

Disciplinary Control System Based on Naïve Bayes Classification Technique using Jsp Servlets



Arpita Roy, Nikhat Parveen, P.Rama Bhargavi, A.Navya, A.Pavan Kumar

Abstract: *The idea behind the task is to create the complaint about the student, who behaves mischievously in any circumstances by not following the dress code which includes things like identity cards, shoes, tuck especially for men, students should not carry mobiles to college if any person/student violates these rules then the disciplinary committee members have the authority to take off that particular student's belongings (I.e. I'd card, mobiles). If that particular student caught more than one time for any condition then then the disciplinary committee will charge a fine and sometimes the mobile phones and I'd cards will be moved into locker (where they head of the committee members will have capacity to keep with them more than 3 days to 1 month, sometimes they will keep with them 1 year too. Students have to spend much time to take their belongings from the disciplinary committee members and also to take permissions, instead of that we are created a website and in that website students, disciplinary committee members including administrators can read, write, create and delete the comments that are created. From that data we can analyse in the way of plotting techniques and classification approach in which year students are not in discipline manner.*

Keywords : Student management, Naïve Bayes classification, JSP Servlets, R Programming, plotting techniques.

I. INTRODUCTION

Now a day, educational systems also well developing their usage of many technologies to interact more with different kinds of students based on their interests. Institutions must maintain their peculiar discipline and ethics on each of their work from both from the student's side as well as faculty side to move ahead in the society. While coming to students in the college/ institution they must possess and carry their the particulars to enter like identity card, must attain in the formal dress code for both boys and girls (including shoes, tuck for boys), not to carry any electronic gadgets I.e. mobiles.

Our project mainly focuses on analysis of the student discipline in the institution in current generation. In order for us to accomplish this we have taken the data from the discipline committee members and also created a link for taking the data from the students and also faculty. This survey conducted with in our college itself. In order to make our process much simpler we were able to use **R-programming** for the analysis of all of our Data.

'R' is open free software which was developed in 1993 for statistical computing and graphics. It is one of the most used software's by data analysts as well as data miners so that they can develop statistical software's. 'R' is a language that comes with many inbuilt libraries that can be used in various scenarios. However this language is well known for the following algorithms; Linear regression, Logistic regression, Decision tree, SVM, Naïve Bayes, KNN, K-means, and Random Forest. There are many kinds of algorithms for different analytical analyzations used by different users. Here we have implemented 'R' within our project to analyze all of our data set both in graphical format as well as prediction of the output. It was also able to plot all of the data using its in-built graphical libraries with in the R environment.

We analyze the relationship of several related matters taken as a whole system including complaint-voicing rate, compliant-dealing ability, satisfaction, repeat purchase rate and loyalty, and build a behavioural model according to the way the system operates. Furthermore we specialize on the tendency of these matters changing during the process enterprises dealing with student's complaints, employing the method of System Dynamics for the first time to simulate the tendency with the data from the markets. Another objective is how to adjust the key factors which can be obtained through analyzing the simulation results for the best possible effects of the system. Finally, an optimum strategy of factors adjusting is derived from experimental data. As a strategic laboratory, the stimulation not only avoids the high cost result from the ex-post analysis, but also ensures the timing and accuracy of forecast result, then offers guidance to managers making decisions.

II. LITERATURE SURVEY

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Arpita Roy, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India.

