JRSSEM 2021, Vol. 02, No. 02, 159 – 166
E-ISSN: 2807 - 6311, P-ISSN: 2807 - 6494

Comparative Analysis of Methods K-Nearest Neighbor, Support Vector Machine and Decision Tree on Prediction Model of Turnover Intention

Syarif Sagaf Adibaji [1]
Onny Marleen [2]
[1] *Gunadarma University, Indonesia*
[2] *Gunadarma University, Indonesia*
*e-mail: assegaf819@gmail.com[1], onny_marleen@staff.gunadarma.ac.id[2]
Correspondence: assegaf819@gmail.com

Submitted: 04 September 2022, *Revised*: 15 September 2022, *Accepted*: 28 September 2022

This study analyzed the comparison of methods on machine learning technique to predict turnover intention, turnover intention refers to intention or possibility of an employee to leave a company or the job that he is currently working on. The analysis with comparing the K-Nearest Neighbor, Support Vector Machine and Decision Tree methods, in an effort to predict turnover intention and reduce the risks of turnover intention in employee. The dataset used is taken from the Kaggle dataset, the dataset file is in the form of human resource (HR) data records with 311 data records with 24 features used out of 36 features. The dataset is obtained by using the K-Nearest Neighbor, Support Vector Machine and Decision Tree methods to calculate the accuracy, precision and sensitivity with a confusion matrix, the results of accuracy, precision and sensitivity from those three methods are compared and the method with the highest average percentage of accuracy, precision and sensitivity will be used as a prediction model.

Keywords: Prediction Model, Turnover Intention, K-Nearest Neighbor, Support Vector Machine, Decision Tree.

1. INTRODUCTION

Employees in a company are one of the components that play an important role in the sustainability of a company. Recently, since the beginning of 2020, at that time cases of the COVID-19 virus have just emerged, there is a new work dynamic that has emerged, namely the number of employees who quit their jobs. This was stated by Professor Anthony Klotz from the University of Texas A&M who predicted that many employees would want to move or leave their jobs in May 2021, which later in America became known as the trend of the phenomenon "The Great Resignation" or "Big Quit". The desire to move or turnover intention is a problem that is widely highlighted because it has a negative impact on the sustainability of projects in the company, company productivity and the sustainability of the company in the long term. Turnover intention refers to the desire or possibility of an employee to leave a company or the work he is doing (Balete, 2018). Turnover intention consists of two types, the first is voluntary, namely employees who want to leave a company or work that they do on their own wishes while involuntary is based on the wishes of the company or the party where the employee works (Perez, 2008).

Predicting the risks that affect turnover intention is one of the keys to addressing this problem. by using the implementation of machine learning techniques to then be able to provide insights for company leaders and human resources (HR) teams. This study was conducted to predict the risks that can affect turnover using machine learning techniques by comparing the accuracy, precision and sensitivity of several methods including K-Nearest Neighbor, Support Vector Machine and Decision Tree. Based on the percentage of accuracy, precision and sensitivity of several methods, the method with the highest average percentage will then be made as a prediction model so that it is hoped that the predictive model can minimize turnover intention and maintain the sustainability of workers in a company in the long term.

2. MATERIALS AND METHODS

Some of the theories that are referenced in this study related to the desire to move or *turnover intention are* described in this section. Penelitian which is started by collecting datasets in the form of human resource *data records* sourced from the Kaggle dataset. The dataset consists of 311 *data records* and 36 features, then the selection or selection of features into 24 features, then the selection or selection of features into 24 features is

carried out . Before being implemented into model *machine learning*, *data transformation* is carried out first by doing *date encode*, *ordinal encode* and *one-hot encode*. Machine *learning methods* compared to create prediction models include the *K-Nearest Neighbor* method, *Support Vector Machine* and *Decision Tree*. The last stage is to test the test data by measuring the level of accuracy, precision and sensitivity using *the confusion matrix*.

Materials

The explanation of some of the theories is explored as follows.

Turnover Intention

Turnover intention is the intention, will or will of the individual himself to exit by itself from the organizationi (Sudarmawan & Suhariadi, 2014). *Turnover intention* is the intention of a person to quit the company because of a reason either voluntarily (originating from within oneself) or not voluntarily (termination of employment from the company)(Sianipar & Haryanti, 2014) .

Turnover intention is the desire of employees to leave the company and try to find another job that is better than before(Waspodo, Handayani, & Paramita, 2017). *Turnover intention consists* of two types, the first is *voluntary*, namely employees who want to leave a company or work that they do on their own wishes while *involuntary* that is, based on the wishes of the company or the party where the employee works (Perez M. , 2008)

Machine Learning

Machine learning is a series of techniques that can help in handling and predicting very large data by presenting these data with learning algorithms(Danukusumo, 2017). *Machine learning can be* defined as an experiential computational method to improve performance or make accurate predictions. The definition of experience here is previous information that is available and can be used as learner data

K-Nearest Neighbor

The *K-Nearest Neighbor* or *KNN* algorithm is a method that uses a *supervised ed* algorithm where the results of the new test sample are classified based on the majority of the categories in the KNN. The propriety of the KNN algorithm is determined by the presence and absence of irrelevant data , or the weight of the feature is equivalent to its relevance to the classification(Nugroho & Wijana, 2015).

