

Prediction of librarian interest in library management in Pamekasan with comparison of SVM and KNN algorithms

¹Lathifah, ²Imam Yuadi

¹Program Studi Magister Sains Informasi dan Perpustakaan, Universitas Airlangga, Indonesia

²Universitas Airlangga, Indonesia

Email: ifalathifah22@gmail.com

received : 25 January 2025, revised : 09 March 2025, accepted : 24 May 2025

ABSTRACT

This study was conducted to determine the prediction of librarian interest in joining a library organization. Using survey data and interviews with librarians that produced 130 test data then divided into two groups of data, namely "interested" and "not interested". Using the Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) models as a comparison of the performance of the two algorithms in classifying librarian interests. The results of the test data were then evaluated using a confusion matrix to assess the accuracy, precision, and recall of each model. The results of the interest predictions tested showed that the use of the SVM model was more consistent in classifying librarian interests with high accuracy, although there were some errors in the "Not Interested" category. While the results of interest predictions using the KNN model tended to dominate the prediction of the "Interested" category, there were more errors in identifying the "Not Interested" category. Both models show their respective advantages and disadvantages in classifying librarian interest predictions. From the results of this study, it can be a picture and insight into the effectiveness of using the two models in classifying librarian interest predictions in joining a library organization and as a guide in choosing the right algorithm in similar research.

Keywords : interest classification; k-nearest neighbors (KNN); library organization; support vector machine (SVM)

1. INTRODUCTION

The library is a form of social organization that is never free from the communication process in its various activities and communication is also needed to facilitate coordination with various parties, both with parties within the library itself and other parties outside the library (Nurul Hayati, 2014). Library organization is a must for librarians to be able to explore information and be involved in the library community. Proper leadership and management practices in libraries are essential to achieving organizational efficiency and librarian involvement, which can have a positive impact on the development of the institution as a whole (Weiner, 2017). Librarians are at the forefront of developing libraries that are in accordance with the needs and interests of their users.

The changing academic library environment has had a significant impact on the knowledge, skill needs, and roles of library leaders, and how organizations can sustain and manage the changing nature of academic libraries and leadership, while effective leadership is essential for library leaders and organizations (Aslam, 2018). The activeness and participation of librarians in library organizations can help improve skills, abilities, innovation, and professional collaboration. Collaboration between librarians and institutions can help bridge institutions whose quality is still relatively underdeveloped. Librarians can also learn about professional leadership through these professional organizations. However, in Pamekasan, not all librarians have joined such organizations. So that librarians are less than optimal in managing libraries and their leadership is less effective, collaborative, and innovative. Research on how librarians' interest in joining library

organizations is carried out so that library organizations can collaborate with librarians in developing their libraries and can collaborate with other institutions that can dominate each other.

This study was conducted by focusing on the use of two modal algorithms, namely K Nearest Neighbor (KNN) and support vector machine (SVM) in classifying librarians' interest in joining a library organization. Both machine learning approaches have been widely used in various studies to classify and predict a research object. The use of data analysis techniques using data mining can be applied in research to determine the classification, prediction and grouping of data. In complex data types, predictions of data results will be very accurate if analyzed using machine learning techniques. The predicted data is analysis data generated based on the observed variables (Helmy et al., 2019) . The K-NN algorithm model can be used in data analysis that is easier to analyze and flexible, while the SVM model is better known as an algorithm model that has the ability to analyze complex data with larger volumes, not only that, SVM can also analyze data quickly and accurately. Both models, both Support Vector Machines (SVM) and K -Nearest Neighbors (k-NN) are algorithm models that are already popular in binary classification analysis (Aslam, 2018) . If these two models are used in research applications that predict librarians' interest in library organizations, then the determinants that can influence the prediction results will be analyzed in a new way by the two models.

