Real Time Data Science Decision Tree Approach to Approve Bank Loan from Lawyer's Perspective

Dr P Rajesh[1], Dr Mansoor Alam[2], Dr Mansour Tahernezhadi[3], C Vikram[4], P N Phaneendra[5]

Associate Professor[1], Researcher[4-5], K L University, Professor Northern Illinois University[2-3],
rajesh.pleti@kluniversity.in[1], malam1@niu.edu[2], mtaherne@niu.edu[3]

Abstract – **Information technology, Digital Services using API, Artificial Intelligence, Data Science, Mobile banking, Automation spawn tremendous contribution in the direction of growing the efficiency, increasing huge number of customers, advance eminence customer services to most banking business application processes. With the enhancement in the banking sector lots of people are applying for bank loans Individual, Mortgage Loans, House loans, Apartment loans, Agriculture loans. These days bank managers and lawyers are facing abounding complications to sanction a loan by perceiving previous 30 years of link documents for specific site for house loan. For the most part from lawyer's perspective, to sanction legal approval, he must have to verify all bulk bundles of Registration documents, link connectivity, Identifying miss matches of sale deeds, owner relationships, family member certificates, death certificates, Re-registrations on same land, verifying any Hostage lands previously. In order to do the total procedure, lawyers may grasp three months span of time to extract conclusions by studying all documents based on project cost and litigations based on lands and it may vary according to project capital of Investment. In this project, we are trying to turn down the risk factor, computational time, ill-assorted Registration documents and Cross verification of all edge communications to make decisions to sanction the loan from lawyers, Bank Mangers, legal perspective and approve loan efficiently within stipulated time using decision tree approach of " Real Time Data Science Decision Tree Approach To Approve Bank Loan From Lawyer Perspective". we incorporated PHP, Python, XAMPP Web server, HTML, Data Science Decision tree approach to make lawyers decisions for legal approval within fraction of two minutes by entering all the document details into a database. we also incorporated execution results and discussed efficiency of proposed system with multiple scenarios.**

Keywords: Decision Tree, Data Mining, Loan Processing, Real Time Analysis, Lawyer Perspective.

I. INTRODUCTION

Distribution of the loans is the core business part of almost every bank. The main profit is directly earned from the banks [1,2]. The main motive in banking environment is to invest their benefits in genuine hands. Today many banks/financial companies approve loan after a regress process of verification and validation but still there is no guarantee whether the chosen applicant is the deserving right applicant out of all applicants [3,4]. In this system we used Decision Tree approach, Decision Tree is one of the part of Data Mining .and have many different applications in Banking Sector. we can predict whether that applicant is safe or not and the whole process of validation of features is automated by machine learning technique [5,6]. The disadvantage of this model is that it emphasizes different weights to each factor but in real life sometime loan can be approved based on single strong factor only, which is not possible through this system [7,8]. Decision Tree algorithm create a tree data structure where each node represents some testing which results on moving to next node and tree-based approach is used for prediction of class label [9,10].

Loan Prediction is very helpful for banks as well as for the user also. The aim of this Paper is to provide fast and easy way to choose the genuine applicants. It can provide special lead to the bank loan processing system. The Loan Prediction System can automatically calculate the weight of each attribute taking part in loan processing and performs new data test, features are processed with respect to their associated weight [11,12]. A time check can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows switch to specific application so that it can be check on priority basis [13,14]. Decision Tree algorithm is used widely in Banking System like Fraud Detection in credit card which decides whether the user is original owner of the card based on Decision Tree rules [15,16]. Many Decision tree projects have done in the recent times like Decision Tree Approach for the Handling of Sequential and Time Series Data, Automatic Webpage Classification by Using Machine Learning Methods [17,18]. The proposed approach exclusively for the managing authority of Bank/finance company, process of prediction is done using automation approach privately, no stakeholders would be able to alter the processing result against Loan Id can be sent to various department of banks so that they can take appropriate action on application. This helps all others department to carried out other formalities. [19,20].

