**1. INTRODUCTION**

**1.1 Introduction:**

The rising technology usage resulted into generation of enormous quantity of digital data which thereby resulted in large storage database. This expansion of database occurred in many prominent areas like government datum, transaction detail of supermarket, mobile phone call details, and record of credit card usage and also in intricate areas like astronomical data records, medical reports and likes. With the expeditious increase in data, it is a sharp need to extract serviceable information from the database which might result into some advantageous information to the user. This task of exploring data and translating into more meaningful patterns/information is known as data mining

Data mining is the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information. Data mining also recognized as knowledge discovery in databases (KDD), where KDD is a conventional method of transfiguring enormous data to consequential interpretation and analysis. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Data mining, to the uninitiated, sounds like the kind of monotonous computational activity that requires a big computer, a mass of information and little human oversight. But in fact it’s a discipline that blurs the lines between artificial intelligence, machine learning, statistics and other cutting-edge disciplines to unearth the golden nuggets that lurk within data.

Weka is data mining software that uses a collection of machine learning algorithms. These algorithms can be applied directly to the data or called from the Java code.

Majority of population in India comes under rural area. The Government has announced so many schemes for rural masses with the aim to bring them out of the folder of poverty. National Rural Employment Guarantee Act 2005 (NREGA) , later renamed as the "Mahatma Gandhi National Rural Employment Guarantee Act", MGNREGA), is an Indian labour law which aims to enhance livelihood security in rural areas by providing at least 100 days of wage employment in a financial year to every household whose adult members volunteer to do unskilled manual work.

Wage per day is set for each worker every year depending upon the work. However, there are many cases where the wage payment is delayed. In case the wages are not paid within 15 days 0.05% of the wage is paid as interest per day as delay compensation. In this case the worker can file a petition for delay compensation. The petition may be approved or rejected by the government.

In our project we try to find out the reasons for the delay of wage payment and the reasons for rejecting the delay compensation in various states of India using data mining technique in Weka.

**1.2 Data Mining**

Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related) in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The

ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining and one that has the most direct business applications.

The process of data mining consists of three stages: (1) the initial exploration, (2) model building or pattern identification with validation/verification, and (3) deployment (i.e., the application of the model to new data in order to generate predictions).

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large

Relational databases.

Although data mining is a relatively new term, the technology is not. Companieshave used powerful computers to sift through volumes of supermarket scannerdata and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

It is no surprise that data mining, as a truly interdisciplinary subject, can be defined in many different ways. Even the term data mining does not really present all the major components in the picture. To refer to the mining of gold from rocks or sand, we say gold mining instead of rock or sand mining. Analogously, data mining should have been more appropriately named “knowledge mining from data,” which is unfortunately somewhat long. However, the shorter term, knowledge mining may not reflect the emphasis on mining from large amounts of data. Nevertheless, mining is a vivid term characterizing the process that finds a small set of precious nuggets from a great deal of raw material. Thus, such a misnomer carrying both “data” and “mining” became a popular choice. In addition, many other terms have a similar meaning to data mining—for example, knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data, or KDD, while others view data mining as merely an essential step in the process of knowledge discovery.

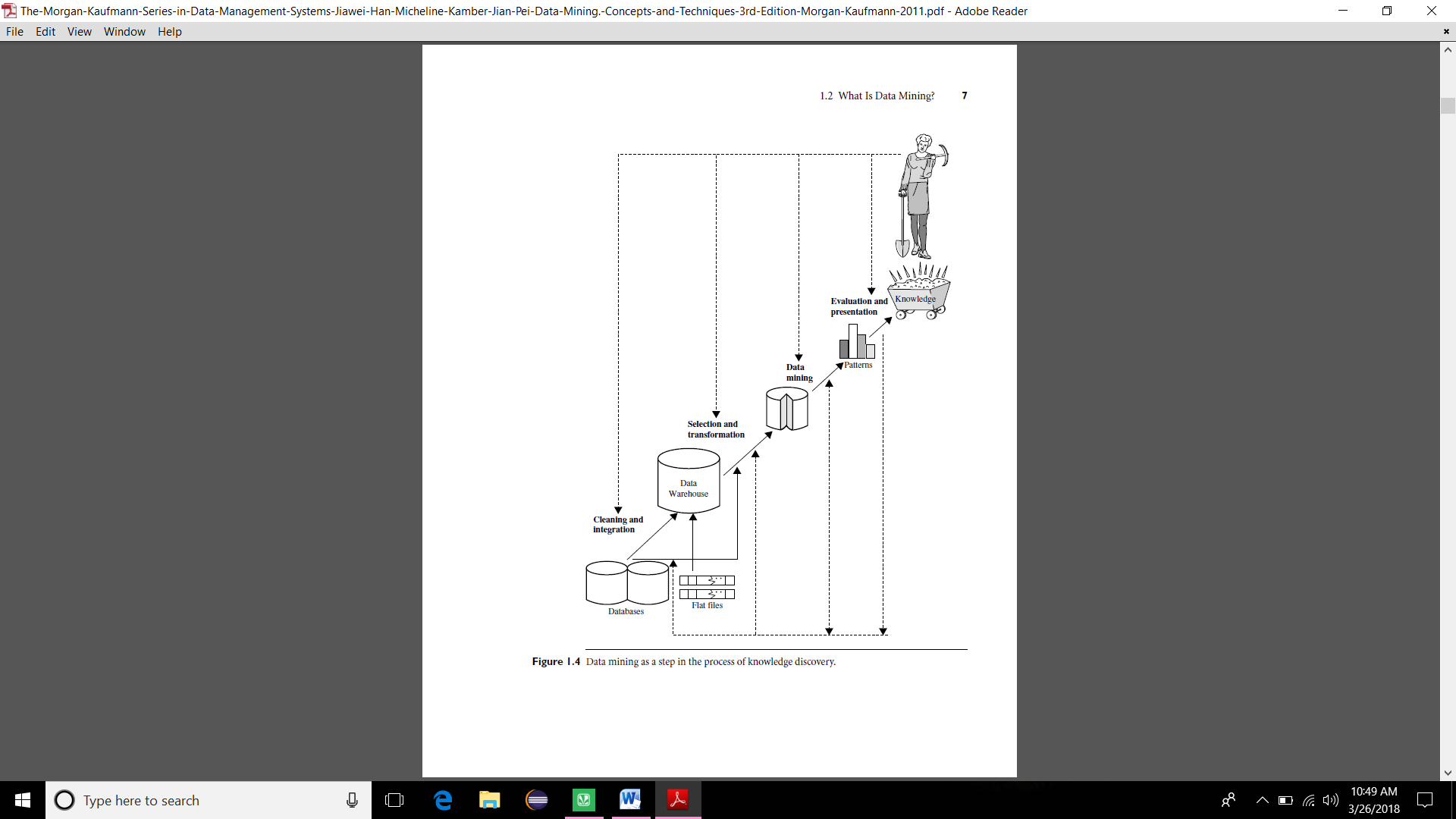


Fig 1.1 Data mining process

The knowledge discovery process is shown in as an iterative sequence of the following steps:

1**. Data cleaning** (to remove noise and inconsistent data)

2. **Data integration** (where multiple data sources may be combined)

3. **Data selection** (where data relevant to the analysis task are retrieved from the database

4. **Data transformation** (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)

5. **Data mining** (an essential process where intelligent methods are applied to extract data patterns)

6. **Pattern evaluation** (to identify the truly interesting patterns representing knowledge)

7**. Knowledge presentation** (where visualization and knowledge representation techniques are used to present mined knowledge to users)

**1.3 Motivation**

Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) is one of the schemes implemented by the government of India in keeping mind that “The soul of India lives in its villages", Most of the people in the rural part of the country are depending on the unskilled manual works for their survival. Many works have been done early in the process of criticizing the positive and negative impacts of this scheme. All works has some limitations such as it lack in the intelligent prediction, and to use suitable technique for the research, that motivates us to take up this specific scheme. This work is analyzing the potholes in this system.

