Analysis of Mahatma Gandhi National Rural Employment Guarantee Act Using Data Mining Technique

Kritika Yadav

CSE/IT Department,
Madhav Institute of Technology and Science, Gwalior, India.

Mahesh Parmar

CSE/IT Department
Madhav Institute of Technology and Science, Gwalior, India.

Abstract

Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) aims at livelihood security of people in rural areas by guaranteeing hundred days of wage-employment in a financial year to a rural household whose adult members volunteer to do unskilled labour work. The Mahatma Gandhi NREGA sponsors various schemes for helping rural people below the poverty-line for creation of wage employment and productive assets. The internal studies conducted on the reasons for the delayed payments pointed out that the delays in release of funds by the Central Government, multi-level release system, continued parking of funds at various levels and the inability of the implementation agencies to get the funds in time for payment - were the main contributory causes for the increased delays. This calls for further steps to improve the system and to assure timely availability of funds as per demand. This paper gives the analysis of the payment of wages to the workers under MGNREG scheme in districts of Rajasthan, using decision tree J48 classification technique.

Keywords: data mining; MGNREGA; delay in payment of wages;decision tree ;J48

I. INTRODUCTION

Data mining field uses many methods to extract the needed hidden data and hidden patterns from big data [1]. Data Mining is one of the disciplines that are used to convert raw data into meaningful information and knowledge [2]. Data mining searches and analyses large quantities of data automatically by discovering, learning and knowing hidden patterns, trends, and structures and it answers questions that cannot be addressed through simply query and reporting techniques [2].

Data Mining is a very crucial research domain in recent research world. The techniques are useful to elicit significant and utilizable knowledge which can be perceived by many individuals. Data mining programs consists of diverse methodologies which are predominantly produced and used by commercial enterprises, Government offices and biomedical researchers. These techniques are well disposed towards their respective knowledge domain. The use of standard statistical analysis techniques is both time consuming and expensive. Efficient techniques can be developed and tailored for solving complex data sets using data mining to improve the effectiveness and accuracy of large data sets [3].

The Government of India has introduced many employment generation programmes to eradicate poverty and unemployment, since in 1980. All these programmes were inadequate and piecemeal in their approach. Therefore, the programmes failed to make any major dent on the problems of poverty and unemployment. With globalization and liberation of the economy, it is always feared that the incidence of poverty and unemployment will increase substantially. In this context, the implementation of National Rural Employment Guarantee Act is the most appropriate course of action.

The MNREGA that aims to cover all of rural India with a potential socio-political significance for the rural poor that are matched only by the 73rd Amendment. One version of the proposed MNREGA bill seeks to provide "at least one hundred days of guaranteed employment at the statutory minimum wage" to adult members of every rural household who volunteer to do casual manual work [4].

As per the Section 22(1) of the Mahatma Gandhi National Rural Employment Guarantee Act 2005, the Central Government is mandated to meet the cost of the wages for unskilled manual work under the Scheme, and upto three-fourths of the material cost of the scheme including the payment of wages to the skilled and semi-skilled workers, and the administrative expenses as decided by the Central Government (currently at 6%).

In order to streamline the system of fund releases and to avoid multiple levels of fund release and thereby do away with the delays and corruption, an electronic Fund Management System (e-FMS), has been introduced in MGNREGA. Under this system, funds are held at one account at the State level (e-FMS Debit account) which

