Optimization of K Value at the K-NN algorithm in clustering using the expectation maximization algorithm

Zulkarnain Lubis\textsuperscript{1,*}, Poltak Sihombing\textsuperscript{2}, Herman Mawengkang\textsuperscript{2}

\textsuperscript{1}Graduate School of Computer Science
\textsuperscript{2}Department of Information Technology, Faculty of Computer Science and Information Technology, University of North Sumatra, Medan, Indonesia

zulkarnainlb@gmail.com

Abstract. Data is the most important thing in a study. The quality of the results of the research will be directly proportional to the quality of the data that will be used in the research is concerned. One of the problems that exist in the data set is the absence of a value in the data for a particular attribute or better known as the missing data. One method that is often used by researchers is the k-nearest Neighbor (KNN). However, this method has several drawbacks, one of which is the selection of appropriate values of k not to degrade the performance of the classification. In the process of calculating the parameters k KNN there that can affect the accuracy of the classification results. To use more than one parameter k then used by majority voting to determine the classification results. If the parameter k in KNN classification used 1 then the result was very tight because it will use the nearest neighbor to the results of the classification. Conversely, if the value of the parameter k used KNN is great then the classification results will blur. This research will optimize the parameters k in the UN tax cluster using the algorithm expectation Maximization (EM). The results of the research in the form of clustering information by using the number of clusters k value optimization and the number of clusters without using the optimization of the value k. Then analysis the results after getting data already clustered. Results from the study showed that k obtained from the optimization algorithm can improve the results of the cluster where the 66% error can be reduced to 64%, yet very close to the best result of the measurement accuracy is tested.

1. Introduction

One way to classify the data is by using Clustering. Data grouping or clustering is a method used to classify into groups or clusters based on similarity, so that related data is placed in the same cluster. There are several clustering algorithms known ie partitional (Expectation-maximization, K-Means) and hierarchical (Centroid Linkage, Single Linkage), overlapping (Fuzzy C-Means) and hybrid. The algorithm can overcome the arbitrary grouping are partitional algorithm. Where in partitional algorithm, a document can be a member of a group or cluster to a process but the subsequent processes such documents can be moved to another cluster. One partitional algorithm that can group documents that have not been labeled is Expectation-Maximization, the algorithm used to find the value of Maximum Likelihood estimation of parameters in a probabilistic model. The characteristics of this algorithm is able to classify the data that has not been labeled or unlabeled data and also the results of the classification will always convergence. This algorithm has two phases: phase and phase Expectation Maximization.
In Expectation step (E-step) use the EM algorithm - Cluster for classifying data based on the model parameters. While on the Maximization step (M-step) will be done peng updates of the model parameters by using Multiple Linear Regression. Phase E-step and M-step is continued until the probability of each cluster achieve convergence. Before performing the necessary process of grouping data pre-processing, namely cleansing, tokenizing, parsing. Labeling of a cluster is done by finding the most actual label appears on a Cluster, and then adopt the label as the label Cluster. With the implementation of the EM algorithm - Cluster in the process of budget clusterisasi it can classify and determine the appropriate number of clusters,

2. Stages of Data Mining

Data mining is actually a part of the process of Knowledge Discovery in Databases (KDD), not as a technology intact and independent. Data mining is an important part of steps in the process of KDD primarily concerned with the extraction and calculation of the data patterns are analyzed, as shown by Figure 1 below:

![Figure 1. Stages in the process of knowledge discovery](image)

a. Data cleaning
To eliminate the data noise (irrelevant data / dealing directly with the ultimate goal of data mining process, eg data mining that aims to analyze the results of the sale, then the data in the collection as "employee name", "age", and so on can -ignore) and inconsistent.

b. Data integration
To combine multiple data sources.

c. Data selection
To retrieve the appropriate data for analysis.

d. Data transformation
To transform data into a form more suitable for mining. Data mining is the most important process in which a particular method is applied to generate the data pattern.
e. **Pattern evaluation**
   To identify whether interesting patterns obtained is sufficient to represent knowledge based on specific calculations.

f. **Knowledge presentation**
   To present the knowledge that has been obtained from the user.

### 2.1 Method of KNN (K-Nearest Neighbor)

The working principle of the K-Nearest Neighbor (KNN) is seeking the shortest distance between the data to be evaluated by K neighbors (neighbor) closest to the training data. This technique is included in the nonparametric classification groups. Here we do not pay attention to the distribution of the data to be grouped. This technique is very simple and easy to implement. Similar to clustering techniques, we classify a new data based on the distance the new data into multiple data / neighbor (neighbor) nearby.

