**WINE QUALITY DETECTION**

**Submitted for**

**Statistical Machine Learning CSET211**

Submitted by:

E23CSEU0468  **PRASANT MAAN**

**E23CSEU0460 JYOTISHMOY HAZARIKA**

Submitted to

**DR. NITIN ARVIND SHELKE**

**July-Dec 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**INDEX**

|  |  |  |
| --- | --- | --- |
| Sr.No | Content | Page No |
| 1. | Abstract |  |
| 2. | Introduction |  |
| 3. | Related Work |  |
| 4. | Methodology |  |
| 5. | Hardwar/software Requirements |  |
| 6. | Experiment Results |  |
| 7. | Conclusion |  |
| 8. | Future Scope |  |
| 9. | Github |  |

1.Abstract: The project essentially uses the techniques of Artificial Intelligence. During this investigation, physicochemical attributes were used to predict wine quality. The dataset consisted of several attributes comprising acidity, sugar content, pH, and alcohol percentage that were used in the determination of a predefined classification scale of machine learning models. Data preprocessing, feature selections, and careful experimentation with various algorithms about machine learning approaches therefore appear. It will act as a predictive system intended to assist wine producers and enthusiasts in making an efficient estimation of wine quality, thereby improving the quality control process and enhancing consumer satisfaction. The results show how AI improves decision-making in the wine industry.

2.Introduction: The significant role of wine quality assessment in the wine industry, in terms of both consumer satisfaction and market demand, can be seen. Traditionally, sensory analysis conducted by experts has been used to assess the quality of wine. Subjective in nature, this process takes long periods of time and is very costly. Modern developments in artificial intelligence and machine learning enable the design of automated systems capable of providing solutions for efficiently, objectively, and scalably predicting wine quality.  
  
This project involves the design of a machine learning system that forecasts the quality of a wine by its physicochemical properties - acidity, sugar levels, pH, alcohol content, and other quantifiable properties. These features will enable the model to classify wine quality in terms of a given standardized scale, thereby making the process more consistent and dependable.  
  
The study will involve pre-processing datasets, feature selection, and experimentation with different algorithms with high accuracy demonstration through various techniques using a set of machine learning algorithms. The project stands out in emphasizing the growth of AI applicability in the food and beverage sector, with data-driven solutions further expediting improvements in quality control and operational efficiency.  
  
This kind of system, in automating wine quality assessment, reduces dependency on manual processes and supports the winemaker by optimizing production and maintaining high standards.

3.RELATED WORK: This remains an active area of research in predicting wine quality using machine learning, with diverse studies focused on the relationship between physicochemical properties of wine and its perceived quality. Most prior work in this domain appears to be developed based on the application of supervised learning algorithms to be applied in developing predictive models that take advantage of public datasets such as the Wine Quality dataset available from the KAGGLE.

Various research studies have proven that proper features, such as alcohol content, residual sugar, and acidity, result in high accuracy for the prediction. Techniques like PCA and RFE are also well-known for selecting feature dimensionality.

Quite recently, deep learning models have started to be used for prediction in the quality of wine. Neural networks, particularly MLPs, capture complex nonlinear relationships between features, which may exist. However, this approach usually requires a more significant amount of dataset and computational powers.

These metrics, namely, accuracy and F1-score, are extensively used in judging the performance of the model. In all regression problems, absolute error (MAE) is a very prominent measure. Most of the articles in the literature are known to have issues with class imbalance in data for testing wine quality measurements, as scores for mid-range qualities are more abundant compared to the extreme high or low ones. Researchers have applied oversampling, under sampling, and weighted loss functions to address this challenge.

This project will expand on the existing literature by applying and comparing numerous machine learning models, addressing dataset imbalances, and fine-tuning hyperparameters for better performance, with the aim of improving predictive accuracy and offering a useful tool to automate wine quality assessment in real-world settings.

4.METHODOLOGY: This is a systematic project on developing a machine learning model predicting wine quality based on physicochemical properties. The methodology is a wide range of stages starting from data collection to preprocessing, training, and evaluating the models. Some of these include the following steps:

1.Data Collection

The dataset used for this project is the Wine Quality dataset from KAGGLE, which contains physicochemical measurements like pH, alcohol, and acidity, and the quality rating of red and white wines. According to sensory analysis by wine experts, this rating ranges from 1 to 10.