is electronically linked to all implementing levels. The implementing agency (Gram Panchayat/ Block), after due verification of the work and the muster rolls, generates an electronic Fund Transfer Order (FTO) to transfer the wages direct into the beneficiary accounts duly debiting the State level account. This electronic advice allows transfer of wages within 2 working days (T+2) into the accounts of the beneficiaries.
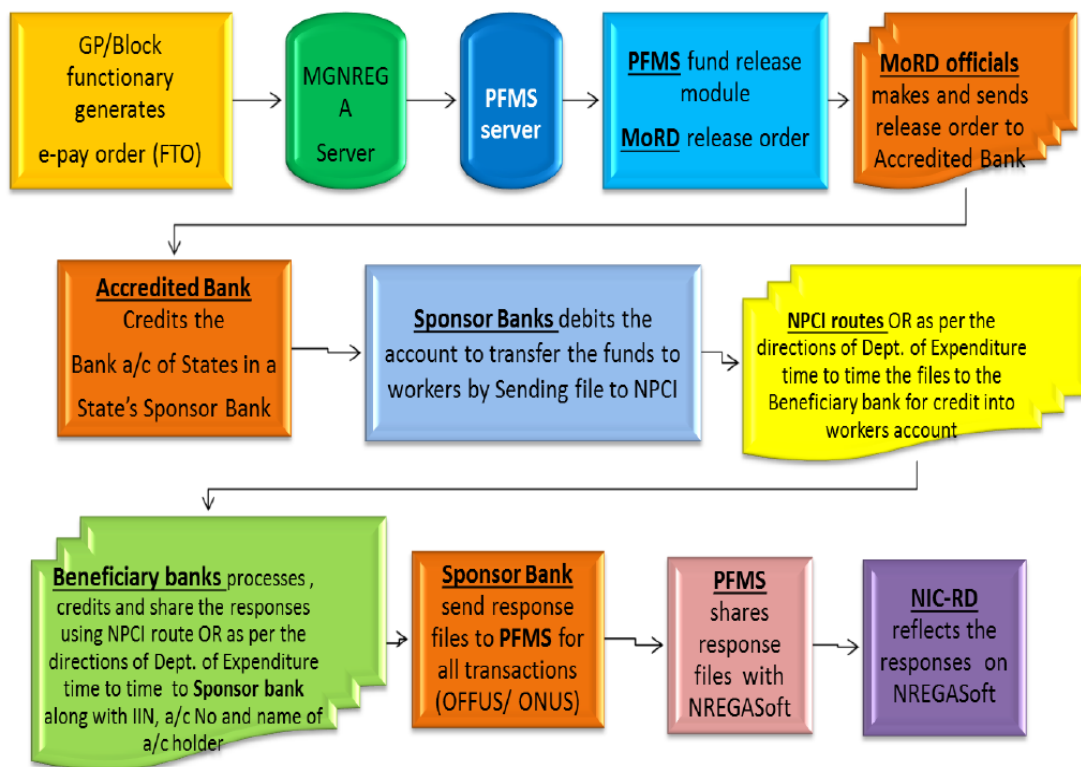

Fig. Detailed workflow of National Electronic Fund Management Systems (NeFMS).

This paper is organized as follows. In Section II, we introduce the dataset and attributes in it, and how the data was collected and pre-processed. It also lists and explains the selected classification algorithms. Section III outlines the results obtained by using two different test methods and also the dataset is analyzed on different criteria's giving us insight on trends and patterns of incidents that have occurred in the due course. Section V concludes the paper.

II. RELATED WORK

P. Sumithra and V. Valli Kumari [2015] analyze the performance of MGNREG scheme in villages of Visakhapatnam district, using distance weighted k-nearest neighbor classification technique. The paper also gives the comparison of previous year statistical data provided by the government.

G. Sugapriyan and S. Prakasam [2015] analyses 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 aim of this work is to analyze the performance and success of this scheme.

G. Chandra [2015] studies the Mahatma Gandhi National Rural Employment Guarantee Act and its impact on the Indian society and analyses the corruption involved in the implementation of the act.

Vrushali Bhuyar [2014] In this paper one of the parameter which is used to increase yield production is considered; that is soil. Different classification algorithms are applied to soil data set to predict its fertility. This paper focuses on classification of soil fertility rate using J48, Naïve Bayes, and Random forest algorithm.

Niketa Gandhi and Leisa J. Armstrong [2016] examines the application of data visualisation techniques to find correlations between the climatic factors and rice crop yield. The study also applies data mining techniques to extract the knowledge from the historical agriculture data set to predict rice crop yield for Kharif season of Tropical Wet and Dry climatic zone of India.

Amit Gupta et al. [2016] highlight the trends of incidents that will in return help security agencies and police department to discover precautionary measures from prediction rates. The classification of algorithms used in this study is to assess trends and patterns that are assessed by BayesNet, NaiveBayes, J48, JRip, OneR and Decision Table. The output that has been used in this study, are correct classification, incorrect classification, True Positive Rate (TP), False Positive Rate (FP), Precision (P), Recall (R) and F-measure (F).

Dr.M.Usha Rani [2012] analyzed the caste-wise households registered and working and out of these the registered households are collected for all 22 districts of Andhra Pradesh from 2006 to 2011. Data mining tools are used to extract the knowledge from the databases created. Data mining tool - Rapid miner is used to discover the interested patterns on the data of caste wise households that are registered and caste wise households that are working in NREGS works. Caste Wise Employee database is created from NREGS data.

III. RESEARCH AREA

A. Study Area

Section 3(2) of Mahatma Gandhi NREGA provides that the disbursement of daily wages shall be made on weekly basis, or in any case not later than a fortnight after the date on which such work is done.