We are motivated to do this project so that our final analysis may result in the proper implementation of this act without any delay of payment,where the workers get to know rejected reasons of delay compensation so that their hard work does'nt go in vain and the officials get to know the reasons behind payment delays and be more careful in future.

**1.4 Problem Statement**

Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) aims to enhance livelihood security in rural areas by providing at least 100 days of wage employment in a financial year to every household whose adult members volunteer to do unskilled manual work. Internal studies conducted on this act resulted in delayed payment of wages. This calls for further steps to improve the system and to assure timely availability of funds.

We strive to derive the average delays in the complete process, sort the states in the order of average delays, percentage share of reason of delay in each district and type of delay in of payment wages.

**1.5 Objective**

In the course of this project we want to achieve a detailed analysis on delayed wage payment in MGNREGA scheme and its compensation using data mining techniques. We tend to focus on the reasons of delay and try to classify them. At the end of this project we try to find out the main reason of delay. We also analyse the different reasons of rejection of delay compensation. The paper also reviews the influence of the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) on the rural India.

**1.6 Organization of dissertation**

**Chapter 2:** The details about the reference paper on this project. Basics of techniques used in this project(Data mining, Weka)

**Chapter 3:** Algorithms used to achieve the objectives of this project.

**Chapter 4:** Analysis of the output along with their screenshots

**Chapter 5:** Conclusions drawnfrom the analysis

**Chapter 6:** Future work that can be done as an extension of this project

**Chapter 7:** References

**1.7 Summary**

Implementation efficiency model to evaluate the rural employment guarantee scheme in India, and evaluates the efficiency within individual states efficiency markings can be done to reflect prevailing local conditions.

**2.RELATED WORK**

**2.1 Literature work**

Most of the work done related to this project is related on how the data is being collected at each level of the country and what are the different methods followed to analyze this project . Many works has been carried out in previous years to study and analyze the success and failure of this scheme. Different methods has been adapted by different authors, each has its own advantages and disadvantages. Analyzing of MGNREGA scheme is a bit tricky work, because so many factors are involved to be considered. Some of the previous works carried out are presented in this section. S. Krishnan et al [2014] [1] analyses this scheme as a India’s policy and programme to eradicate poverty by 2015. Since there are some problems, this is the most successful programme to eradicate poverty in India. It provides money directly in the hands of poor, especially women without any middlemen. This scheme creates visible and invisible effects on women and their saving capacity is increased. Majority of women workers are now able to participate in social activities. Soumya Mohanty [2012] [2] study says that MGNREGA is a good scheme for eradicating unemployment and poverty, and how this scheme has provided alternative source of livelihood which reduces the migration, and makes rural people in villages self sustained. This study examines the impact created in the tribal livelihoods and to what extent this scheme provides justice in developing the livelihoods of poor tribal people in sundargarh district, Odisha. The result of this study shows that a small amount of impact has been created on tribal livelihoods. The core concept of the programme is ruined by its wrong implementation method. This study pointed out that major hurdles for this programme are religion, street biasness, improper co-ordination and defective leadership. Jyoti poonia [2012] [3] this study shows that how most of the schemes until 90’s in India concentrates on the organized sector. MGNREGA stands out of all the previous schemes because it potentially covers the whole country to provide a minimum income for household and also stimulates the local development. Initial findings shows that there has been shift out agriculture in the MGNREGA in Kerala, mainly in respect of women workers; many new persons are started to work under this scheme. This study shows that this scheme has the potential to stimulate the local development and labour market. Abhishek Thakur[2011] [4] The objective of this study is to understand the source of livelihood and impact of this scheme in Seoni District of Madhya Pradesh. This study also tries to understand the daily changing private wage rate due to the implementation of MGNREGA and compare it to the time before the implementation of the scheme. In addition to this, it also attempts to find the change in relationship between labourers and formers by the impact of MGNREGA. The study shows that private wage has been increased, agriculture pattern has changed and relationship between labourers and formers has been altered by the implementation of MGNREGA. Vikas Chaurasia [2014] [5] study shows that many information’s are hidden in the data available in various fields which can be converted into valuable information and knowledge. It is a complicate process to find the meaningful information from it. Data mining is a important method for doing this kind of process. In the recent years, detecting breast cancer using data mining is one of the important research topics taken for research purpose. For calculating the outcome or to analyze the behavior of tumor the breast cancer classification data can be used. It compares three decision trees, for interpreting the result. Retika Khera and Nandini Nayak (2008)[6] in his work on “Women Workers and Perceptions of the National Rural Employment Guarantee Act” examined the socio economic consequences of the NREGA on women workers. He also attempts to understand the perception of the NREGA legislation as reported by workers. Gundeti Ramesh, Dr. T.Krishna Kumar (2009)[7] in their article on “Facet of rural women Empowerment: A study in Karimnagar District in Andhra Pradesh” finds that NREGP has become a beacon of light in the empowerment of rural women and contributed substantially for the increased living and economic conditions by creating equal wages to male and female workers and increasing the minimum wages. The study reveals that 51.6% of 48 the workers are backward class communities, and 46.6% of workers are from S.C Category and the rest are from ST and O.C communities. Ashok Pankaj, Rukmini Tankha (2010) [8]in their paper examines the empowerment effects of the National Rural Employment Guarantee Scheme on rural women in Bihar, Jharkhand, Rajasthan and Himachal Pradesh. The authors argue that women workers have gained from the scheme primarily because of the paid employment opportunity, and benefits have been realised through income-consumption effects, intra-household effects, and the enhancement of choice and capability. Naganagoud S.P. and H.H. Uliveppa (2010)[9] in their article on “Employment Guarantee and Human Rights: Some Observations” considers that It will not be possible to achieve the full potential of NREGA unless the structure for its implementation is more adequately strengthened. Sugapriyan G. and S. Prakasam, (2015)[10] analyzes the Success of MGNREGA in Kanchipuram District, using Data Mining Technique along with the comparison of previous year statistic data provided by the government. The author analyzes the performance and success of this scheme. The study shows that total number of person days generated for Women in the financial year 2012-13 are 14994815, for the year 2013-14 the number of days counts to 15826218 and for the financial year 2014-15 till November the number came to 10588415. Bindiya Narang (2014)[11] made a livelihoods analysis of MGNREGA. The paper based on a field study conducted in Mewat, a backward district of Haryana, analyses the livelihoods context in selected villages and determines the effectiveness of this Act within wider livelihood strategies of rural poor. Farooq Ahmad Ganiee (2014)[12] made an attempt has been made to comprehensively understand the development effort to rebuild the rural life and livelihood on the basis of various secondary data. The author felt that there is a need to critically examine the implementation process of this programme and its impact on livelihood of the rural people. The author concluded that the success of this Act depends upon its proper implementation and in this scenario, the community participation is very important to make this programme more effective. Ragendra Narayan[13] made an attempt to analyze how the delay analysis are being processed in the fianancial year 2016-2017.he conducted his survey among ten states.he came to know about the flawe methods of delays compensation calculation ,lack of clarity in the accountability of payments.methodology used by him is of simple strategy and data procurement and data anmalysis are involved,his outcomes from this project consists of findings in the partial delays captured and rejected payment. Neha Tiwari and Rajshree Upadhyay (2012)[14] conducted a study o find out constraints faced by the women beneficiaries under Mahatma Gandhi National Rural Employment Guarantee Act. The sample consisted of 100 randomly selected respondents from two panchayat samities. The author used Personal interview technique for collecting data. NIRD (2013) [15]considered that The National Rural Employment Guarantee Act is a landmark initiative in providing 100 days guaranteed employment. MGNREGA has several gender-sensitive features that are attractive for women workers. But even then the national average number of days of employment in MGNREGS is still less than 50 days and it varies 58 across different regions and states.