Support Vector Machine

The Support Vector Machine (*SVM*) is a set of *supervised learning* methods that analyze data and recognize patterns, used for classification and regression analysis. This classification is done by looking for *hyperplanes* or *decision boundaries* that separate one class from another. *SVM* seeks to find the best hyperplane

by maximizing the margins/distances between classes(Hadna, Santosa, & Winarno, 2016) .

Decision Tree

Decision tree or pohon decision is a very powerful and well-known method of classification and prediction. The *decision tree* method can be described as a decision tree because if visualized the structure is similar to a tree where the decision tree turns a very large fact into a decision tree that represents rules. Rules can be easily understood in natural language. In addition it can be expressed in the form of a database language such as *Structure Query Language* (SQL) to search for records in a specific category(Nasrullah, 2021).

Data Transformation

Methods in *data mining* or *machine learning* often require special data formats or structures before they can be implemented. The *process of data transformation* is the process of changing existing data from one format or structure to another format or structure that is ready to be processed. Through the transformation process, it allows *data mining* or *machine learning* thatcan beobtained more effectively and efficiently. Not only that, but the patterns found are also easier to understand (Leolianto, Thayf, & Angriani, 2020).

Confusion Matrix

Confusion matrix is a table consisting of many rows of test data that are predicted to be correct and incorrect by a classification or prediction model, this table is needed to determine the performance of a classification or prediction model (Wijayanto, 2015) .

Table 1. Table *Confusion Matrix*

		Prediction Class	
	Class	Positive	Negative
Actual Class	Positive	TP	FP
	Negative	FN	TN

TP: The predicted class is positive while the actual class is positive.
FP: The predicted class is positive while the actual class is negative.
FN: The predicted class is negative while the actual class is positive.
TN: The predicted class is negative while the actual class is negative.

Based on *the confusion matrix*, accuracy, precision and sensitivity of a classification or prediction model can be calculated. Accuracy determines how accurately the model is in classifying test data correctly, precision describes between the positive correct prediction results and the entire positive class in the actual class while sensitivity describes between the prediction results true positive with the entire positive class in the prediction class. Theequations for calculating accuracy,

precision and sensitivity using the confusion matrix table are as follows:

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Presisi = \frac{TP}{TP + FP}$$

$$Sensitivitas = \frac{TP}{TP + FN}$$

2. MATERIALS AND METHODS

Thedata set collected is sourced from Kaggle. Kaggle is an online community that gathers experts in the field of *data science*, Kaggle was built by Goldbloom in 2010 and already has more than 1000 datasets. human power. The record dataset is 311 data with 36 features, sourced from https://www.kaggle.com/datasets/rhuebner/human-resources-data-set. The features of the collected dataset include Employee_Name, EmpID, Married ID, Marital Status ID, Gender ID, EmpStatus ID, Dept ID, Perf Score ID, From Diversity Job Fair ID, Salary, Termd , PsitionID, Position, State, Zip, DOB, Sex, Marital Desc, Citizen Desc, Hispanic Latino, Race Desc, Dateof Hire, Date of Termination, Term Reason, Employment Status, Department, Manager Name, ManagerID, Recruitment Source, Performance Score, Engagement Survey, Emp Satisfaction, Special Projects Count, Last Performance Review_Date, Days Late Last 30 and Absences. Furthermore, selecting or selecting features from the dataset used. feature selection is carried out by choosing which f-features have a major effect on *turnover intention* and are relevant. If there is a feature that has no connection at all, it can be removed from the feature. The features removed in the feature selection process totaled 12 features including Employee_Name, EmpID, Marital Status ID, Sex, PositionID, DeptID, Perf ScoreID, Employment Status , Emp Status ID, Date of Termination, TermReason & Manager ID. Data transformation is carried out by encoding *date encode*, *ordinal encode* and *one-hot encode*.

3. RESULTS AND DISCUSSION

The result of creating a prediction model created using the *K-Nearest Neighbor, Support Vector Machine* and *Decision Tree methods.* To see the implementation process of each model, a simulation of the calculations of each method is carried out against the data sample.

Table 2. Sample Data

Em	Date	Perform	Recruitm	Ter
10	7/5/	Exceeds	LinkedIn	0
10	3/30	Exceeds	Indeed	1
10	7/5/	Fully	LinkedIn	1

From the data sample, processing is then carried outwith the K-Nearest Neighbor method, Support Vector

Machine and Decision Tree. After that, the trial is carried out by measuring the level of accuracy, precision and sensitivity to the dataset used so that the method used can be compared and tested to be concluded and specified in the creation of the prediction model. By utilizing *the confusion matrix* of measuring the level of accuracy, precision and sensitivity of *the K-Nearest Neighbor* method, *the Support Vector Machine* and *Decision Tree* are as follows:

Acuration, precision and sensitivity of the K-Nearest Neighbor method in predicting turnover intention with the following equations of calculation of accuracy, precision and sensitivity.

$$Akurasi_{KNN} = \frac{55+17}{55+5+17+17} = \frac{72}{94} = 0.76 = 76\%$$

The result of the calculation of the accuracy value shows a figure of 76%.

$$Presisi_{KNN} = \frac{55}{55+5} = \frac{55}{60} = 0.91 = 91\%$$

The result of the calculation of the precision value shows a figure of 91%.

$$Sensitivitas_{KNN} = \frac{55}{55+17} = \frac{55}{72} = 0.76 = 76\%$$

The result of the calculation of the sensitivity value shows a figure of 76%.