KNN algorithm purpose is to classify the new objects based on attributes and sample training. Classifier not use any model to be matched and only based on memory. Given query point, will find a number of objects or k (training points) closest to the query point. Classification using the voting majority among the classification of k objects. KNN classification algorithm uses adjacency as the predicted value of the new query instance. Algorithm KNN method is simple, operates on the shortest distance from the query instance to the training sample to determine its KNN.

K best value for this algorithm depends on the data. In general, a high k value will reduce the effect of noise on klasifikasi, but draw the line between each classification is becoming increasingly blurred. Nice k value can be selected by optimization of parameters, for example by using cross-validation. The special case where the classification is based on the training data diprediksi closest (in other words, k = 1) is called Nearest Neighbor algorithm.

**Excess KNN (K-Nearest Neighbor):**
1. Resilient to training data that has a lot of noise.
2. Effective if training data is huge.

**Weakness of KNN (K-Nearest Neighbor):**
1. KNN need to determine the value of the parameter k (the number of nearest neighbors).
2. Training based on distance is not clear on what kind of distance that must be used.
3. Which attributes should be used to get the best results.
4. The computational cost is high because the necessary calculation of the distance of each query instance in the whole training sample.

### 2.2 KNN algorithm

1. Determine the parameter K
2. Calculate the distance between the data to be evaluated with all the training
3. Sort range formed (ascending)
4. Determine the shortest distance to the order of K
5. Pair the corresponding class
6. Find the number of classes from the nearest neighbor and set the class as a class data to be evaluated
KNN formula:

\[ d_i = \sqrt{\sum_{j=1}^{p} (x_{2i} - x_{1j})^2} \]

Information:
- \( x_1 \) = Sample Data
- \( x_2 \) = Data Test / Testing
- \( i \) = Variable Data
- \( d \) = Distance formed
- \( p \) = Dimension Data

Below is a flowchart of the method KNN:

![Flowchart of KNN Method](image)

**Figure 2.** Flowchart of KNN Method

3. Methodology

Broadly speaking, the stages in this study is illustrated in Figure 3.
Figure 3 above is a research methodology that will be done by the author. The research methodology aims to outline all the activities carried out during the course of the study. From the picture above, it is known that there are three stages to be done to resolve the case at this research that includes: data collection, pre-process the data, data transformation, optimization of the value of k and cluster results. The preparation process includes three main things:

3.1 Data Selection

Select the data that will be used in the data mining process. In the process of the election is done also attributes that are tailored to the data mining process. In this study, the data used is in the form of data-ready, meaning that the data obtained has been the form of the target data. At this stage the problem to be faced is noisy data and missing values. Data cleaning process done to clean data from duplicate data, the data is inconsistent, or typographical errors. So the data that has been through this process are ready to be processed in data mining. In this study, the data used is data that has been consistent, so that the data cleansing process is only performed on any data missing value.
### 3.2 Data transformation

This study procedures carried out as in figure 3.1, namely, the data obtained from the database of the UN tax revenue Deli Serdang. Data will be modified. Data in the form of Excel 2016 spreadsheet files (.xls) as input to the Weka open source software. Before the data is transformed into ARFF, the data is converted first into the .csv format. Weka transform data from .csv be ARFF. The result of the transformation is preliminary data that will be used for optimization processes with k values. The results of the data transformation xls, csv, ARFF can be seen in the picture below.
Table 2. The data as a .csv