The SVM and K-NN models can be considered for use in comparing the results of research data predictions, considering the efficiency of algorithms based on complex data (Manimaran, Priyadharshini, and Yamuna Devi 2022) . In addition, in a study that predicted the interest of PKBM students using data from package C graduates in 6 TPN seats, using the SVM and K-NN models, the results showed that the use of data analysis results with the SVM model was superior to using the K-NN algorithm model with a value of 97.8% (Anin Nazura and Budi Susanto 2023) . This was also stated in a study that predicted customer interest in savings products, showing that the use of the SVM algorithm model was appropriate and produced accurate prediction data (Saputra and Ary 2020) . In addition, a comparison of the SVM and K-NN algorithm models was also used in predicting patients with heart disease. In the test data results using both normalization and without normalization, the use of the SVM model showed better results and higher accuracy compared to using the K-NN model. (Anggoro and Permatasari 2023) . However, it is different from previous studies. The use of the K-NN model produces more accurate data compared to using the SVM algorithm model analysis, this can be seen from the implementation time, accuracy, precision and classification factors (Parul Sinha and Poonam Sinha 2015) .

A systematic methodology is used in this study to determine the results of librarian interest predictions in library organizations. It begins with the collection of librarian information data in Pamekasan, then continues with data preprocessing so that the data is of higher quality. After the data is processed, data testing will be carried out using the SVM and K-NN algorithm models. The performance results of the model are then evaluated using confusion and curation matrices in order to produce a more comprehensive data comparison. The use of this model not only looks at the results of data classification and prediction, but can also provide an understanding of the application of machine-based data analysis models that can be applied to cases that occur in the main library.

The results of this study can ultimately provide a real contribution to the library world regarding the application of machine based learning in analyzing various problems that occur in the library world. By analyzing the interests of librarians in library organizations using the SVM and K-NN algorithm models, it is hoped that it can be a basis for library leaders and library organizations to take better policies for librarians, especially in Pamekasan. By using both models, it can increase knowledge in comparing the two models in predicting librarian interests and add references in showing the advantages and disadvantages of the K Nearest Neighbor (KNN) and support vector machine (SVM) algorithm models.

Previous studies have shown that the use of SVM and K-NN algorithm models has been widely used in analyzing prediction data in various types of research. As in the study using the results of

a comparison of two algorithm models, K Nearest Neighbor (KNN) and support vector machine (SVM), in classifying Arabic texts, the results showed that the use of the SVM model showed a better prediction time compared to the use of the K-NN model (Hmeidi et al., 2008). Not only that, the K-NN algorithm model is also used to create a book recommendation system in libraries that follows comparable library visitor preference data. The K-NN algorithm model can also be applied well to small or simple and medium data. This makes the use of the K-NN model appropriate for application in library cases with a small amount of library data or simpler systems (Sharma et al., 2023).

In other classification studies, several studies have tried to combine both SVM and K-NN algorithms to obtain more optimal results. In conducting comparative data research, both SVM and K-NN algorithms are model options that can be combined with recommendations based on (Tabrizi & Cavus, 2016). The SVM model can be used to identify categories or groups of user data in specific interest prediction research, while the K-NN model can recommend or predict requests based on similarities with other users. By using both SVM and K-NN analysis approach models, it has advantages in terms of accuracy and efficiency in predicting interests such as predicting librarian interests in library organizations.

In terms of predicting the number of students who graduate on time which is affected by student inactivity, using three algorithm models, one of which is the SVM model, the results of the study showed that the SVM algorithm has the best accuracy in making predictions with a precision value of 95% (Wiyono & Abidin, 2019). While the SVM and K-Nearest Neighbor models are most widely used for classification & regression analysis (Raj Kumar, 2016). When compared to other cases, the KNN, SVM and Decision tree classifications are reported to be the leading classifiers in producing data with high accuracy. However, only a few studies have compared the performance of these models on different sample data (Thanh (Noi & Kappas, 2017).

