Journal of Theoretical and Applied Information Technology
31st October 2017. Vol.95. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPLEMENTATION OF K-NEAREST NEIGHBOR ALGORITHM ANALYSIS IN PREDICTING REGULAR HAJJ APPLICANT FAILURE

[1]EVA PURNAMANINGTYAS, [2]EMA UTAMI

[1,2] Magister of Informatics Engineering of Universitas AMIKOM Yogyakarta,
Ring Road Utara ST, Condong Catur, Depok Sleman Yogyakarta Indonesia
Email : [1]tyasavka1@gmail.com , [2] emma@nrar.net

ABSTRACT

Growth of number hajj applicant is always increasing so that it produces very large dataset. Hajj office department on Local Religion Ministry Office Bantul is the only official government's instance that serving hajj in Bantul Regency trying to always provide the best service for people by providing accurate information and data as quickly as possible. Business intelligence become one of chosen solution to solve various need on providing and processing information and data accurately, effectively, and up to date so that can produce good report. One of key information is to predict hajj attendance that will fail so that it can be useful for others hajj attendance which are on the waiting list to prepare themself, this prediction is using k-nearest neighbor (KNN) algorithm. Hajj and Umrah Department has various services accommodating hajj applicants end-to-end including registrations and cancellations of each applicants. Currently the applicants that cancel their applications is unpredictable and only have very short notice period so that the applicants that is in waiting list only have very limited time to prepare themself. Considering this, system that can predict the number of cancelled applicants is very important and highly needed to solve the problem. On this prediction, preprocessing are used by giving weightage for each factors and then calculation have been done by using KNN algorithm so that the output is the nearest distance of the new case from the old case in the dataset, and after that that new case can predict the result of the new case as the Nearest Neighbor's result. Measurement of this algorithm is using cross validation. The highest accuracy in this thesis is 96.2% by using value of k optimum=9. All cases will be classified based on the data set (data training) in this KNN algorithm even though the distance will not be 100% for each all cases. Setting up the weightage variable and weightage of distance on each factors being done manually, this make the values of those weightage vary based on user's assumptions. However those weightage is highly affecting the end value of distance on each case in the result. The determination of variable weights and the comparative weight of manually assigned variable ratios allows different values according to the user's perception, whereas the value of the variable weights and the weight of the variable values greatly affects the proximity between one case and the other.

Keywords: *Predict Hajj Attendance, Report, K-Nearest Neighbor, Cross Validation*

1. INTRODUCTION

The number of hajj applicant is growing very rapidly over time so that it makes very long queue in the waiting list of departure. The waiting list already reached 55,666 applicants, according to the official website of Ministry of Religious Affairs Indonesia [12]. Hajj office department on Local Religion Ministry Office Bantul is the only official government's instance that serving hajj in Bantul Regency trying to always provide the best service for people by providing accurate information and data as quickly as possible. Those data is needed for the analysis and to create reporting in the organizations. Moreover the accountable data is very much needed in the process of hajj services which involving many other government instances such as Ministry of Health, Ministry of Transportation, etc. Problems that being faced today is that the existing system could not provide real time data so that the operator need to manually calculate and create the report using Ms Excel. The purpose of this analysis are:

a. Understanding how Business Intelligence can provide the best solutions for analysis and reporting for Local Ministry of Religious

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Affair in Bantul Regency.

b. Understanding how Business Intelligence help to predict the number of failed hajj applicants for the upcoming years which the same information can be used for helping to prepare applicants on the waiting list.

Data training that being used for the prediction of number failed applicant on upcoming year is the data of hajj applicant from 2013 to 2016 which include :

a. Manual data, such as log of registration file.
b. Soft data, exported from the database of information system and from other applications like Ms Excel.

Business Intelligence become one of chosen solution to solve various need on providing and processing information and data accurately, effectively, and up to date so that can produce good report. Business Intelligence as one of the result of development in technology, in the beginning being used for analyze decision making process [8]. Business Intelligence is mainly to increase the preciseness and quality of the information and for the manager to understand market condition better [9].

Data Warehouse is collection of data that subject oriented, integrated, time-variant, and has the fix value in the process of supporting decision making management. This process is subject-oriented, integrated, time-variant and permanent [10]. K-Nearest Neighbor being done by find the nearest (most similar) group of k object in the data training with the object in the new dataset or testing data [11]. To validate a sample point in the train set, the H nearest neighbors of the point is considered. Among the H nearest neighbors of a train sample x, validity(x) counts the number of points with the same label to the label of x. Eq. 1 is the formula which is proposed to compute the validity of every points in train set [6].

$$\text{Validity}(x) = \frac{1}{H}\sum_{i=1}^{H} S(lbl(x), lbl(Ni(x))) \quad (1)$$

Where H is the number of considered neighbors and lbl (x) returns the true class label of the sample x. also, Ni(x) stands for its nearest neighbor of the point x. The function S takes into account the similarity between the point x and its nearest neighbor. Eq. 2 defines this function.

$$S(a,b) = \begin{cases} a = b \\ a \neq b \end{cases} \quad (2)$$

Predicting accuracy is needed since data training in each fold is not exactly same or represent the full data training. 5 or 10-fold cross validation are mainly being recommended and agreed. Calculating accuracy value can be done with the equivalent [1] :

$$\text{Accuracy} = \frac{\text{The number of correct classifica}}{\text{The number of test data}} \text{x100\%} \quad (3)$$

The number of regular pilgrims who canceled or resigned during this time is known at the time of repayment, so for pilgrims who occupy the waiting list afterwards felt very narrow time in preparation. Under these circumstances it requires a prediction system for the number of regular pilgrims who resign so that it can be used for early preparation for prospective pilgrims who are waiting list in the year after.

