

Weather Forecasting in Bandung Regency based on FP-Growth Algorithm

Farida Nur Khasanah ^{#1}, Fhira Nhita ^{*2}

#School Of Computing, Telkom University

** School Of Computing, Telkom University*

Telekomunikasi Street, Terusan Buah Batu, Bandung, 40257, Indonesia

¹ faridankh@student.telkomuniversity.ac.id

² fhiranhita@telkomuniversity.ac.id

Abstract

Weather change is one of the things that can affect people around the world in doing activities, including in Indonesia. The area of Indonesia, especially in Bandung regency has a high intensity of rainfall, compared with other regions. The people of Bandung Regency mostly have livelihoods in the fields of industry and agriculture, both of which are closely related to the effects of the weather. Weather prediction is used for reference, so the future of society can prepare all possible weather before the move. In this study, predicted rainfall for the next month, with monthly data from BMKG in Bandung regency. One method of data mining used to predict weather is the association rule method. In this method there is Frequent Pattern Growth (FP-Growth) algorithm, this algorithm is used to know pattern of linkage between attribute of weather with rainfall. The result of the FP-Growth algorithm is the rule association, the result of the algorithm is then used as a reference for data entry in the classification process, where the process is performed to obtain estimates based on the rainfall category to obtain maximum accuracy. The highest performance FP-Growth algorithm based on confidence value of the rule result is 92% and the highest accuracy J48 classification algorithm based on all scenarios show result 83,3% for all weather attributes.

Keywords: Rainfall, Association Rule, FP-Growth, Classification

I. INTRODUCTION

Indonesia is one of the tropical countries, which has high rainfall. Rainfall is a weather condition that needs to be considered, because it is very influential on human life in various fields, including culture, social, economic, tourism, etc. The region of Indonesia, especially in the regency of Bandung has a high intensity of rainfall, compared with other regions. The people of Bandung Regency mostly have livelihood in the field of industry and agriculture, both of which are very closely related to the influence of the weather. Lately, the weather changes are very irregular, making the weather unpredictable. Weather changes can disrupt community activities in general if not well anticipated, so weather prediction is required.

There are many studies that discuss about weather prediction, including in the field of soft computing and data mining. In previous research, has done a lot of weather prediction using soft computing approach (Nurchahyo, Nhita, & Adiwijaya, 2014 ; Nhita & Adiwijaya, 2013 ; Adiwijaya, Wisesty & Nhita 2014) with Genetic Algorithm (GA). There are several methods in Data mining that can be used for predicting the weather including association and classification. Association method is a method used to determine the correlation between attributes to be predicted on a data (Zhenhai, Dezhong, Jiw, & Wenyu, 2014), while classification method is a learning method where class attributes are involved in the classification to predict unknown sample data (Sudhakar & Manimekalai, 2014). In the method of data mining, there are several studies on the prediction of meteorological data such as: research (Folorunsho & Adesesan, 2015) explaining about weather prediction in Ibadan area with neural network algorithm and decision tree with 70% accuracy,

(Valmik & Meshram, 2013) explains weather prediction with Bayesian network on Indian Meteorological Data, (Ramzan et al., 2017 ; Fahad et al., 2016) weather prediction with Decision Tree algorithm. In the Association rule method itself there are several studies related to the prediction, including: (Amruta, Taksande, & Mohod, 2013) explained about weather prediction in Nagpur Station area India, using Frequent Pattern Growth (FP-Growth) algorithm got more than 90% accuracy and research (Zhenhai, Dezhong, Jie, & Wenyu, 2014) explaining about wind speed prediction strategy in Hexi Corridor area of China, using Apriori algorithm obtained result is more effective and shows that an a priori algorithm with its association rule can predict wind speed values for abnormal cases.

The association classification is the classification of a data based on the results of the association rules obtained, the result is used for prediction and maximizing accuracy (Sudhakar & Manimekalai, 2014). In this case, weather data with all attributes will be done by association process to get the rules by using FP-Growth algorithm, then classified by J48. The J48 classification is a simple algorithm of C4.5 for the construction of a tree which is then used for classification (Tina & Sherekar, 2013 ; Neeraj, Girja, Ritu, & Manish, 2013). In this journal, the data classified is rainfall. This study conducted weather prediction with data from BMKG Bandung Regency in 2005-2016 with Algorithm Frequent pattern Growth and J48.

