## **Artificial Intelligence and Machine Learning**

Project Report

Semester-IV (Batch-2022)

**PREDICTING WINE QUALITY**

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Yashika khanna-2210990978

Vanshita Garg-2210990949

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

**ABSTRACT**

* Data Collection:
  + The dataset includes a wide range of features that describe the chemical properties of wines, such as acidity levels, residual sugar, alcohol content, and pH.The data collection process involved ensuring data integrity, quality, and relevance to the problem at hand.Attributes related to wine characteristics and quality was gathered from reputable sources such as the UCI Machine Learning Repository or Kaggle.
* Data Preprocessing:
  + Handle missing values: Identify and handle missing values in the dataset. This might involve techniques like imputation (replacing missing values with a suitable estimate) or removal of rows or columns with too many missing values.
  + Encode categorical variables: Convert categorical variables (such as gender or port of embarkation) into numerical format using techniques like one-hot encoding or label encoding.
  + Feature scaling: Scale numerical features to ensure that they have a similar range and distribution, which can help improve the performance of certain machine learning algorithms.
* Exploratory Data Analysis (EDA):
  + Statistical summary of the data.
  + Data visualization (histograms, box plots, correlation matrices) to understand the distribution and relationships between features.
  + Identification of any outliers or missing values.
* Feature Engineering:

Feature selection techniques used.

Creation of new features if applicable.

Transformation of features (scaling, normalization).

* Model Selection:
  + Selection of appropriate machine learning algorithms for regression (e.g., Linear Regression, Decision Trees, Random Forest, Support Vector Machines, etc.).
  + Justification for the selected algorithms.
* Model Training:
  + Splitting the data into training and testing sets.
  + Hyperparameter tuning using techniques like GridSearchCV or RandomizedSearchCV.
  + Cross-validation to ensure the generalization of the model.
* Model Evaluation:
  + Evaluation of each model using appropriate metrics.
  + Visualization of model performance metrics (e.g., bar plots, ROC curves).
  + Interpret results: Analyze the model's predictions and errors to gain insights into its strengths and weaknesses.
* Results and Discussion:
  + Summary of the findings.
  + Interpretation of the model's predictive capabilities.
  + Insights into which features contribute most to predicting wine quality.

**CONCLUSION**

In conclusion, the endeavor to predict wine quality using machine learning techniques has yielded valuable insights and practical applications for the wine industry. Through rigorous data analysis, feature engineering, and model training, we have developed predictive models capable of estimating wine quality with notable accuracy. Our findings underscore the importance of leveraging machine learning approaches to assist winemakers in assessing and improving wine quality based on intrinsic characteristics. Moreover, the project has demonstrated the significance of data-driven decision-making in optimizing production processes and enhancing overall product quality. While our models have shown promising results, there remains room for refinement and further exploration, particularly in incorporating additional data sources or refining feature engineering techniques. Overall, this project contributes to the growing body of knowledge in the intersection of machine learning and winemaking, paving the way for future advancements in wine quality prediction and production optimization.