2 RELATED WORK

The researches in [7] used star schema to implement functional data and optimize data. Dimension table also being used depend on the number of languages needed. Each different language stored in different database schema. This method has limitation which will make the ETL become very complex when the number of database schema is growing.

Researches being done by [2] is using K-Nearest Neighbor algorithm in the process of face recognition on ARM processor. Proposed algorithm tested on three datasets which is Olivetti Research Laboratory (ORL), Yale face and the proposed algorithm being implemented on ARM11 700MHz. 10-fold cross-validation showing that K-NN face recognition detect 91.5% face with k=1. Overall the research showing that algorithm proposed able to recognize face on 2.66 seconds ARM processor.

The researchers in [3] using K-Fold Cross Validation to find best predictor for extracting information process. However not being explained in detail how to find the best predictor.

The researchers in [4] explain that K-nearest neighbor (KNN) was applied to classify the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

monogenean specimens based on the extracted features. 50% of the dataset was used for training and the other 50% was used as testing for system evaluation. Our approach demonstrated overall classification accuracy of 90%. In this study Leave One Out (LOO) cross validation is used for validation of our system and the accuracy is 91.25%.

Researches being done by [5] built A Texture of Feature Extraction and morphology model that can be used to identify the type of malaria parasites along with the stadium on the image of a thin blood smear. Samples are a number of preparations that have been given indications of malaria and have given a Giemsa staining. Image acquisition process is done by using a digital microscope with 1000 times of magnification. The process of segmentation used thresholding method of Otsu. Selection feature used sequential forward selection (SFS). Classification technique used artificial neural network of learning vector quantization (LVQ) with K-fold cross validation to identify patterns of types of parasites and their life stages in order to get different types and stages of malaria disease.

The research will be conducted using K-Nearest Neighbor to predict the hajj regular attendant who resign while for accuracy testing by finding the optimum k value by the 10-Fold Cross Validation method.

From the previous research, the author tries to find k-optimal with K-Nearest Neighbor method and by testing K-Fold Cross Validation to assist PHU Section of the Office of the Ministry of Religious Affairs of Bantul Regency in predicting the number of regular pilgrims useful for early preparation for pilgrims regulars that occupy the waiting list in the year after.

3 RESEARCH METHODS

There are two parts on this analysis and research. They are external environment business analysis and internal environment business analysis. These analysis based on interviews and observation being done by author. There are two categories of questions need to be addressed, those are strategic and operations. Strategic category produces external business analysis of company which showcased on five forces model, while internal business analysis with the output as Strength, Weakness, Opportunities, Threats on the SWOT analysis, and the main activity as well as supporting activity will be on Value chain analysis. For the analysis of value chain depicted by Fig. 1.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

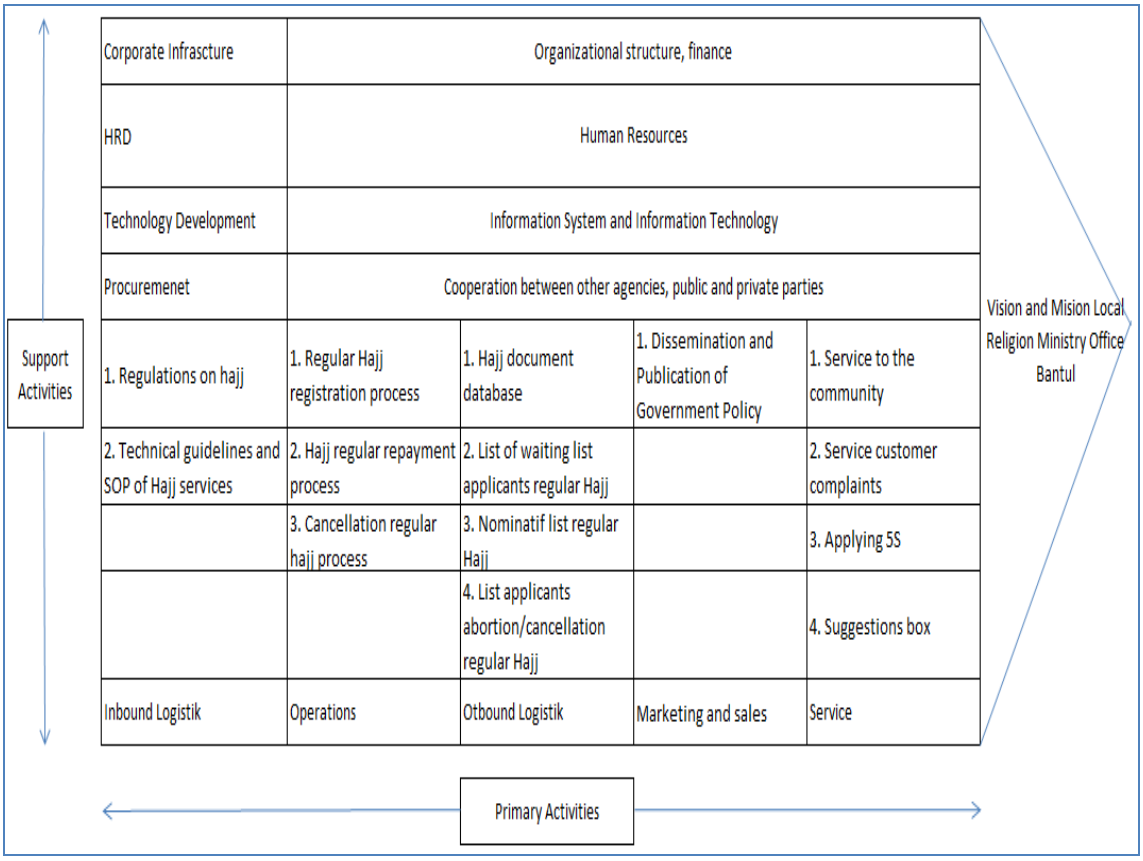

Fig. 1. Value Chain Analysis

For the analysis of five forces depicted by Fig. 2 below :