**3.PROPOSED METHOD**

**3.1 General Decision tree algorithm**

The algorithm is called with three parameters: D, attribute list, and Attribute selection method. We refer to D as a data partition. Initially, it is the complete set of training tuples and their associated class labels. The parameter attribute list is a list of attributes describing the tuples. Attribute selection method specifies a heuristic procedure for selecting the attribute that “best” discriminates the given tuples according to class. This procedure employs an attribute selection measure such as information gain or the Gini index. Whether the tree is strictly binary is generally driven by the attribute selection measure. Some attribute selection measures, such as the Gini index, enforce the resulting tree to be binary. Others, like information gain do not, therein allowingmultiway splits (i.e., two or more branches to be grown from

a node).

The tree starts as a single node, N, representing the training tuples in D (step 1).

**Algorithm**: Generate decision tree. Generate a decision tree from the training tuples of

data partition, D.

Input:

Data partition, D, which is a set of training tuples and their associated class labels; attribute list, the set of candidate attributes; Attribute selection method, a procedure to determine the splitting criterion that “best” partitions the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split-point or splitting subset.

Output: A decision tree.

Method:

STEP 1: create a node N;

STEP 2: if tuples in D are all of the same class, C, then

STEP 3: return N as a leaf node labeled with the class C;

STEP 4: if attribute list is empty then

STEP 5: return N as a leaf node labeled with the majority class in D; // majority voting

STEP 6: apply Attribute selection method(D, attribute list) to find the “best” splitting criterion;

STEP 7: label node N with splitting criterion;

STEP 8: if splitting attribute is discrete-valued and

multiway splits allowed then // not restricted to binary trees

STEP 9: attribute list attribute list splitting attribute;

// remove splitting attributes

STEP 10: for each outcome j of splitting criterio

// partition the tuples and grow subtrees for each partition

STEP 11: let Dj be the set of data tuples in D satisfying outcome j; // a partition

STEP 12: if Dj is empty then

STEP 13: attach a leaf labeled with the majority class in D to node N;

STEP 14: else attach the node returned by Generate decision tree(Dj , attribute list) to node N; endfor

STEP 15: return N;

If the tuples in D are all of the same class, then node N becomes a leaf and is labeled with that class (steps 2 and 3). Note that steps 4 and 5 are terminating conditions. All terminating conditions are explained at the end of the algorithm.

Otherwise, the algorithm calls Attribute selection method to determine the splitting criterion. The splitting criterion tells us which attribute to test at node N by determining the “best” way to separate or partition the tuples in D into individual classes (step 6). The splitting criterion also tells us which branches to grow from node N with respect to the outcomes of the chosen test. More specifically, the splitting criterion indicates the splitting attribute and may also indicate either a split-point or a splitting subset. The splitting criterion is determined so that, ideally, the resulting partitions at each branch are as “pure” as possible. A partition is pure if all the tuples in it belong to the same class. In other words, if we split up the tuples in D according to themutually exclusive outcomes of the splitting criterion, we hope for the resulting partitions to be as pure as possible. The node N is labeled with the splitting criterion, which serves as a test at the node (step 7). A branch is grown from node N for each of the outcomes of the splitting criterion. The tuples in D are partitioned accordingly (steps 10 to 11). There are three

possible scenarios. Let A be the splitting attribute. A has v distinct values, fa1, a2, : : : , avg, based on the training data.

1. A is discrete-valued: In this case, the outcomes of the test at node N correspond directly to the known values of A. A branch is created for each known value, aj , of A and labeled with that value Partition Dj is the subset of class-labeled tuples in D having value aj of A. Because all the tuples in agiven partition have the same value for A, A need not be considered in any future partitioning of the tuples. Therefore, it is removed from attribute list (steps 8 and 9).

2. A is continuous-valued: In this case, the test at node N has two possible outcomes,corresponding to the conditions A \_ split point and A > split point, respectively,where split point is the split-point returned by Attribute selection method as part of the splitting criterion. (In practice, the split-point, a, is often taken as the midpoint of two known adjacent values of A and therefore may not actually be a preexisting value of A from the training data.) Two branches are grown from N and labeled according to the previous outcomes The tuples are

partitioned such that D1 holds the subset of class-labeled tuples in D for which A \_ split point, while D2 holds the rest.

3. A is discrete-valued and a binary tree must be produced (as dictated by the attribute selection measure or algorithm being used): The test at node N is of the form“A 2 SA?,” where SA is the splitting subset for A, returned by Attribute selection method as part of the splitting criterion. It is a subset of the known values of A. If a given tuple has value aj of A and if aj 2 SA, then the test at node N is satisfied. Two branches are grown fromN By convention, the left branch out of N is labeled yes so that D1 corresponds to the subset of class-labeled tuples in that satisfy the test. The right branch out of N is labeled no so that D2 corresponds to the subset of class-labeled tuples from D that do not satisfy the test.

The algorithm uses the same process recursively to form a decision tree for the tuples at each resulting partition, Dj , of D (step 14).

The recursive partitioning stops only when any one of the following terminating conditions is true:

All the tuples in partition D (represented at node N) belong to the same class (steps 2 and 3).There are no remaining attributes on which the tuples may be further partitioned (step 4). In this case, majority voting is employed (step 5). This involves converting node N into a leaf and labeling it with the most common class in D. Alternatively, the class distribution of the node tuples may be stored.There are no tuples for a given branch, that is, a partition Dj is empty (step 12). In this case, a leaf is created with the majority class in D (step 13).

The resulting decision tree is returned (step 15).

The computational complexity of the algorithm given training set D is O.n\_jDj\_ log.jDj//, where n is the number of attributes describing the tuples in D and jDj is the number of training tuples in D. This means that the computational cost of growing a tree grows at most n\_jDj\_log.jDj/ with jDj tuples. The proof is left as an exercise for the reader.

Incremental versions of decision tree induction have also been proposed. When given new training data, these restructure the decision tree acquired from learning on previous training data, rather than relearning a new tree from scratch.

A decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. The internal nodes of a decision tree denote the different attributes, the branches between the nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes tell us the final value (classification) of the dependent variable.

The attribute that is to be predicted is known as the dependent variable, since its value depends upon, or is decided by, the values of all the other attributes. The other attributes, which help in predicting the value of the dependent variable, are known as the independent variables in the dataset.

Classification is the process of building a model of classes from a set of records that contain class labels. Decision Tree Algorithm is to find out the way the attributes-vector behaves for a number of instances. Also on the bases of the training instances the classes for the newly generated instances are being found [15]. This algorithm generates the rules for the prediction of the target variable. With the help of tree classification algorithm the critical distribution of the data is easily understandable [5]. J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java implementation of the C4.5 algorithm. The WEKA tool provides a number of options associated with tree pruning. In case of potential over fitting pruning can be used as a tool for précising. In other algorithms the classification is performed recursively till every single leaf is pure, that is the classification of the data should be as perfect as possible. This algorithm it generates the rules from which particular identity of that data is generated. The objective is progressively generalization of a decision tree until it gains equilibrium of flexibility and accuracy

The J48 Decision tree classifier follows the following simple algorithm. In order to classify a new item, it first needs to create a decision tree based on the attribute values of the available training data. So, whenever it encounters a set of items (training set) it identifies the attribute that discriminates the various instances most clearly. This feature that is able to tell us most about the data instances so that we can classify them the best is said to have the highest information gain. Now, among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for the target variable, then we terminate that branch and assign to it the target value that we have obtained.