Nikhat Parveen, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

P.Rama Bhargavi, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

A.Navya, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

A.Pavan Kumar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

S.No	Paper Name	Author Name	Techniques Used	Results
1	A performance comparison between classification techniques with CRM application[1].	Dalia Ahmed Refaat Mohamed Dr. Mohammed Sakre	1. Centroid-based classifier 2. K-Nearest centroid neighbor classifier 3. And also applied all pre-processing techniques like normalization, splitting, filtering etc.	Analysis on the dataset based on the complaints that are entered and also finding the comparison based on the distance parameter.
2	Public Complaint Service Engineering based on Good Governance Principles[2].	Suhardi Novianto Budi Kurniawan, Deni Prayitno, Jaka Sembiring	Software Engineering Overview and using website creation	Count of query how many times asked by the customer
3	The study of customer complaints management based on system dynamics: modelling and simulation[3].	JIN-LAN LIU, JIAN KANG, YIN BAI, XIN ZHANG	System Dynamics	It gives result in the form of relationship between the variables that are defined.
4	Smart Complaint Management System[4].	Pattamaporn Kormpho, Panida Liawsomboon, Narut Phongoen, Siripen Pongpaichet	SVM Decision Tree Naïve Bayes Sequential Minimal Optimization(SMO)	Here the results will be in both classification and regression models too, visualization is also applicable.
5	Improving user complaint management system and satisfaction level via reader-friendly linguistic features[5].	Zurah Binti Abu Fadilah Ezlina Binti Shahbudin Mastura Binti Mansor Nurul Zahirah Binti Abd Rahim Nur Aqilah Binti Norwahi	Creation of dynamic website of the	Displaying the list of complaints the are received from others based on the customer problems.
6	Rule-base data mining systems for customer queries[6].	S Sangeetha Ravichandran, D.Sathya R. Shanmugapriya, G. Isvariya	Means of Classification Data Clustering Apriori Algorithm	Displaying clustering graph by using queries

III. PROPOSED APPROACH

Existing Process

In existing process, limitations of existing process are Less efficiency, More time consumption, Highly rely on human effort, Ineffective for large no of complaints, It does not provide any security about the details who complained

Proposed Idea

Creating a website link or a google form will be much easy and everyone can post their complaint and also comment for their response to the higher authorizes & analysis part will be easy by taking the data from the database which will be under controlled by particular institution member. Here student number is unique so that with the help of the student number it is very easy to find no of times that the student was caught by the disciplinary committee members. By taking that data we are performing the machine learning techniques like firstly we used Naïve Bayes algorithm where it deals with the posteriori probability, secondly we used Plotting Techniques like box plot, ggplot2, histogram, finally at last we used the DataExplorer for generating the report in the form of all the missing value graphs, getting graphs for memory storage etc.

In our project there are 3 steps of getting the result they are firstly Gathering of data from all the connected databases via using link, second importing the data in to the r studio and applying the suitable algorithm to get accurate result better than the existing process, finally third step is visualizing the data into the graphical format.

The technologies that we used in our project are

- Servlets
- JSP
- Machine learning using R programming
- The algorithms that are used
- Naive Bayes
- Plotting techniques
- DataExplorer

Servlets

Servlets are used to create web applications. Basically Servlets internally works on Java language to create web applications. Web applications are helpful for connecting other connected applications which are occupied at server because of these additional features dynamic web pages are connected to servlets. As servlet is a dynamic web page it means that anything that we change in the alignment or even if we change the variable also it will reflects on the connected screen or connected page too. Servlets works on Java, the web applications which are created based on Servlets are much faster, strong and secure.

JSP

The abbreviation of JSP is Java Server Pages. It is a client to server or server to client automation. It is able to applicable mainly for creating active web based application and also used for creating the active web content whenever an end user is requested and also able to access any kind of programming language.

These JSP labels are used to insert JAVA code into HTML or XML pages and also able to write directly with in the JSP code into the suitable labels. It is also an advancement step of Servlets. It is a Web oriented technology that helps us to create dynamic and platform independent web pages. In this, Java code can be inserted in any web based languages like HTML/ XML pages or both. At first JSP converts the connection into servlet by using JSP storage space before processing the client's query. JSP handles the problem from the two way direction. Instead of inserting the code in other html pages to redirect new HTML pages, JSP itself able to insert the scripted code into HTML pages. Java is the default scripting language of JSP, but the JSP specification allows for other languages as well, such as JavaScript, Perl, and VBScript.