The internal studies conducted on the reasons for the delayed payments pointed out that the delays in release of funds by the Central Government, multi-level release system, continued parking of funds at various levels and the inability of the implementation agencies to get the funds in time for payment - were the main contributory causes for the increased delays. This calls for further steps to improve the system and to assure timely availability of funds as per demand.

delay payment in FY 2016-17
(% of transactions with delay)
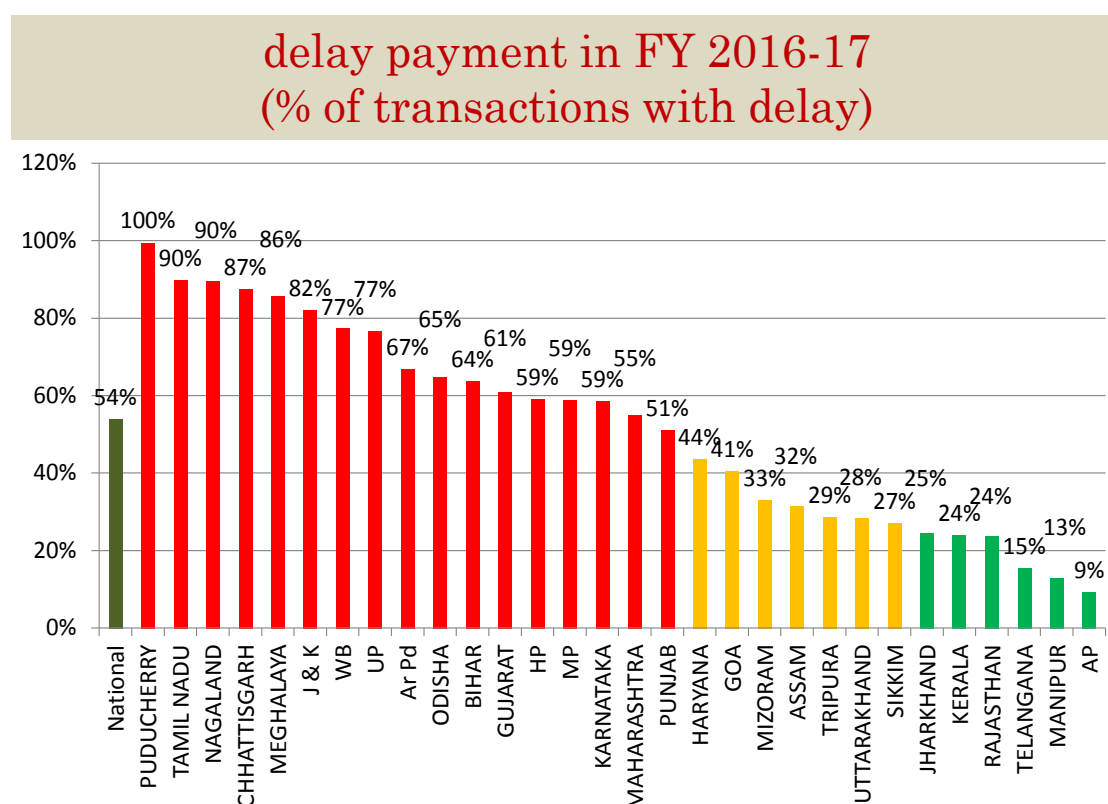
As per MIS no wage payment has been done in A & N and Lakshdweep in current FY

Fig. MGNREGA Performance Review Committee (PRC), MoRD, 17th January, 2017.

B. Data Set Used

For the present study all the data sets were sourced from the MGNREGA offices of Rajasthan, **MGNREGA MIS Portal, NIC and sponsor Bank i.e. SBI, Jaipur and National Informatics Centre, Jaipur.** Only few factors having effect on the payment of wages to the workers were selected for the present study. All the factors were considered for the duration of one financial year from 2016 to 2017. The factors considered are payment file generation date, payment due date, processing date and file received date.

From the data of a year for a particular district, the percentage share of various reason of delay for every district was calculated quarterly.

The average delay for each district was calculated from data for the quarter of the financial year 2016-2017.

C. Methodology Used

For the present study, data was extracted from live systems of MGNREGASoft, PFMS and Sponsor Bank, imported in test database and then datasets were created . The further analysis on the extracted sub-datasets were visualised using Microsoft Office Tools and Database IDE. Scatter plots were used to show the different factors influencing the payment of wages to the workers. The raw data set consisted of the following fields in database: (CPSMSMSGID, *FTO Generation DATE*, Transaction Type, BANK NAME, Batch Number, Scheme Name, FTO Name (contains District Code and FTO date), Payment Status, *Payment Date*, Payment Reference Number, Reason of Failure, *File Received Date*, Number of Transactions, PAYMENT FILE STATUS, PAYMENT FILE NAME, Amount). The following steps shown in Figure 1 were followed for processing and preparing the data for applying data mining techniques.