Both models have their own advantages and disadvantages in classifying librarian interest predictions. The results of this study will provide insight into the effectiveness of using SVM and KNN in similar studies and be a guide in choosing the right algorithm to predict librarian interest in joining a library organization. The background description above leads to the formulation of the research problem: how do the SVM and KNN algorithms compare in determining librarian interest in joining a library organization in Pamekasan?

Therefore, the purpose of this study is not only to compare the results of prediction performance using SVM and K-NN algorithm model analysis, but also to explore the collaborative tendencies of librarians as reflected in their interest in joining library organizations. This analysis represents the potential for professional collaboration between individual librarians and library organizations that can support institutional development. By understanding these interests, library organizations can design more targeted, data-driven participation strategies that can increase librarian involvement in institutional activities.

2. METHODS

This study uses a combination approach between quantitative techniques based on machine learning and descriptive qualitative methods to strengthen scientific validity through data triangulation and analysis. The types of data used are semi-structured interview data and questionnaires from librarians in various types of libraries (schools, Islamic boarding schools, public, and universities). Construct validity tests were carried out on the questionnaire instrument. Data were analyzed using the Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms, which compared their performance in predicting librarian interest. Model evaluation using Confusion Matrix, accuracy, precision, recall, and reliability through Cronbach's Alpha.

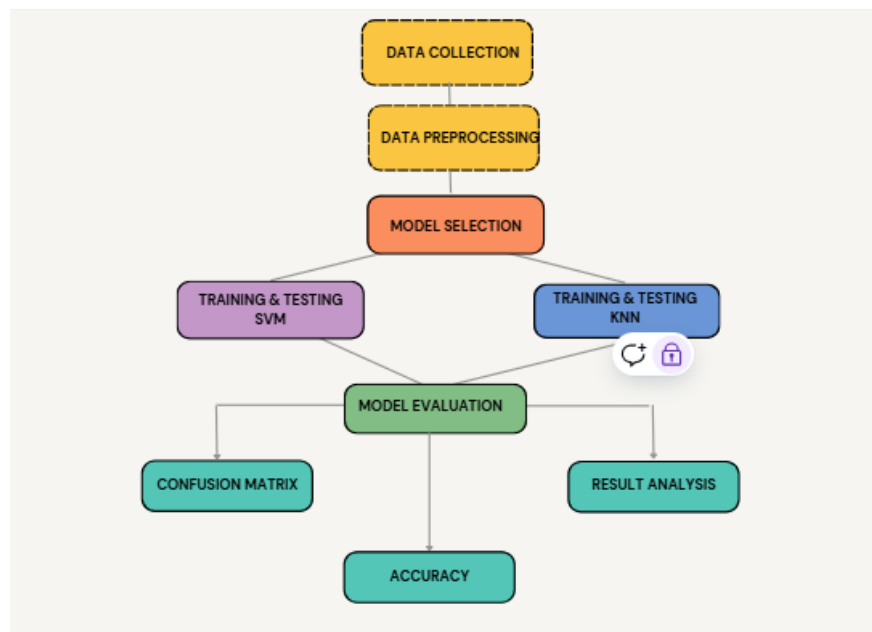


Image 1 research method flowchart

This research was conducted through several stages, including data collection, data preprocessing, data processing, algorithm model selection, algorithm model comparison, and data result evaluation. At the data collection stage, the data were in the form of interview results and survey responses from librarians in Pamekasan. The type of interview used was a semi-structured interview. The data collection method which included interviews and questionnaires had gone through construct validity testing. The librarians involved in this study worked in various types of libraries, including school libraries, Islamic boarding school libraries, public libraries, and college libraries. A total of 49 librarian data were successfully collected, consisting of 26 school librarians, 4 public librarians, 4 Islamic boarding school librarians, and 15 college librarians. However, not all of them were members of a library organization. Only 21 librarians were members of a library organization, while 28 librarians were not yet members of a library organization. Based on interview and survey data, analysis can be carried out to determine the level of librarian interest in joining a library organization.

