Prediction Analysis of Employee Resignation Level using K-Nearest Neighbor Algorithm

Dea Andini Andriati
SKom
Gunadarma University
Depok, Indonesia

Rodiah ST, PhD
MMSI
Gunadarma University
Depok, Indonesia

ABSTRACT

PT. Micro Madani Institute is a company engaged in the training and recruitment of employees, the company is given the mandate to recruit Mekaar employees. In 2018 along with the rapid growth of the Mekaar Branch Office from 1,100 units to 2,500 units, the number of HR management will automatically increase by 27,500 people. With this increase in HR management targets, PT. Micro Madani Institute takes a strategic step to accelerate employee fulfillment.

The level of fulfillment and resignation of employees is highly considered by the company especially with the number of employee resignations every year continues to increase constantly hencethe process of achieving targets for employee fulfillment is hampered as many employees resigns. One way to deal with this is to implement data mining to assist in decision making. In this study the author implemented the K-Nearest Neighbor method to assist analysts in predicting employee resignation rates. The K-NN analysis process in this study uses 4 variables: position, distance traveled, years of service, and resignation reason. The study applied the euclidian distance to calculate the distance between training data and test data sorted according to the smallest distance value. The study used the 5 nearest neighbors (k) namely 1,3,5,7 and 11 which were used to see the cluster patterns of the neighbors. Reduction of employee resignation analysis implemented with the K-Nearest Neighbor method can generate resign reasoning patterns from the results of euclidian distance. The results expected to facilitate the company's management in the process of analyzing the rate of employee reesignation by examining previous data in order to create the right decision in minimizing employee resignations.

Keywords

Data Mining,Employee Resignation, K-Nearest.

1. INTRODUCTION

PT. Micro Madani Institute (PT.MMI) is a company that was founded with the purpose of helping PT. Permodalan Nasional Madani (PT. PNM) in their development of Human Resources management, Training and Certification. As business development accelerates within PT. PNM (Persero), especially with the establishment of the Mekaar Program in 2015, PT. Micro Madani Institute is given the opportunity to contribute by recruiting for Mekaar Employees, especially for Account Officer (AO), Finance Administration Officer (FAO) and Senior Account Officer (SAO) positions as these position is the spearhead of the company's business success

In 2017 with a total of 1,400 Mekaar Branch Offices the Mekaar's HR management managed by PT. Micro Madani Institute had as many as 13,000 people under their management. In 2018, along with the rapid growth of the Mekaar Branch Office from 1,100 units to 2,500 units, the

number of HR management targeted will automatically increase by 27,500 people. With this increase, PT. Micro Madani Institute has taken strategic steps to accelerate the fulfillment of Account Officers (AO) and Finance Administration Officers (FAO) including collaborating with BBLK, BLK, SMK, SMU, LPK, BPSDM, Disnaker and holding simultaneous tests and recruitments.

The level of fulfillment and resignation of employees is highly considered by the company especially along with the high number of employee resignations every year continues to increase so that the process of achieving targets for employee fulfillment is hampered as many employees prefer to resign. A number of employee changes can really improve the company's performance, infuse new energy and new ideas into the company. Each time an employee leaves, the company loses the hard-won knowledge and (often expensive) skills acquired.

2. LITERATURE REVIEW

Table 1 Related studies

Year	Author's name	Title	Method	Results
2018	Supria, Lidyawati, Sri Mawarni	Prediction System for Resignation of Prospective New Students Using C45 algorithm	Algoritma C4.5. Desicion tree	the combined data for 2016 & 2017 has an accuracy of 68.93%, the data for 2016 has an accuracy of 66.53% and the data for 2017 has an accuracy of 71.46%
2018	I Gst Ayu Sri Melati, Linawati, I A Dwi Giriantari	Discovery Academic Data For Prediction Of Prospective Students Resignation	Decision Tree	The test results show, knowledge analysis academic data discovery for prediction of resignation. Prospective students use classification techniques and frequent pattern can help the

				registration program for new students in reducing the risk of not obtaining enough students
2019	Mardiani Putri Agustini, Ahmad Afif Supianto, Welly Purnomo	Data Mining Application to Predict Students with Drop Out Potential using the K-Nearest Neighbor (K-NN) Algorithm	K-Nearest Neighbor	The results of the evaluation and validation of the K-NN algorithm for predicting the potential drop out of students using a confusion matrix of 99.2337%. While the evaluation of the ROC curve obtained an AUC value of 0.8918, which means the classification is categorized as ''Good classification" as it is at an AUC value of 0.80 - 0.90