|   | A  | B  | C  | D  | E  | F  | G  | H  | I  | J  |
|---|----|----|----|----|----|----|----|----|----|----|
| 1 | 5PTTARGET,5PT Y Banyar,Tercapai,5PT Belum Banyar,Target Tok Capa,Perenerimaan,Pencapaian,Kategori | 997.111.60.711.85.450.90.120.96.5.547.21.558.133.80.70.TERCAPAI.NAIK |
| 2 | 673.58.806.882.316.49.427.433.35.7.9.469.449.83.92.TERCAPAI.TURUN |
| 3 | 2.106.100.493.314.935.72.531.398.1.173.33.901.910.66.11.TERCAPAI.NAIK |
| 4 | 1.400.57.434.795.535.23.445.960.890.11.985.850.44.30.TIDAK TERCAPAI.TURUN |
| 5 | 498.32.945.562.233.18.769.761.265.15.23.741.56.79.TIDAK TERCAPAI.TURUN |
| 6 | 473.52.273.687.175.34.890.876.285.17.583.321.60.36.TIDAK TERCAPAI.TURUN |
| 7 | 942.30.941.684.524.12.563.782.416.13.379.902.56.76.TIDAK TERCAPAI.TURUN |
| 8 | 718.25.903.318.390.13.759.256.258.12.143.048.53.12.TIDAK TERCAPAI.TURUN |
| 9 | 385.26.294.360.169.10.269.476.216.15.024.833.35.06.TIDAK TERCAPAI.TURUN |
| 10 | 336.10.101.119.173.5.799.489.103.4.301.630.57.41.TIDAK TERCAPAI.TURUN |
| 11 | 991.70.296.473.414.45.907.888.577.21.295.185.69.71.TERCAPAI.TURUN |
| 12 | 941.156.416.754.489.138.097.411.452.18.319.343.88.29.TERCAPAI.TURUN |
| 13 | 1.113.147.920.197.598.106.572.381.223.41.347.810.72.05.TERCAPAI.NAIK |
| 14 | 2.903.144.972.498.1.429.75.793.680.1.483.67.768.812.54.63.TIDAK TERCAPAI.TURUN |
| 15 | 288.54.674.689.169.21.335.751.119.23.334.938.47.77.TIDAK TERCAPAI.TURUN |
| 16 | 2.725.305.400.622.2.537.204.438.216.3.188.125.018.412.62.65.TERCAPAI.TURUN |
| 17 | 1.1.106.668.1.1.166.668.0.0.100.00.TERCAPAI.TETAP |
| 18 | 1.194.161.668.1.1.166.668.0.0.100.00.TERCAPAI.TETAP |
| 19 | 544.23.659.035.225.16.708.075.121.1.150.961.70.03.TERCAPAI.TURUN |
| 20 | 1.196.43.499.225.789.31.428.523.410.12.040.332.72.32.TERCAPAI.TURUN |
| 21 | 1.482.11.070.838.387.16.191.750.1.105.34.879.686.11.70.TIDAK TERCAPAI.TURUN |
| 22 | 1.658.70.956.662.856.42.877.844.802.28.088.820.60.42.TERCAPAI.TURUN |
| 23 | 597.22.847.140.207.10.037.302.300.12.293.844.46.19.TIDAK TERCAPAI.TURUN |
| 24 | 1.974.311.421.537.530.57.231.401.1.444.60.190.172.48.74.TIDAK TERCAPAI.TURUN |
| 25 | 1.456.93.092.242.338.36.377.254.921.56.714.951.39.08.TIDAK TERCAPAI.TURUN |
| 26 | 437.23.355.82.196.12.465.996.241.10.925.826.53.30.TIDAK TERCAPAI.TURUN |
| 27 | 673.34.245.85.578.17.610.111.265.9.635.783.72.03.TERCAPAI.TURUN |
| 28 | 1.433.42.586.720.820.26.287.982.613.16.296.738.61.73.TERCAPAI.NAIK |
| 29 | 725.20.110.320.381.9.477.328.344.10.632.992.47.13.TIDAK TERCAPAI.TURUN |
| 30 | 1.367.40.779.944.852.76.035.515.505.1.744.471.63.84.TERCAPAI.TURUN |
| 31 | 2.465.33.499.805.1.872.49.500.323.590.13.590.482.78.59.TERCAPAI.TURUN |
| 32 | 683.61.402.741.398.47.953.448.284.13.488.293.78.10.TERCAPAI.TURUN |
| 33 | 874.33.368.374.443.17.988.106.431.1.370.274.53.94.TIDAK TERCAPAI.TURUN |
| 34 | 1.095.111.948.767.328.53.875.448.771.56.073.319.48.13.TIDAK TERCAPAI.TURUN |
| 35 | 1.543.91.371.016.795.34.217.988.748.37.852.038.29.21.TIDAK TERCAPAI.TURUN |
| 36 | 1.506.34.355.422.857.19.945.262.649.14.449.160.57.99.TIDAK TERCAPAI.TURUN |
| 37 | 5.064.219.555.560.2.577.108.554.531.5.087.111.401.029.49.33.TIDAK TERCAPAI.TURUN |
| 38 | 1.338.110.809.551.785.72.628.801.573.38.180.750.65.54.TERCAPAI.TURUN |
| 39 | 1.087.39.854.178.858.54.81.175.879.26.287.502.57.12.TIDAK TERCAPAI.TURUN |
Table 3. The data in ARFF format