II. FREQUENT PATTERN GROWTH (FP-GROWTH) ALGORITHM

A. *FP-Growth*

FP-Growth is one of the data mining algorithms used to determine frequent itemset that often arises from a data. The FP-Growth algorithm is the development of the Apriori algorithm, so the algorithm does not generate candidate because FP-Growth uses the concept of tree development in search of frequent itemset (Amruta, Taksande, & Mohod, 2013). The pattern growth algorithm, commonly used to find the full set of periodic partial patterns of frequent patterns in a database (Han & Kamber, 2001). The reasons for the FP-Growth algorithm being more efficient than other algorithms are:

1. Divide and Conquer: data mining is decomposed into sub-datasets according to the pattern that is often identified. This leads to the search being focused on smaller databases
2. No generation of candidates
3. No repetition on scan of entire database (Sidhu, Kumar & Nawani, 2014)

Each transaction on the method association rules with FP-Growth algorithm, using the value of support and confidence value to know the percentage of the number of transactions.

B. *FP-Tree*

FP-Tree is the problem solving structure of FP-Growth algorithm. The structure of the FP-Tree itself consists of a root that is initialized with null, then a set of prefix subtrees as the root branches, etc. Each subtree consists of three fields, namely:

The item label, to inform the type of item represented on the node

- Support count, to represent the number of transactions passing through the node, and
- Node-link, to connect vertices with the same item label between paths, indicated by dashed arrow lines. (Sidhu, Kumar & Nawani, 2014)

C. *Classification J48*

The association classification is a classification of data based on the results of the association rules obtained, then used for forecasting and maximizing accuracy (Sudhakar & Manimekalai, 2014). In this case, weather data with all attributes will be done by association process to get the rules by using FP-Growth algorithm, then classified by J48. The J48 classification algorithm is a simple algorithm of C.45 for the development of a decision tree which is then used for the classification process. Development of decision tree on J48 algorithm for classification there are several stages, among them (Neeraj, Girja, Ritu, & Manish, 2013):

1. Calculation of Entropy

Entropy is a measure of data interruption or it can also be said to be the size of the uncertainty of a random variable. Entropy can be calculated by the formula:

$$Entropy(P) = - \sum_{i=1}^n b_i \times \log_2 b_i \tag{1}$$

Where, n is the number of classes and b_i represents the number of samples in class i .

2. Calculation of Gain

Gain is the result of the calculation of Entropy (P) subtracted by the result of the number of samples for the value of A divided by the sum of the sample data multiplied by entropy for the sample of value A. Gain can be calculated by the formula:

$$Gain (P, Q) = Entropy(P) - \sum \frac{|P_A|}{|P|} Entropy (P_A) \tag{2}$$

III. RESEARCH METHOD

Weather prediction in Bandung Regency done by doing some processing stages to get the best prediction result. Here is a design system of this study:

TABLE I
DESIGN SYSTEM FP-GROWTH ALGORITHM AND J48

Design System	
Input:	Monthly weather dataset from BMKG Bandung Regency in 2005-2016
Output:	Rules FP-Growth Prediction and Accuracy Classification J48
Preprocessing:	Discretization Data Partition (training-testing)
FP Growth Algorithm (Han & Kamber , 2001):	if <i>tree</i> contains single path then for each combination of the node in the path Z Generate pattern $X \cup Y$ with $sup_count = min_sup_count$ of nodes in X; else for each iteration of tree { generate pattern $X = i_1 \cup Y$ with $sup - count = i_1.sup_count$; construct X's conditional pattern base and then X's conditional FP_Tree <i>treeX</i> ; if $treeX \neq 0$ then call FP_Growth(X,Y);
Classification J48:	Performance Calculation eq. (7), (8), (9)
Analysis Performance:	Confusion Matrix detail in V.D

A. Preprocessing

1.) Discretization

The weather data consists of 6 attributes, including: humidity (K), solar radiation(PLM), wind velocity (A), temperature (T), evaporation (U), and Rainfall, all attributes are numerical values which then converted to categorical data by using equal width. The parameters contained in the dataset are divided into three categories (Nhita & Adiwijaya, 2013) for all parameters except rainfall. Equal width is the difference between the maximum value and the minimum value divided by the number of categories. Here is the equation for equal width:

$$equal\ width = \frac{\max\ value - \min\ value}{number\ of\ categories} \quad (3)$$