Acuration, precision and sensitivity of the *Support Vector Machine* method in predicting *turnover intention* with the following equations of accuracy, precision and sensitivity calculations.

$$Akurasi_{SVM} = \frac{59+26}{59+1+8+26} = \frac{85}{94} = 0.90 = 90\%$$

The result of the calculation of the accuracy value shows a figure of 90%.

$$Presisi_{SVM} = \frac{59}{59+1} = \frac{55}{60} = 0.98 = 98\%$$

The result of the calculation of the precision value shows a figure of 98%.

$$Sensitivitas_{SVM} = \frac{59}{59+8} = \frac{59}{67} = 0.88 = 88\%$$

The result of the calculation of the sensitivity value shows a figure of 88%.

the accuracy, precision and sensitivity of the *Decision Tree* method in predicting *turnover intention* with the following equations of accuracy, precision and sensitivity calculations.

$$Akurasi_{DT} = \frac{60+33}{60+0+1+33} = \frac{93}{94} = 0.98 = 98\%$$

The result of the calculation of the accuracy value shows a figure of 98%.

$$Presisi_{DT} = \frac{60}{60+0} = \frac{60}{60} = 1 = 100\%$$

The result of the calculation of the precision value shows a figure of 100%.

$$Sensitivitas_{DT} = \frac{60}{60+1} = \frac{60}{61} = 0.88 = 98\%$$

The result of the calculation of the sensitivity value shows a figure of 98%.

CONCLUSIONS

Basedonthe results of the trials carried out, several conclusions can be drawn, this research can analyze the features needed and can be used to make prediction models. This study successfully implemented *machine learning* to predict *turnover intention.* This study can determine and select the best prediction method among *K-Nearest Neighbor, Support Vector Machine* and *Decision Tree*. The best method is *Decision Tree* because the accuracy value is 98%, then the precision value is 100% and the sensitivity value is 98%. The accuracy, precision and sensitivity values of the *Decision Tree* method are the highest compared to the other two methods.

REFERENCES

Balete, AK. (2018). *Turnover Intention Influencing Factors of Employees: An Empirical Work Review.* Journal of Enterpreneur & Organization Management, 1-7.

Danukusumo, KP. (2017). *Deep Learning Implementation Using Convolutional Neural Networks for GPU-Based Classification of Temple Images.* Yogyakarta : Atma Jaya University.

Hadna, NMS. , Santosa, PI. and Winarno, WW. (2016). *A Literature Study of Comparative Methods for sentiment analysis processes on Twitter.* National Seminar on Information and Communication Technology, 1-8.

Leolianto, I. , Thayf, MSS. and Angriani, H. (2020). *Implementation of Naive Bayes Theory in the Classification of Prospective New Students of STMIK Kharisma Makassar.* Science and Information Technology Journal, 110-117.

Mohri, M. , Rostamizadeh, A. and Talwalkar, A. (2018). *Foundations of Mechine Learning.* Cambridge : MIT Press.

Nasrullah, AH. (2021). *Implementation of the Decision Tree Algorithm for the Classification of Best-Selling Products.* Scientific Journal of Computer Science, Faculty of Computer Science, Al Asyariah Mandar University, 45-51.

Nugroho, RS. and Wijana, K. (2015). *Auxiliary Program to Predict Sales of Goods.* Journal of EXISTENCE, 83-93.

Perez, M. (2008). *Turnover Intent Diploma Thesis.* Basılmamış Yüksek Lisans Tezi : University of Zurich.

Sianipar, ARB. and Haryanti, K. (2014). *The Relationship between*

Organizational Commitment and Job Satisfaction with Turnover Intentions in Employees in the Production Sector of CV. X. Jurnal Psychodemensian, 98-114.

Sudarmawan, SH. and Suhariadi, F. (2014). *The Effect of Employee Perceptions of Organizational Fairness on Turnover Intentions in PT. ENG Gresik.* Surabaya : Airlangga University.

Waspodo, AA. , Handayani, NC. and Paramita, W. (2017). *The Effect of Job Satisfaction and Work Stress on Turnover Intention on PT. Unitex in Bogor.* Indonesian Journal of Science Management Research, 97-115.

Wijayanto, H. (2015). *Batik Classification Using the K-Nearest Neighbour Method Based on Gray Level Co-Occurrence Matrices (GLCM).* Fik UDINUS Journal, 1-6.