For the other cases, we then look for another attribute that gives us the highest information gain. Hence we continue in this manner until we either get a clear decision of what combination of attributes gives us a particular target value, or we run out of attributes. In the event that we run out of attributes, or if we cannot get an unambiguous result from the available information, we assign this branch a target value that the majority of the items under this branch possess.

Now that we have the decision tree, we follow the order of attribute selection as we have obtained for the tree. By checking all the respective attributes and their values with those seen in the decision tree model, we can assign or predict the target value of this new instance.

Differences in decision tree algorithms include how the attributes are selected in creating the tree and the mechanisms used for pruning . The basic algorithm described earlier requires one pass over the training tuples in D for each level of the tree. This can lead to long training times and lack of available memory

when dealing with large databases. Improvements regarding the scalability of decision tree induction are discussed in presents a visual interactive approach to decision tree construction. A discussion of strategies for extracting rules from decision trees is given in regarding rule-based classification.

**3.2 FLOW CHART**

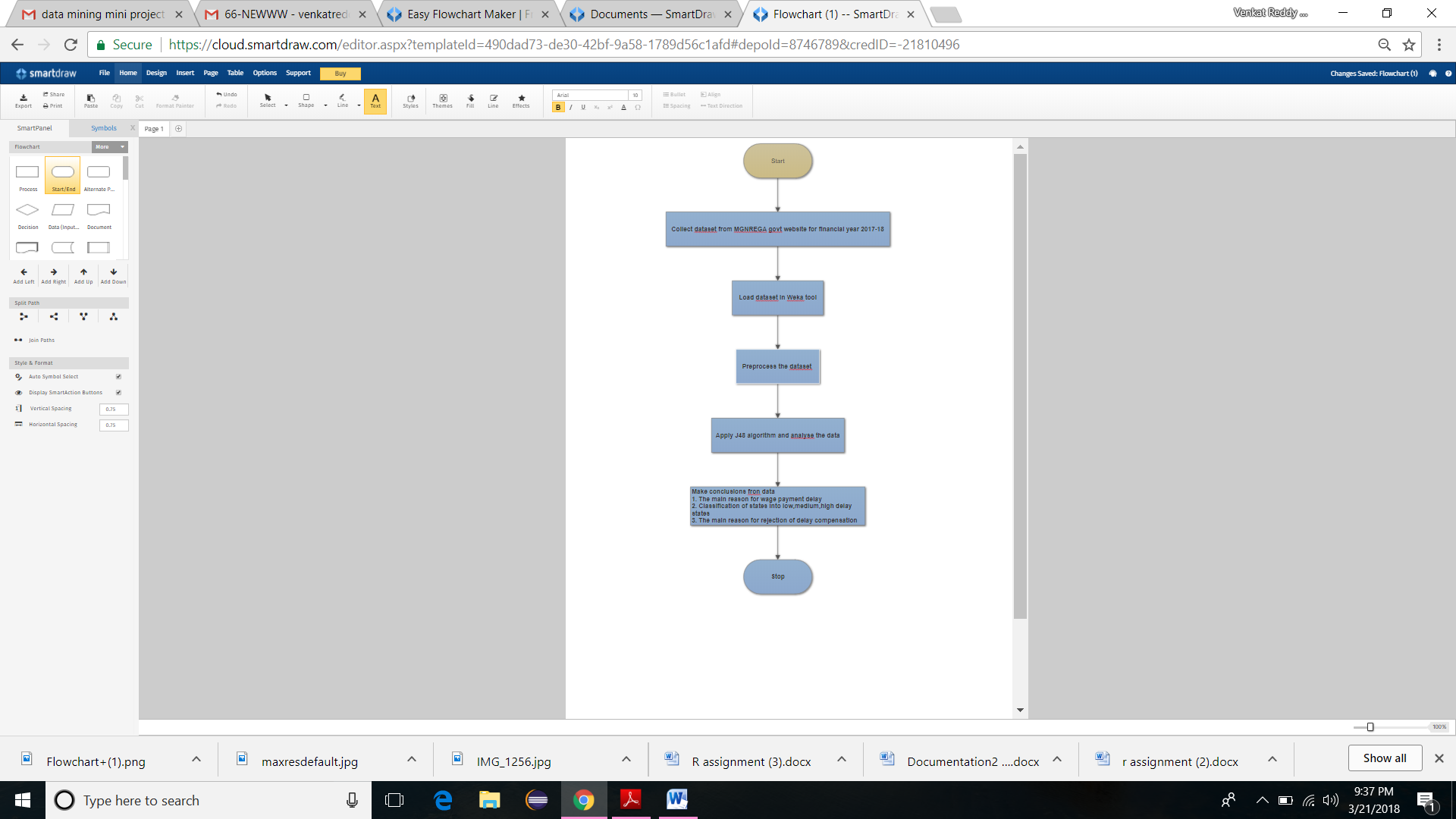


Fig 3.1 Process flow of flowchart

In the reference with above figure 2.1 deals with the brief process flow of our project.firstly we have to take the input dataset from government officials from their website and then comes dealing with the dataset that is loading the data into weka tool,which is the basic tool for data mining and then we have to preprocess the data so as to make it ready to do classification and other visualisations.For doing the preprocessing we will come across missing values and other aspects then we have to use filters to remove these missing values .the type of filter we use depends on the attribute type .After applying all the filters then missing values will be removed and we get a clean data.

For classification purpose we have to choose a algorithm to make our data set into understandable structure,we are using J48 decision tree algorithm to classify the dataset .Here we have to choose the training and testing percentages and based on the percentage and also on the algorithm we use we will have a classified output and when we visualize we will get a classification tree .Here we are using J48 algorithm because of the properties to obtain a classification tree.

**3.3 ALGORITHM INVOLVED IN DOING THE PROJECT:**

STEP 1: Start the project by taking the dataset from MGNREGA official website for the financial year 2017-18.Actually we took three datasets related to delay payments in the corresponding years.first one is related on delay payments in the respective year in the states off the country,second one relates to the delay reasons for the payments and the last dataset is related to reasons of rejection of delay .

STEP2: Now,load the dataset in weka tool,which is the generally used for all the data mining purposes,data can be loaded either in CSV or ARFF format.there are many ways to load the data like directly opening or by using URL .As we know that weka tool is the best in applying data mining algorithms ,weka provides many facilities in opening a file and the loading of the dataset can also be done in different formats ,here in our project we used csv format an opened our respective datasets in the explorer tab of the weka tool where we have predefined methods for data mining.

STEP3:Preprocess the data using various filters to fill the missing values and totally preprocess the raw data for relailable classification. Different tasks are involved in data preprocessing ,there are mainly five steps included mainly :

* **Data cleaning:**fill in the missing values,smooth noisy data,identify or remove outliers,and resolve inconsistencies.
* **Data integration**:uses multiple databases or files.
* **Data transformation**:normalization and aggregation.
* **Data reduction**:reducing the volume but producing the same or similar analytical results.
* **Data discretization**:part of data reduction,replacing numerical attributes with nominal values.

In our project we loaded the data and data cleaning ,data reduction and data discretization techinques are followed to preprocess the raw data into reliable data set.

STEP4:In the classification step ,we apply the algorithms suitable for the dataset.here we are using desicion tree algorithm and using j48 algorithm to classify the dataset to tree which will be in understandable structure.Classification can be done on different attributes,it depends on our dataset.we can apply different algorithms on the dataset but in our project we are using j48 decision tree method so that our complex data set can be analysuzed into a clearly understandable tree.this result can also be visualized in different graph forms also.

STEP5:We have to arrive at three conclusions mainly the first one is to find the major reason for the delay of wage payment to the employess working under this act in the res[pective time period.The second one is tp find the major reason for rejection for delay compensation .The third conclusion is to classify the states in high ,low and medium delay payment states.