In this case, three categories have been set based on equal width. L1 for low-level, L2 for medium-level, and L3 for high-level. The following is the result of the calculation of equal width, for the range of values of each class:

TABLE II
 RESULT OF DISCRETIZATION WITH EQUAL WIDTH

	Solar Radiation	Humidity	Wind Velocity	Temperature	Evaporation
L1 (Low-Level)	19-23,3	61-70	1-2,7	22,5-23,3	2,7-3,7
L2 (Medium-Level)	23,4-65,7	71-79	2,8-4,3	23,4-24,03	3,84,7
L3 (High-Level)	65,8-89	80-88	4,4-6	24,04-24,8	4,8-5,7

Based on sources from BMKG, rainfall is classified into five and four categories (Nhita & Adiwijaya, 2013), the following table is an explanation of each division:

TABLE III
 CLASSIFICATION OF RAIN CHAIN TO 5 CATEGORIES

Class	Range	Category
A	Rainfall <= 0.1mm	No Rain
B	0.1 < Rainfall <= 20 mm	Light Rain
C	20 < Rainfall <=50 mm	Moderate Rain
D	50 < Rainfall <= 100 mm	Heavy Rain
E	Rainfall > 100 mm	Very Heavy Rain

Categorization of rainfall into 4 categories is a combination of class A and B so that for the range of no rain / light rain into rainfall <= 20 mm. The division of rainfall classification into 4 and 5 categories aims to know the difference as well as to compare the results in the classification process. Thus, the rainfalls were clarified as determined by BMKG.

2.) *Data Partition*

The J48 classification algorithm divides data into training data and testing data. In the training data, learning J48 to generate prediction model and then from the results of the training data to test the system so that the predicted results obtained. At this stage, the data is divided into two parts, namely training data and testing data, with each part has been determined as follows:

TABLE IV
 DATA PARTITION AND PERCENTAGE OF DATABASE

Data Partition	Data	Data Distribution				
		1	2	3	4	5
Training	Jan 2005-Dec 2013	2,78%	4,63%	7,41%	15,74%	69,44%
Testing	Jan 2014-Dec 2016	0,00%	8,33%	11,11%	5,56%	75,00%

Data partition aims to determine the dissemination of data from each attribute based on training data and testing data.

B. FP-Growth Algorithm

The input data in FP-Growth process is the entire weather data, without any division into training data and testing data. This is because the FP-Growth process aims to obtain linkages on each attribute to rainfall. The following is an explanation of each process:

1. Min-Support

The Support value is used to determine how many rules can be applied to the data set. The value of support (X, Y) is the result of chance transaction event X combined Y. Here is the value formula support (Han & Kambr, 2001):

$$Support (X, Y) = P(X \cup Y) \tag{4}$$

In the calculation phase of minimum support, the entire data is generated for each item. Results from the generate data, determined the value of support for each item, then done the determination of minimum support to eliminate items that are not in accordance with the minimum support that has been determined. The items that have met the minimum support value are then sequenced from the largest frequent items to the smallest and then the FP-Growth process to produce association rules.

2. Min-Confidence

Confidence value is used to determine how often an item in Y appears in a transaction containing X. The value of trust is the result of the division of the number of X transactions combined with Y number of transactions containing X. The following is the formula for calculating the value of confidence (Han & Kamber, 2001):

$$Confidence (X, Y) = \frac{Support (X \cup Y)}{Support (X)} \tag{5}$$

The result of FP-Growth is an association rule showing the pattern relationship between the weather attribute. Each of these rules, seen the value of his belief to know how strong the influence of the rule on the overall results that appear.

3. Lift Ratio

Lift Ratio is a simple correlation measure used to determine the strength level of the association rules. The higher the value of the lift ratio, the rule is also more influential. The lift ratio is derived from the percentage distribution of transactions on the database containing A and B with the number of transactions containing B. Here is the formula of the lift ratio (Han & Kamber, 2001):

$$Lift (X, Y) = \frac{Confidence (X, Y)}{Support Y} \tag{6}$$

C. Analysis Performance

The result of classification has been obtained by using J48 algorithm, then analyzed its performance in terms of accuracy, precision, and recall. This is done to show that the predicted results with the J48 classification algorithm produce the best performance. Confusion matrix is a picture of accuracy arranged with matrix pattern to facilitate in solving classification problem. The pattern represents the actual class with the prediction class. Here is an overview of confusion matrix (Han & Kamber, 2001):

TABLE V
CONFUSION MATRIX

		Prediction Class	
		TRUE	FALSE
Actual Class	TRUE	TT	FF
	FALSE	FT	TF

With TT is the actual class true that is predicted to be true, TF is the actual class false which is predicted to be true, FT is the actual class false which is predicted to be true, FF is the actual class true which is predicted to be false.

