**Wine Quality Prediction**

**Pg. 1**

Pinot noir is a light red wine with scents of game, leather, mushroom/vegetal, violets, cherry, plum, and raspberry. It is produced in countries like New Zealand, Australia, the United States, Switzerland, and Romania. Pinot noir cultivation is complex due to its soil needs and chilly environment. It has earliest bud break and harvest dates, making winemakers cautious. High-quality wines are made from low yield and tiny fruit sizes. Growers manage this by monitoring water supply, planting in low-nutrient soil, and trimming vines to avoid overproduction. The grape is genetically complicated and susceptible to point mutations, leading to the production of different clones. New Zealand Pinot noirs are well-known worldwide, with the majority produced in the South Island of New Zealand. The output of Pinot noir in New Zealand reached a new height in 2019, surpassing Sauvignon Blanc.

**Pg. 2**

Wine quality is a crucial issue in the wine industry, with experts defining it based on their understanding of wine production and chemical composition. The chemical makeup of wine, including fragrance, flavor, and color, is influenced by factors like grape type, environmental conditions, fermentation strains, and viticulture practices. Physiochemical laboratory tests are used to describe wine characteristics, such as pH, alcohol content, total sulphur, and anthocyanin levels. Machine learning, a branch of artificial intelligence, has gained traction since the 1950s due to its efficiency and the abundance of high-quality datasets. Research in machine learning focuses on addressing real-world problems by breaking problems down into manageable chunks using algorithms. Studies have shown that machine learning models can predict red wine quality, with some models having higher accuracy rates than others. Comparisons of classification algorithms have also been conducted, with decision tree classifiers and machine learning models based on RF and KNN algorithms being used to determine wine quality.

This research aims to predict wine quality using machine learning techniques incorporating physio-chemical and chemical characteristics. However, challenges include limited sample size, data leaking, and a large number of features. To overcome these, synthetic data was created with comparable characteristics to original data, and six samples were set aside for model testing. Several feature selection techniques were applied to address the issue of a large number of features, and the findings were compared with 54, top 10, and six key features.

**Pg. 3**

General Information about the info we are going to read on the paper.

**Pg. 4-5-6-7**

This study analyzed chemical and physicochemical data from 18 New Zealand Pinot noir wines, focusing on their quality. The study found that 54 features were related to physiochemical data and 47 to chemical data. The wine was assessed by 22 wine experts, with an average experience of 18.2 years. The researchers also conducted a sensory analysis to understand the relationship between perceived quality and complexity. The study found that tannin complexity, harshness, bitterness, and astringency were negatively associated with wine quality, while softness was positively correlated with black glass. The study concluded that perceived varietal typicality, wine quality, and complexity are intertwined for wine experts.

The study utilized the Synthetic Minority over Sampling Technique (SMOTE) to generate enough samples for machine learning training. The dataset was split into two pieces, with one having 12 samples and the remaining six used as a testing dataset to protect data leakage. The SMOTE method is used to balance data by creating minority class samples that can be matched against the majority class. In this case, 18 samples were received, all of which were Pinot noir wines. A dummy data set was created with 1400 rows and 55 columns, with 12 rows from original samples and 1388 rows containing 0 values. The remaining rows were encoded with 0 (less and equal to 5.77) and 1 (greater than 5.77) values. The SMOTE algorithm used the KNN method to generate new samples, setting the minority class as set A and constructing a new minority class set equivalent to the majority class.

The study utilized various machine learning classifiers, including SVM, RF, DTC, GNB, XGB, KNN, AdaBoost, and SGDC, to predict wine quality using their default parameters. Model evaluation metrics we can use in order to check which is the better ML algorithm.

**Pg. 7-8-9**

Four Methods to train the dataset 🡪 XGB, Extra trees classifier, RF and Gradient Boosting Classifier. Six essential variables in Pinot noir wine were identified, with two being significant by all four methods. Ethyl octanoate, a member of the ester family, contributes to the sweet and fruity qualities of the wine. 4-ethyl-2-methoxyphenol, a volatile phenol, enhances the spicy and flowery qualities. The remaining four variables, Geraniol, ethyl-3-phenyl propanoate, Nerol, and 𝛽-Ionone, contribute to the wine's aroma, including caramel, apple-sweet, and flowery notes, honey, and violet and black berry notes.

AdaBoost classifier performed exceptionally well using XGB method features and essential variables post-machine learning analysis, while RF classifiers improved performance when trained on essential variables.

**Pg. 9-10**

This study aims to predict wine quality rating using machine learning techniques. The researchers used a SMOTE algorithm with 12 samples and tested it on various features related scenarios. The AdaBoost and RF classifier were found to be the best model for predicting wine quality after training and testing. The study highlights the importance of feature selection and the influence of essential variables on model performance. Machine learning techniques have been used to advance wine studies, with studies showing their ability to classify 4898 samples of Portugal white wine and predict quality based on high, normal, and poor wine. The study also validates results with oenological theory and uses a Recursive Feature Elimination approach.

Machine learning has been shown to accurately predict wine quality, with studies proving its effectiveness in various aspects. For instance, a study on sensory profiles and colour of Australian Pinot noir showed a correlation coefficient score of 0.96. Another study on red wine data showed a 91.04% accuracy rate. Additionally, a RF classifier was used to classify wine with 100% accuracy, and grapes with an overall 88.9% accuracy rate.