**4.OUTPUTS**

**4.1 Classification of states based on percentage delayed**

Fig 4.1-percentage delays of different states

1. Andhra Pradesh 17. Meghalaya

2. Arunachal Pradesh 18. Mizoram

3. Assam 19. Nagaland

4. Bihar 20. Odisha

5. Chhattisgarh 21. Punjab

6. Goa 22. Rajasthan

7. Gujarat 23. Sikkim

8. Haryana 24. Tamil Nadu

9. Himachal Pradesh 25. Telangana

10. Jammu and Kashmir 26. Tripura

11. Jharkhand 27. Uttar Pradesh

12. Karnataka 28. Uttarakhand

13. Kerala 29. West Bengal

14. Madhya Pradesh 30. Andaman and Nicobar

15. Maharashtra 31. Lakshadweep

16. Manipur

The above graph is a barplot between the different states and the percentage delay in each state. States like Chattisgarh, Himachal Prades, Jharkhand, Rajasthan, Tamil Nadu, Uttarakhand. have very low delay percentage i.e. <10% . States like jammu and kashmir, Mizoram, Nagaland, Andaman and Nicobar have a very high percentage delay i.e. >80% .

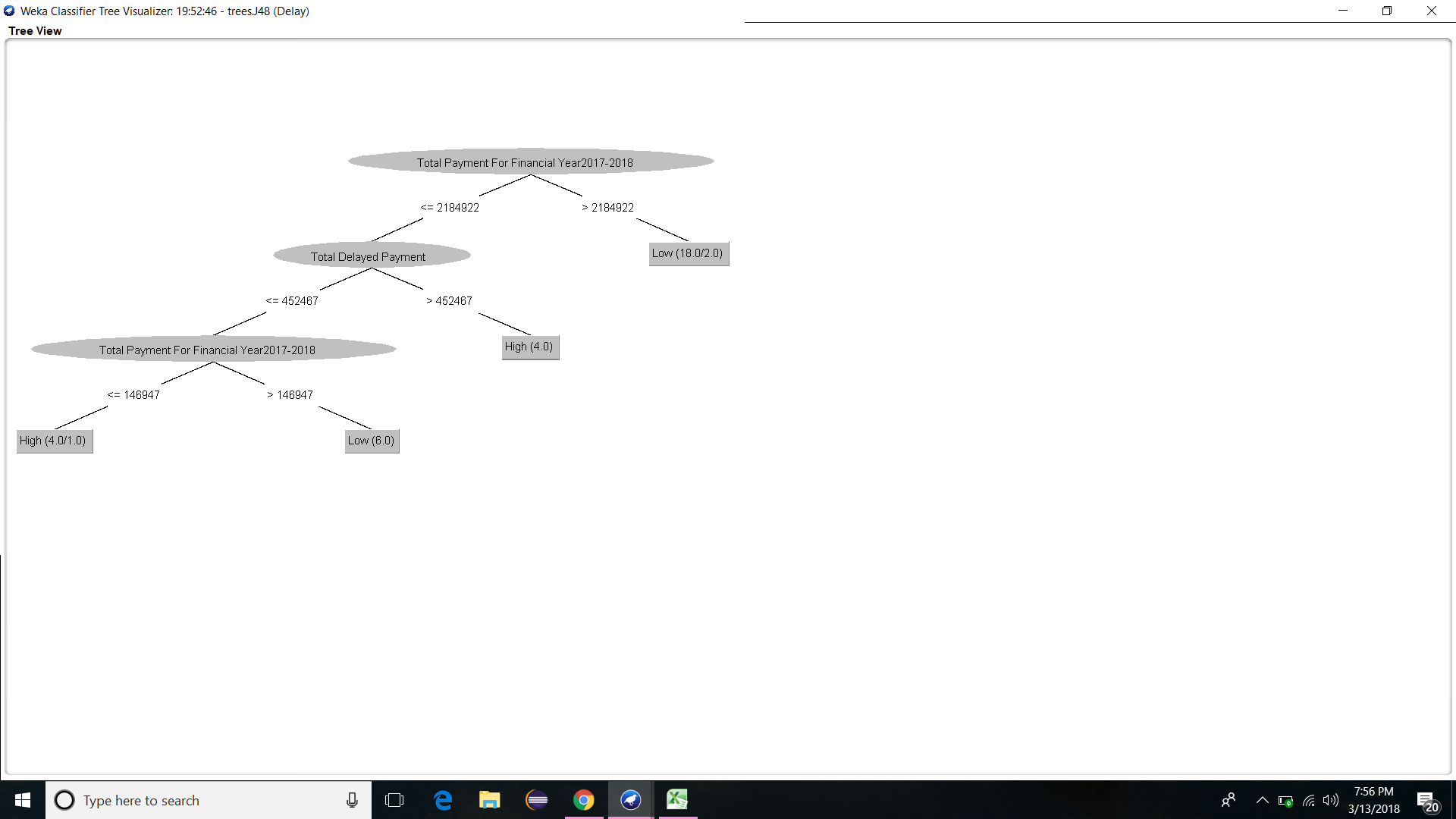


Fig 4.2 classification tree of delay payments

The first data set that we are mining is to find the basis of classification of states into low and high delay states on the basis of Total delayed payment (The amount that has been delayed and have to be payed) and Total payment

( The wages to be given/have been given to the workers).

After classification of data using J48 algorithm the above figure is the generated tree. It is easily understood from the figure that if the total payment of the state is greater than Rs21,84,922 then the state is more likely to have low delayed payment perentage. The next observation is that if the total payment of the state is less than the above amount and total delayed payment is greater than Rs4,52,467 then the state is likely to have high delay.

By using the above classification tree we can try to anticipate the delay in the state by looking at the above details.

**4.2 Reason for delay**

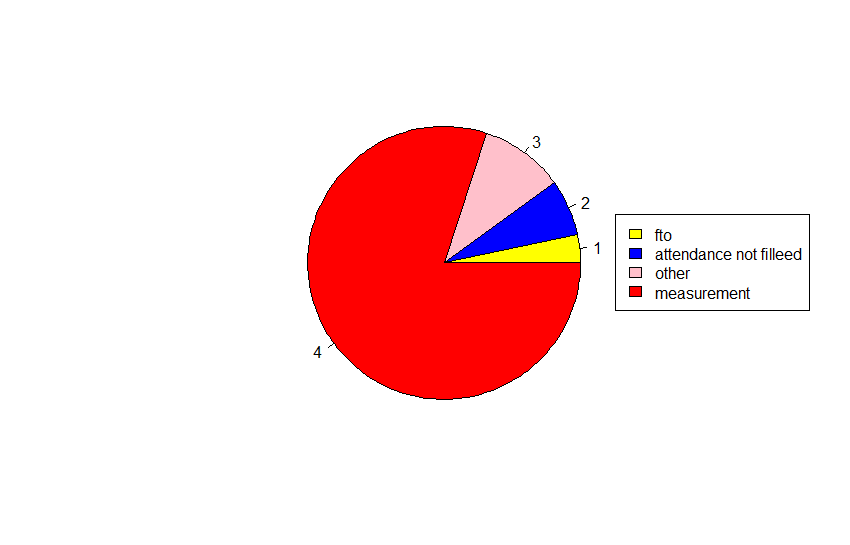


Fig 4.3frequencies of reasons of delay

The next data set that we are mining is to find out the reasons of delay. the first figure is a pie chart showing the frequency of each reason. The reasons include Measurement book not being filled in time, Attendance not being signed properly by the workers and fund transfer order delay. Measurement book records the progress of the work on a daily basis by technically qualified officials which on not being filled on time will cause delay in accepting and sending of fund by the government. Attendance book records the presence of the worker on daily basis which on not being filled the government cannot accept the wages. Fund transfer order must be given from higher officials to release the wages which when delayed causes delay of the wage payment.