R Programming

R is a language of programming developed in 1993 by Ross Ihaka and Robert Gentleman. R has a robust numerical and visual system collection. This contains algorithm for machine learning, linear regression, time series, and numerical inference to name only a few. Most R libraries are written in R, but C, C++ and FORTRAN codes are used for heavy computational functions. R used for Statistical representations, Data analysis, Machine learning algorithm. Including artificial knowledge and it able to communicate with the Integrate codes, graphs, and outputs to a report with R Markdown or build Shiny apps to share with the world The machine learning algorithms are

1. Naive Bayes
2. Plotting Techniques
3. Data Explorer

Naive Bayes Algorithm

Naive Bayes is a Bayes Theorem-based Supervised Machine Learning algorithm that uses a probabilistic approach to solve classification problems. It is based on the idea that the predictor variables are independent of each other in a Machine Learning system. This implies that a model's outcome depends on a number of independent variables which have nothing to do with each other. The principle behind Naive Bayes is the Bayes theorem also known as the Bayes Rule/ Bayesian rule. The Bayes theorem is used to calculate the conditional probability,

$$P(A|B) = (P(B|A) P(A)) / P(B)$$

Here for Naive Bayes the package "e1071" is used. "e1071" is the conditional a-posterior probability of the categorical class variable (attribute name), here can mention the threshold values, the values are of 2 types 1.apriori it is a dependent variable.

2.tables it generates or stores the mean, standard deviation for each target class for each categorical variable.

Plotting Techniques:-

There are other ways for displaying results in R can graphically display format. This is a brief guide to some of the basic commands for plotting. The plotting will be varied for different attributes with in the dataset by using different commands for visualization.

1. Histogram
2. Boxplot
3. Pie-chart

Histogram:-

Histogram is a simple plotting graph. It maps the frequencies within certain ranges the data exists. Here we give examples using the data frame w1 described at the top of

this page, and w1\$vals is the one column of data. It is able to display the ranges of the particular attribute

```
> hist(usedcars$price, main = "Histogram of Used Car Prices", xlab = "Price ($)")
```

Boxplot:-

A boxplot provides a visual view of a data set's mean, quartile, average, and low. Using two different data sets, we give explanations here. It is able find the mean, maximum and minimum values in the given dataset.

```
> boxplot(pro$fine_amount, main="Boxplot for fine collection", ylab="fine (rs)", xlab="year(int)", col="pink")
```

Pie-chart:-

To construct charts and graphs, R programming language has various libraries. A pie chart is a representation of values of different colors as slices of a ring. The slices are numbered and the map also displays the numbers relating to each slice. The pie chart is created in R using the pie() (function that takes positive numbers as an input of the vector. To control labels, color, title etc., the additional parameters are used.

```
> pie(x, labels, radius, main, col, clockwise)
```

Data Explorer

Data Explorer is used to simplify and automate the Exploratory Data Analysis (i.e. it can be able to validate any assumptions that are present in the dataset). The data explorer will be processed in the way of analysing the data and building the model in the pictographically. Here in data explorer there will not be any statistical approach, the process will be repeated until the dataset should not contain any outliers, null values and missing values within the dataset.

The functions that are present in the data explorer are

- **Package name "Data Explorer"**
- **Plot_str("reference_string")** it gives result what are the attributes that are present
- **Create_report("reference_string")** here by using this command we are able to generate a report to see complete data
- **Plot_missing("reference_string")** it displays where the missing values are present in the dataset
- **Introduce("reference_string")** it shows output as the attribute names with the row count
- **Plot_intro("reference_string")** it shows the graph with the attributes, rows and their frequencies of the storage capacity.

IV. METHODOLOGY

In our project initially for taking the data we are created a dynamic link type like which is similar to website where the data will be stored directly with in the database. In next step the collected data will be exported into the R studio for analyzing the data based on the machine learning algorithm while exporting the data we looked for finding any null or missing values with in the data set later in final step the algorithm result will be able to display the result either in graphical or in text format.

V. RESULTS AND ANALYSIS

After the data imported into the R studio particular program can begin the process of analyzing the data for analysis.

This step our process will not only contribute in studying the data but we will also look much deeper into how accurate as well as how volatile our data sets are Naïve Bayes algorithm gives the high probability of the value in each attribute. The result is Fig1. This result explains the more frequent complaint that is noted about the student.