| No | No   | Target | No   | Target | No   | Target | No   | Tercapai | Naik |
|----|------|--------|------|--------|------|--------|------|----------|------|
| 1  | 977  | 111    | 450  | 80     | 1288 | 547    | 21569 | 133      | 89.7  |
| 2  | 673  | 58     | 390  | 60     | 427  | 357    | 94699 | 449      | 83.92 |
| 3  | 208  | 106    | 930  | 72     | 521  | 1173   | 33919 | 916      | 68.11 |
| 4  | 140  | 57     | 530  | 25     | 445  | 865    | 31688 | 855      | 44.3  |
| 5  | 498  | 32     | 945  | 18     | 758  | 265    | 14255 | 741      | 56.75 |
| 6  | 473  | 52     | 273  | 175    | 499  | 299    | 17583 | 321      | 66.36 |
| 7  | 942  | 20     | 415  | 524    | 175  | 418    | 13379 | 302      | 56.76 |
| 8  | 718  | 20     | 390  | 13     | 729  | 328    | 12143 | 308      | 53.12 |
| 9  | 386  | 20     | 168  | 162    | 216  | 216    | 1604  | 303      | 39.08 |
| 10 | 338  | 10     | 1730 | 1830   | 4869 | 183    | 4301  | 630      | 57.41 |
| 11 | 991  | 70     | 296  | 414    | 497  | 577    | 21283 | 365      | 69.71 |
| 12 | 941  | 156    | 415  | 489    | 138  | 392    | 18319 | 343      | 88.29 |
| 13 | 1111 | 147    | 520  | 596    | 160  | 572    | 41347 | 676      | 72.05 |
| 14 | 293  | 144    | 972  | 1420   | 79   | 203    | 65768 | 812      | 54.63 |
| 15 | 288  | 44     | 674  | 169    | 213   | 119    | 2334  | 308      | 47.77 |
| 16 | 275  | 329    | 450  | 1537   | 2044 | 138    | 1205  | 1418     | 62.05 |
| 17 | 10   | 1318   | 686  | 1.0     | 3188 | 686    | 0     | 0        | 100   |
| 18 | 344  | 23     | 869  | 223    | 17   | 378    | 7155  | 961      | 70.63 |
| 19 | 1596 | 43     | 469  | 780    | 31   | 458    | 12040 | 332      | 72.32 |
| 20 | 1462 | 51     | 7070 | 387    | 16   | 191    | 34873 | 696      | 31.7  |
| 21 | 1658 | 700    | 850  | 827    | 42   | 872    | 29888 | 820      | 60.42 |
| 22 | 507  | 22     | 947  | 207    | 16   | 553    | 12293 | 844      | 46.19 |
| 23 | 1974 | 117    | 421  | 630    | 87   | 531    | 60100 | 172      | 48.74 |
| 24 | 1459 | 33     | 052  | 538    | 35   | 777    | 58714 | 391      | 39.08 |
| 25 | 437  | 23     | 3985 | 519    | 12   | 465    | 10925 | 826      | 53.3  |
| 26 | 6730 | 24     | 425  | 574    | 17   | 610    | 6535  | 743      | 72.63 |
| 27 | 1633 | 42     | 295  | 820    | 26   | 287    | 16299 | 738      | 61.73 |
| 28 | 725  | 20     | 110  | 381    | 81   | 472    | 10632 | 902      | 47.13 |
| 29 | 1367 | 40     | 770  | 862    | 26   | 635    | 14744 | 478      | 63.94 |
| 30 | 2489 | 63     | 400  | 1977   | 49   | 900    | 13590 | 482      | 78.59 |
| 31 | 693  | 61     | 402  | 590    | 47   | 853    | 13449 | 293      | 78.1 |
| 32 | 914  | 33     | 396  | 440    | 17   | 896    | 15700 | 214      | 53.94 |
| 33 | 1909 | 111    | 948  | 328    | 53   | 875    | 59073 | 319      | 48.13 |
| 34 | 1543 | 81     | 517  | 750    | 54   | 217    | 37353 | 028      | 59.21 |
| 35 | 1500 | 34     | 953  | 857    | 19   | 946    | 14449 | 165      | 79.99 |
| 36 | 5664 | 21     | 9368 | 577    | 106  | 554    | 11401 | 029      | 49.35 |
| 37 | 1539 | 110    | 063  | 785    | 72   | 626    | 38180 | 756      | 65.54 |
| 38 | 338  | 95     | 585  | 570    | 50   | 338    | 46245 | 756      | 52.12 |
| 39 | 9220 | 21     | 1177 | 449    | 16   | 054    | 5071  | 953      | 75.98 |