S.no	Survey Number	Document Number	Seller Name	Buyer Name	Date of Reg.	Reg. Office	Extent
1	137/1	2322/2010	PALEM SIVAIAH,PALEM VENKATESWARLU	KOCHRLA ANTHONY TARAKA BABU	10/12/2011	GUNTUR R.O.,	100
2	137/1	3312/2010	PALEM SIVAIAH,PALEM VENKATESWARLU	PULLISULA SURENDRA BABU	10/1/2012	GUNTUR R.O.,	100
3	137/1	3428/2010	KOCHRLA ANTHONY TARAKA BABU	PULLISULA SURENDRA BABU	31/01/2015	GUNTUR R.O.,	100
4	137/1	1234/2012	PULLISULA SURENDRA BABU	GOTTIPATI NAGA MANI	22/02/2015	GUNTUR R.O.,	50
5	137/1	123/2015	PULLISULA SURENDRA BABU	KOKKIRALA SRINIVASU	2/04/2015	GUNTUR R.O.,	50
6	138/1	1123/2015	PULLISULA SURENDRA BABU	KAMEPALLI JYOTHI	4/4/2015	GUNTUR R.O.,	100

Table 2.2: 1:N and N:1 Records of documents

S.no	Survey Number	Document Number	Seller Name	Buyer Name	Date of Reg.	Reg. Office	Extent
1	244/A	3535/2015	SANIVARAPU RAMANJANEYA REDDY,NEELA LAKSHMI PADMAVATHI,VELAGAPUDI BHAVANI PRASAD	VELAGAPUDI RADHIKA,PALEM SIVAIAH,KOKKIRALA JYOTHI	31/01/2017	GUNTUR R.O.,	4840
2	244/A	3304/2000	CHALLA SIVAIAH	SANIVARAPU RAMANJANEYA REDDY,NEELA LAKSHMI PADMAVATHI,VELAGAPUDI BHAVANI PRASAD	22/02/2002	GUNTUR R.O.,	4840
3	244/A	3429/2004	ANANTHA INDIRA	KALARI APPA RAO	26/07/2007	GUNTUR R.O.,	500
4	244/A	5322/2007	ANANTHA INDIRA	YETIGADDA SURESH	21/11/2010	GUNTUR R.O.,	300
5	244/A	4022/2000	ANANTHA INDIRA	CHALLA SIVAIAH	01/02/2001	GUNTUR R.O.,	100
6	244/A	3923/2007	KALARI APPA RAO	NEELA SRIRAM	19/08/2010	GUNTUR R.O.,	100

Table 2.3, N:N Records of Document

II. DESCRIPTION OF THE DATASET

S.no	Survey Number	Document Number	Seller Name	Buyer Name	Date of Reg.	Reg. Office	Extent
1	1999/C	4	TUMMALA GEETHA	TUMMALA GOVIND	06-06-2003	VIJAYAWADA	1200
2	1999/C	4	PUVVADA NARAYANA	PALEM RAM	06-06-2005	VIJAYAWADA	1200
3	2000/B	6	PALEM RAM	PALEM SHANKAR	06-06-2006	VIJAYAWADA	1000
4	2000/B	7	PALEM SHANKAR	SANIVARAPU SUJATHA	06-06-2007	VIJAYAWADA	1000
5	2000/B	9	PALEM RAM	PALEM RAVI	06-07-2006	VIJAYAWADA	100
6	2000/B	8	SANIVARAPU SUJATHA	SANIVARAPU RAGHU	06-06-2008	VIJAYAWADA	800

Table 2.1 : one to one Records of Client Document

S.no	Survey Number	Document Number	Seller Name	Buyer Name	Date of Reg.	Reg. Office	Extent
1	137/1	2923/2009	KOCHRLA ANTHONY TARAKA BABU	PULLIRUSULA SURENDRA BABU	10/1/2012	GUNTUR R.O.,	123
2	137/1	2322/2000	KESARI SUBBA RAO	KOCHRLA ANTHONY TARAKA BABU	2/04/2001	GUNTUR R.O.,	200
3	137/1	1123/2000	KOCHRLA ANTHONY TARAKA BABU	PALEM SIVAIAH	2/04/2010	GUNTUR R.O.,	77
4	137/1	3428/2010	PULLIRUSULA SURENDRA BABU	NEELA LAKSHMI PADMAVATHI,MADALA BRAHMANAMDAM	10/11/2012	GUNTUR R.O.,	50
5	137/1	9852/2012	MADALA BRAHMANAMDAM	ADUSUMALLI RADHA MRUDULA	22/02/2015	GUNTUR R.O.,	25
6	137/1	5322/2010	PALEM SIVAIAH	PALEM VENKATESWARLU	02/01/2012	GUNTUR R.O.,	50