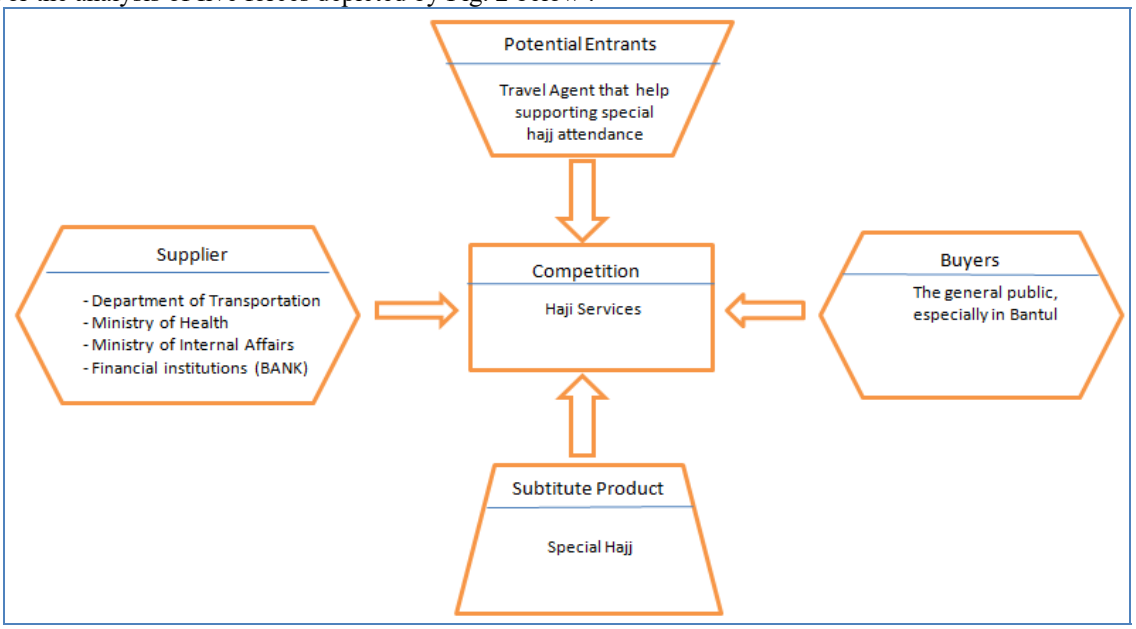

Fig. 2. Five Forces Analysis

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Flow in this research depicted by Fig. 3 below :

