermawan, Y.D., Hariadi, V. and Amaliah, B., 2017. Implementasi Algoritma K-Nearest Neighbors dengan Particle Swarm Optimization dalam Klasifikasi Trouble pada Base Transceiver Station (BTS). *Jurnal Teknik ITS*.

[14] Nofriansyah, D. and Nurcahyo, G.W., 2015. *Algoritma data mining dan pengujian*. Deepublish.

[15] Fahlevi, R., 2020. *PENERAPAN GENETIC MODIFIED K-NEAREST NEIGHBOR DALAM KLASIFIKASI PENERIMA BERAS SEJAHTERA* (Thesis, Universitas Islam Negeri Sultan Syarif Kasim Riau).

[16] Iku, M.H., Mustofa, Y.A. and Kumala, I.S., 2019. Metode k-nearest neighbor untuk memprediksi harga eceran beras di pasar tradisional gorontalo. *Jurnal Cosphi*, *3*(2).

[17] Villacampa, O., 2015. *Feature selection and classification methods for decision making: a comparative analysis* (Thesis, Nova Southeastern University).

[18] Sharif, A., 2019. Data mining untuk memprediksi itemset promosi penjualan barang menggunakan metode market basket analysis (mba)(studi kasus: toko sentra ponsel). *Jurnal Mantik Penusa*, *3*(2, Des).

[19] Angriani, H., 2017. SISTEM MANAJEMEN PERSEDIAAN BARANG PADA RETAILER MENGGUNAKAN METODE SINGLE EXPONENTIAL SMOOTHING. *JTRISTE*, *4*(1), pp.60-71.

[20] Hutabarat, S.M. and Sindar, A., 2019. Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, *2*(2), pp.126-132. DOI: https://doi.org/10.32672/jnkti.v2i2.1555

Copyright © 2023 IJSECS