The major role for the dataset is the user. User must keep his full attention while filling the data because any mistake leads to the infelicitous data ingression to the database. In this Project we used phpMyAdmin database to store and manipulate the data. and the user will fill the data and submit the information of the documents. In this Project we come to know that after observing and analysing a lot of documents, there exists four types of relational documents that are 1:1, 1: N, N:1, N: N Documents to transfer from one to one . Each Document contains the details of **Survey Number, Document Number, Seller Name, Buyer Name, Date Of Registration, Name Of Registration Office, Land Extent with Option Units (Acres, Cents, Yards, Feet).** The Dataset may contain N number of Documents as it is a data of Past 30 years link documents of a Land it depends on the selling and purchasing of the Land and it can make an

un-imaginary number of Documents from 20 to 30 by depending on the landowner's profits and losses.

III.   PROPOSED METHODOLOGY

Step 1: - Get Input Link documents dataset bulk bundles of Registration documents, link connectivity. Identifying miss matches of sale deeds, owner relationships, family member certificates, death certificates, Re-registrations on same land, verifying any Hostage lands previously. Data Attributes are seller, buyer, document number, date of registration, place of registration, extends of Sq . Yards, Survey No etc.

Step 2: - Sort the data according to Date of registration.

Step 3: - Get all Sellers and Buyers Names.

Step 4: - Make them as nodes and Connect them like Seller be the parent node and Buyer be the child node and add colour to green.

Step 5: - for each in tree except root node

   Step 5.1: - check if the seller is owner of that document if not check relation between original owner and seller.

   Step 5.2: - If relation does not satisfy

      Step 5.2.1: - make child node colour as red and all nodes belonging as red colour by recursion.

   Step 5.3: - If node colour is Green check land extent of that owner if selling more then.

      Step 5.3.1: - make child node colour as red and all nodes belonging as red colour by recursion.

Step 6: - For each node in tree.

   Step 6.1: - Add details of nodes like Date of registration, Place of registration, Land Extent by using input data.

Step 7: - Create Graphviz.

Step 8: - Add these Tree nodes to Graphviz Make Graphviz node colour as node colour.

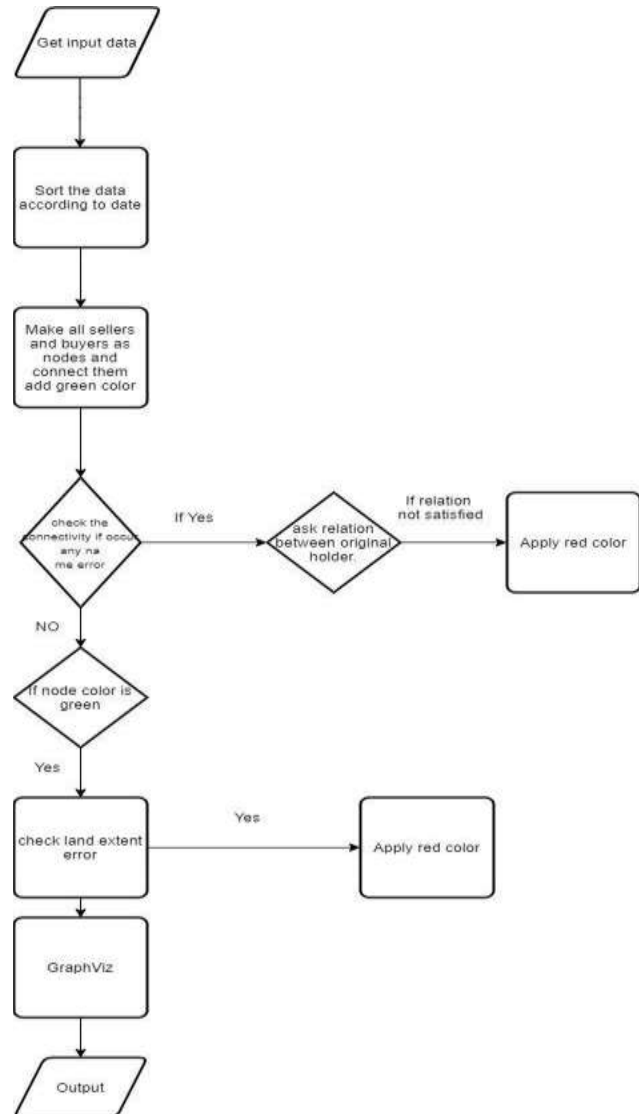**Step 9: -** Get Output of prediction of decision Tree to make decision.