Precision is the ability of the system to call data relevant to what is desired, the recall is the proportion of the amount of relevant data to be found and F-measure is an evaluation of the result of a combination of precision and recall. Precision, recall, f-measure can be formulated as follows (Han & Kamber, 2001):

$$Precision = \frac{TT}{TT + FT} \quad (7)$$

$$Recall = \frac{TT}{TT + FF} \quad (8)$$

Accuracy is determined based on the confusion matrix, because accuracy is the percentage of the accuracy of the results of the classification. Here is the formula of accuracy (Han & Kamber, 2001):

$$Accuracy = \frac{TT + TF}{TT + FT + FF + TF} \quad (9)$$

IV. RESULTS AND DISCUSSION

A. Test Scenario

Testing begins with processing the entire data with FP-Growth algorithm to get the rules, rules show the attributes associated with rainfall, after obtained the rules results, it then classified by J48 to get prediction results and accuracy values. In this experiment we will do some test scenarios. Here are the scenarios of each algorithm used:

1.) FP-Growth Algorithm

Scenarios performed on the FP growth algorithm are to use the overall attribute on the weather data that has been discretized, except on the rainfall attribute. On the attribute of rainfall is used classification of 4 categories and 5 categories as in Table II and Table III. The overall data on the FP-Growth algorithm is used to generate linkages between weather attributes and the effects of rainfall.

2.) J48

The J48 classification algorithm scenario is a continuation of the FP-Growth scenario that has resulted in association rules. Input of J48 is rules from result FP-Growth algorithm. The result of the rule shows that all attributes are closely related to each other, and then re-examined the interrelationship between attribute and rainfall. The result of the rule that affects rainfall or in other words that have a close relationship with the rainfall attribute that will be used for input in the classification using J48 algorithm. Once identified, the classification is done with J48 algorithm on training data and data testing. The above scenario is done to find out how much influence of each attribute on the weather data with the classification process to predict rainfall.

3.) Result and Discussion

Tests conducted for weather prediction in Bandung Regency using FP-Growth Algorithm which then classified with J48 get mixed results. Here are the results of test scenarios that have been done:

TABLE VI
RESULT OF FP-GROWTH SCENARIO WITH 5 CATEGORIES RAINFALL CLASSIFICATION

Antecedent	Consequent	Confidence	Lift Ratio
Evaporation=Low	Very Heavy Rain	0,80	1,13
Temperature=Medium	Very Heavy Rain	0,79	1,12
Solar Radiation=Medium	Very Heavy Rain	0,89	1,25
Humidity=High	Very Heavy Rain	0,92	1,30
Evaporation=Low	Humidity=High	0,71	1,36
Humidity=High	Evaporation=Low	0,82	1,36
Wind Velocity=Medium	Very Heavy Rain	0,74	1,04
Very Heavy Rain	Wind Velocity=Medium	0,85	1,04
Temperature=Medium	Wind Velocity=Medium	0,88	1,07
Solar Radiation=Medium	Wind Velocity=Medium	0,82	1,01
Solar Radiation=Medium	Wind Velocity=Medium, Very Heavy Rain	0,76	1,26
Wind Velocity=Medium, Solar Radiation=Medium	Very Heavy Rain	0,99	1,31
Very Heavy Rain, Solar Radiation=Medium	Wind Velocity=Medium	0,86	1,05
Humidity=High	Wind Velocity=Medium	0,82	1,01
Evaporation=Low	Wind Velocity=Medium	0,85	1,04
Wind Velocity=Medium, Evaporation=Low	Very Heavy Rain	0,82	1,16
Very Heavy Rain, Evaporation=Low	Wind Velocity=Medium	0,87	1,06