The database connectivity was established with Weka Tool for further analysis by applying data mining technique. Different parameters were set before applying technique. Each of the parameter used and set in each of the techniques used is described in detail in section IVB below.

IV. DATA ANALYSIS

A. Data Visualization

This section presents the data visualization of MGNREGA in thirty three districts of Rajasthan state for the financial year 2016-2017 with percentage share of the reason of delay in the payment of wages to the workers and average delay in each district quarterly for the year 2016-17. The data is analyzed quarterly.
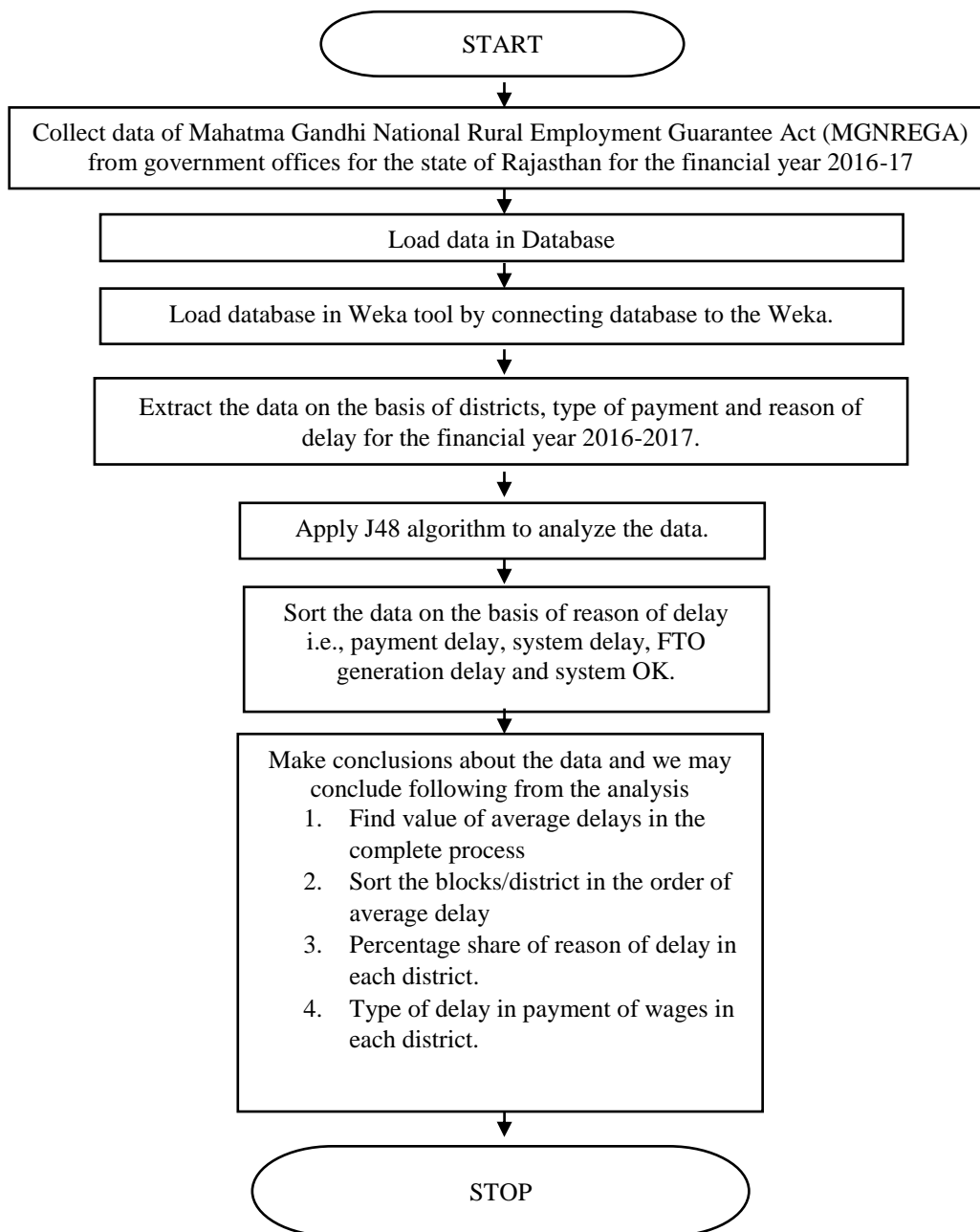
```
                    ┌─────────────────┐
                    │     START       │
                    └────────┬────────┘
                             │
┌────────────────────────────┴────────────────────────────┐
│ Collect data of Mahatma Gandhi National Rural Employment │
│ Guarantee Act (MGNREGA) from government offices for the  │
│ state of Rajasthan for the financial year 2016-17        │
└────────────────────────────┬────────────────────────────┘
```

START

Collect data of Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) from government offices for the state of Rajasthan for the financial year 2016-17

Load data in Database

Load database in Weka tool by connecting database to the Weka.

