**Project Name:** Classifying Wine Using Logistic Regression and SVM Models

**Course:** PROG39051 Machine Learning Techniques

**Team Members:** Abira Esther Demello, Xiaohong Deng

## Project Proposal

### Domain Problem:

In the world of winemaking, where each bottle is like a piece of art, figuring out and guessing wine quality is a tricky and multi-layered challenge. "Wine is bottled poetry," a fancy way to say wine is special and complex, captures this idea well. However, deciding how good a wine is often requiring expensive and personal opinions from experts. Our project focuses on using machine learning to classify wine quality based on its chemical makeup. This could provide a more efficient method for assessing wine quality.

### Choice of Logistic Regression and SVM Models:

The selection of logistic regression and Support Vector Machine (SVM) models for our project was considered based on their effectiveness in classification tasks and their suitability for the wine dataset. Logistic regression is a widely used algorithm for binary and multi-class classification, providing a clear probabilistic interpretation of the results. In our case, where we have three classes of wine types (1, 2, and 3), logistic regression offers a straightforward approach to predicting the class labels.

On the other hand, SVM is known for its versatility in handling both linear and non-linear classification tasks. SVM works well in high-dimensional spaces, making it a suitable choice for datasets with multiple features, such as the wine dataset with its 13 features. Additionally, SVM is effective in dealing with noisy data and is less prone to overfitting, which is crucial when working with real-world datasets like ours.

By using logistic regression and SVM models, we aim to create a robust classification system that can accurately predict the types of wines based on their chemical attributes. These models will not only provide insights into the relationships between the features and the wine types but also offer a practical approach for winemakers and enthusiasts to assess and differentiate wines.

### Dataset - Wine:

The dataset we are using for this project is the 'wine' dataset, obtained from OpenML. This dataset contains the results of a chemical analysis of wines originating from the same region in Italy. These wines are produced from three distinct cultivars, resulting in three classes of wine types (1, 2, and 3). Within this dataset, we have measurements of 13 chemical constituents present in each wine, offering a diverse set of features for our analysis and classification.

The 13 features in the dataset include:

1. Alcohol

2. Malic acid

3. Ash

4. Alcalinity of ash

5. Magnesium

6. Total phenols

7. Flavanoids

8. Nonflavanoid phenols

9. Proanthocyanins

10. Color intensity

11. Hue

12. OD280/OD315 of diluted wines

13. Proline

These features offer a comprehensive view of the chemical composition of the wines, allowing us to explore the relationships between these attributes and the wine types. By training our logistic regression and SVM models on this dataset, we aim to develop a classification system that can accurately predict the wine types based on these chemical characteristics, which could be helpful for winemakers and others interested in wine.

### References

A. Trivedi and R. Sehrawat, "Wine Quality Detection through Machine Learning Algorithms," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018, pp. 1756-1760, doi: 10.1109/ICRIEECE44171.2018.9009111. keywords: {Forestry;Logistics;Decision trees;Radio frequency;Data models;Machine learning algorithms;Data mining;Data Mining;Data pre-processing;Logistic Regression;Random Forest Classifier},

Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G. M., Mei, T., Bertini, M., Escalante, H. J., & Vezzani, R. (2021). A Hybrid Wine Classification Model for Quality Prediction. In ICPR Workshops (4) (Vol. 12664, pp. 430–438). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-68799-1\_31

Koranga, M., Pandey, R., Joshi, M., & Kumar, M. (2021). Analysis of white wine using machine learning algorithms. Materials Today : Proceedings, 46, 11087–11093. https://doi.org/10.1016/j.matpr.2021.02.229

Tingwei, Z. (2021). Red wine quality prediction through active learning. Journal of Physics. Conference Series, 1966(1), 12021-. https://doi.org/10.1088/1742-6596/1966/1/012021

Portinale, L., Leonardi, G., Arlorio, M., Coïsson, J. D., Travaglia, F., & Locatelli, M. (2017). Authenticity assessment and protection of high-quality Nebbiolo-based Italian wines through machine learning. Chemometrics and Intelligent Laboratory Systems, 171, 182–197. https://doi.org/10.1016/j.chemolab.2017.10.012

Mao, T., Zhou, L., Zhang, Y., & Sun, Y. (2022). Classification algorithm for class imbalanced data based on optimized Mahalanobis-Taguchi system. Applied Intelligence (Dordrecht, Netherlands), 52(9), 10674–10691. https://doi.org/10.1007/s10489-021-02929-8

de Andrade, B. M., de Gois, J. S., Xavier, V. L., & Luna, A. S. (2020). Comparison of the performance of multiclass classifiers in chemical data: Addressing the problem of overfitting with the permutation test. Chemometrics and Intelligent Laboratory Systems, 201, 104013-. https://doi.org/10.1016/j.chemolab.2020.104013

Mani, S., Krishnankutty, R. A., Swaminathan, S., & Theerthagiri, P. (2023). An investigation of wine quality testing using machine learning techniques. IAES International Journal of Artificial Intelligence, 12(2), 747-754. doi:https://doi.org/10.11591/ijai.v12.i2.pp747-754

P. Shruthi, "Wine Quality Prediction Using Data Mining," 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2019, pp. 23-26, doi: 10.1109/ICATIECE45860.2019.9063846. keywords: {Classification algorithms;Data mining;Training;Logistics;Bellows;Wine industry;Decision trees;data mining;classification;attributes},

Dhaliwal, P., Sharma, S., & Chauhan, L. (2022). Detailed study of wine dataset and its optimization. International Journal of Intelligent Systems and Applications, 10(5), 35. doi:https://doi.org/10.5815/ijisa.2022.05.04