Project Charter

Vision: Conventional methods of evaluating wine quality are primarily based on the subjective experience of tasters. This project not only explores the relationship between physiochemical properties of wines and their taste, but also intends to utilize machine learning methods to figure out which physiochemical properties make a wine taste "good."

The vision is clearly stated. From the vision statement, the project will be attractive to all wine-lovers especially those who work in winery industry. I think the stakeholder can understand the meaning of the project from the vision statement here.

Mission: Enable users to predict quality of a wine based on physiochemical properties of the wine.

The mission statement matches the goal of vision. According to my knowledge about wine quality prediction, I believe this project should be able to get finished within the time. Maybe it could be clearer about the outcome from the web app.

Success criteria:

- (Machine Learning) The accuracy of the selected model on testing dataset should be at least 80%.
- (Business) User retention rate should be at least 60%.
- (Business) Those users who have tasted the wine they try to predict via the web app will be asked whether they think the prediction is accurate. The project is considered successful if the satisfaction rate is greater than 70%.

What's the metric that should be used to measure "accuracy"? The threshold for each performance metrics is clearly specified and all metrics seem to be reasonable for the project.

Backlog

Develop Themes:

Wine lovers always hope to spend their money on wines they will actually enjoy. However, without actually tasting the wine, it is often challenging for people to predict wine quality ahead of time. Even though a few famous wine critics do provide ratings on wine, those scores are still based on their subjective experience.

The aim of this project is to investigate how the taste of wines are related to their measurable physiochemical properties (e.g. pH, volatile acidity, residual sugar, sulfur

dioxide, etc.). Through various machine learning methods, the project will help to identify the most physiochemical properties that make a wine taste good as well as allow users to predict the quality of wines based on their physiochemical properties. In this way, users will be able to assess wine quality from a more objective perspective and make better buying decisions.

The final web app would allow user to input values of the most important physiochemical properties of wine (either white or red) and return the predicted wine quality.

The develop theme is clearly stated with project propose and final outcome of the web app.

Epic 1: Data Preparation

- Story #1: Downloading the data files (red wine & white wine) from UCI Machine Learning Repository (1 pt) **(Planned for the next 2 weeks)**
  - **Backlog**
  - Link to data source:[https://archive.ics.uci.edu/ml/datasets/wine+quality]
- Story #2: EDA and Data Cleaning (2 pts) **(Planned for the next 2 weeks)**
  - **Backlog**
  - Conducting exploratory data analysis (variable structures, distribution of variables, etc.)
  - Performing data cleaning (null values, outliers, etc.)

Epic 2: Modeling and Model Selection

- Story #1: Building classification models on randomly selected training data using Random Forest, XGBoost, and Neural Networks (4 pts) **(Planned for the next 2 weeks)**
  - **Backlog**
- Story #2: Comparing modelings based on performance metrics (Accuracy) on testing data (2 pts) **(Planned for the next 2 weeks)**
  - **Backlog**
- Story #3: Exploring additional models (e.g. CNN, SVM)
  - **Icebox**
- Story #4: Reviewing models with QA partner (4 pts)
  - **Backlog**

Epic 3: Online Deployment and Testing

- Story #1: Web app UI design (8 pts)
  - **Backlog**
- Story #2: Deploying web app (Flask) on AWS (8 pts)

    - **Backlog**
- Story #3: Creating an RDS instance (4 pts)
    - **Backlog**
- Story #4: Testing (Unit tests and Configured reproducibility tests) (8 pts)
    - **Backlog**

Epic 4: Final Presentation

- Story #1: Presentation slides (4 pts)
    - **Backlog**

There is no major task missing in the epics and stories and each story is actionable with clear expectation of the result. The priority of each task seems to be reasonable at this point of time.