Evidently the major reason of the delay in the payment of wages is untimely filling of Measurement books. this may be due to neglegence or unavailability of qualified technical persons.

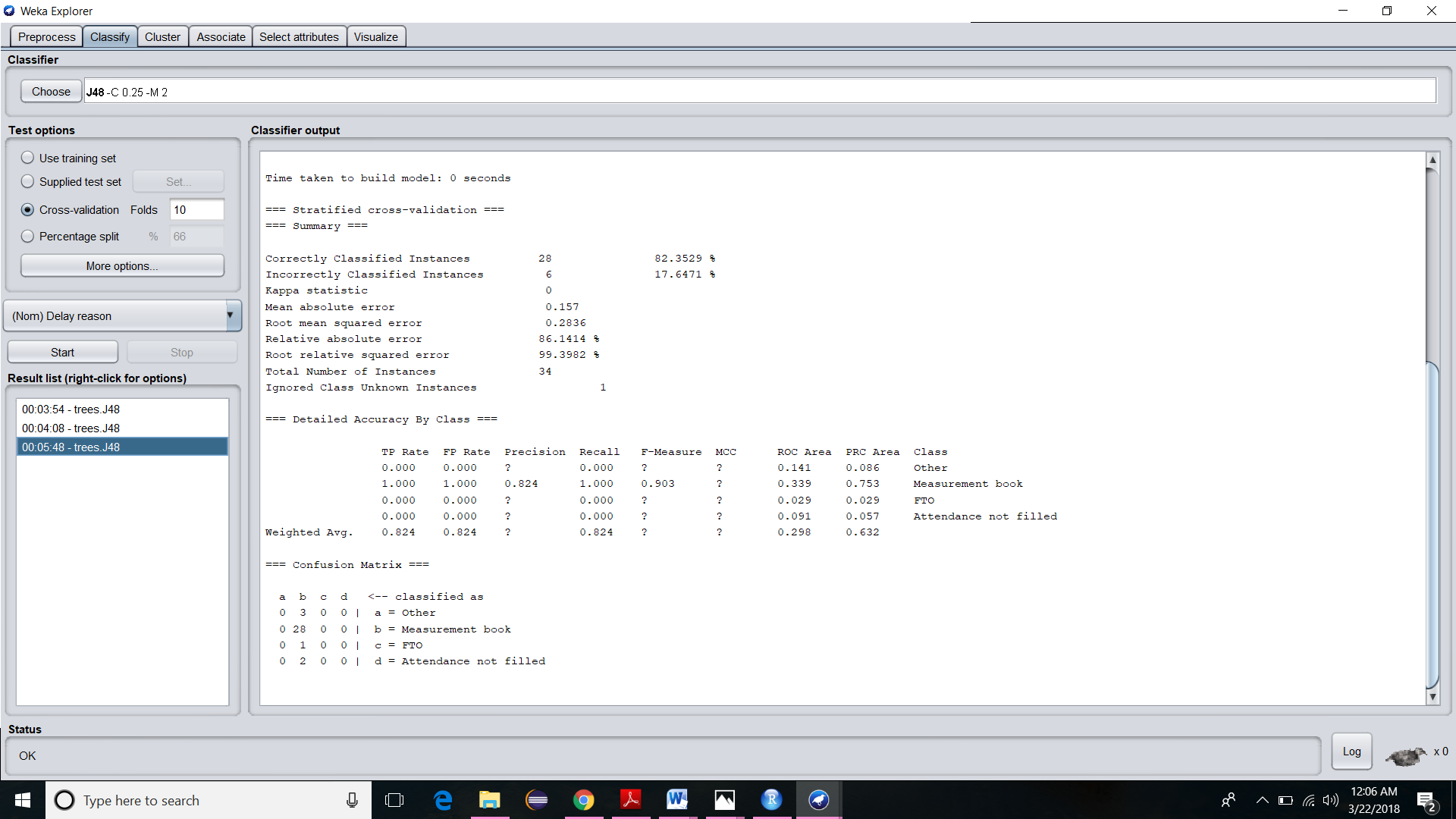


Fig 4.4 Confusion matrix for classification

The data set on classification using J48 algorithm produces the above confusion matrix. As it clearly understood, Measurement book, FTO, Attendance not filled and Others are the class labels and the states are classified under the respective class labels having frequency 28,1,2 and 3.

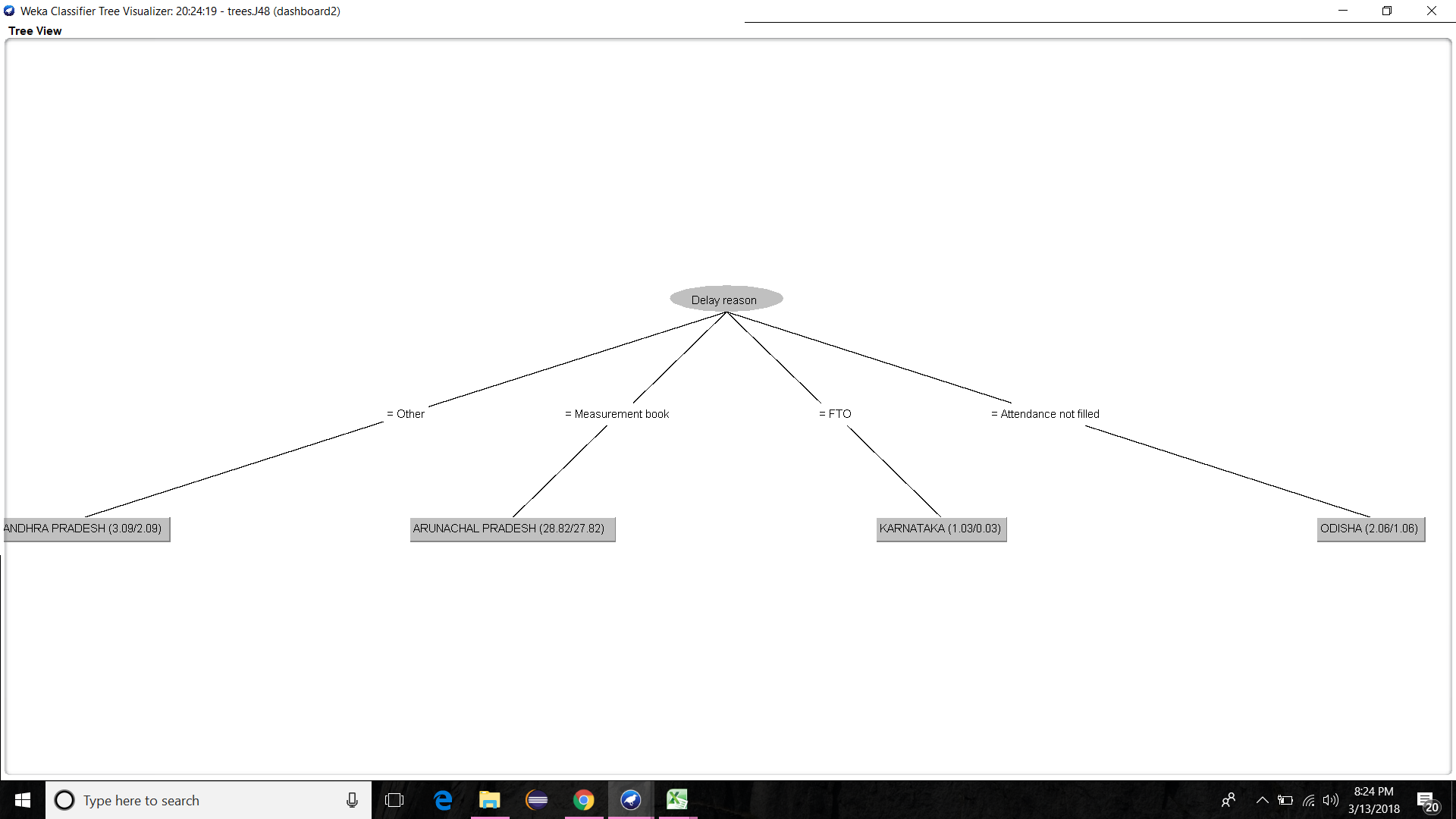


Fig 4.5 delay reasons classification tree

The classification tree generated while classifying the states with respect to the various reasons for delay is shown above