The next stage is data preprocessing, which involves adjusting the data by handling missing values and duplicate data to ensure that it is ready for analysis using the appropriate algorithm. The data must be completely transformed into a usable format. Once data cleaning is completed in the preprocessing stage, the next step is data processing, which includes selecting an algorithm model that aligns with the research objectives and data analysis needs. This study uses the Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN) algorithm models, adopting a comparative study approach. These two models were chosen because they are suitable for binary classification, which aims to classify librarians into the category of “interested” or “not interested” in joining a library organization. Once the analysis model is determined, the next step is data testing. The data is divided into two sets: a training data set and a testing data set. This division allows the model to learn patterns from a portion of the data (training set) and then validate its performance on unseen data (test set). Through this data processing stage, new data will be generated as a result of the analysis using the two models.

After the data is processed using the SVM and K-NN models, the next stage which is the final step of data processing is model evaluation. The concept of reliability used is Cronbach's Alpha. Cronbach's Alpha is a measure of internal consistency or reliability of a set of test items or scales (Tavakol and Dennick 2011) . This can help provide knowledge about the performance of the model in producing new data that is in accordance with the objectives and needs of the study. In the evaluation of this model, Confusion Matrix and Accuracy are used. Confusion Matrix will

show the results of true positive, false positive, true negative, and false negative data, so that it can provide detailed knowledge about the results of the model's performance in classifying and distinguishing librarian data that are "Interested" and "Not Interested". However, Accuracy can produce the right and correct percentage that can combine the advantages of each model as a whole.

Metrics are evaluated at the final analysis stage to produce conclusions regarding the effectiveness of the performance of using the SVM and K-NN models. This stage can help in determining the use of the most appropriate algorithm model to be used in analyzing research data with the expected goals and needs. The use of this research model can ensure that the data analysis is comprehensive and the data results from the selected model are supported by strong evidence.

3. RESULTS AND DISCUSSION

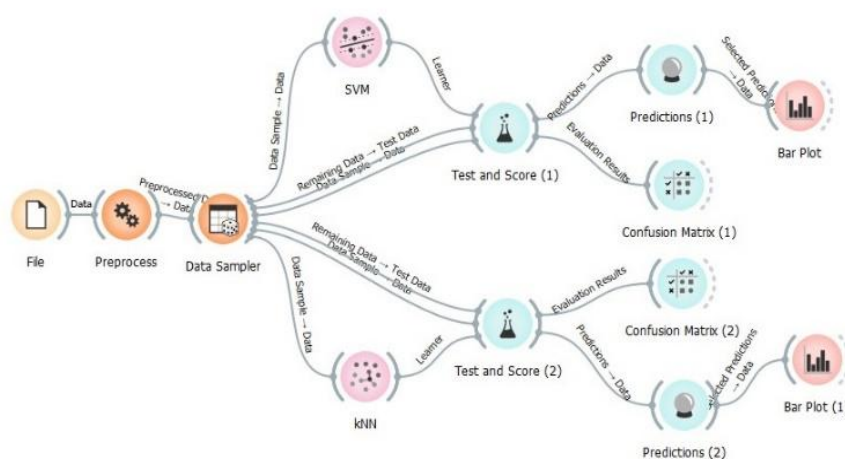


Image 2 data processing using orange

Data processing in Image 2 uses the Orange Data Mining application and produces new data for librarian data who are interested and not interested in library organizations. In predicting librarian interest in library organizations, this study uses two classification algorithm models as a comparison of their performance, namely the SVM (Support Vector Machine) and KNN (K-Nearest Neighbors) models. This processed data uses the orange data mining application and produces new data as data for librarians who are interested and not interested in library organizations. The results of the two models can be compared because they show different classification patterns. It can be seen that the SVM model looks more consistent in predicting librarian data with the "interested" category, although it shows some inaccurate data results in predicting "not interested" librarian data.