Fig 3.1: Flow chart diagram for proposed approach.

IV. EXPERIMENTAL RESULT ANALYSIS

Python 3.1:

We Used Python as a main body of our project that takes all the information from database and checks all litigations utilizing the DECISION TREE approach and converts the data to graph. And it is one of the backend software of this project.

Graphviz:

After shaping the data from the decision tree approach the data will convert to graphical representation with green/red colour to denote the document status and for a better understanding of the solution by the utilizer, we used Graphviz.

Xampp:

To introduce a facile controllable software to users we had taken the option available which is an open source software Xampp, with this we can,

- enable the apache server to inscribe the PHP files which is responsive to HTML pages which generates UI to make things facile for user.

- enable the PHP my admin database to store the data utilizing PHP commands.

Result Analysis:

1. **Error rate = Number of mismatches/total no of records**
2. **Execution time = Avg time taken by the python model**

Here,

- The mismatches indicates the record of a document is duplicated or unconnected with the previous records and perchance the data entered is in the type of 1:N , N:1 , N:N Format.
- Execution time will be considered as the Average of 3 execution times of a python model.

(1 : 1) Document of Fig 2.1

Analysis: -

Graph: -
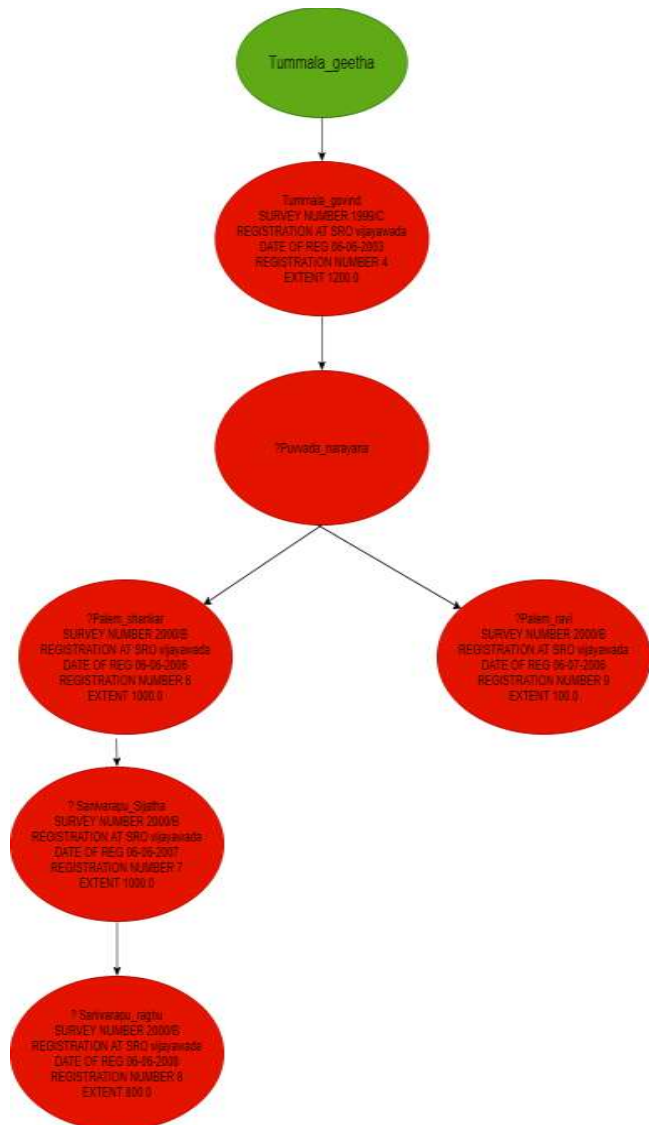
Results: -

Mismatches = 0

Total No of Nodes = 8

1) Error Rate = 0/8 = 0

   % Error = Error rate * 100 = 0 * 100 = 0%

2) Execution time = (0.4006+0.4039+0.4018)/3

   = 0.4021

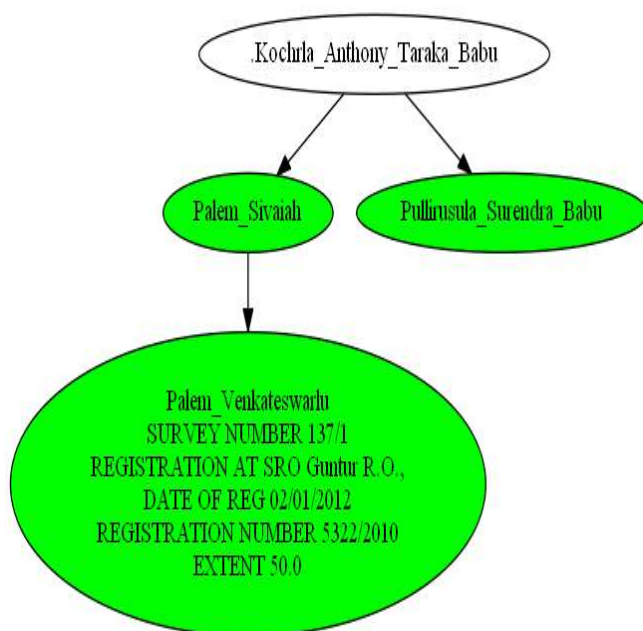**(1 : N) Document of Fig 2.2:**

Analysis: -
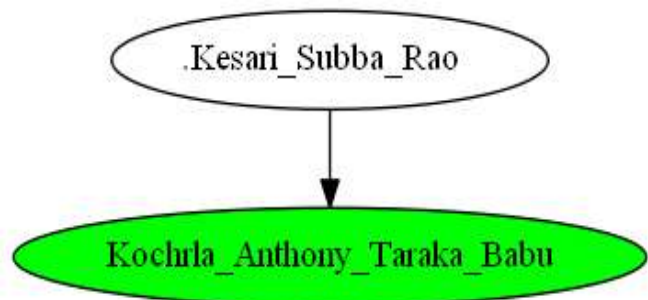

Graph: -

Mismatch 1: -

Mismatch 2: -

Mismatch 3: -

Results: -

Mismatches = 3

Total No of Nodes = 8 (Duplicate records are treated as one record)

925

1)Error Rate = 3/8 = 0.375

% Error = Error rate * 100 = 0.375 * 100

= 37.5%

2)Execution time = (0.7525+0.5685+0.7352)/3

= 0.6854

(N : 1) Document of Fig 2.3:

Analysis: -

Graph: -
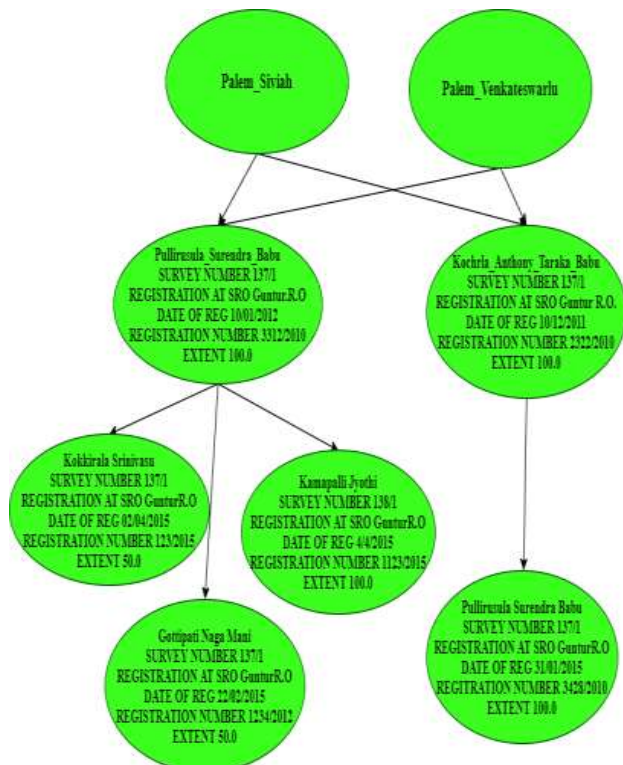
Results: -

Mismatches = 1 (One main Record is Divided into two main records)

Total No of Nodes = 8 (Duplicate records are treated as one record)

1)Error Rate = 1/8 = 0.125

% Error = Error rate * 100 = 0.125 * 100

= 12.5%

2)Execution time = (0.8921+0.7523+0.7124)/3

= 0.7856

(N : N) Document of Fig 2.4

Analysis: -

Graph: -

926
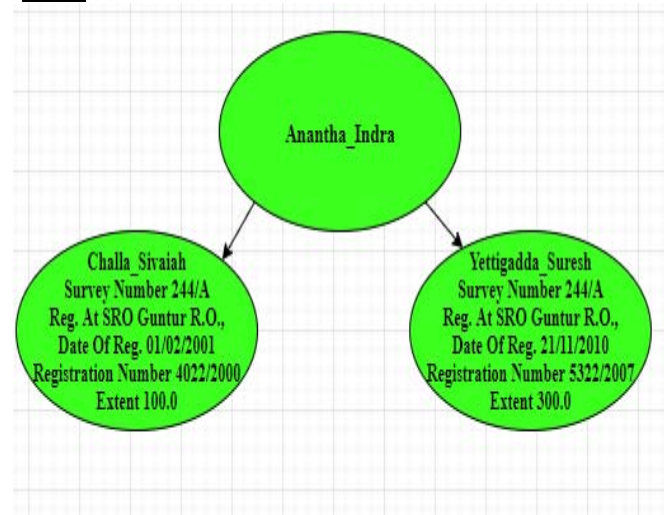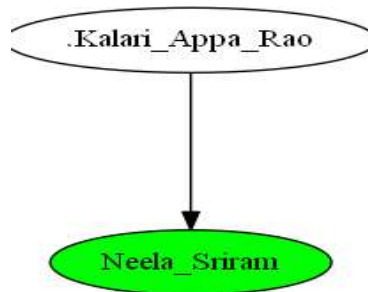
Mismatch 1: -

Mismatch 2: -

Mismatch 3: -

Mismatch 4: -

Results: -

Mismatches = 4

Total No of Nodes = 7 (Duplicate records are treated as one record)
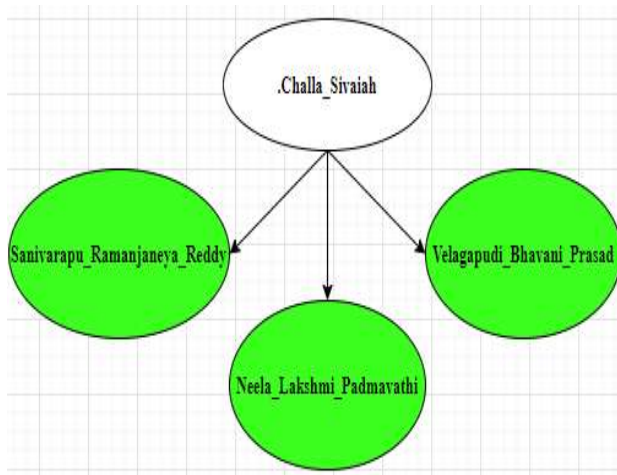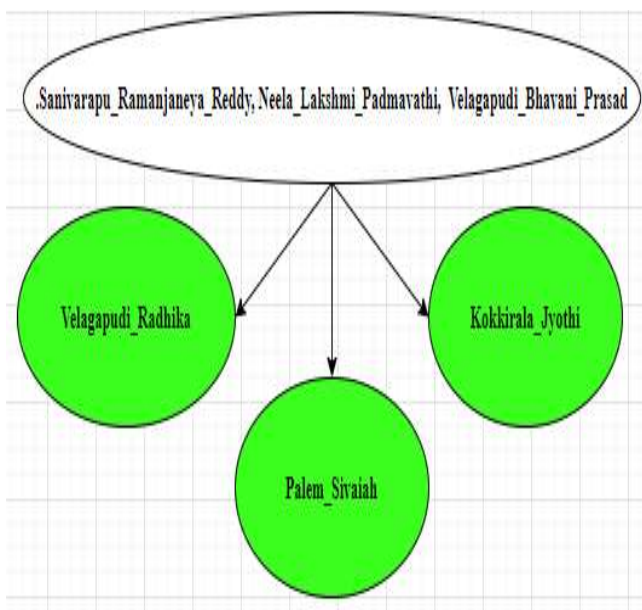
1)Error Rate = 5/7 = 0.5714