Extract the data on the basis of districts, type of payment and reason of delay for the financial year 2016-2017.

Apply J48 algorithm to analyze the data.

Sort the data on the basis of reason of delay i.e., payment delay, system delay, FTO generation delay and system OK.

Make conclusions about the data and we may conclude following from the analysis
1. Find value of average delays in the complete process
2. Sort the blocks/district in the order of average delay
3. Percentage share of reason of delay in each district.
4. Type of delay in payment of wages in each district.

STOP

Fig. 1 Steps for collecting, processing and analyzing the data

B. Proposed Algorithm

The data which was collected from different Government offices was imported in the database which contains many fields of which only few were considered for our research work. The data set was created with the help of Toad software. Toad

Software is a database management toolset from Quest that database developers, database administrators and data analysts use to manage both relational and non-relational databases using SQL.

A SQL script is written in SQL Editor window in order to create dataset from raw data collected from Government offices which is then executed. Thereafter, the data set is created for the research.

Now, the data mining tool − Weka is used for the analysis of the data set. Weka is connected to the database using user id and password. SQL script is again inserted in the query block of the tool which makes data set ready for the analyzing purpose. Then, the classification algorithm is used for the analysis. Decision tree algorithm J 48 is used for our study.

J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. C4.5 is a program that creates a decision tree based on a set of labelled input data. This decision tree can then be tested against unseen labelled test data to quantify how well it generalizes. This algorithm was developed by Ross Quinlan. It is an extension of Quinlan's earlier ID3 algorithm. C4.5 uses ID3 algorithm that accounts for continuous attribute value ranges, pruning of decision trees, rule derivation, and so on [3].

The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier [3].

The different parameters set used for J48 algorithm were as follows: binary splits on nominal attributes when building the trees = false; the confidence factor used for pruning = 0.25; debug = false; the minimum number of instances per leaf = 2; the amount of data used for reduced error pruning = 3, ten fold is used for pruning, the rest for growing the tree; reduced pruning error is used = false; to save the training data for visualization = false; the seed used for randomizing the data when reduced error pruning is used = 1; to consider the subtree raising operation when pruning = true; unpruned = false; use Laplace = false.

C. Results of Data Analysis

- *Analysis of percentage share of delay quarterly for the financial year 2016-17:*

  The data is analyzed and it was found that mainly four types of delay in the payment of wages were noticed. The percentage share of reason of delay was calculated quarterly for the financial year 2016-17. Below are the graphs for the percentage share of reasons of delay
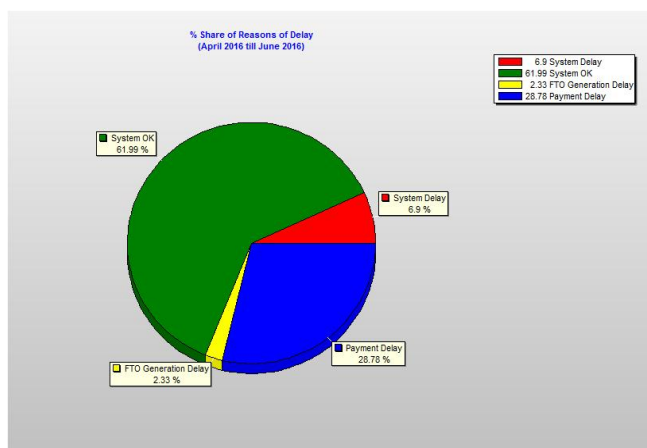
Fig. Percentage share of Delay (April 2016-June 2016)

For the first quarter of the year (April 2016-June 2016), the delay in payment of wages was mainly due to Payment delay which was 28.78%. Other reasons include System delay 6.9% and FTO Generation delay 2.33%. Overall system was found to be working fine for 61.99% transactions.
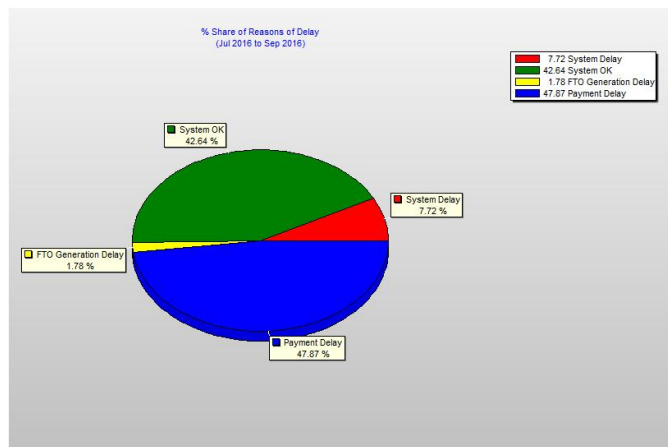

