Name: - Haixiao Lu

ID: 113029434

Semester: - FALL– 2021

Course: - DSA5900 Professional Practicum (4 credit hours)

Faculty Supervisor: - Dr. Audrey Reinert

Mail: hx.lu@ou.edu

**Project Title**

1. Introduction

A central challenge facing the Federal Aviation Administration (FAA) is how to determine if an airman will be flight ready. Advancements in machine learning technology and improvements in text mining capabilities allow for researchers and federal authorities to examine a variety of data in new ways. This project proposal will focus on data mining medical records to recommend who should get their pilot’s license renewed and for how long based on features in the data.

2. Objective

Project Objectives: -

* Analyze the existing data to process medical into useful features by using text analysis (Natural language processing).
* Work with Subject Matter Experts to determine how they make this assessment
* Train a machine learning models based on processed medical record data.
* Identify which model produce the best predictions
* Visualizing the outputs.

Learning Objectives: -

* Become more familiar working with text data using different natural language processing techniques and other algorithms to learn more about the relationship between extracted features and output labels.
* Acquire experience by working with real-world data to know how it affect people’s day to day life.
* Carry out a complete modeling process, starting with the problem understanding, through data analysis, model training and testing, and ending with the model’s validation.
* Finally, to get experience with a professional team that is working towards making peoples life better using technology, in my case data science.

3.Plan

* The dataset will be provided by DISC team. It will contain medical records.
* The data format will be like collection of general words may be in any language.
* To achieve the objectives, the textual data will be thoroughly preprocessed using different algorithms like natural language algorithms and analyzed to get efficient model.
* Once features from the transformed data are extracted, selected, and analyzed, several machine learning models will be trained with an iterative process.
* Lastly, visualizations will be made to make it understandable to an average user.

4. Deliverables: -

The student is expected to submit the following deliverables.

|  |  |  |
| --- | --- | --- |
| Task | Description | Due dates |
| Attend Weekly meetings | Attend weekly Zoom meeting with Dr. Reinert to discuss project status | Established meeting time |
| Submit Weekly timesheet | Submit a weekly "Timesheet" detailing which tasks were accomplished | End of work week |
| Midterm Presentation | Present a midterm status presentation to the DISC Team | October 15th |
| Midterm Report | Submit a midterm status report to the DISC Team | October 15th |
| Final Presentation | Present a final presentation to the DISC Team | Week of December 6th |
| Final Report | Submit a final report to the DISC Team | Week of December 6th |
| Functional prototype | Functional Python prototype which classifies medical records | Week of December 6th |

5. Schedule: -

|  |  |
| --- | --- |
| Week of | Tasks |
| 8/30/2021 | General team introduction/paperwork |
| 9/6/2021 | Access code repository, read project documentation |
| 9/13/2021 | Coding |
| 9/20/2021 | Coding / Work with SMEs |
| 9/27/2021 | Coding |
| 10/4/2021 | Developing Initial Results |
| 10/11/2021 | Coding |
| 10/18/2021 | Coding |
| 10/25/2021 | Algorithm Refinement |
| 11/1/2021 | Algorithm Refinement |
| 11/8/2021 | Present second formal prototype to team |
| 11/15/2021 | Coding data visualization |
| 11/22/2021 | Coding data visualization |
| 11/29/2021 | merge into main Repository/Delivery to SME |
| 12/6/2021 | Submission and demo of final functional prototype |