% Error = Error rate * 100 = 0.5714 * 100

= 57.14%

2)Execution time = (0.6243+0.5552+0.5794)/3
= 0.5863

Types of Mismatches: -

- Duplicate Records.

- New Record which is not Relevant.

Reasons: -

1. If a record having more than one person in seller position our algorithm will create duplicate record to rectify the conditions that relates to an existing record.

2. If a record having more than one person in buyer position our algorithm will divides each person to an individual record to

927

identify each person selling the land in the given limits of their own land or not in the future case and this may lead to splitting of a record.

3. If a record entered is not relevant to the previous buyers or the main record the tree may deviates from main root and split its node to form a new root.

Success Percentage (%): -
The success rate of the algorithm for each type of document as follows:
For,

$$(1 : 1) = 100\% - \%Error$$
$$= 100\% - 0\%$$
$$= 100\%$$

$$(1 : N) = 100\% - \%Error$$
$$= 100\% - 37.5\%$$
$$= 62.5\%$$

$$(N : 1) = 100\% - \%Error$$
$$= 100\% - 12.5\%$$
$$= 87.5\%$$

$$(N : N) = 100\% - \%Error$$
$$= 100\% - 57.14\%$$
$$= 42.86\%$$

By observing the results and analysis the output and execution of a document is differ from one type of document to another type of document. As the given set of records are equal number in total but the execution time is varying to each type of documents that indicates the time of execution doesn't depend on the number of records and may depends on the number of sellers and buyers in a record. By the analysis of above success percentages, it specifies that the algorithm is perfectly working under the condition of (one:one) types of documents and the decision tree approach of our algorithm execution time is less as compared to the other type of documents. In future we will try to increase the success rate of our algorithm with other types of documents.

V. CONCLUSION AND FUTURE SCOPE

From an opportune checking of positive points and control on the component that the product is a highly efficient part from lawyer's perspective. This application is functioning felicitously and meeting to all Banker requisites. This component can be easily plugged in many other systems. There have been numbers cases of computer glitches, errors in content and most paramount weight of features is fixed in automated prediction system, "REAL TIME DATA SCIENCE DECISION TREE APPROACH TO APPROVE BANK LOAN FROM LAWYER PERSPECTIVE" could be made more secure, reliable, dynamic weight adjustment and reduces computational overhead, time consumed by attorney to make decision to sanction loan or not and it also represents kind of obligations available at each and every node of edge link communications. This kind of Proposed Project will be helpful to lawyers, Bank managers, Clients, system Operators, Registration Documents people, Revenue Officials, Common people to purchase lands, Bank field officers in terms of Rural Urban People. It can be also useful to avoid re- registrations on same lands and hurdles faced by the common people in society. In near future this factor can be combine with the module of automated processing system and work with increased success rate of datasets like N:N (many to many), N:1(many to one node) communications with different data sets. The system is trained on old training dataset in future software product can be made early testing of date attribute additionally take part in training data after some fix time. we are also planning the data set to run on dynamical data sets representations.