Fig. Percentage share of Delay (July 2016-September 2016)

In the second quarter (July 2016-September 2016), the delay in system was due to Payment delay which increased to 47.87% and System OK decreased to 42.64%. System Delay increased to 7.72% and FTO Generation delay 1.78%.
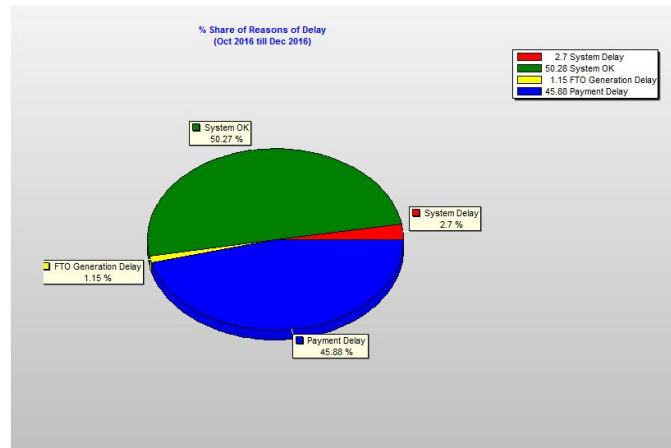
Fig. Percentage share of Delay (October 2016-Dec. 2016)

In the third quarter (October 2016-December 2016), Payment delay decreased marginally to 45.88%. Other reasons viz. System delay 2.7% and FTO Generation delay 1.15% also reduced marginally. System correctness increased to 50.27%.
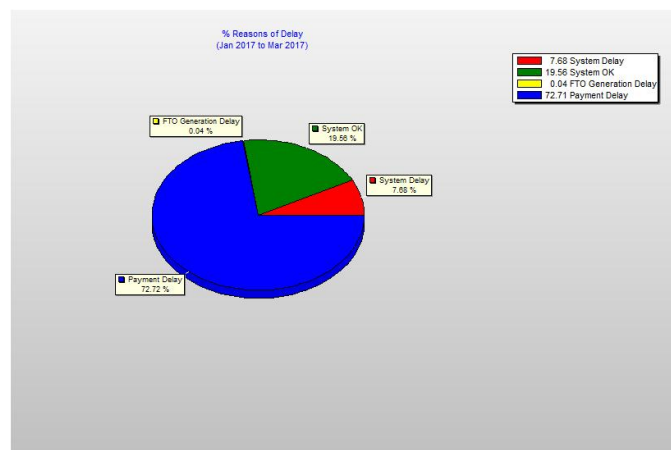

Fig. Percentage share of Delay (January2017-March 2017)

In the fourth quarter (January 2017-March 2017), there is tremendous increase in the payment delay with 72.72% and System promptness decreased to as low as 19.56%, System delay also increased to 7.68% when FTO Generation delay was as low as 0.04%.

From above, it can be concluded that the main reason of delay in the payment of wages is the Payment delay. This is on account of various factors such as 'Timely releases of Funds for making payments by Central Government', 'Delays in

processing of transactions by the Bank', 'Delays in the intermediate systems of the Payment Gateways', 'Capacity issues of handling the volume of transactions by the systems' etc.

- *Analysis of average delay for each district of Rajasthan for FY 2016-17:*

The average delay is calculated for each district of Rajasthan for the financial year 2016-2017. The average delay is calculated quarterly. The graph is drawn between average delay and each district. Below are the graphs for average delay.
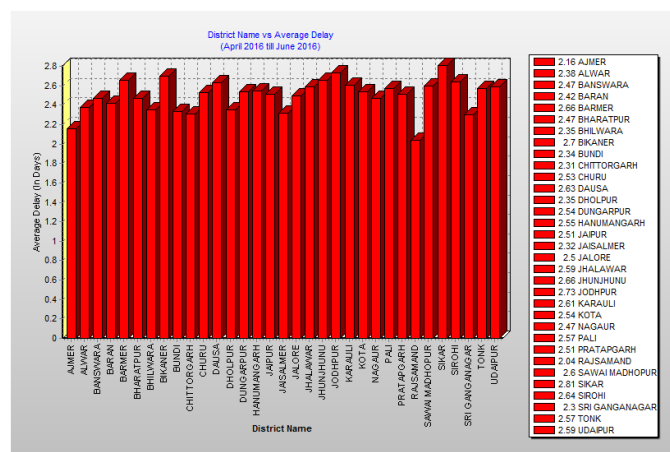