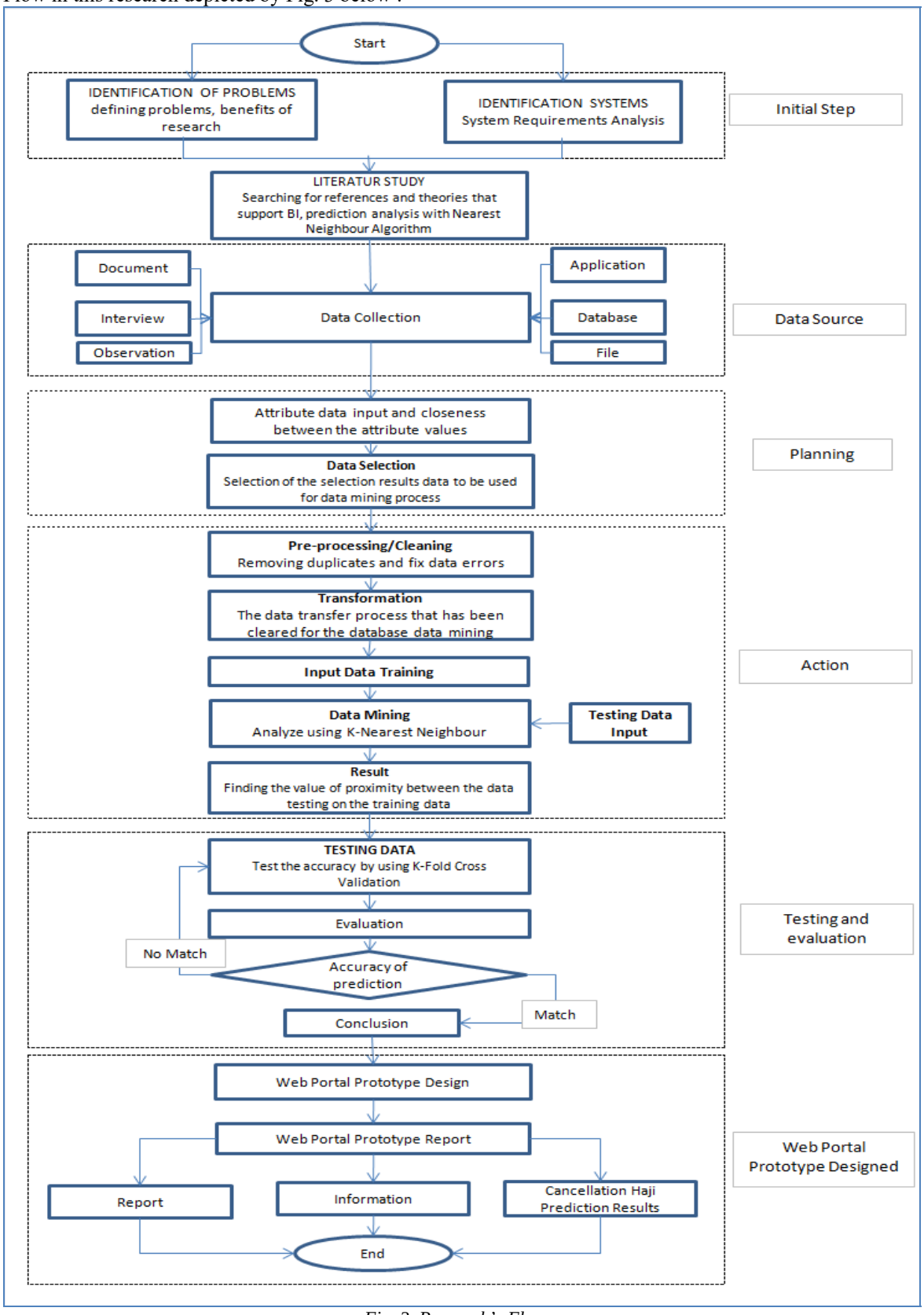

Fig. 3. Research's Flow

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Pre-processing on the following ETL will explain about the data source used in research, the data staging, extract process, transformation, loading, data mining process is used for prediction needs are :

a. Source data
   Hajj applicant data in Local Ministry of Religious Affairs Bantul for the years of 2013-3016 consist of 8 files with excel format which each containing 24 attributes : portion code, KBIH (hajj private agency), group, team, name, father's name, address, village's name, sub-district, notes, district, gender, active type, payment type, occupation, education, hajj status, registration date, payment date, user code, bank name, bank accounts number, amount submitted, and age. Other source of data come from Local Ministry of Health Bantul that consist of 4 excel file which each containing 20 attributes : departure date, check up date, input, portion code, name, height, weight, age, gender, education, occupation, address, district, province, check up place, check up district, check up province, high risk, conclusion, doctor's name.
b. Merger and data cleansing
   All the data of hajj attendant both from PHU Ministry of Religious Affairs Bantul and Ministry of Health Bantul from years of 2013-2016 combined using Microsoft Excel. Only needed attributes still remain on the combined data.
c. Choosing the data
   Final data generated consist of 2437 records and 13 attributes namely the portion of code, sub-districts, banks, KBIH, risk istithaah, assembly name, gender, age, occupation, type of education, years of leaving, and waiting list time.
d. Transformation data
   Date format need to be transformed for easier calculation.

4 RESEARCH RESULT AND DISCUSSION

In data mining process, this research used 2437 total records. Based on interviews that have been conducted, the authors used data consists of five attributes, of which 4 attributes (risk, istithaah, age, occupation) and a predictor (conclusion) shown in table 1 below:

Table 1. Value Of Attribute

Attribute	Value of attribute
Risk	White
	Green
	Yellow
	Red
Istithaah	Meet the Requirement
	Need guidance
Age	< 21
	21 – 40
	41 – 60
	61 – 70
	> 70
Occupation	Farmer
	Retirement
	House wife
	Entrepreneur
	Private sector
	Student
	Army/Police/TNI
	Civil servant/PNS
Conclusion	Depart
	Fail

To measure the distance between attributes, will be given weight in the attribute. This distance weighting given value between 0 to 1. Value 0 means that if the attribute has no effect and instead value 1 if the attribute is very influential as shown in Table 2 below:

Table 2. *Weights Of Attribute*

Attribute	Weight of attribute
Risk	0.8
Conclusion	0.6
Age	0.5
Occupation	0.4

Attribute value is determined by the proximity effect to the failure applicants. Here the proximity between the value of each attribute:

a. Risk attribute
   Values and weights of risk attribute shown in Table 3 below :

Table 3. *Values And Weights Of Risk Attribute*

Risk	Value	Weight
Red	100	0.8
Yellow	80	
Green	60	
White	40	

The distance between the values of risk attribute can be seen from Table 4 below:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4. *Distance For Risk Attribute*

Risk	Red	Yellow	Green	White
Red	1.00	0.80	0.60	0.40
Yellow	0.80	1.00	0.75	0.50
Green	0.60	0.75	1.00	0.67
White	0.40	0.50	0.67	1.00

b. Istithaah attribute
Values and weights of istithaah attribute are shown in Table 5 below :

Table 5. Values And Weights Of Istithaah Attribute

Istithaah	Value	Weight
Need guidance	100	0,8
Meet the requirement	50	

The distance between the values of istithaah attribute can be seen from Table 6 below :

Table 6. Distance For Istithaah Attribute

Istithaah	Need guidance	Meet the requirement
Need guidance	1	0,5
Meet the requirement	0,5	1

c. Age attribute
Values and weights of age attribute are shown in Table 7 below:

Table 7. Values And Weights Of Age Attribute

Code	Age (year)	Value	Weight
A1	>70	100	0,5
A2	61 – 70	90	
A3	41-60	70	
A4	21 – 40	50	
A5	<21	30	

The distance between the values of age attribute can be seen from Table 8 below :

