## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Project- Predicting Wine Quality using Wine Quality Dataset**

[Url:-](about:blank)

JUPYTER NOTEBOOK-

[Wine.ipynb](https://drive.google.com/file/d/1epu5xAvIASrifNGbAfRiZnAOcH1Z5vgJ/view?usp=drive_link)

CSV FILE-

[Wine csv](https://drive.google.com/file/d/1BZ0KWXqSjqGEds4nKiggc17xAfLxgeWN/view?usp=drive_link)

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Yajat Dureja, Yajur, Vipash

Arora

Roll Number: -2210990970

2210990971

2210990964

Group - 14

**Department of Computer Science and Engineering,** Chitkara University Institute of Engineering & Technology

**Wine Quality Classification Documentation**

**Overview:**

**This document provides documentation for the wine quality classification project, including data preprocessing, exploratory data analysis (EDA), model building, and feature importance analysis.**

**Importing necessary libraries:**

* **numpy: Library for numerical computations in Python.**
* **pandas: Library for data manipulation and analysis.**
* **matplotlib.pyplot: Library for creating static, animated, and interactive visualizations in Python.**
* **seaborn: Library for statistical data visualization based on matplotlib.**

**1. Exploratory Data Analysis (EDA):**

* **Histograms: Visualized the distribution of each feature in the dataset using histograms.**
* **Kernel Density Estimation (KDE) Plots: Used KDE plots to estimate the probability density function of each feature.**
* **Boxplots: Generated boxplots to visualize the distribution of numerical data and identify outliers.**

**2. Data Preprocessing:**

* **Feature and Label Separation: Separated the features (X) and the target label (y) from the dataset.**
* **Train-Test Splitting: Split the dataset into training and testing sets using the train\_test\_split() method from sklearn.model\_selection.**

**3. Model Building and Evaluation:**

* **Standard Scaling: Applied standard scaling to the features using StandardScaler() from sklearn.preprocessing.**
* **Pipeline Construction: Constructed pipelines using make\_pipeline() from sklearn.pipeline to chain preprocessing steps and classification models.**
* **Classification Algorithms: Implemented various classification algorithms including:**
  + **Decision Tree**
  + **Random Forest**
  + **AdaBoost**
  + **Gradient Boosting**
  + **Naive Bayes**
  + **Support Vector Classifier (SVC)**
  + **Logistic Regression**
* **Model Evaluation: Evaluated each model's performance using classification metrics such as precision, recall, F1-score, and accuracy. The classification\_report() function from sklearn.metrics was used to generate detailed classification reports.**
* **Model Comparison: Compared the performance of different models based on their accuracy scores.**

**4. Feature Importance Analysis:**

* **Feature Importance Calculation: Calculated feature importances using the feature\_importances\_ attribute of tree-based models like Random Forest.**
* **Visualization: Visualized feature importances using horizontal bar plots.**

**Conclusion:**

* **Key Insights: Identified alcohol content as the most important feature in predicting the quality of wine based on feature importance analysis.**
* **Model Selection: Random Forest achieved the highest accuracy among the models evaluated.**