**4.3 Reasons for rejection of delay compensation**

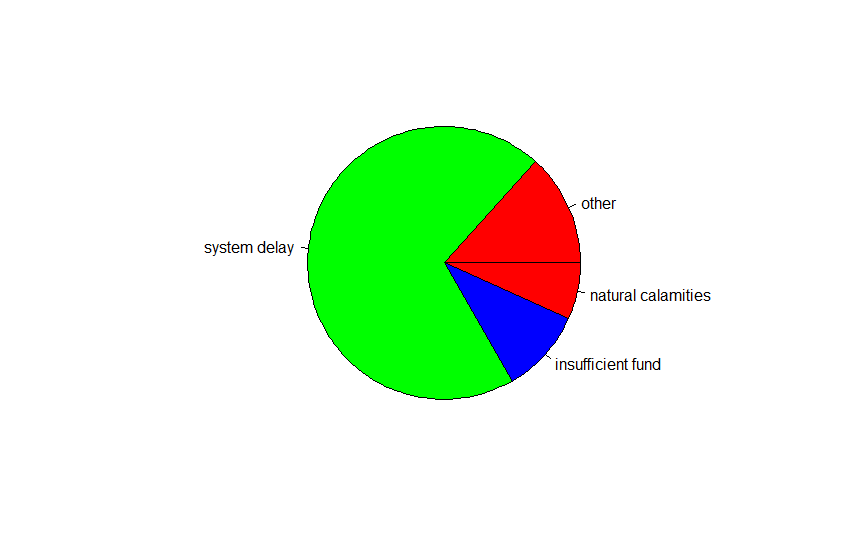


Fig 4.6frequency of reason of rejection

The next data set that we are mining is to find out the reasons for rejection of delay compensation. The first figure is a pie chart showing the frequency of each reason. The reasons include natural calamities, insufficient fund and system delay. When natural calamities occur the government cannot be held responsible for delay of wage payment so the compensation is rejected. When there is a breakdown in server or the database(System delay) the delay caused in payment may be exempted causing the rejection of compensation. When there is no sufficient fund to pay the wages the delay is not compensated.

Evidently the major reason for rejection of delay compensation is system delay

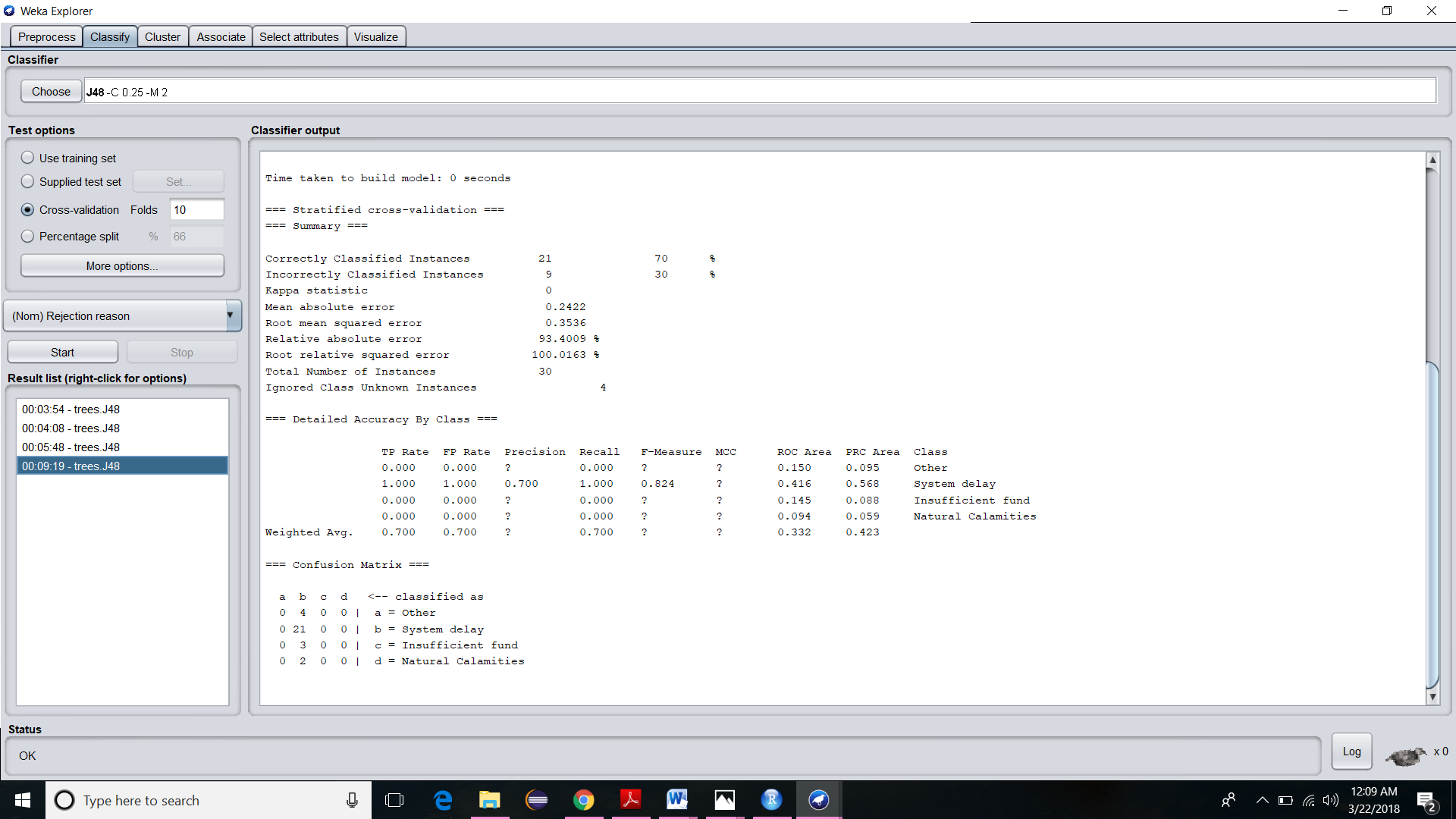


Fig 4.7Confusion matrix for the classification

The data set on classification using J48 algorithm produces the above confusion matrix. As it clearly understood, System delay, Insufficient fund, Natural calamities and Others are the class labels and the states are classified under the respective class labels having frequency 21,3,2and 4.

