Lab 07 – ITAI 1371 November 01, 2025

Prof. Viswanatha Rao

Group 04:

Kim Nguyen

Omar Alhakeem

Maria Jose Viveros

Janice Underwood

REFLECTIVE JOURNAL

The lab began by establishing the difference between regression (continuous prediction) and classification (discrete category assignment, like stay/leave). Using a dataset rich in features like salary, satisfaction, and age, a significant amount of time was spent on essential data preprocessing, including scaling and encoding, to prepare the data.

We then learned how to train and compare two foundational models:

- Logistic Regression: This model was explored to understand the mathematical core of classification, predicting the probability of attrition and relying on a decision threshold to assign the final class. Crucially, it provided a clear pathway to quantify feature importance, explaining the linear relationship between predictors and the probability of an employee leaving.
- Decision Trees: This provided a contrasting, rule-based approach that is inherently more interpretable. By generating a series of logical splits, the decision tree allowed for direct and easy explanation of the rationale behind individual predictions, a non-negotiable requirement for business users and management in a sensitive domain like HR. Comparing these models deepened the appreciation for the trade-off between the mathematical simplicity and feature insight of Logistic Regression and the structural flexibility and clarity of Decision Trees.

A key segment of the lab centered on moving beyond the deceptive simplicity of accuracy as a sole metric. This was particularly relevant given the high likelihood of class imbalance in an attrition dataset (most employees stay). A deep dive into robust evaluation metrics was essential:

- Precision, Recall, and F1-score: These metrics were used to gain a balanced view of model performance, especially regarding the ability to correctly identify the minority class (attrition cases). The F1-score, as the harmonic mean of precision and recall, served as the primary target metric for optimization.
- Visualization Tools: Confusion matrices provided a clear, quantitative breakdown of True Positives, False Positives, True Negatives, and False Negatives. The ROC curve offered a visual representation of the model's performance across all possible classification thresholds.

Overall, the lab emphasized that building a model is only half the battle; correctly evaluating and interpreting it is equally critical. The experience also highlighted the importance of ethical considerations when deploying predictive models in sensitive contexts like HR, ensuring they are used responsibly to support employees.