**Final Project draft Paper: Wine Quality Analysis**

**Abstract**

By examining physicochemical characteristics and their effect on quality ratings, this study investigates the variables affecting wine quality (Cortez et al., 2009).The study explores prediction models to improve quality evaluation and discovers critical characteristics that determine wine quality using machine learning techniques(Banjade & Gaire, 2021).In order to extract insights, the study uses datasets from reliable sources and utilizes statistical and analytical approaches(Zaza et al., 2023).

**Introduction**

Expert opinion, which is subjective and subject to variation, has historically been used to evaluate the quality of wine (Cortez et al., 2009). The goal of this research is to use machine learning models to provide a data-driven method for predicting wine quality (Banjade & Gaire, 2021). The study aims to ascertain the importance of chemical compositions in assessing quality by examining factors including pH levels, alcohol concentration, and acidity (Zaza et al., 2023).

**Literature Review**

Previous work has examined a number of approaches to wine quality evaluation, such as machine learning and statistical modeling (Cortez et al., 2009). Cortez et al. (2009), for example, used data mining techniques to estimate wine preferences based on physicochemical characteristics, emphasizing the importance of volatile acidity and alcohol concentration in establishing wine ratings. Alcohol was also found to be a key factor impacting wine quality in a research by Banjade and Gaire (2021) that used machine learning algorithms to predict wine quality. Further research on the influence of feature significance on quality prediction is conducted by Zaza et al. (2023).

**Methodology**

The UCI Machine Learning Repository provided the dataset used in this work, which includes wine quality ratings and physicochemical characteristics (Cortez et al., 2009). In data preparation, characteristics were normalized and missing values were handled (Banjade & Gaire, 2021). Metrics including accuracy, precision, and recall were used to train and assess machine learning models, such as logistic regression, decision trees, and neural networks (Zaza et al., 2023).

**Results**

According to initial research, there is the strongest association between wine quality and citric acid, volatile acidity, and alcohol concentration (Cortez et al., 2009). With an accuracy of 85%, the Random Forest model exceeded both Neural Networks and Logistic Regression (Banjade & Gaire, 2021). The results of the study support earlier studies showing that one important factor influencing wine quality assessments is alcohol concentration (Zaza et al., 2023).

**Conclusion**

According to Cortez et al. (2009), the study shows that machine learning may be used to predict wine quality, with alcohol concentration appearing as a significant predictor. To improve prediction accuracy, future research will investigate deep learning techniques (Banjade & Gaire, 2021; Zaza et al., 2023).

**References**

Banjade, H., & Gaire, S. (2021). Prediction of Wine Quality Using Machine Learning Algorithms. Open Journal of Statistics, 11, 278–289. https://doi.org/10.4236/ojs.2021.112015

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4), 547–553. https://doi.org/10.1016/j.dss.2009.05.016

Zaza, S., Atemkeng, M., & Hamlomo, S. (2023). Wine feature importance and quality prediction: A comparative study of machine learning algorithms with unbalanced data. arXiv preprint arXiv:2310.01584. https://arxiv.org/abs/2310.01584