M6 Project

Johnathan Wang, Nicholas Charles Vitellaro, Makenna Shae Owens, and Marc Mahanna   
COSC523: Artificial Intelligence, November 10, 2023

***Abstract* - This project focuses on use machine learning approaches to implement models predicting which passengers survive the Titanic shipwreck. The format of this project is that of a Kaggle competition** [1]**. Goals of this project are to 1) implement a baseline score and 2) try to improve on the baseline score.**

**DESIGN CHOICES AND IMPLEMENTATION**

The approach selected was to implement a baseline and an improved model intended to improve score beyond the baseline. RandomForrestClassifier is the default implementation and is provided in the Kaggle competition tutorial [2]. The improved model…

For each implementation, data was loaded, data was cleaned, features were explored and selected, models were trained, and test results were produced.

The baseline model followed the Kaggle tutorial. Data cleaning and preparation was minimal as the selected features ‘Pclass’, ‘Sex’, ‘SibSp’, and ‘Parch’ do not have missing values. Feature exploration in the baseline model produced interesting results of 74% of women surviving and 19% of men. The RandomForrestClassifier model was initialized per the tutorial with n\_estimators=100, max\_depth=5, random\_state=1. Resulting performance when submitted to Kaggle was 77.511%.

The improved model … Data cleaning and preparation … Feature exploration in the improved model produced … The … model was initialized … Resulting performance when submitted to Kaggle was …%.

**CHALLENGES AND OBSTACLES**

The project group members are familiar with machine learning concepts and have prior experience with all three selected models. Use of standard practices of data exploration, cleaning, and normalization were employed making exploration of model performance achievable. Overall, the project was not fraught with any significant challenges or obstacles.

**SUMMARY**

This project provided the opportunity to compare the performance of different machine learning models. A consumable data set and a challenge issued in the form of a competition provided an enriching and rewarding experience for the project group.

**REFERENCES**

|  |  |
| --- | --- |
| [1] | "Titanic - Machine Learning from Disaster," [Online]. Available: https://www.kaggle.com/competitions/titanic/overview. |
| [2] | "Titanic Tutorial," [Online]. Available: https://www.kaggle.com/code/alexisbcook/titanic-tutorial. |