REFERENCES

1. Josip.M, Dario Sebalj (2016)."Decision trees for predicting the academic success of students". Croatian Operational Research Society.
2. Dr. Mohammad Miyan (2017). "Applications of Data Mining in Banking Sector", International Journal of Advanced Research in Computer Science.
3. Andrea B, Enrico M, Angelo M, Guido M (2019)."A Novel Decision Tree Approach for the Handling of Sequential and Time Series Data ", www.mdpi.com/journal/computers.
4. IqbalH.Sarker, Alan Colman, JunHan, Asif Irshad Khan, YoosefB, Abushark, KhaledSalah (2019)."A Behavioural Decision Tree Learning to Build User-Centric Context-Aware Predictive Model ", Springer Science Business Media, LLC, part of Springer Nature.

5. Hammed, Mudasiru1 and Soyemi, Jumoke (2020)."An implementation of decision tree algorithm augmented with regression analysis for fraud detection in credit card ", International Journal of Computer Science and Information Security (IJCSIS), Vol. 18, No. 2.

6. Makoto Tsukada, Takashi Washio, Hiroshi Motoda, "Automatic Web-Page Classification by Using Machine Learning Methods".

7. M. R. Islam and M. A. Habib, "A Data Mining Approach to Predict Prospective Business Sectors for Leading in Retail Banking Using Decision Tree", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.2, March 2015, pp. 13-22. DOI: 10.5121/ijdkp.2015.520213.

8. Quinlan RJ (1986) Induction of decision trees. Mach Learn 1(1):81–106.

9. Y. Sahin, E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines", Proceeding of International Multi-Conference of Engineering and Computer Statistics, Vol. 1, 2011.

10. V. Bhambri, "Application of Data Mining in Banking Sector" International Journal of Computer Science and Technology, Vol. 2, Issue 2, June 2011, pp.199-202.

11. M. L. Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October 2006.

    (http://www.agricerp.com/mts/data_mining/ Banking/banking% 20data%20mining.pdf.)

12. Sonia, A. Arora, "Review on Use of Data Mining in Focusing Bank Frauds and Enhancing Business", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 3 Issue VI, June 2015, pp. 175-178.

13. B. Siddhartha, K.T. SanjeevJha, and W.J. Christopher, Data mining for credit card fraud: A comparative study. Elsevier, Decision Support Systems, Vol. 50 Pp. 602–613, 2011).

14. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont California: Wadsworth, Inc.; 1984. [Google Scholar].

15. M. P. Thapliyal, "Data Mining: A Tool for Banking Industry, International Journal of Emerging Research in Management Technology "ISSN: 2278-9359 (Volume-4, Issue-4), April 2015.

16. P.Rajesh, Dr.G.Narsimha, " Cerebration Of Privacy Preserving Data Mining Algorithms" in International conference on machine learning and data analysis ICMLDA_ 2014, USA in association with springer, IEEE Explore, DBLP.

17. P.Rajesh, Dr.G.Narsimha, "Fuzzy based privacy preserving classification of data streams." in ACM conference (CUBE), Pune, PP: 784-788, 2012, DBLP, ISBN: 978-1-4503-1185-4.

18. P Sravanthi, Dr. P.Rajesh "An Exploration of Prediction of Heart Disease using Machine Learning Classification" has been published in open access International Journal of Scientific & Technology Research (IJSTR) Volume 9 - Issue 3, March 2020 Edition.

19. V.I.S. RamyaSri, Ch. Niharika, K. Maneesh, Mohammed Ismail "Sentiment Analysis of Patients&; Opinions in Healthcare using Lexicon-based Method" International Journal of Engineering and Advanced Technology Volume-9 Issue-1, October 2019.

20. K.Srinivas, Mohammed Ismail.B "Testcase Prioritization With Special Emphasis On Automation Testing Using Hybrid Framework" Journal of Theoretical and Applied Information Technology 96(13) 4180-4190 July 2018.

21. Dr.P.Rajesh, Sai Prasanna " A Forensic Approach To Perform Android Device Analysis " at National Women Conference on Technological Innovations", Dec 2018. Scopus, with Best paper award of conference.

929