2.1 Company Profile

PT. Micro Madani Institute was founded in 2015 as an affiliated company of PT. PNM (Persero) which is engaged in Training, Certification, Recruitment and Management of Human Resources (HR) and committed to helping customers manage their human resources optimally. Human Resources has an important role in the success of an organization. Therefore companies consistently pay attention to the effectiveness of their human resources. If an organization really pays attention to its human resource, then all the foundation, organizational culture, and strategic steps are for realizing the company's goals

PT. Micro Madani Institute Specializes in the development and management of human resources with a variety of products that is superior and offering them to all industrial sectors. Human resource development is the main point of a company's business, because we recognize the importance of the position of human resources as a determining factor for the success of a company. Since the company was founded, it has gained the trust of companies and prestigious institutions in their field.This shows that the development of our business is very impressive in the future. This is an invaluable trust for us in pioneering a better future. PT. Micro Madani Institute always strives to improve the quality of service and professionalism of our business which is supported by professionals and experts in their fields continue to be committed in creating great quality and satisfying services.

In 2016, PT. Micro Madani Institute is trusted to manage the Human Resources (HR), for the Fostering Family Welfare (Mekaar) program. The Mekaar Program is a program launched to improve Economic and Social Welfare, especially for the underprivileged. The target of this program is for women in rural and urban environments with the target of improving economic and social living standards.

2.2 Data Mining

Data Mining can be defined as the process of finding patterns in data. The process must be automatic or (more usually) semi-automatic. The patterns found must be meaningful as they produce several benefits, usually economical. Data is always present in large numbers. (Ian H.Witten, Eibe Frank, Mark A.Hall.2011)

Data Mining is the extraction or collection of useful information from a data sets. Information that is usually collected is hidden patterns in data, relationships between data elements, or modeling for data forecasting purposes (Sigit Adinugroho & Yuita Arum Sari, 2018)

Data Mining is a series of processes to raise the value added in the form of information that has not been known manually from a database. The information generated is processed by extracting and recognizing important patterns or pulling data contained in the database. Data Mining is mainly used to search for certain information contained in large databases thus it is often refered to as Knowledge Discovery Databases (KDD) (Retno Tri Vulandari, 2017)

2.3 Employee Recruitment

Recruitment is a series of activities that convert inputs into outputs. This process collects information of job requirements from various candidates, consideration from decision makers, and leads to employees will be hired. When going to search for candidates the thing to consider is determining the job requirements themselves. To define a job and its requirements, an HRD needs to understand the main responsibilities and tasks involved in the job, background characteristics needed to do the tasks such as education and experience, personal characteristics for example, does the individual need to have strong interpersonal skills ?, key features of organizational culture for example, team orientation, level of conformity, reward system, and managerial style for example, authoritative, coercive, democratic as well as its implications for effective working relationships.

Professional recruiters will be able to save time. Even though the price is expensive, when the company employs competent people, the company's money will be well spent. Specialized companies have an active network of important people in the industries they serve and can deliver news quickly and confidentially to qualified people. They also screen respondents so that only candidates who meet the requirements are presented for evaluation.
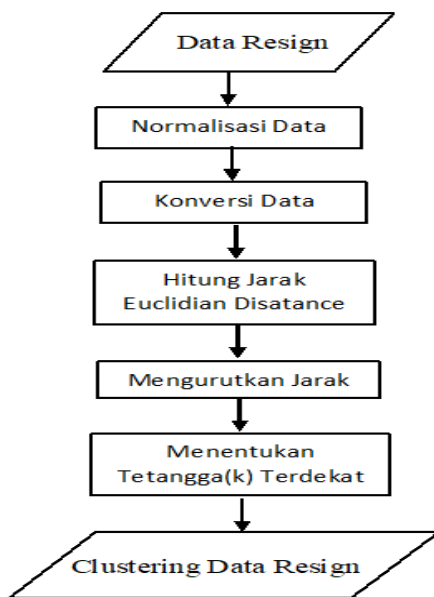
3. METHODS

Fig.1 Research Methods

Figure 1 is the research method used where there are several steps to process the resgination data. The first step is to normalize the data, then convert the data, after the data has been normalized and converted, the next process is entered into the K-Nearest Neighbor algorithm calculation by calculating the euclidian distance, then sorting the distance, determining the closest distance, then searching for class from the nearest neighbor so that clustering can be determined from the resign data.

