Team Self-Evaluation Report

Aniruddh Gupta
Dept. of Computer Science and Information Systems
BITS Pilani, Hyderabad Campus
Hyderabad, India
f20170149@hyderabad.bits-pilani.ac.in

Anushray Mathur
Dept. of Computer Science and Information Systems
BITS Pilani, Hyderabad Campus
Hyderabad, India
f20171570@hyderabad.bits-pilani.ac.in

Anuj Kharbanda
Dept. of Computer Science and Information Systems
BITS Pilani, Hyderabad Campus
Hyderabad, India
f20171508@hyderabad.bits-pilani.ac.in

Vidish Bharadwaj
Dept. of Mechanical Engineering
BITS Pilani, Hyderabad Campus
Hyderabad, India
f20171391@hyderabad.bits-pilani.ac.in

We started the work on this project as soon as the team was finalised, with our first team meeting dating back to 21st January 2020. For the first submission due on 30th January which required us to provide the selection of our dataset along with a data report, we finished the work well on time by 26th January and kept the last 4 days to refine, recheck and fine tune the details of our submission.

Our next team meet happened on Monday, 3rd February and we shifted our focus on the midterm submission which was due on 16th February. We divided tasks amongst ourselves and started working on them right away. We held a meet every 3-4 days, updating each other on the progress made and giving/taking suggestions if required. All the tasks for our midterm submission which included – Data Cleaning, Data Transformation, Data Visualisation and Exploratory Data Analysis, were completed by 14th February. Thereafter we rechecked all the codes and graphs made and made some minor changes as deemed fit by us as a team. After finalizing the codes and graphs finally the midterm report was made following which our midterm submission was ready.

After the midterm submission we took a break to prepare for the midsem examination. The college closed and the lockdown started before we could hold another meet and we were forced to continue our work on the project from our homes. We divided work to be done individual and have been constantly in touch with each other to suggest and improve each other's work.

Before the project all four of us were not completely comfortable and experienced with working on python and the complete data mining project. But working on the project has made each one of us confident up to a certain degree in both the complete process and the grasp over the language to perform data mining tasks. The project also helped us understand the importance of team-work as well as time management.

INDIVIDUAL CONTRIBUTIONS:

Anushray Mathur (2017A7PS1570H) – I, initially worked on Exploratory Data Analysis and created the program to perform the chi-square test on the data and find out relationships between different attributes along with which I also drafted and edited the midterm report. After the midterm submission, I started work on performing Association Rule Mining on the dataset using different combinations of minimum support, minimum confidence and lift. Finding this perfect combination of the three was an arduous task and required some time. Finally found an optimum combination and discovered 12 Association Rules in the data set. I also worked on compiling the final report and editing it into the IEEE format as well as creating and editing the self-evaluation report.

Aniruddh Gupta (2017A7PS0149H) - I applied KNN model to predict the correct call drop category. Major challenge was to identify whether to apply classification or clustering. But we decided to apply classification as there were no visible clusters when we plotted the data. Next problem I faced to get the optimal value of K in the algorithm. I used a sample testing set and ran the algorithm for all the primes between 1 to 1001 and selected the prime with highest accuracy. Using a random small sample was necessary because of the time constraint. It took me a lot of time to maximise the accuracy as even a small mistake resulted in wastage of 3-4 hours because of executing the faulty algorithm.

Vidish Bharadwaj (2017A4PS1391H) - I worked on the data visualisation part of pre-processing. 23 individual plots for each month's data were created to get an insight into the data through visual cues. Deciding which attributes to visualize to get more prominently visible patterns was difficult and required a considerable amount of attempts. I also helped with the documentation of the final submission, the demo video and some fine tuning of the association rule mining parameters to get the optimal result.

Anuj Kharbanda (2017B4A71508H) - Initially I worked on Exploratory Data Analysis and wrote a program to find Pearson and Spearman correlation between rating and call drop category and find out the relationships between them. After the midterm submission, I started working on Operator Classifier problem. I applied the Decision Tree classification model (created Decision Tree and Random Forest) to predict the best operator. One of the main challenges was hyperparameter tuning, i.e., find the optimal values of hyperparameters like Max. Depth in the decision tree and No. of forests in the random forest. These are found by hit and trial method. I found the accuracy by varying Max. Depth and creating a tree and repeating the process again and again. Similar process was done by varying no. of trees in the random forest and calculating accuracy for different values.
Another challenge that I faced was that initially the algorithm was run generally without giving any preference to any feature, but due to that it was taking a lot of time to execute even for a single month.
But later I found that if we fix state as base criteria for the decision tree then it has to search in a small subset of the dataset and the algorithm was very much faster than the original one. So, till we match with a state at some node we recursively run the algorithm with only state name as the feature. Once we find a state, we run the algorithm on all the remaining features. This was done to reduce time complexity while getting almost the same result with almost the same accuracy. I also maintained the GitHub repository for the code.