TABLE VII
RESULT OF FP-GROWTH SCENARIO WITH 4 CATEGORIES RAINFALL CLASSIFICATION

Antecedent	Consequent	Confidence	Lift Ratio
Evaporation=Low	Very Heavy Rain	0,80	1,13
Temperature=Medium	Very Heavy Rain	0,79	1,12
Solar Radiation=Medium	Very Heavy Rain	0,89	1,25
Humidity=High	Very Heavy Rain	0,92	1,30
Humidity=High	Evaporation=Low	0,81	1,36
Wind Velocity=Medium	Very Heavy Rain	0,74	1,04
Very Heavy Rain	Wind Velocity=Medium	0,84	1,04
Temperature=Medium	Wind Velocity=Medium	0,88	1,08
Solar Radiation=Medium	Wind Velocity=Medium	0,81	1,00
Solar Radiation=Medium	Wind Velocity=Medium, Very Heavy Rain	0,75	1,26
AL2, Solar Radiation=Medium	Very Heavy Rain	0,92	1,30
Very Heavy Rain, Solar Radiation=Medium	Wind Velocity=Medium	0,85	1,04
Humidity=High	Wind Velocity=Medium	0,82	1,01
Evaporation=Low	Wind Velocity=Medium	0,84	1,03
Wind Velocity=Medium, Evaporation=Low	Very Heavy Rain	0,82	1,16

In Table VI and Table VII we can see the test results for Rainfall classification into 4 and 5 categories. Both tables, the results show the interrelationship between rules, the results show the overall attribute of the weather affects rainfall, including: low-level evaporation, medium temperature, medium solar radiation, medium wind speed, high humidity and very heavy rain.

The value of confidence describes the value of the belief of the rule on the result of FP-Growth, the higher the confidence value the higher the strength of the rule. In both tables it is seen that the highest confidence

value is 0.99 in the rainfall classification with 5 categories and 0.9 in the rainfall classification with 4 categories, both of which are medium wind velocity and medium solar radiation.

Lift Ratio Test is a correlation measure used to determine the strength of the association rules. The overall attribute of the weather affects rainfall, but from those 2 tables when viewed from the value of its lift ratio, the factors that most influence the rainfall are medium wind velocity and medium solar radiation with the largest lift ratio value of 1,3.

Rules result shows the highest confidence and lift ratio is in the same attribute, both have a very close relationship in affecting rainfall. The FP-Growth scenario that classifies rainfall into 4 categories and 5 categories has no significant effect, both of which show similar results. The association process (FP-Growth) produces a rule, which results can be seen as a link between weather attributes and rainfall. To make predictions of rainfall, the process of classification is done by using J48 classification algorithm. Here is a scenario that is done on J48 classification:

TABLE VIII
 SCENARIO J48

Scenario	Attribute of Input Data
1	All Weather Attributes
2	Rainfall and Humidity
3	Rainfall and Evaporation
4	Rainfall and Wind velocity
5	Rainfall and Solar radiation
6	Rainfall and Temperature

The above scenario is done to find out how much influence of each attribute on the weather data with the classification process to predict rainfall. J48 classification data entry is a combination of each attribute related to the weather attribute.

In testing with FP-Growth algorithm has been obtained the overall attribute closely related to Rainfall, from the overall attributes are then used as input in the classification process using several scenarios that have been set as in Table VIII, the following tables are the result of J48 algorithm scenario:

TABLE IX
 RESULTS OF SCENARIO J48 WITH 5 CATEGORY RAINFALL CLASSIFICATION

Scenario	Training			Testing		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
1	0,856	0,973	80,56%	0,844	1	83,30%
2	0,733	0,977	70,80%	0,625	1	62,50%
3	0,75	1	74,20%	0,625	1	62,50%
4	0,725	1	72,50%	0,625	1	62,50%
5	0,725	1	72,50%	0,929	0,867	70,80%
6	0,694	1	69,40%	0,75	1	75,00%

TABLE X
RESULTS OF SCENARIO J48 WITH 4 CATEGORY RAINFALL CLASSIFICATION

Scenario	Training			Testing		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
1	0,859	0,973	82,41%	0,867	0,963	83,30%
2	0,785	0,966	75,00%	0,625	1	62,50%
3	0,75	1	74,20%	0,625	1	62,50%
4	0,725	1	72,50%	0,625	1	62,50%
5	0,725	1	72,50%	0,929	0,867	70,80%
6	0,75	1	75,00%	0,69	1	69,40%