Meanwhile, in the K-NN algorithm model, the tendency of most data shows classification data as "interested" data. Meanwhile, librarian data that is predicted to be "not interested" appears sporadically. Both of these data result patterns show a comparison of data that can provide knowledge about the advantages and disadvantages of the two analysis models. This can also be seen from the data pattern and accuracy in classifying data so that it can be used as a reference and consideration in choosing an algorithm model that is in accordance with the objectives and characteristics of the research analysis.

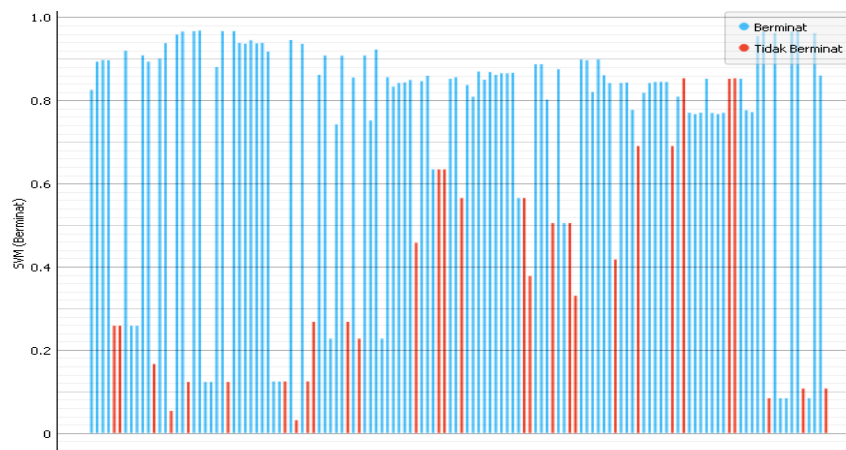


Image 3 Graph of data analysis results using svm

In the graph of the results of data analysis using SVM shown in Image 3, to predict librarian interest data in library organizations, it is shown using a bar chart. In the diagram shown in the vertical line is the probability value of librarian interest prediction which has a value of 0 to 1. The more confident the librarian is to join a library organization, the closer the value will be to 1, this belief is included in the "interested" prediction category. While the lower value approaching 0 is librarian data that shows librarian data that tends to be "not interested" in library organizations. On the horizontal line is data that presents librarian data that is sorted by providing an overview of the prediction results for each librarian data analyzed.

Next, the diagram showing the category of librarian data that is "interested" is visualized with a blue line. While the red line is the data of librarians who are included in the "not interested" category. However, the diagram shows that most of the librarian data is included in the "interested" data. This is indicated by the number of blue lines compared to the number of red lines. The red lines spread between the blue lines and show a lower probability value compared to the probability value of the blue line which is close to 1. From the results of this data comparison, it can be concluded that data analysis using SVM shows more data with predicted classification of interested librarians or can also indicate a bias in the data that causes the model to tend to predict interest in the majority of cases.

		Predicted		
		Berminat	Tidak Berminat	Σ
Actual	Berminat	85	15	100
	Tidak Berminat	4	26	30
Σ		89	41	130

Image 4 SVM confusion matrix

Image 4 on the SVM confusion matrix shows the performance of the prediction model on the librarian data that is "interested" and "not interested" is shown in the SVM model confusion matrix table above. With the actual data in this matrix shows the comparison data from the SVM model prediction results. The results of the study show that there are four categories of prediction values in this matrix, namely "True Positive" (Interested predicted Interested) totaling 85, "False Negative" (Interested predicted Not Interested) totaling 15, "False Positive" (Not Interested predicted Interested) totaling 4, and "True Negative" (Not Interested predicted Not Interested) totaling 26. Overall there are 130 data with 100 cases of "Interested" and 30 cases of "Not Interested".

