## horizontal line

Wine Quality Predictor

**─**

ENEL 682: Machine Learning Project Proposal

Ahmad Masri

UCID: 30115844

**Proposal**

I am writing this proposal to the owners of the winery “Masri Wine Ltd.”, where I believe my project can be of great benefit to your winery. As you know, ensuring consistent and high-quality wine is critical to maintain a positive reputation in the industry. Manually evaluating every bottle to ensure it meets your high standards can be a tedious process considering the large amount of wine produced on a yearly basis. For that, I am proposing a wine quality classifier that uses machine learning algorithms to assess the quality of your wine.

­­­**Methodology**

By analyzing key parameters and features, which are fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol, the classifier will be able to quickly and accurately identify bottles that meet your quality criteria and those that do not. The term “quality” will be the target label. To develop the classifier, I will gather data from your winery (Kaggle dataset) which will be used to train the machine learning algorithms to accurately assess the quality of your wine. Once the classifier has been developed, it can be integrated into your production process to automatically assess the quality of each bottle. This will save time and resources by reducing the need for manual evaluation and will help to ensure that only the highest quality wine is released to the market.

3 machine learning models will be trained and then compared based on accuracy scores, to find the best classifier. Depending on the data in hand and the ML model selected, preprocessing may include scaling, encoding, cleaning, transforming the original data and tuning different hyperparameters do get the best model.

**Dataset**

I will be using a public winery dataset to test my models.

If I happen to use other datasets later, I will include them in the project report instead of here. Kaggle: Red Wine Quality; Source: https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009

**Models**

The machine learning models suggested for such a project would be supervised ML classifiers such as Random Forest Classifier, deep feed-forward neural networks, and SVM classifier. The libraries to be used would be sklearn, tensorflow, pandas, numpy, alongside plotting libraries like matplotlib and seaborn to display our results.

**Framework**

The main framework suggested for each ML model, although it may vary during implementation: Preprocessing -> Data splitting -> Hyperparameter tuning -> Model

implementation -> Validation -> Visualization

**Components**

Types of preprocessing: Encoding Categories, Feature Scaling, Replacing missing data

Hyperparameter Tuning: Depending on Model, for ex. For Random forest: num of trees

Models: Random Forest Classifier, SVM Classifier, Deep feed-forward Neural Network

Validation Metrics: Validation Score, Accuracy score, F1-Score, Precision, Recall

Visualization: Using python plotting libraries like matplotlib or seaborn