2.Data Preprocessing

Missing Value Handling: Missed or null values are examined in the data. Techniques for imputation or removal are appropriate when they are found.

3.Normalization and Scaling:

Feature normalization or standardization is used in the case of physicochemical attributes because the measurements are available on different scales; consistency is performed, and model performance is improved.

4.Class Balancing:

Techniques like SMOTE (Synthetic Minority Oversampling Technique) or class weighting during training of the model are utilized to handle the imbalanced ratings of quality for the class of wine.

5.Exploratory Data Analysis (EDA)

EDA was performed to determine the relationship of features, and which among the features will have more influence on wine quality. This has all been done by using visualization techniques in the form of scatter plots, heatmaps, and box plots on trends, correlations, and outliers if any existed.

6.Feature Selection:

Feature selection was done to zero down to the most influential attributes in terms of wine quality. Techniques that were involved in this include:

Correlation analysis to avoid major correlation of features.

Recursive Feature Elimination (RFE) to rank and select optimal features for model training.

Model Selection

A lot of machine learning models are tested to predict wine quality. The handful that are used here include:

Logistic Regression

Decision Trees

Random Forests

Each model will be trained and tested with comparisons based on accuracy, F1-score, and precision.

7.Model Training and Validation

Cross-Validation:

K-fold cross-validation. This is used to test the model and avoid overfitting.

8.Model Evaluation

Accuracy: It calculates total percentage of correct predictions.

Precision, Recall and F1-score: In case of imbalanced classes

Confusion Matrix: To understand true positives, false positives, true negatives, and false negatives.

Deployment and Testing

Lastly, the best model out of the selected ones is tested on a new unseen data set to prove the reliability of the system. The model is then wrapped up in a convenient form for practical application, like a REST API or web interface.

Conclusion and Insight

The results from the project give insights into the most important physicochemical features of a wine related to the quality while underpinning the efficacy of using machine learning on the domain. This structured methodology will ensure that the best practices in data science and machine learning are followed, thus permitting the proper development of an accurate and reliable wine-quality predictor.

5. Hardware/Software Requirements:

1. Hardware Requirement In order to ensure effective smooth development, training, and deployment of

the wine quality prediction model, the recommended hardware resources include:

Processor Intel Core i5/i7 or AMD Ryzen 5/7 (Quad-Core or more)

A stronger use of GPUs like NVIDIA GeForce GTX 1650 and above will be useful for increasing computation speed to handle large models for training.

Memory (RAM)

Minimum 8 GB

Recommended 16 GB or more

Storage

Minimum of 256 GB SSD

Recommended: 512 GB SSD and above (for fast read/write operations on data).

Graphics Card (Optional):

A dedicated GPU such as NVIDIA CUDA-enabled, which is necessary for deep learning frameworks like TensorFlow.

Internet:

To download datasets, libraries, and frameworks.

2. Software Requirements

Operating System: Windows 10/11, macOS, or Linux; and Ubuntu 20.04 and above, which is still recommendable for better compatibility to the Python libraries.

Programming Environment: Python 3.8 and above; it is the principal programming language of the project.

IDE/Code Editors: Visual Studio Code or PyCharm for development.

Libraries and Frameworks: Data Processing and Analysis, Pandas, NumPy Data, Visualization Matplotlib, seaborn, Machine Learning Algorithms.

Model Evaluation and Optimization

SciPy for statistical tests.

Version Control and Collaborationist for version control, GitHub/GitLab/Bitbucket for code repository and collaboration.

Tools: Anaconda: a distribution platform for dealing with various Python libraries and dependencies

Virtual environments, e.g. vend or Conda for isolated development.

Such hardware and software tools would make sure that the development and deployment of the wine quality prediction model would be very efficient and scalable.

6.EXPERIMENT RESULTS: The wine quality prediction model was developed using a dataset containing physicochemical properties and their corresponding wine quality ratings. The experiments involved data preprocessing, exploratory data analysis, feature importance analysis, and model evaluation. Below are the results of the experiments:

Here are the experimental results extracted from our project:

**Dataset Summary**:

The dataset contains 1599 samples and 12 attributes.