The table data shows that the use of the SVM model has a fairly high accuracy in predicting librarian data that is "Interested" in the library organization. However, the data still shows prediction errors in the data of librarians who are "Not Interested" in the library organization. There are some data that should show data on librarians who are "interested" or "not interested" are predicted incorrectly and are included in the "False Negative" and "False Positive" categories . From the number of correct predictions (85 and 26) , it can be seen that most predictions using this SVM model are quite accurate. In analyzing the advantages and disadvantages of the SVM model, a confusion matrix can be used. This can help in identifying deficiencies and errors in the process of analyzing "not interested" prediction data.

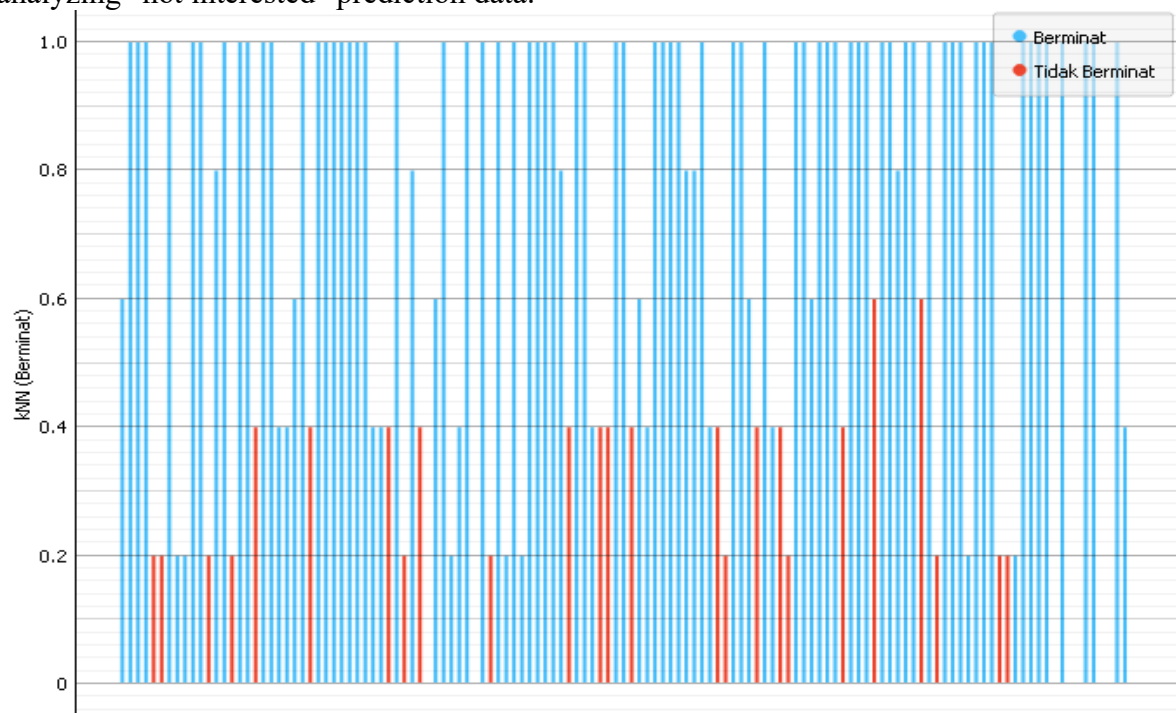


Image 5 graph of data analysis results using knn

Image 5 shows a graph of the results of data analysis using KNN. The results of predicting librarian interest in library organizations using the K-NN (K-Nearest Neighbors) algorithm model are illustrated in the bar chart above. The chart shows that most of the librarian data is classified as "interested". This is indicated by the number of blue bars being greater than the number of red bars. The red bars are spread among the blue bars and show a lower probability value compared to the probability value of the blue bars, which is close to 1.