Table 8. Distance For Age Attribute

Age	A1	A2	A3	A4	A5
A1	1.00	0.90	0.70	0.50	0.30
A2	0.90	1.00	0.78	0.56	0.33
A3	0.70	0.78	1.00	0.71	0.43
A4	0.50	0.56	0.71	1.00	0.60
A5	0.30	0.33	0.43	0.60	1.00

d. Occupation attribute
Values and weights of occupation attribute shown in Table 9 below:

Table 9. Values And Weights Of Occupation Attribute

Code	Occupation	Value	Weight
W1	Farmer	100	0.4
W2	Retirement	90	
W3	House wife	80	
W4	Entrepreneur	70	
W5	Private sector	60	
W6	Student	50	
W7	Army/Police/TNI	40	
W8	Civil servant/PNS	30	

The distance between the age attribute value can be seen from Table 10 below:

Table 10. Distance For Occupation Attribute

Work	W1	W2	W3	W4	W5	W6	W7	W8
W1	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30
W2	0.90	1.00	0.89	0.78	0.67	0.56	0.44	0.33
W3	0.80	0.89	1.00	0.88	0.75	0.63	0.50	0.38
W4	0.70	0.78	0.88	1.00	0.86	0.71	0.57	0.43
W5	0.60	0.67	0.75	0.86	1.00	0.83	0.67	0.50
W6	0.50	0.56	0.63	0.71	0.83	1.00	0.80	0.60
W7	0.40	0.44	0.50	0.57	0.67	0.80	1.00	0.75
W8	0.30	0.33	0.38	0.43	0.50	0.60	0.75	1.00

5 IMPLEMENTATION

The implementation phase is the continuation of the processing data. This also phase to put the system to be ready be operated and can be seen as an attempt to concretize the system that already designed and can be described more details as follows :

a. Reporting
Information obtained by using the data of applicant departing for 4 (four) years: 2013, 2014, 2015 and 2016 are as follow:
   i. KBIH (hajj private agency)
   KBIH as the servicing agency for hajj being used to measure number of hajj attendance which joined to independence program or joined any particular KBIH in one period of time affecting on the group formation. These factors depicted in Fig. 4 below :

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

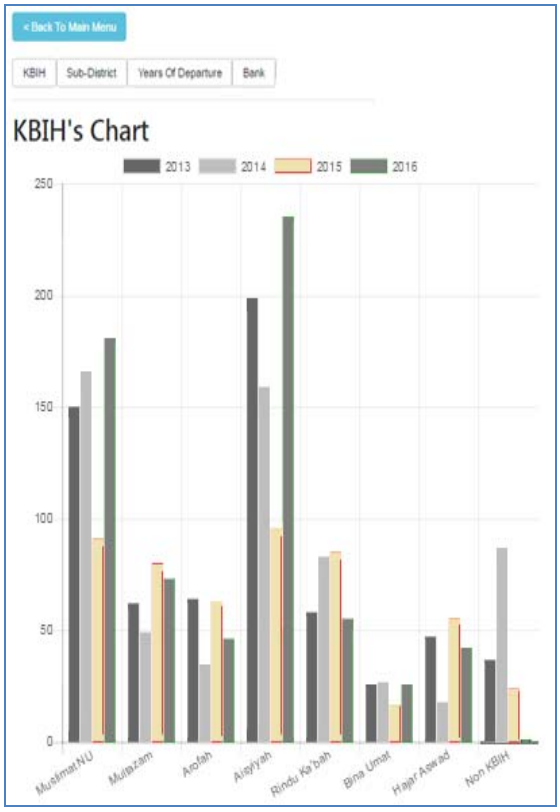

Fig. 4. KBIH Chart

From the chart above, we can know the largest number of applicants in 2016 followed the Aisyiyah KBIH.

ii. Sub-District

Region can be used to measure number of hajj attendance based on sub-district on some period of time which also affecting the group formation. This depicted in Fig. 5 below :

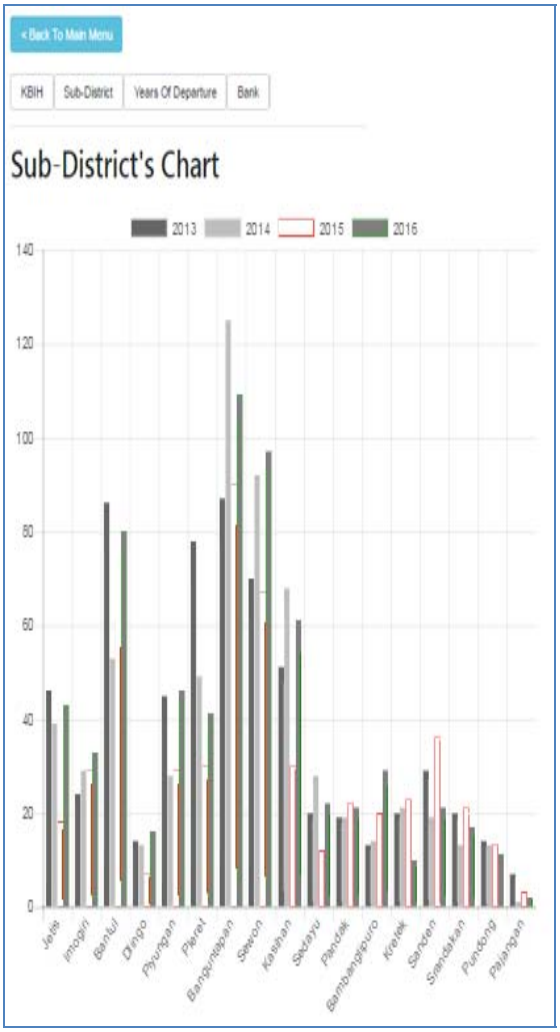

Fig. 5. Sub-District Chart

From the chart above, we can know the number of applicants in 2016 mostly came from the Banguntapan Sub-District.