In Table IX and Table X it is clear that the highest accuracy is in scenario 1, where all attributes are included in the classification. The overall precision value reaches > 0.6 which means the actual data and the predicted results show the correct result. To see the performance analysis on the classification result described in the confusion matrix, the following is a one of confusion matrix for the highest accuracy result from classification in training data:

TABLE XI
CONFUSION MATRIX WITH THE LARGEST ACCURACY

Rainfall Class		Prediction Class			
		Very Heavy Rain	Heavy Rain	Light Rain	Moderate Rain
Actual Class	Very Heavy Rain	27	0	0	0
	Heavy Rain	2	0	0	0
	Light Rain	1	0	2	0
	Moderate Rain	2	0	1	1
Total		32	0	3	1

Analysis performance describes the level of accuracy in forecasting performed using J48 classification algorithm. The four confusion matrix tables above show that the most widely classified in the rainfall class is very heavy, due to uneven distribution of data. The results of the J48 classification indicate that the output of this month will be the reference for the coming months.

IV. Conclusion

In this journal, shows that weather factors that affect rainfall are humidity, wind speed, solar radiation, and evaporation. For classification results, in addition to 4 attributes that affect the value of precipitation, precision and recall will also affect the accuracy of the results of rainfall classification. Results obtained from FP-Growth algorithm testing, obtained the highest performance results are at the highest confidence value, because the higher the value of confidence the greater the strength of the rule. The highest performance FP-Growth algorithm based on confidence value of the rule result is 92% and accuracy J48 classification algorithm based on all scenarios show result 83,3% for all weather attributes in testing data.

VI. ACKNOWLEDGMENT

The authors would like to thank the School of Computing and Computational Science, for providing facilities and providing guidance until this journal can be completed properly.

REFERENCES

- Nurchahyo Septian, Nhita Fhira, Adiwijaya. (2014). *Rainfall Prediction in Kemayoran Jakarta Using Hybrid Genetic Algorithm (GA) and Partially Connected Feedforward Neural Network (PCFNN)*. ICoICT.
- Nhita Fhira, Adiwijaya. (2013). *A Rainfall Forecasting using Fuzzy System Based on Genetic Algorithm*. ICoICT.
- Adiwijaya, U.N. Wisesty, Nhita Fhira. (2014). *Study of Line Search Techniques One The Modified Backpropagation for Forecasting of Weather Data in Indonesia*. FJMS.
- Zhenhai Guo, Dezhong Chi, Jiw Wu, Wenyu Zhang. (2014). *A new Winnd Spees Forecasting Strategy basen on the ChaoticTimeseries Modelling Technique and the Apriori Algorithm*. Elseiver.
- K.Sudhakar, Dr. M. Manimekalai.(2014). *Study of Heart Disease Prediction using Data Mining*. IJARCSS.
- Folorunsho Olaiya, Adesesan Barnabas Adeyemo. (2015). *Application of Data Mining Techniques in Weather Prediction and Climate Change Studies*. Information Engineering and Electronic Business.
- Valmik B Nikam, B.B.Meshram. (2013). *Modeling Rainfall Prediction Using Data Mining Method : A Bayesian Approach*. IEEE.
- M. Ramzan Talib, Toseef Ullah, dkk. (2017). *Application of Data Mining Techniques in Weather Data Analysis*. IJCSNS.
- Fahad Sheikh, S.Karthick, dkk. (2016). *Analysis of Data Mining Techniques for Weather Prediction*. Indian Journal of Science and Technology.
- Anruta A., Taksande, P.S. Mohod. (2013). *Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm*. International Journal of Science and Research (IJSR).
- Zhenhai Guo, Dezhong Chi, Jie Wu, Wenyu Zhang. (2014). *A New Wind Speed Forecasting Strategy Based on the Chaotic Time Series Modelling Technique and the Apriori algorithm*. Elsevier, vol. 84, pp 140-151.
- Tina R. Patil, Mrs. S. S. Sherekar. (2013). *Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification*. International Journal Of Computer Science And Applications.
- Dr. Neeraj Bhargava, Girja Sharma, Dr. Ritu Bhargava, Manish Mathuria. (2013). *Decision Tree Analysis on J48 Algorithm for Data Mining*. IJARCSSE.
- J. Han and M. Kamber. (2001). *Data mining : Concepts and Techniques*. Morgan Kaufmann
- S. Sidhu, U. Kumar, A. Nawani, dkk. (2014). *FP Growth Algorithm Implementation*. International Jurnal of Computer Applications.