Fig . District versus average delay (April 2016-June 2016)

For the first quarter (April 2016-June 2016), the delay is maximum for Sikar district. However, there is marginal difference in the average delay for the districts. The least average delay is noticed for the district Rajasmand.
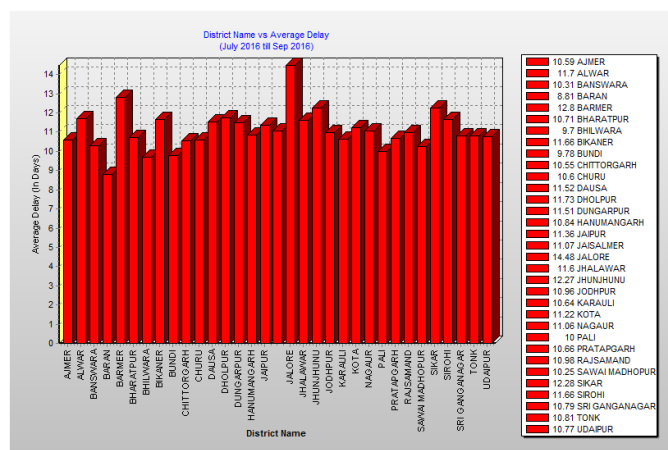

Fig . District versus average delay(July 2016-September 2016)

In the second quarter (July 2016-September 2016) , there was remarkable increase in the average delay as compared to the first quarter in which the delay was 2-3 days for every district which increased to 10-12 days for the districts. The maximum delay was accounted for Jalore.
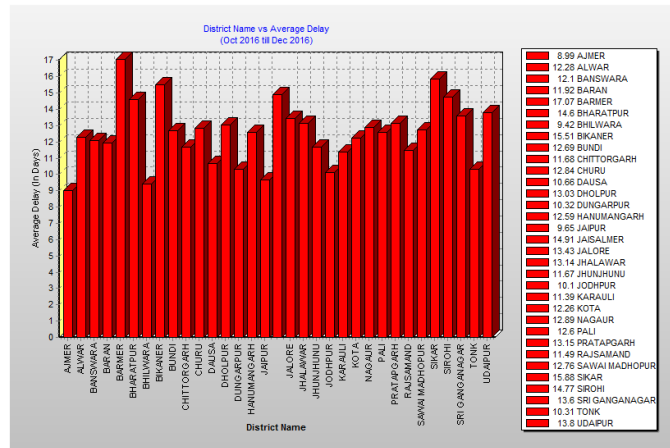

Fig. District versus average delay(October 2016-December2016)

In the third quarter (October 2016-December 2016), the average delay increased marginally between 8-17 days. In this quarter, Barmer has high average delay of 17 days and Ajmer has least average delay.
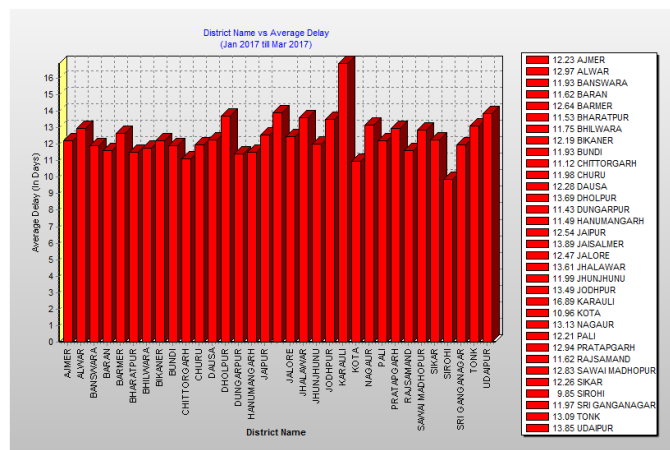

Fig . District versus average delay(January 2017- March 2017)

In the fourth quarter (January 2017-March 2017), the average delay decreased marginally but in Karauli it increased considerably with 16.89 days.

V. CONCLUSION

This research shows that data extracted from the bank and National Informatics Centre, processed and analyzed with data mining tool-Weka provide useful information to the Government which could be used for further improvement in the scheme.

From the obtained results several conclusions can be drawn:

Districts in the central Rajasthan have comparatively less average delay.

Districts in the western Rajasthan have high delay in the payment of wages as compared to all other districts of Rajasthan.