PT. Micro Madani Institute is one of the companies in the field of employee recruitment to fulfill the demand of HR in the company. Employees are widely spread throughout Indonesia. The data used in this study is a private data set that is data obtained from within the organization, with the amount of original data of 10,447 obtained from the Human Resource Information System (HRIS) and is the result of every employee's resignation from 2016 to 2018 The original data of Employee Resignation can be seen in table 3.1 where the data has 146,272 records.

3.1 Data Normalization

Table 2 Normalized Data

Posistion	Distance (Km)	Years of Service	Resignation Cause
Account Officer	3	2	Getting Married
Account Officer	4	1	Finding Other Job
Account Officer	2	1	Finding Other Job
Account Officer	1	3	Health Issues
Account Officer	2	12	Going to College
Account Officer	3	2	Going to College
Account Officer	3	5	Internal Family Problems
Senior Account Officer	3	8	Internal Family Problems
Account Officer	3	6	Not being promoted into permanent position
Account Officer	12	4	Moving to other city
Account Officer	4	6	Unsuitable Salary
Finance Account Office	13	4	Internal Family Problems

The results obtained from the normalization process are Position, Gender, Distance, Working Period and Reason for Resign. As seen in the example in Table 2 in the green parts there are Account Officers as Position, 3 as Distance in kilometers, 2 as Work Years in Months, and Want to Study as Resign Reasons. 4 These fields will be the supporting parameters for this research.

3.2 Data Conversion

3.3.1 Position Conversion

Table 3 Rating for Positions

No	Position	Score
1	Account Officer	100
2	Finance Account Officer	80
3	Senior Account Officer	60

The position parameter is then converted into a score form as in table 3 where the score for the Account Officer position is 100, the score for the Finance Account Officer position is 80, and the score for the Senior Account Officer is 60.

3.3.2 Distance Conversion

Table 4 Convertion of Distance

Distance	Conversion
2	2
3,5	4
1,5	2
1,4	1
2,5	3
3,75	4
4,2	4
1,5	2
5,3	5
1,5	2

In this conversion example the numbers is round up to make it easier to calculate. Example shows the employee distance that is colored green is 2.5 which then rounded up to 3, while example for employee distance of 3.75 is rounded up to 4, and for 4.2 is rounded down to 4.

3.3.3   Conversion of Years of Service
Table 5 Conversion Years of Service

Years of Service		Tgl Efektif Kerja	Tgl Efektif Resign
10	10months2days	20-Feb-16	22-Dec-16
0	0months11days	13-Sep-16	24-Sep-16
1	0months16days	29-Aug-16	14-Sep-16
1	0months28days	3-Aug-16	31-Aug-16
1	0months22days	19-Aug-16	10-Sep-16
0	0months7days	26-Sep-16	3-Oct-16

As in the example of table 5 which is colored green, the years of service is scored as 0 with the calculation of 0 Months 11 Days because it is less than 15 days , which creates the given score for Working Period of 0 (months), whereas in the Working Period 1 with a calculation of 0 Months 16 Days as it is more than 15 days they are given a score of 1 (month) Work Period

3.3.4   Conversion of Resignation Reasons
Table 6 Score/Ratings For Resignation Reasons

No	Resignation Cause	Score
1	Finding other jobs	100
2	Going to College	90
3	Moving to other cities	80
4	Internal Faily Problems	70
5	Health Issues	60
6	Getting Married/Marriage	50
7	Not being promoted into permanent position	40
8	Unsuitable Salary	30

As shown in table 6 the reason for resignation has been filtered and grouped then which the converted to obtain scores. Where Looking for Other Job is worth 100, Wanting to Study is 90, Moving to Other Cityis worth 80, Internal Family Problems is valued at 70, Health issues is valued at 60, Marriage Value is valued at 50, Not getting promoted to permanent employee is valued at 40, and Unsuitable Salary Value is valued at 30.

