CS7CS4 Research Assignment 1

# Team ID : 25 Task ID : 104 - Data Amount

|  |  |  |
| --- | --- | --- |
| Bhaskar Rao raob@tcd.ie 18300829 | Bhavesh Mayekar mayekarb@tcd.ie 18301407 | Hamid Hassani hassanih@tcd.ie 15338952 |

**Contribution**

All the team members have contributed equally during the project. The team met every second day at the Lloyd building to brainstorm, plan project and discuss the results. Team finalized the 3 data sets to be used for the research question (available on the Github repo). Team then distributed data sets between themselves (1 for each member) & started their analysis. Individual work contribution has been detailed below:

**Bhaskar Rao:** He worked on the public “Bike Sharing Dataset” from UCI machine learning repository. He did EDA, data cleaning, feature engineering on the data set before implementing machine learning models to predict the count of bikes rented. He the analyzed the effect of training data size on the accuracy of the model. He discussed his results & findings with the team and asked for help whenever required.

**Bhavesh Mayekar:** He worked on the publicly available “Census Income Data Set” from UCI machine learning repository. He performed various data cleaning and feature engineering operation on the dataset and implemented various machine learning models. The aim was to predict the whether an individual has a salary above 50k based on various input parameters. He tested various models with varying sizes of training datasets and analyzed the accuracy. Later the results and findings were discussed with the team.

**Hamid Hassani:** He worked on the publicly available “Bank Marketing Data set” from UCI machine learning repository. He did detailed EDAs to understand the data and shared insights with other team mates. He implemented multiple classification models to predict if a customer will subscribe to a term deposit based on marketing campaigns ran by the bank. He used this data set to test the effects of data amount on accuracy of the models. He presented his results and incorporated team’s feedback to further improve his analysis.

Apart from the individual analysis, team was also reviewed each other’s code and methodology. Team also collaborated to complete the research paper. Each individual prepared his own points for the paper and sat together to discuss and consolidate the key points using which the paper was finalized.

**Work Count**

Word count excluding cover sheet, title, author names, tables and figures, references, acknowledgement, and appendix is **XXXXX**

**URL to source code repository**

<https://github.com/bhaskarrao511/CS7CS4--task-104--team-25/tree/master/sourceCode>

**URL to source code repository activity**

<https://github.com/bhaskarrao511/CS7CS4--task-104--team-25/graphs/contributors>

**commit activity of team members**

Add picture here

*We would like to thank Dr Joeran Beel and Dr Douglas Leith for their teachings without which this research would have been impossible*