



Assignment

M Tech AIML WIPL (Birla Institute of Technology and Science, Pilani)



Scan to open on Studocu

M Tech (AIML)_Machine Learning _Lab Assignment

(October 2025)

Problem Statement

In the agricultural industry, distinguishing between different varieties of dried grapes is crucial for quality control, pricing, and consumer preferences. This assignment aims to leverage machine learning techniques to predict the variety of dried grapes based on their physical properties. Specifically, we will focus on classifying them into two distinct categories.

1. Import Libraries/Dataset

- a. Download the dataset.
- b. Import the required libraries.

2. Data Visualization and Exploration [1M]

- a. Print 2 rows for sanity check to identify all the features present in the dataset and if the target matches with them.
- b. Provide appropriate data visualizations to get an insight about the dataset.
- c. Do the correlational analysis on the dataset. Provide a visualization for the same. Will this correlational analysis have effect on feature selection that you will perform in the next step? Justify your answer. **Answer without justification will not be awarded marks.**

3. Data Pre-processing and cleaning [2M]

- a. Do the appropriate pre-processing of the data like identifying NULL or Missing Values if any, handling of outliers if present in the dataset, skewed data etc. Mention the pre-processing steps performed in the markdown cell.
- b. Apply appropriate feature engineering techniques. Apply the feature transformation techniques like Standardization, Normalization, etc. You are free to apply the appropriate transformations depending upon the structure and the complexity of your dataset. Provide proper justification. **Techniques used without justification will not be awarded marks.** Explore a few techniques for identifying feature importance for your feature engineering task.

4. Model Building [5M]

- a. Split the dataset into training and test sets. **Answer without justification will not be awarded marks.** [1M]
 - i. Train = 80 % Test = 20%
 - ii. Also, try to split the dataset with different ratios of your choice.
- b. Build the classifiers [4 M]
 - i. Tune hyperparameters (e.g., number of trees, maximum depth) using cross-validation. Justify your answer.

5. Performance Evaluation [2M]

- a. Compare the performance of the Logistic Regression and Decision Tree models using appropriate evaluation metrics.
- b. Provide insights into which model performs better and why. **Answer without justification will not be awarded marks.**