3.3 Calculating the Euclidian Distance

$$d(x,y) = \sqrt{(\sum_{i=1}^{n}(x_{1-}y_1)2)}$$

Where :

xi = sample data

yi = test data

i = data variable

d (x, y) = dissimilarity / distance

n = data dimension

Tabel 7 Case Study

Position	Distance (km)	Years of Service (months)	Resignation Cause
80	2	9	?

1st Example of Euclidien calculation:

$= \sqrt{(100\text{-}80)^2 + (2\text{-}2)^2 + (10\text{-}9)^2}$

$= \sqrt{400 + 0 + 1}$

$= \sqrt{401}$

$= 20,024$

2nd Example of Euclidien calculation :

$= \sqrt{(100\text{-}80)^2 + (4\text{-}2)^2 + (1\text{-}9)^2}$

$= \sqrt{400 + 2 + (\text{-}8)}$

$= \sqrt{468}$

$= 21,633$

3.4 Sorting Distance
The results of the euclidian distance have not been arranged neatly from the smallest largest to the.

The distance is sorted from the Euclidian Distance calculation results from the smallest as follows:

Euclidian Distance1 = 12,083

Euclidian Distance2 = 20,025

Euclidian Distance3 = 20,050

Euclidian Distance4 = 20,125

Euclidian Distance5 = 20,224

Euclidian Distance6 = 20,248

Euclidian Distance7 = 20,322

Euclidian Distance8 = 20,421

Euclidian Distance9 = 20,905

Euclidian Distance10 = 21,213

Euclidian Distance11= 21,541

Euclidian Distance12 = 21,633

Euclidian Distance13 = 21,749

Euclidian Distance14 = 22,913

Euclidian Distance15 = 24,920

4.   RESULTS AND DISCUSSION
As in the previous discussion in chapter 3 that using the K-Nearest Neighbor method has several steps to finally find the closest k value. This chapter is the last stage of the method proposed in analyzing the rate of employee resignation using the K-Nearest Neighbor method.

4.1 Results of Data Conversion

Table 8 Results of Data Conversion

Posisti on	Distance (km)	Years of Service (months)	Resign ation Cause
100	2	10	50
100	4	1	100
100	2	1	100
100	1	3	60
100	2	12	90
100	3	2	90
100	3	5	70
60	3	8	70
100	3	6	40
80	16	4	70

As in the example of Table 8 which is colored green where the position with the contents of data 60 (Senior Account Officer), Mileage with data content of 3 (km), Working Period with a value of 8 (months), and Reason for Resign with data content of 60 (sick) .

4.2 Results of Euclidian Distance

Table 9 Results of Euclidian Distance

Posistion	Distance (km)	Years of Service (months)	Euclidian Distance
100	2	10	20,025
100	4	1	21.633
100	2	1	21,541
100	1	3	20,905
100	2	12	20,224
100	3	2	21,213
100	3	5	20,421
60	3	8	20,050
100	3	6	20,248
100	3	7	20,125
100	12	4	22,913
100	4	6	20,322
100	5	1	21,749
60	16	4	12,083
80	13	4	12,083