While in the diagram showing the category of librarian data that is "interested" is visualized with a blue line. While the red line is the data of librarians who are in the "not interested" category. However, in the diagram it can be seen that most of the librarian data is included in the "interested" data. This is indicated by the number of blue lines compared to the number of red lines. The red lines spread between the blue lines and show a lower probability value compared to the probability value of the blue line which is close to 1. From the results of the analysis, it can be seen that the use of the K-NN model in predicting the interest of librarians who are in the "interested" category has a strong tendency compared to data in the category of librarians who are "not interested" in the library organization. So it can be concluded that the use of predictive analysis using the K-NN model has a tendency prediction pattern and can be used when analyzing interest predictions in certain studies.

		Predicted		
		Berminat	Tidak Berminat	Σ
Actual	Berminat	80	20	100
	Tidak Berminat	2	28	30
Σ		82	48	130

Image 6 knn confusion matrix

The data generated in the confusion matrix in Image 6 shows data in two categories, namely librarians who are “interested” and “not interested” in the library organization. There is actual prediction data generated, in the data display column is data that shows librarian data that has been successfully predicted, while in the data row is actual data that has been successfully predicted. Between the actual data and the prediction data applied to the table is the result of a combination of analyzed data. There are 80 accurate librarian data in the “interested” category and 28 librarian data in the “not interested” category to join the library organization. However, there are 20 data that show errors in the classification of “interested” librarian data into the “not interested” librarian category, and 2 librarian data that are predicted to be “not interested” into the “interested” librarian data category to join the library organization.

Comparative analysis using SVM and KNN models in predicting librarian interest in joining a library organization is very relevant to previous research. The application of these two models can be used to predict interest in various types of research related to libraries, such as predicting user interest in library services or librarian interest in library organizations. The SVM and KNN models can be compared to find out the most effective results, so that they can be used as a relevant further research method. In several previous studies on the application of the SVM and KNN models, the results showed that both machine learning models were quite effective in predicting interest in a study. However, when comparing the two models, the SVM model has a better prediction time than the KNN model. This finding is in accordance with research conducted by Hmeidi, et al. 2008

In the study of librarian interest prediction in joining a library organization using SVM and KNN analysis, it is necessary to pay attention to its use in determining the prediction of interest in the library organization. Previous research can be a guideline and foundation in deeper knowledge about when the use of the SVM and KNN algorithms can be applied in a study. Meanwhile, the combination of the two algorithm models can enrich the research results in comparing more accurate and effective prediction results that can be adjusted to library research. As in the study conducted by (Tabrizi and Cavus 2016) which suggested combining both SVM and KNN algorithms in a study, to find out the most accurate algorithm model in determining the results.

In previous studies, the SVM model algorithm showed different character recognition accuracy results in different studies. Meanwhile, the KNN model is very easy to use and apply in data processing. This is an advantage in itself if using the KNN model (Sharma et al. 2023) . So it is important to be able to choose the use of an algorithm that can handle complex data effectively and precisely. Comparing the use of the two models can enrich the research results on the effectiveness of using the KNN and SVM algorithms in predicting interest in a study related to libraries. Several previous studies have also found that KNN can provide easy and fast prediction results, while the SVM model is superior in handling more complex data. From the results of data

analysis using the two models, it can be seen that the use of the SVM algorithm model is superior to the K-NN algorithm model in predicting librarian interest in library organizations. This is shown in the results of the prediction analysis and also the evaluation results using the confiding matrix on the results of the two models. However, both models are quite effective in analyzing certain data according to the objectives of the research being carried out.

4. CONCLUSION

Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) models in comparing data analysis results show data results with their respective patterns, this can illustrate the differences in advantages and disadvantages of each model. More consistent prediction data results on more complex data sets are shown in the use of the SVM model, because the model has the ability to optimally separate class data using hyperplanes, the model is also known to be robust and can handle high-dimensional data. However, on simpler data sets with smaller sizes and more uniform distributions, the K-NN model can be applied and shows good and accurate performance. Although the K-NN model is more susceptible to overfitting on unstructured data. This is a weakness in itself and is a consideration when using the K-NN analysis model.