iii. Years of Departure

Years of departure being used as measurement number of hajj attendance on some period of time, this shown in Fig. 6 below :

Journal of Theoretical and Applied Information Technology
31st October 2017. Vol.95. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

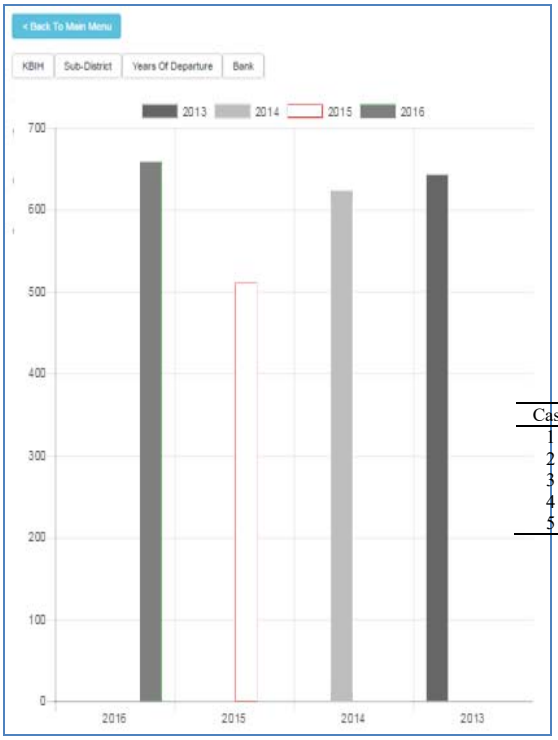

Fig. 6. Years Of Departure Chart

From the chart above, we can know the highest number of applicants was in 2016.

iv. Bank

Bank accounts being used as measurement number of hajj attendance in some period of time as depicted in Fig. 7 below :

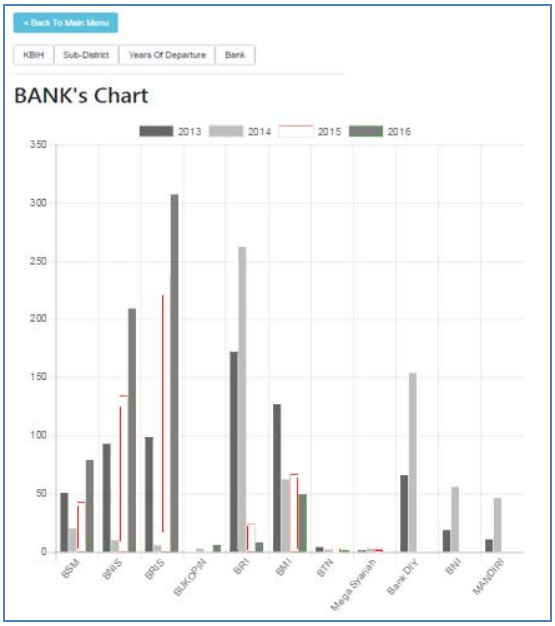

Fig. 7. Bank Chart

From the chart above, we can know the number of applicants who save in BRIS are the highest in 2016.

b. Data Mining

Applications that can predict hajj applicants failure for Bantul district, using test data history and training to produce prediction and accuracy for the classification by using K-Nearest Neighbor algorithm. Table 11 illustrate some of applicants data.

Table 11. Sample Existing Applicants Data

Case	Risk	Istithaah	Age	Type of Work	Conclusion
1	White	Meet the requirement	41-60	Civil servant/PNS	Depart
2	Yellow	Need guidance	61-70	Retirement	Depart
3	Green	Meet the requirement	61-70	Entrepreneur	Depart
4	Red	Need guidance	>70	House wife	Fail
5	Red	Meet the requirement	>70	Retirement	Fail

There is new case with category is shown in Table 12 below :

Table 12. Sample Of New Case

Case	Risk	Istithaah	Age	Type of work	Conclusion
1	White	Need guidance	61-70	Private sector	?

Distance value and weight for each attributes relatives to the new case is shown in Table 13 below:

Table 13. Distance Value And Weight For Each Attribute Relative To New Case

No	Criteria	C1	C2	C3	C4	C5
a	Risk distance value	1.0	0.5	0.67	0.4	0.4
b	Risk weight	0.8	0.8	0.8	0.8	0.8
c	Istithaah distance value	0.5	1	0.5	1	0.5
d	Istithaah weight	0.6	0.6	0.6	0.6	0.6
e	Age distance value	0.78	1.0	1.0	0.9	0.9
f	Age weight	0.5	0.5	0.5	0.5	0.5
g	Occupation distance value	0.5	0.67	0.86	0.75	0.67
h	Occupation weight	0.4	0.4	0.4	0.4	0.4
	Sum (Distance*Weight)	1.69	1.768	1.68	1.67	1.338
	Sum of Weight	2.3	2.3	2.3	2.3	2.3
	Distance	0.735	0.769	0.730	0.726	0.582

Calculating distance proximity:

Total Distance = $\frac{(a*b)+(c*d)+(e*f)+(g*h)}{b+d+f+h}$

From Table 13 can see that nearest distance for the new case is 0.8248 with case 4 which have conclusion "fail" so for this new case the prediction will be "fail". Number of applicants being used in this research is 2437 so if being done manually it will take very long time, hence the calculation being done using applications below :

a. Main menu

Journal of Theoretical and Applied Information Technology
31st October 2017. Vol.95. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Main menu is the front page for the app after user input the username and password, screenshot is shown in Fig. 8 :