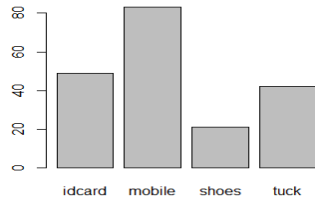


Fig 1 result for naïve Bayes classification for each occurrence of the complaint type

The fig 2, 3 indicates the result after smoothing the attributes smoothing means removing the outliers, null values from the imported dataset even it removes all the irregularities in the dataset, by using the smoothing the accuracy of the data will be increased

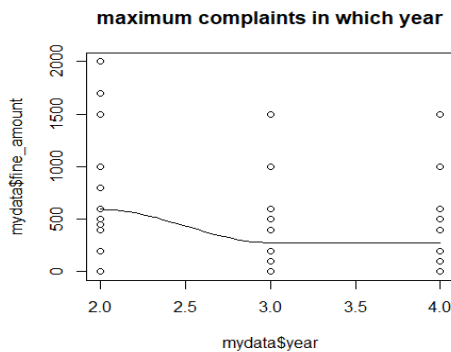


Fig 2 result for naïve Bayes classification for smoothing of the data based on fine amount & year

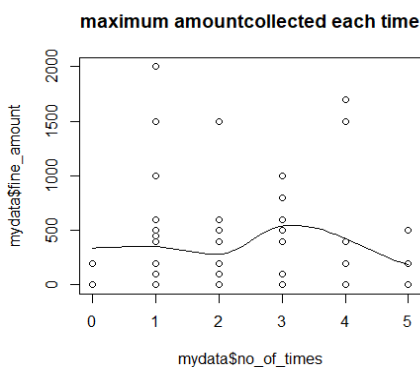


Fig 3 result for naïve Bayes classification for smoothing of the data based on fine amount & year

The fig 4 indicates the percentage or count that each complaint type based on the students and it displays counts no of students involved in that complaint.

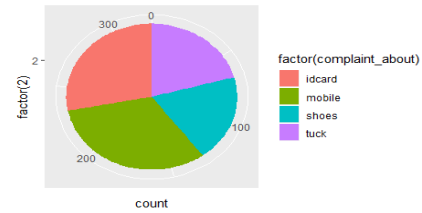


Fig 4 pie-chart representation of the complaint time with the count based on handover to the committee.

By DataExplorer the first function that we will be using is to learn about the structure of our Data-set. This displays all the questions asked in the form of nodes in a tree. All the questions that are represented as nodes are all connected to the root node which holds the integral structure of our tree. In order for us to identify the structure we will be using the following command in 'R'...

```
plot_str(data)
```

On execution of this command we will get the following output Fig 5. The structure of data is:

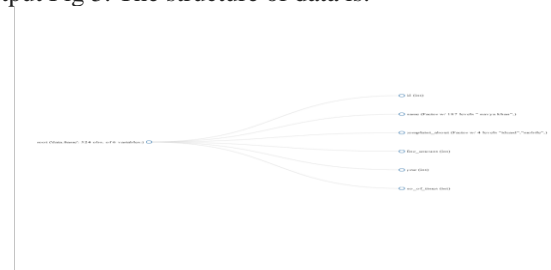


Fig 5 indicates the structure of the data in the dataset in the form of the tree with the nodes as attribute names

Fig 6 displays the plot after constructing tree for the dataset we are finding the missing values in the dataset

