## Artificial Intelligence and Machine Learning

Project Abstract

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

**Title: Predicting Wine Quality using Wine Quality Dataset**

**Abstract:**

This project presents a comprehensive analysis of wine quality classification using machine learning techniques. Here's a breakdown of the key aspects covered:

1. Data Exploration and Understanding:
   * The dataset consists of 1599 samples with 11 physicochemical attributes and a quality rating.
   * Exploratory data analysis (EDA) techniques, including histograms, KDE plots, boxplots, and correlation heatmaps, are utilized to understand data distribution, identify outliers, and explore feature relationships.
2. Data Preprocessing:
   * The dataset is divided into feature variables (X) and the target variable (y).
   * Data preprocessing steps involve splitting the data into training and testing sets and scaling features using StandardScaler.
3. Model Building:
   * Various classification algorithms are implemented:
     + Decision Tree Classifier
     + Random Forest Classifier
     + AdaBoost Classifier
     + Gradient Boosting Classifier
     + Naive Bayes Classifier
     + Support Vector Classifier
     + Logistic Regression Classifier
   * Each model is trained on the training data and evaluated using metrics such as precision, recall, F1-score, and accuracy.
4. Model Evaluation:
   * Performance evaluation is conducted using classification reports, which provide a comprehensive overview of each model's performance across different quality categories.
   * Models are compared based on their accuracy scores to determine the most effective classifier for predicting wine quality.
5. Feature Importance Analysis:
   * Feature importance is assessed using the Random Forest Classifier to identify the most influential attributes in predicting wine quality.
   * Alcohol content emerges as the most crucial feature for predicting wine quality, as indicated by the feature importance analysis.
6. Conclusion and Recommendations:
   * The study concludes that machine learning algorithms can effectively predict wine quality based on physicochemical attributes.
   * Insights derived from the analysis offer valuable information for winemakers to optimize production processes and ensure wine quality.
7. Future Work:
   * Future research could explore more advanced machine learning techniques or incorporate additional features to further enhance prediction accuracy.
   * Moreover, integrating sensory data or expert evaluations could improve the model's ability to predict wine quality more accurately.