The result of the 1st Euclidien Distance calculation is 20,025

The result of the 2nd Euclidien Distance calculation is 21,633

The result of the 3rd Euclidien Distance calculation is 21,541

The result of the 4th Euclidien Distance calculation is 20,905

The result of the 5th Euclidien Distance calculation is 20,224

The result of the 6th Euclidien Distance calculation is 21,213

The result of the 7th Euclidien Distance calculation is 20,421

The result of the 8th Euclidien Distance calculation is 20,050

The result of the 9th Euclidien Distance calculation is 20,248

The result of the 10th Euclidien Distance calculation is 20,125

The result of the 11th Euclidien Distance calculation is 22,913

The result of the 12th Euclidien Distance calculation is 20,322

The result of the 13th Euclidien Distance calculation is 21,749

The result of the 14th Euclidien Distance calculation is 24,920

The result of the the Euclidien Distance calculation is 12,083

4.3 Results of the Nearest Neighbors

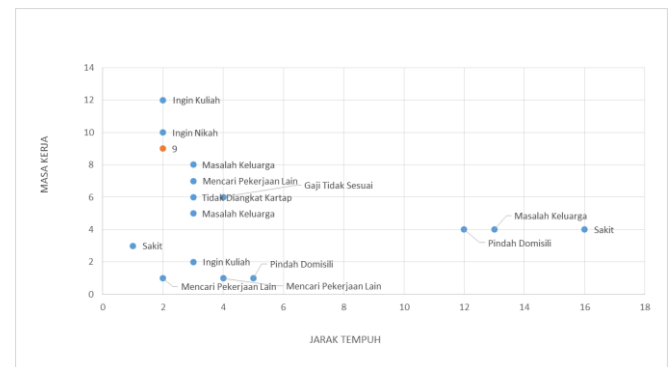Figure 2 circled in red shows the location of the coordinates of the sample test data case

Fig.2 Scatter Training Data Chart and Test Data

4.4 Results of the Resignation Data Clustering

Table 10 Results of Resignation Data Clustering

No	Resignation Cause	Clustering
1	70	70
3	70 , 50 , 70	70
5	70,50,70,100,90	70
7	70,50,70,100,90,40,30	70
9	70,50,70,100,90,40,30,70,60	70
11	70,50,70,100,90,40,30,70,60,90,100	70

Explanation of Table 10 is as follows:

For K = 1 the first value is worth 70 with a description of family problems, thus the results of clustering from K-1 are 70 because only 1 neighbor is seen.

For K = 3 because the value of 70 is more dominant than the other values thus the clustering result of K-3 is 70, which is family problem

For K = 5 because the value of 70 is more dominant than other values therefore the results of clustering from k-5 is family problem.

For K = 7 because the value of 70 is more dominant than other values therefore the results of clustering from k-7 is family problem.

For K = 9 because the value of 70 is more dominant than other values so the results of k-9 clustering is also family problem.

For K = 11 because the value of 70 is more dominant than other values therefore the results of clustering k-11 is family problem.

5. CONCLUSIONS AND SUGGESTIONS
5.1 Conclusions
1. Some conclusions drawn in this study include:

2. Employee resignation factors based on the data are determined by position, distance to workplace and years of service. The highest reason for resignation is to find another job.

3. K-Nearest Neighbor has been successfully implemented and clustered with data resign.

5.2 Suggestions
Suggestions for the development of this study, future researcher shoulf add more supportive variables, and experimenting with other algorithms, so as to increase the subjectivity of the decision results.

6. REFERENCES

[1] Biau Gerard, Luc Devroye (2010). Lecture On The Neirest Neighbor Method. London: Springer

[2] Bramer MAx (2016). Principles Of Data Mining. London: Springer.

[3] Gorunescu Florin (2011). Data Mining Concepts, Model, dan Techniques. India: Springer.

[4] Harvard Business Essentials (2003). Keep Hiring and Keeping the Best People. United States of America : Harvard Business School Publishing

[5] Hastie Trevor, Robert Tibshirani, Jerome Friedman (2008). Data Mining, Inference, and Prediction. California: Springer

[6] Hurwitz Judith, Daniel Kirsch Khaeruddin (2018). Machine Learning For Dummies. Hoboken : IBM Limited Edition

[7] Jayanti Ni Ketut Dewi Ari, Ni Kadek Sumiari (2018). Teori Basis Data. Yogyakarta : Penerbit Andi

[8] Kusrini (2007) . Strategi Perancangan Dan Pengelolaan Basis Data.. Yogyakarta : Penerbit Andi

[9] Larose Daniel T (2005) . Discovering knowledge in data : an introduction to data mining. Canada : Wiley Interscience

[10] Nofriansyah Dicky, Gunadi Widi Nurcahyo (2015). Algoritma Data Mining dan Pengujian. Yogyakarta: CV Budi Utama.

[11] Suryadharma, Triyani Budyastuti (2019). Sistem Informasi Manajemen. Ponorogo : Uwais Inspirasi Indonesia

[12] Widodo Agus Wahyu, Diva Kurnianigtyas (2017). Sistem Basis Data. Malang: UB Press.

[13] Witten Ian H, Frank Eibe (2005). Data Mining : Practical Machine Learning Tools and Techniques. San Francisco : Elsevier

[14] Xindong Wu, Vipin Kumar (2009). The Top Ten Algorithms In Data Mining. London : CRC Press