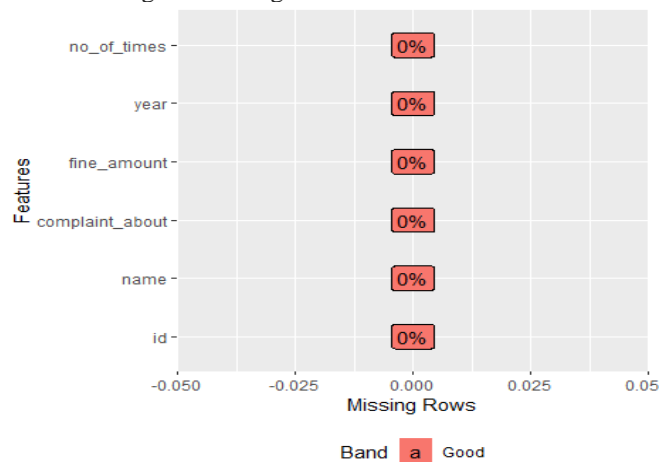


Fig 6 indicates no null values and no missing values with in the dataset

By the DataExplorer package we can able to create a report and plot the graph by generating the percentages for memory usage, data structures in tree format, correlations, histograms, bar charts based on attribute, Principal Component Analysis specifications. From fig 7,8,9,10,11 these plotting are occurred after creating the report with our dataset and visualizing each attribute and results will be in a pictographically.

Basic Statistics

Raw Counts

Name	Value
Rows	324
Columns	6
Discrete columns	2
Continuous columns	4
All missing columns	0
Missing observations	0
Complete Rows	324
Total observations	1,944
Memory allocation	22.6 Kb

Fig 7 displays the memory usage of the each every row and columns including missing values in rows and columns

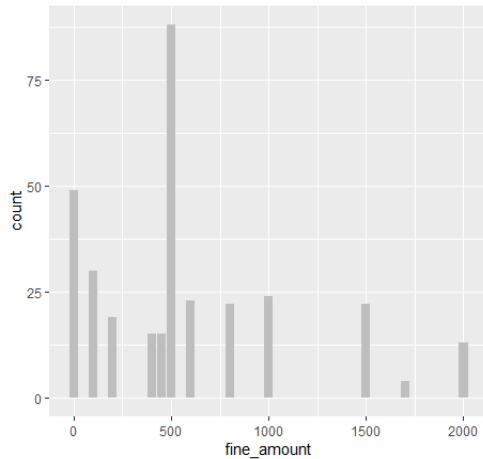


Fig 8 displaying the count of same amount collection

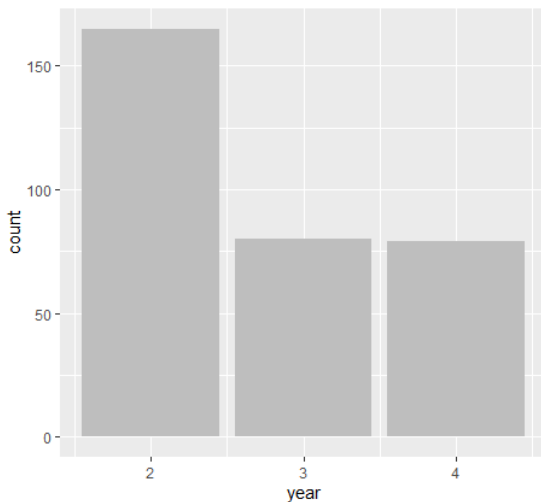


Fig 9 Displaying the count of students based on the year

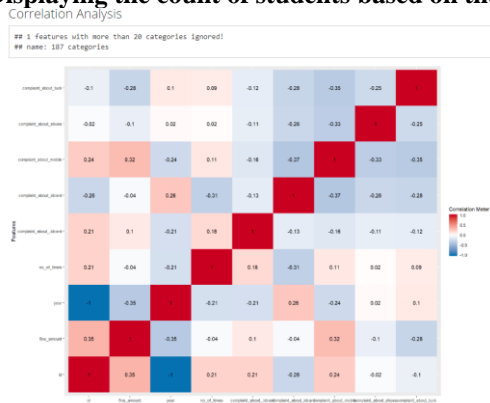


Fig 10 correlation analyses on each feature in the dataset

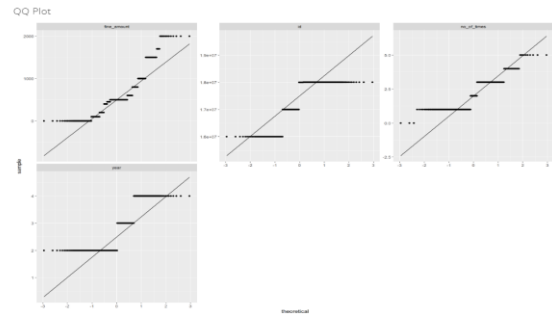


Fig 11 Quantile Quantile plot by considering 2 different attributes each time both theoretical and sample data

