**Individual Self-Assessment- Capstone project**

* **Neeraja Jayaraman**

**Summary of Project:**

The Project selected was an analysis of heart disease risk in female patients based on demographics, pre-existing conditions, personal habits and health parameters. A binary classification machine learning model was built to predict the likelihood of a female patient to get heart disease, by training the model on the existing dataset of pre-existing conditions and parameters.

The analysis explored the relationship of the different parameters with heart disease risk. It was discovered that a person’s risk increased with age and even more so when the age factor was combined with Hypertension, high cholesterol, and high BMI. Hypertension was found to be the most significant factor influencing heart disease risk. As for machine learning, it was discovered that with the current dataset, a Decision Tree classification model was the most optimal model in predicting heart disease risk with accuracy as well as with good precision and recall scores. It was concluded that with availability of more data and additional features, it would be possible to build a more robust machine learning model to predict heart disease.

**Self- Assessment:**

My roles and contribution during the various segments of this project were as follows:

* **Segment 1: Triangle role**: Machine Learning Model: Identifying the value to be predicted, planning the approach, performing exploratory data analysis, building a preliminary machine learning model, attempting different approaches to gaining a good accuracy score. Preliminary Machine Learning model proposal document was prepared.
* **Segment 2: Square role**: Machine learning Model: Since there was a change in project topic, all the steps performed in Segment 1 were repeated. Preliminary exploratory data analysis was done, a simple machine learning model was built and good accuracy scores were generated. The Machine Learning Model documentation was updated.
* **Segment 3: X Role and Machine Learning Model:**  Full Exploratory Data Analysis was completed. Multiple combinations of machine Learning Models were attempted along with resampling, feature selections and train-test splits to arrive at optimal model with balance between accuracy, precision and recall scores. The Machine Learning documentation was completed. The X role was assumed by all team members as we split and tested the modules we owned.
* **Segment 4: Square role, X role and Team roles:** Roles were shared among the members as all worked together on practicing the presentation. My primary shared duties were to clean up the README as well as reviewing the rubric to ensure project meets the requirements.

Besides my individual roles, my contribution to the team included:

1. Strategizing the approach to be followed and technologies to be used for each deliverable via team discussions.
2. Project planning with timelines and action items via team discussions.
3. Ensuring the rubric was followed each week.
4. Coordinating with the team to ensure check-ins and merges on GitHub are not conflicted.
5. Contribution towards planning the flow and content of presentation.

**Personal Challenge:**

The biggest personal challenge encountered was researching various reasons and possible solutions to address a low recall score in machine learning model prediction. Multiple approaches were followed such as various feature selection techniques, resampling techniques, class weights and moving threshold methods to address class imbalances. The most optimal module that balanced the accuracy score and precision and recall scores was finally recommended along with recommendations for future analysis, for cleaner and larger datasets.

**Team Assessment:**

The following strategies were followed by the team:

* Project Planning and assignment of tasks based on individual strengths
* Team involvement in every deliverable when any challenges were faced by the person in charge.
* Regular communication and meetings helped with status updates and issue resolutions.
* Planned and Coordinated check ins and branch merges on GitHub so that latest versions are not lost.
* Prompt and timely completion of tasks helped with weekly submissions on time and also handover of tasks from one team member to another.
* Sharing of roles: All team members shared roles during one week or the other especially when the deliverable was on the critical path. This way the deliverable did not hold up any of the other member’s deliverables and tasks were completed on time.

**Team Challenges:**

The challenge faced by the team was when the project topic had to change after week 1. But the team took ownership of the challenge together instead of placing blame. Proper planning and prompt action in completing the deliverables from week 1 and week 2 helped the team get back on track in spite of the setback. Lessons were learnt on how to pick datasets and what actions to perform first to help with choosing the dataset. The flow of tasks and what should be done first and next was not known earlier but learnt as part of the experience.

**Team Strengths and Lessons learnt:**

1. Regular Communication
2. Thoroughness in planning
3. Promptness and speed in completion of tasks especially when deliverable is on critical path.
4. Division of roles, responsibilities, and action items.
5. Helping each other when there were issues.
6. Learning the flow of tasks – ie. What needs to be done first and what follows next.
7. Learning how to pick a dataset – our experience taught us to perform an Exploratory data analysis first and create a simple Machine Learning Model to see if it would yield a good accuracy score, before delving deep into all the project deliverables.