3.3 Optimization Rated K

k-Nearest Neighbor (KNN) is a method using supervised algorithms where the results of the new query instance is classified based on the majority of categories on KNN. The purpose of this algorithm is to classify a new object attributes and training Based on the sample. Classifier does not use any model to be matched and only based on memory. Given query point, will find a number of objects or K (training points) closest to the query point.

3.4 Expectation Maximization Clustering

Expectation maximization algorithm is an algorithm unsupervised learning that has the ability to perform searches darisekumpulan knowledge of data that do not have labels or targets a particular class, by seeing the value of any instances distributed into the Gaussian distribution, more tepatnyaadalah Gaussian mixture, then do iterations ascending to seek the highest likelihood value for each instance (see proximity to each cluster instances). Expectation Maximization algorithm (EM algorithm) is an algorithm that utilizes the mixture of Gaussian mixture.
Basically EM algorithm consists of two steps, i.e., expectation and maximization. Calculating expectation to a likelihood probability value, then the second step of fixing the value of the probability of the stretcher by changing parameters on Gaussian mixture so as to achieve maximum likelihood. There some things that need to be emphasized in the EM algorithm Algorithm namely:

1. Maximum Likelihood Estimation (MLE)
2. Mixtures of Gaussians
3. Estimation-Maximization (EM)

But the EM algorithm using Gaussian mixture or words of a Gaussian lainlebih used or seeking mixture of yangdidapatkan distribution. EM Algorithm has the task of finding each Gaussian yangterdapat on Gaussian mixture distribution and develop each Gaussian yangditemukan at the optimum condition (so the model is more fit) that's called maximization, and the clustering process.

3.5. Interpretation / Evaluation

At this stage of the evaluation and interpretation of the patterns obtained based on the results of clustering data using EM-cluster method. If the results obtained are not appropriate, then the process would be repeated to the stage of the clustering process data. Knowledge of this stage is the final part of the KDD process where possible to investigate whether a pattern or information found in conflict with the facts. Pattern information generated from the data mining process should be presented in a form easily understood by the parties concerned.

4. Result and Discussion

Furthermore, from the data of the parameter with a k-nn algorithm, the data in the pull to get the Weka application cluster also using two parts of the cluster with no parameters and cluster k-nn-nn with parameter k. Both parts are in the cluster by using an algorithm expectation Maximization (EM). This algorithm is already available in Weka and can be directly used. The output from these two different parts and will be compared. For the results of the cluster with no parameters can be viewed as in Table 4.
Table 4. Cluster results with the original data

| No. cluster | Total Instant |
|-------------|---------------|
| 1           | 0             |
| 2           | 1             |
| 3           | 2             |
| 4           | 3             |
| 5           | 4             |
| 6           | 5             |
| 7           | 6             |
| 8           | 7             |
| 9           | 8             |
| 10          | 9             |
| 11          | 10            |

Table 5. Results of Cluster Without Parameter K-NN

| No. cluster | Total Instant |
|-------------|---------------|
| 1           | 15            |
| 2           | 31            |
| 3           | 10            |
| 4           | 14            |
| 5           | 12            |
| 6           | 6             |
| 7           | 19            |
| 8           | 11            |
| 9           | 14            |
| 10          | 25            |
| 11          | 48            |
The second phase of testing is done with the data optimization results using the k value of K Nearest Neighbor with cluster model validation is performed on the original data. When implemented generate data as in Table 6.

Table 6. The results of Cluster with parameter k

| No. | cluster Total Instant |
|-----|----------------------|
| 1   | 0                    |
| 2   | 1                    |
| 3   | 2                    |
| 4   | 3                    |
| 5   | 4                    |
| 6   | 5                    |
| 7   | 6                    |
| 8   | 7                    |
| 9   | 8                    |

4.1. Influence Selection of Parameter Values k

In the test will be analyzed the effect of optimization parameters k value the success rate with algorithms cluster expectation Maximization, The k value is the number of nearest neighbors for use as consideration in determining the number of cluster decision.