Using Decision tree algorithm (J 48) the payment of wages for the state of Rajasthan can be discovered. Results from our analysis show that most of the districts of Rajasthan have payment delay in the payment of wages. The analysis also shows that other reasons for the delay are negligible as compared to the Payment Delay.

Further detailed analysis may be carried out on each reason of Payment Delays. However, this is out of scope of this analysis as the objective is mainly to identify the reasons which are responsible for Delay in MGNREGA DBT system.

ACKNOWLEDGMENT

REFERENCES

[1] Yahya M. Tashtoush, Majd Al-Soud, Manar Fraihat, Walaa Al-Sarayrah, and Mohammad A. Alsmirat, "Adaptive E-learning Web-based English Tutor Using Data Mining Techniques and Jackson's Learning Styles," 8th International Conference on Information and Communication Systems (ICICS) 2017.

[2] Amit Gupta, Ali Syed , Azeem Mohammad , Malka N. , A Comparative Study of Classification Algorithms using Data Mining: Crime and Accidents in Denver City the USA, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 7, 2016.

[3] Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade, Vahida Attar , "Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction" , unpublished.

[4] G. Chandra, " A Study on Mahatma Gandhi National Rural Employment

Guarantee Act Opportunity and the Corruption (MGNREGA)," International *Journal of Management Research and Social Science (IJMRSS) , Volume 2, Issue 1, January - March 2015*

[5]   Dr.M.Usha Rani, "Expenditure Analysis Through Data Mining Techniques on NREGS(National Rural Employment Guarantee Scheme) Data of Andhra Pradesh," IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.2, No. 4, August 2012.

[6]   G. Sugapriyan, S. Prakasam, "Analyzing the Performance of MGNREGA Scheme using Data Mining Technique," International Journal of Computer Applications (0975 – 8887) Volume 109 – No. 9, January 2015.

[7]   Nisar Ahmad Shiekh,Mushtaq Ahmad Mir,"Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA): A Right Based Initiative towards Poverty Alleviation through Employment Generation", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611.

[8]   Evaluation of National Rural Employment Guarantee Act in Tamil Nadu, RTBI, Indian Institute of Technology, Madras.

[9]   MGNREGA Sameeksha An Anthology of Research Studies on the Mahatma Gandhi National Rural Employment Guarantee Act, 2005.

[10]   MGNREGA Sameeksha   United Nations, AN ANTHOLOGY OF RESEARCH STUDIES (2012-2014).

[11]   M. Ravindar, "EMPOWERMENT OF WOMEN THROUGH MGNREGS: A STUDY IN WARANGAL DISTRICT OF TELANGANA STATE", International Journal of Multidisciplinary Research and Modern Education (IJMRME) ISSN (Online): 2454 - 6119 (www.rdmodernresearch.org) Volume II, Issue I, 2016

[12]   RAHUL BAHUGUNA, AKHILESH CHANDRA PANDEY, VISHAL SOODAN, "A STUDY ON SOCIO ECONOMIC IMPACT OF MGNREGA ON BENEFICIARIES IN RUDRAPRYAG DISTRICT OF UTTARAKHAND INDIA", International Journal of Management and Applied Science, ISSN: 2394-7926| Volume-2, Issue-10, Oct.-2016.

[13]   Moksha Shridhar, Mahesh Parmar: Survey on Association Rule Mining and Its Application. International Journal of Computer Science and Engineering (JCSE), March 2017, Volume-5, Issue-3,pp. 129-135, E-ISDN: 2347-2693.

[14]   Mala Bharti, Vineet Richhariya and Mahesh Parmar: An Implementation of IDS in a Hybrid Approach and KDD CUP Dataset. International Journal of Research Granthaalayah (IJRG) Indore, Dec-14, Vol. 2, pp. 2-12, ISSN-2350-

0530(o), ISSN-2394-3629(p).

[15] Mala Bharti, Vineet Richhariya and Mahesh Parmar: A Survey on Data Mining based Intrusion Detection Systems. International Journal of Computer Networks and Communications Security, Dec-14, Vol. 2, No.12, pp. 485-490, ISSN 2308-9830.

[16] MGNREGA Sameeksha An Anthology of Research Studies on the Mahatma Gandhi National Rural Employment Guarantee Act, 2005.

[17] MGNREGA Sameeksha United Nations, AN ANTHOLOGY OF RESEARCH STUDIES (2012-2014).

[18] Kritika Yadav, Mahesh Parmar: Review Paper on Data Mining and its techniques and Mahatma Gandhi National Rural Employment Guarantee Act. International Journal of Computer Science and Engineering (JCSE), April 2017, Volume-5, Issue-4,pp. 68-73, E-ISDN: 2347-2693.

[19] http://nrega.nic.in

[20] http://rural.nic.in