VI. CONCLUSION AND FUTURE SCOPE

This project is very much useful as it contains all the student details that who were caught under disciplinary actions. After paying fine or submitting a proper explanation from the student we remove his record and maintain the status as resolved. In further we can also use neural networks for checking the errors with the help of backtracking and feedforward method. It is easy to analyze the data in the form of pictographically manner. Online process of sending or generating the data via Internet is much easy and time reducing process and will not be take much man power for writing the data into the books. The automated storing of the complaint into the database without man power will increases the work efficiency and decreasing the mistakes that are occurred usually, and also helps for the further process of analysis in deep learning or other learning techniques too.

REFERENCES

1. Dalia Ahmed Refaat Mohamed, Dr.Mohammed Mahmoud Sakre A Performance Comparison between Classification Techniques with CRM Application, SAI Intelligent Systems Conference 2015 November 10-11, 2015 | London, UK
2. Suhardi, Novianto Budi Kurniawan, Deni Prayitno, Jaka Sembiring, Purnomo Yustianto, Public Complaint Service Engineering based on Good Governance Principles, A Case Study at the Government of Cimahi – West Java, Indonesia, 978-1-5090-6255-3/17/\$31.00 ©2017 IEEE
3. Jin-lan liu, jian kang, yin bai, xin zhang, the study of customer complaints management based on system dynamics: modeling and simulation, proceedings of the fifth international conference on machine learning and cybernetics, dalian, 13-16 august 2006.
4. Pattamaporn Kormpho, Panida Liawsomboon, Narut Phongoen, Siripen Pongpaichet , Smart Complaint Management System, Faculty of Information and Communication Technology Mahidol University Nakhon Pathom, Thailand, 2018 Seventh ICT International Student Project Conference (ICT-ISPC).
5. Zurah Binti Abu,Fadilah Ezlina Binti Shahbudin ,Mastura Binti Mansor ,Nurul Zaherah Binti Abd Rahim, Kampus Jasir Melaka Nur Aqilah Binti Norwahi ,MARA Melaka, Kampus Jasir IMPROVING USER COMPLAINT MANAGEMENT SYSTEM AND SATISFACTION LEVEL VIA READER-FRIENDLY LINGUISTIC FEATURES, 2015 International Symposium on Mathematical Sciences and Computing Research (ismsc).
6. S.Sangeetha Ravichandran, D.Sathya, R.Shanmugapriya, G.Isvariya, RULE-BASE DATA MINING SYSTEMS FOR CUSTOMER QUERIES, I CCCNT' 12 26 -28 July 2012, Coimbatore, India, IEEE-2018.

AUTHORS PROFILE



Ms. Arpita Roy is an Assistant Professor in Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, A.P. She has 02 years of teaching experience and pursuing PhD. She has completed her Mtech from Maulana Abdul Kalam Azad University, Durgapur, W.B. and Btech from West Bengal University of Technology, W.B. Her area of interest is Artificial Intelligence, Machine Learning. She is currently working in the area of Soft Computing. She has also published & presented papers in refereed journals and conferences. She is a member of CSI and IAENG.



Dr. Nikhat Parveen is an Associate Professor in Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, A.P. She has more than 10 years of teaching experience and 6 years of research experience. Her area of interest is Security Software, Security Testing Software Engineering, and Requirement Engineering. She is currently working in the area of Soft Computing and Big Data Security Optimization. She has also published & presented papers in refereed journals and conferences. She is a member of ACM, CSI and IAENG.



P.Rama Bhargavi, Student of Computer Science and Engineering in Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur, Andhra Pradesh, India.



A.Navya, Student of Computer Science and Engineering in Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur, Andhra Pradesh, India.



A.Pavan Kumar, Student of Computer Science and Engineering in Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur, Andhra Pradesh, India.