Distance parameter used to optimize the use of simulation data that euclidean distance and the Hamming distance, while the value of k used is k = 13. Based on the above data processing results, when using early data without any additional parameters obtained by the
number of clusters found and incorrect as many as 11 clusters of 66% then when using the optimization parameters obtained by the number of clusters k sebnayak 9 and can minimize incorrect cluster to 64%.

5. Conclusion

By using clustering algorithms can mengdentifikasi Cluster EM-attainment status and budget plans in the coming year. In the process of this grouping K-NN with k = 13 an algorithm and can be used for the type of data berimensi high. Determination of parameter k in K-NN algorithm can affect and improve the number of clusters in advance.

References

[1] Connolly, Thomas, C. B 2010, Database Systems: A Practical Approach to Design, Implemantation, and Management Fifth Edition: Pearson Education Inc.
[2] Cuzzocrea, Alfredo 2011, warehousing Data and Knowledge Discovery, Springer-Verlag Berlin Heidelberg, London.
[3] Hermawati, Fajar Astuti, 2013. Data Mining. ANDI Publisher: Yogyakarta.
[4] Kimball, R, Margy R, Warren T, Joy M and Bob B 2008, The Data Warehouse Lifecycle Toolkit, Wiley Publishing Inc., Canada.
[5] Indrajani 2009 Database System In Package Five In One, PT.Elex Media Komputindo, Jakarta.
[6] Ponniah, Paulraj 2010, Data warehousing. Canada: John Wiley & Sons Inc.
[7] Tantra, Rudi 2012, Project Management Information Systems, Andi Offset, Yogyakarta.
[8] Jiawei Han and Micheline Kambar 2010, Data Mining Concepts and Techniques, Verlag Berlin.
[9] Oded Maimon and Lior Rokach, 2010, Data Mining And Knowledge Discovery Hanbook, Springer Science.
[10] Ernastuti, S. &. (2010). Graduation Prediction of Gunadarma University Students Using Algorithm and C4.5 Naive Bayes Algorithm.
[11] Gunadi, G., Sensuse, D., I., 2012, Application of Data Mining Methods Market Basket Analysis to book the product sales data by using algorithms Apriori and Frequent Pattern Growth (FP-Growth), MKom TELEMATIKA Journal, Vol. 4, No. 1, 118-132.
[12] Gorunescu, F. (2011). Data Mining Concepts and Techniques Models. Craiova: Springer.
[13] Hastuti, K. (2012, June). ANALYSIS COMPARISON OF CLASSIFICATION OF DATA MINING ALGORITHM National V.Seminar Applied Information & Communication Technology (979 - 26 - 0255-0), 241 249.
[14] Ian H. Witten, f. E. (2011). Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.). (ASBurlington, Ed.) United States of America: Morgan Kaufmann.
[15] Kalyankar, Q. &. (2010). Drop Out Feature of Student Performance Data Using Decision Tree forAcademic techniques. Global Journal of Computer Science and Technology, 2-4.
[16] Mardiani, 2014, Comparison Algorithm K-Means and EM for Clusterisasi Value Based Home School Students, CITEC Journal, Vol. 1, No. 4, 316-325
[17] Oyelade, O. &. (2010). Application of Kmeans Clustering algorithm for predicting of Students AcademicPerformace. International Journal of Computer Science and Information Security, 292-295.
[18] Syaifullah. (2010). Implementation of Data Mining Algorithm Apriori Sales System, Amikom, Yogyakarta.

[19] Tahyudin, I. (2013, December). Comparing Classification Of Data Mining Algorithm to Predict the Graduation Students on Time. Information Systems International Conference (ISICO).

[20] Vercellis. (2009). Business Intelligence: Data Mining and Optimization for Decision Making Decision Making. John Wiley & Sons Inc: Southern Gate.

[21] Vrettos, K. &. (2009). Sentivity Analysis of Neural Network for Identifying the Factors for Success College Students. World Congress on Computer Science and Information Engineering. (978-0-7695-3507-4).

[22] Yanto. R, Khoiriah. R., 2015, Implementation of Data Mining with Apriori Algorithm Method in Determining Drug Purchasing Patterns, CITEC Journal, Vol. 2, No. 2, 101-113.