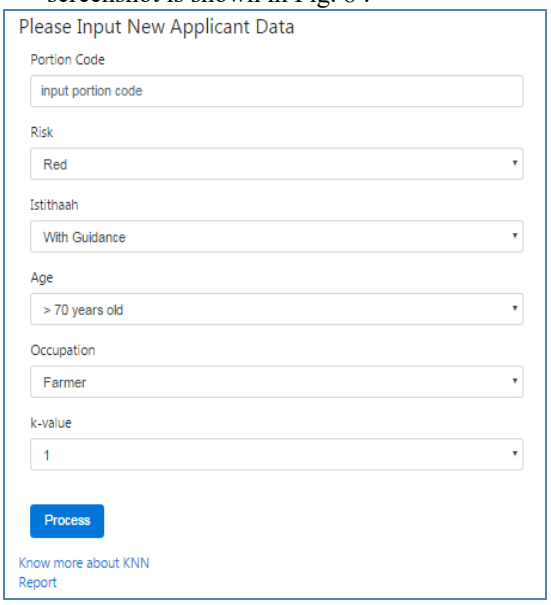

Fig. 8. Main Menu

b. Input Form for new applicant
This is the input form for user to input the data of new applicant. System will process the data and show the prediction after user click process button. After user click process button, it will shows the prediction in Fig. 9 below :

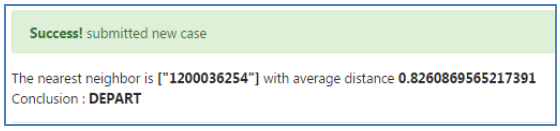

Fig. 9. Prediction Result For New Applicant

6 RESULT ANALYSIS

Both manual calculation and by using applications on the same new case produce the same conclusion which is "depart" however the nearest distance is different, the reason is because number of data training is different. Manual calculation only used 5 datasets while on applications used 2437 datasets. K-Fold Cross Validation is being used to validate the accuracy by calculated the distance value from k-nearest neighbor for each value of k (1, 3, 5, 7, 9, 11, 13, 15, 17, 19). From total 2437 records, 1497 being used as data training and 940 as data testing. The result of the validation is shown in Fig. 10 below :

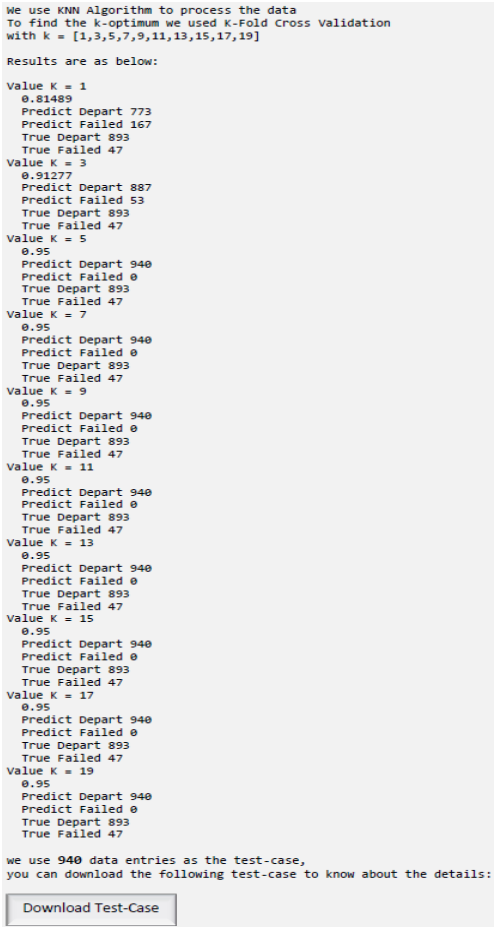

Fig. 10. Validation using K-Fold Cross Validation

To calculate the accuracy used 10-Fold Cross Validation that is by dividing data 2437 records into 10 sets consisting of training data with data testing. This is because the predicted results in this study are only for predictions of regular hajj resignations only in the current year, which are described in Table 14, 15, and 16 below:

Table 14. Distribution Of Data Test With10-Fold

10-fold usage	1	2	3	4	5	6	7	8	9	10
Count of data test	244	244	243	243	243	244	244	244	244	244

Table 15. Test Results With 10-Fold With K = 1, 3, 5, 7, 9

No	Data		Value of k				
	Training	Testing	1	3	5	7	9
1	2193	244	0.94672	0.96311	0.95902	0.96311	0.96311
2	2193	244	0.93033	0.96721	0.96721	0.96721	0.96721
3	2194	243	0.91358	0.95473	0.95885	0.95885	0.96296
4	2194	243	0.93827	0.95885	0.95885	0.95885	0.95885
5	2194	243	0.93827	0.9465	0.94239	0.94239	0.94239
6	2193	244	0.94672	0.95082	0.96721	0.96721	0.96721
7	2193	244	0.94262	0.96721	0.97131	0.97131	0.97131
8	2193	244	0.93852	0.95902	0.96721	0.97131	0.97131
9	2193	244	0.94262	0.94672	0.94672	0.94672	0.94672
10	2193	244	0.88934	0.96721	0.97131	0.97131	0.97131
	Average value		0.932699	0.958138	0.961008	0.961827	0.962238
	Maximum value		0.962238				