Overall, SVM is superior in the context of predicting complex librarian interests. For library organizations, it is recommended to develop a membership system based on predictive data analysis to increase librarian participation. In addition, librarian interest-based training and integration of collaborative activities into the organization's work program can encourage active involvement. Understanding the characteristics of algorithms and their uses will help in determining the right approach for predictive analysis in the library field and the development of librarian professional organizations.

REFERENCES

- Afrash , M.R., Shafiee , M., & Kazemi-Arpanahi , H. (2023). Building a machine learning model to predict early risk of gastric cancer based on lifestyle factors. *BMC Gastroenterology* , 23 (1).
- Anggoro , D.A., & Permatasari , D. (2023). Comparison of Kernel Performance of Support Vector Machine Algorithm for Diabetes Mellitus Classification. *International Journal of Advanced Computer Science and Applications* , 14 (1), 580–585.
- Anin Nazura, I., & Budi Susanto, A. (2023). *Analysis of Student Satisfaction Package C PKBM Siliwangi Pamulang Accepted in State Universities Using Support Vector Machine Algorithm and K-Nearest Neighbor Method (Case Study: PKBM Siliwangi School)* . 1 (4), 39–49.
- Aslam, M. (2018). Current trends and issues affecting academic libraries and leadership skills. *Library Management* , 39 (1–2), 78–92.
- Helmy, Y., Khedrb , A.E., Koliefa , S., & Haggag, E. (2019). An enhanced business intelligence approach to improve customer satisfaction using mining techniques. *International Journal of Computer Science and Information Security (IJCSIS)* , 17 (4), 159–175.
- Hmeidi , I., Hawashin , B., & El- Qawasmeh , E. (2008). *Performance of KNN and SVM classifiers on full Arabic articles* . Volume 22 (Issue 1), Pages 106-111.
- Manimaran, P., Priyadharshini, G., & Yamuna Devi, N. (2022). Predicting placement eligibility for students using data mining. *International Journal of Health Sciences* , 6 (May), 460–467.
- Massis, B. E. (2012). The use of predictive analytics in libraries. *New Library World* , 113 (9), 491–494.
- Nurul Hayati. (2014). Communication in library organizations. *Communication in library organizations Nurul* , 2 (1), 49–59.
- Parul Sinha, & Poonam Sinha. (2015). Comparative Study of Chronic Kidney Disease Prediction Using KNN and SVM. *International Journal of Engineering Research and Development* , V4 (12).
- Raj Kumar, MAS (2016). *Comparative analysis of SVM & KNN for academic prediction of*

students . Volume -- 1 , 15–19.

- Saputra, D., & Ary, M. (2020). Prediction of Customer Interest in Savings Products Using Polynomial Kernel Parameter SVM Algorithm. *E-Proceedings Information System (POTENSI)* , 1 (1), 162–171.
- Sharma, N., Appukutti, S., Garg, U., Mukherjee, J., & Mishra, S. (2023). Analysis of Students' Academic Performance Based on Time Spent on Extracurricular Activities Using Machine Learning Techniques. *International Journal of Modern Education and Computer Science* , 15 (1), 46–57.
- Tabrizi, S. S., & Cavus, N. (2016). Hybrid KNN-SVM Model for Vehicle License Plate Recognition in Iran. *Procedia Computer Science* , 102 (August), 588–594.
- Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classification for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)* , 18 (1).
- Weiner, SG (2017). Academic Library Leadership: A Literature Review. *Educational Libraries* , 26 (2), 5.
- Wiyono, S., & Abidin, T. (2019). Comparative Study of KNN, SVM, and Decision Tree Machine Learning Algorithms to Predict Student Learning Achievement. *International Research Journal - GRANTHAALAYAH* , 7 (1), 190–196.