Springboard

You've passed your capstone project!

SUMMARY

The impact of this project is significant and it can contribute in understanding risk factor for traffic violations and accidents using statistical analysis, machine learning and data science. This project uses Maryland's Montgomery County Traffic Stop Database in order to show the applications of Data Science in forecasting and future prediction. Experiments are designed and executed in order to generate actionable insights. You have tried all the standard machine learning algorithms which is a good practice to learn and evaluate them. Experiments have shown good prediction results using Logistic Regression and Decision Trees algorithms which can predict whether certain data points and information can lead to traffic accidents with an accuracy of 77%. Overall you did a great job completing this project.

COMPLETION

All the deliverables are available on GitHub including Initial Project Ideas, Project Proposal, Data Wrangling, Exploratory Data Analysis, Data Story, Milestone Report and In-depth Analysis on Python Notebook.

PRESENTATION

The report and presentation are clear and comprehensive.

UNDERSTANDING AND PROCESS

You selected a problem which has practical application in understanding risk factor for traffic violations and accidents. The data utilized in the project shows understanding of how to acquire, merge, wrangle and clean data. Maryland's Montgomery County Traffic Stop Database is a well-chosen data source relevant to the problem. The technical framework you designed covers all the main components of Data Analytics such as (1) Select and utilize appropriate algorithms for this application (2) Apply standard EDA techniques such as histograms, scatter plots and hypothesis testing (3) Utilized machine learning algorithms for prediction (4) and confusion matrix classification accuracy as evaluation metric/technique for your algorithm. Your project demonstrates skill in communication of the project results that includes (1) Utilizing data visualization (2) Developing a story to present your hypotheses (3) and outcome analysis in a well articulated data story relevant to the target audience. Your code is well documented, clean and contains output for each step. Each step has its own code file which is good for code usability.