**Data Distribution**:

The quality attribute distribution (using sns.catplot) reveals the class frequencies for wine quality, indicating imbalance in the dataset. Most wines have quality ratings of 5 and 6.

**Data Statistics**:

**Fixed Acidity**:

Mean: 8.32

Min: 4.60, Max: 15.90

**Volatile Acidity**:

Mean: 0.53

Min: 0.12, Max: 1.58

**Alcohol**:

Mean: 10.42%

Min: 8.4%, Max: 14.9%

**Train-Test Split**:

Training data: 1279 samples

Test data: 320 samples

**Model Training**:

RandomForestClassifier was used to train the model.

**Model Performance**:

**Accuracy**: 94.06% on the test dataset, indicating strong model performance.

**Prediction Example**:

For a specific input, the model predicted **"Bad Quality Wine"** (class: 0).

**Limitations and Challenges**

* **Class Imbalance**:
  + The dataset's uneven distribution of quality ratings affected the model's ability to generalize across all quality levels.
* **Subjective Nature of Wine Quality**:
  + The dataset only included physicochemical properties, potentially overlooking subjective sensory factors.

7.CONCLUSION: Now, this new AI model developed for predicting wine quality may turn out to display excellent machine learning capabilities in the assessment of wines. The systems proved to predict wine quality with an encouraging level of accuracy after taking on considered datasets of wine attributes-chemical composition and sensory features were among those included. From the results, it therefore reflects that acidity, alcohol content, and sugar levels are some of the factors considered in determining wine quality based on expert knowledge in the domain.

The primary focus of the project is that the benefits of AI prove fully to automate the evaluation of quality wine, which therefore provides an economic and efficient alternative to the traditional methods of tasting. Improvement of the model can be achieved through feature engineering, going into more complex machine learning techniques, and combining data sources with sensory feedback or regional origins.

Ultimately, this AI solution could be a real treasure for wine producers, consumers, and researchers. It could help wine makers make better decisions and improve the quality control process.

8.FUTURE SCOPE: The wine quality prediction AI project opens several promising avenues for further development and analysis. Some of the areas to be developed include:

* Model Enhancement with Advanced Algorithms: The use of more advanced machine learning techniques such as deep learning or ensemble methods may improve the model's accuracy as well as the generalization of the model. Techniques such as neural networks might express much more difficult relationships between features and quality.
* Adding Sensorial Data, it adds professional wine tasters ratings to enhance the performance of the model. This would allow the model to use both the chemical and sensory aspects of wine evaluation, thus increasing its reliability in prediction.
* Feature engineering and data augmentation: More diverse samples can be added to the dataset or extra features such as terroir or environmental conditions during the harvest leading to stronger predictive power. Data augmentation techniques can also be applied to synthesize data when the real-world data is sparse.
* Real-Time Prediction Systems. Developing an AI-based quality prediction system for wine with real-time data input, such as online wine shops or production facilities, may immediately assess the quality of wines. It will guide the producers to optimize their production processes and assist consumers in better buying decisions.
* Multimodal Learning: Combining different sources of data, such as images-bottle labels or the color of the wine-and geographical data-region, climate, and even customer review data-might give a more holistic model of wine quality considering all the dimensions involved.
* Expanding the scope from quality prediction, building on individual preferences combined with wine characteristics shall open the door for personalized use. There is an opportunity for recommendations in line with a user's past preferences by the combination of wine quality prediction and user taste profiles as possible inputs to the AI system.
* Industry Integration: The AI model can be integrated into the wine industry to assist in quality control at production facilities. It would keep batches of wine consistent, thereby allowing early detection of changes in quality during the manufacturing process.
* Sustainability and Environmental Impact: Integration of ecological and sustainable considerations into the model could involve organic farming practices and carbon footprint so that producers will not only be concerned about the quality of the wine but also about the ecological footprint that a certain product might have in the future.
* With continuous innovations and improvement of the approach, this project will be capable of revolutionizing the wine industry as far as making more precise, efficient, and accessible quality assessments are concerned.

9.GITHUB LINK: https://github.com/PrashantMaan/SML-project/blob/main/Untitled1.ipynb