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Tabel 16. Test Results With 10-Fold With K = 11, 13, 15, 17, 19

No	Data		Value of k				
	Training	Testing	11	13	15	17	19
1	2193	244	0.96311	0.96311	0.96311	0.96311	0.96311
2	2193	244	0.96721	0.96721	0.96721	0.96721	0.96721
3	2194	243	0.96296	0.96296	0.96296	0.96296	0.96296
4	2194	243	0.95885	0.95885	0.95885	0.95885	0.95885
5	2194	243	0.94239	0.94239	0.94239	0.94239	0.94239
6	2193	244	0.96721	0.96721	0.96721	0.96721	0.96721
7	2193	244	0.97131	0.97131	0.97131	0.97131	0.97131
8	2193	244	0.97131	0.97131	0.97131	0.97131	0.97131
9	2193	244	0.94672	0.94672	0.94672	0.94672	0.94672
10	2193	244	0.97131	0.97131	0.97131	0.97131	0.97131
Average value			0.962238	0.962238	0.962238	0.962238	0.962238
Maximum value			0.962238				

From table 15 and 16 it is seen that the maximum value of the average value of each k is 0.962238 obtained from k = 9, 11, 13, 15, 17, and 19. Taken with a minimum k value of k = 9 so that k-optimally obtained for k = 9.

7 CONCLUSIONS

Based on the results an discussion, then get the following conclusion :

a. The resulting system can also help predict the number of regular pilgrims who resign for the coming year by entering the portion code of regular pilgrims that can be used for early preparation for prospective pilgrims who are waiting list in the following year as evidenced by a letter of acceptance The result of recommendation of Designing System of Regular Hajj Resignation from Section of Hajj and Umrah Organizer of Religious Affairs Office of Bantul Regency (attached).
b. Data mining process using K-Nearest Neighbor algorithm, while the test using 10-Fold Cross Validation get the greatest accuracy of 96.2% with k = 9 as k-optimal with 4 variables that is risti, istithaah, age, and type work.
c. Determination of variable weights and the weight of the comparisons of manually entered variable values allows there are values that vary according to the perception of the user, whereas the value of the variable weight and the weight of the variable value greatly affects the proximity between one case to another.
d. It is recommended that the system of regular pilgrimage resignation generated does not enter one-by-one portion code data of regular pilgrims because it will take a long time.

8 FUTURE WORK

Here are some suggestions that can be given to the development of this research include:

a. Periodic evaluation will be suggested on the system as any adjustment may needed in the future.
b. Setting up the weight variable and value of distance on each factors being done manually, this make the values of those weighting vary based on user's assumptions. However those weightage is highly affecting the end value of distance on each case in the result.
c. It is recommended that the system of regular pilgrimage resumption generated does not enter one-by-one portion code data of regular pilgrims because it will take a long time.
d. Research using K-Nearest Neighbor to predict the resignation of hajj pilgrims in the next 30 years should use greater data testing than training data.
e.

REFERENCES

[1] Rodiyansyah, Sandi Fajar, Winarko Edi. "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification". IJCCS, Vol.6, No.1 (2012) pp. 91~100.
[2] Setiawan Eko, Muttaqin Adharul. "Implementation of K-Nearest Neighbors Face Recognition on Low-power Processor". TELKOMNIKA, Vol.13, No.3.ISSN: 1693-6930 (2015), pp. 949~954.
[3] Widyantoro Dwi Hendratmo, Kuspriyanto Kurnia Muludi,. "Winner-Takes-All based Multi-Strategy Learning for Information Extraction". TELKOMNIKA Indonesian Journal of Electrical Engineering Vol. 12, No. 11 (2014), pp. 7935 ~ 794.
[4] Elham Yousef Kalafi, Wooi Boon Tan, Christopher Town and Sarinder Kaur Dhillon, Automated identification of Monogeneans using digital image processing and K Nearest Neighbour approaches. The Author(s) BMC Bioinformatics 2016, 17(Suppl 19):511 DOI 10.1186/s12859-016-1376-z (2016), pp 260-295.
[5] Rika Rosnelly, Sri Hartati, Edi Winarko, Sri Mulatsih, Identification Of Malaria Disease And Its Stadium Based On Digital Image Processing. Journal of Theoretical and Applied Information Technology Vol.95. No.3 (2017), pp 700-710.
[6] Yang Yiming and Liu Xin, "Re-Examination of Text Categorization Methods," Proc. SIGIR "99, August 1999, pp. 42-49.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] Dedić, Nedim., Stanier, Clare. "An Evaluation of the Challenges of Multilingualism in Data Warehouse Development". Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS) - Volume 1, ISBN: 978-989-758-187-8, 2016, pages 196-206

[8] Sahay, Sanja Kumar., "Data Analysis of Gravitational Waves", Royal Natioanl Library of Sweden, Stockholm. August 2008.

[9] Khan, Rafi Ahmad. & Quadri, S. M. K., "Business Intelligence: An Integrated Approach. Business Intelligence" Journal, 5(1), pp. 64-70, January 2012.

[10] Inmon, William., "Building the Data Warehouse", edisi 3, Canada: John Wiley & Sons, Inc,. 2002

[11] Chang Fen Chang, Wu Yi Ling., Hou, Sheng Shu. "Preparation and Characterization of Superparamagnetic Nanocomposites of Aluminosilicate/Silica/Magnetite", Coll. Surf. A336: 159,166 (2009).

[12] (http://haji.kemenag.go.id/v2/basisdata/waiting-list, accessed date October 24th (2016).