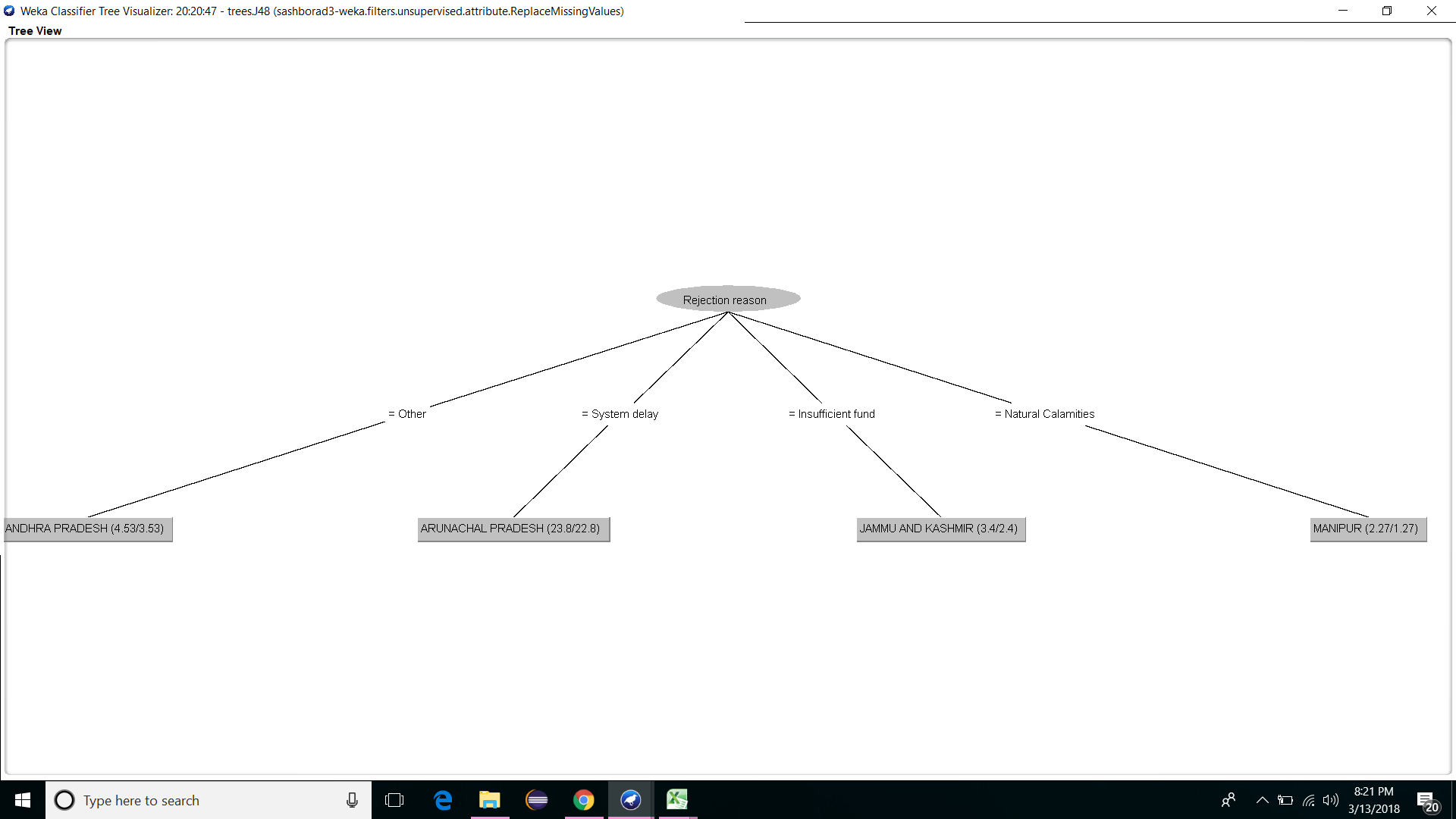


Fig 4.8clasification tree of rejection reason

The classification tree generated while classifying the states with respect to the various reasons for rejection of delay compensation is shown above.

**5. CONCLUSION**

From the analysis of the outputs we arrive at three different conclusions. The first conclusion is the delay payments of different states are analyzed using a classification tree. The second conclusion is the major reason of delay is found out to be delayed authentication of measurement book. Regular measurement and supervision of works has top be done by qualified technical personal on time and they need to recorded in measurement books. There may be many reasons in delay recording of the measurements due to carelessness, lack of time, lack of professionals. Due to the delayed recording the payments are also delayed. The third conclusion is that the main reason for the rejection of delay compensation is found out to be system delay. System delay is primarily due to crashing of server ion the government office, incorrect account numbers in the system, mismatch of names in the account and aadhar card.For such payments fund transfer has to be regenerated which could take several days causing delay of payments. The current payment system is completely centralized and the state governments cannot pay the workers even if they intend too.

It is the time the government digs deeper into how it csn better utilize its workforse budget and show accountability for the payments it is rolling out in an effort to create value from money and human resources.

**6. FUTURE WORK**

In the project done we are dealing with the statistics of the states considered as one. But as an extension we strive to analyze the delay payments from mandal level to district level. The reasons have to be brought to limelight so that the officials and workers will get awareness of the loopholes of the system. Firstly all those who register do not come to work and the quantum of work is not measured properly. That’s why we insist working in the presence of an officer and maintaining regular attendance and measurement records. There is undoubtedly a great value that MGNREGA is able to add in the country in terms of providing jobs to desperately needy people. But it will be a good idea if the daily wage workers can also be trained in using computers so that they can be hired to provide services like photo copy, printing, scanning.

**REFERENCES**

[1] S.Krishnan, Dr. A.Balakrishnan (2014), MGNREGA Marching towards achieving the millennium development goals-an analysis. Journal of International Academic Research for Multidisciplinary. Volume 2.

[2] Soumya Mohanty, Dr. Nihar Ranjan Mishra(2012), Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) and Tribal Livelihoods: A Case Study in Sundargarh District of Odisha.

[3] Jyoti poonia(2012), Critical Study of MGNREGA: Impact and women’s participation. International Journal of Human Development and Management Sciences, Vol. 1 No. 1.

[4] Abhishek Thakur, Dr. Manish K. Jha(2011), A Study on MGNREGA and its impact on wage and work relation.

[5] Vikas Chaurasis, Saurabh Pal (2014), A Novel Approach for Breast Cancer Detection Using Data Mining Techniques. Internation Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1.

[6]Retika Khera and Nandini Nayak, “Women Workers and Perceptions of the National Rural Employment Guarantee Act” Economic and Political Weekly, Vol. XLIV, No.43, October 24, 2009, PP. 49-57.

[7] Dutta A.K., A.K and Prosty. S (2009), “An appraisal of NREGA in the States of Meghalaya and Sikkim” http://www.nrega.net/pin/ reportsand-resources/reports-submitted-to-the-ministry-of-ruraldevelopment /reports-28-jan-2010/NREGA-IIMShillong-Final%20Report.pdf.

[8] Gundeti Ramesh, Dr T. Krishna Kumar, “Facet of Rural Women Empowerment: A Study in Karimnagar District in Andhra Pradesh” Kurukshetra, Vol. 58, No.2, December, 2009, pp.29-30.

[9] Ashok Pankaj and Rukmini Tankh, “Empowerment Effects of the NREGS on Women Workers: A Study in Four States”, Economic & Political Weekly, Vol. XLV, No 30, July 24, 2010, pp.45-55.

[9] Richard Mahapatra (2010), ‘How Women Seized NREGA’ http:// www.downtoearth.org.in/ node/2282.

[10]Sugapriyan G. and S. Prakasam, “Analyzing the Performance of MGNREGA Scheme using Data Mining Technique”, International Journal of Computer Applications, Vol. 109, No. 9, January 2015, pp.11-15.

[11] Bindiya Narang, “Ensuring Rural Livelihood Security through MGNREGA: A Study in District Mewat, Haryana”, International Journal of Social Science: Vol. 3. No. 2, 133-154, June 2014, pp.133-154.

[12]Farooq Ahmad Ganiee, “ A Case Study of Rural Development Programmes In India”, International Journal of English Language, Literature and Humanities, Vol.1, Issue V, February 2014, pp.40-47.

[13]Rajendran Narayanan1 , Sakina Dhorajiwala2 , and Rajesh Golani3” Analysis of Payment Delays and Delay Compensation in NREGA: Intermediate Findings of an Ongoing Study Across Ten States for Financial Year 2016-17”

[14] Neha Tiwari and Rajshree Upadhyay , Constraints Faced by the Women Beneficiaries under Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), Stud Home Com Sci,Vol. 6, No.2, 2012, pp. 99-102 .