ITAI 1371

Professor Viswanatha Rao

October 01, 2025

Group 4. Integrants

Kin Nguyen

Janice Underwood

Omar Alhakeem

Maria Jose Viveros

Reflective Journal

Our collective reflections show a robust and shared understanding of the core machine learning paradigms. The group demonstrated a clear grasp of the fundamental differences among Supervised Learning, Unsupervised Learning, and Reinforcement Learning. Supervised Learning was consistently identified as requiring labeled data for training, often connected to the idea of dependent and independent variables in statistics or described as the computer being "taught what is correct." In a similar way, Unsupervised Learning was understood as working with unlabeled data to uncover hidden patterns or structures, which can be related to clustering in statistics. Reinforcement Learning was defined across reflections as a process of learning through trial and error, where an agent receives rewards and penalties in order to improve its strategy, often illustrated through the example of a game. Beyond these definitions, several reflections emphasized the importance of the overall ML workflow and particularly the critical step of data understanding, which shows that the group is already connecting theoretical distinctions to applied methodology.

The main differences appeared when analyzing the wine classification project, particularly in the choice of the best-performing model. Some favored Logistic Regression, highlighting its effectiveness with linearly separable data, its stability, and its interpretability, while others chose Random Forest, pointing to its ability to consider multiple features and combine decision paths to capture complex relationships. This divergence is valuable because it illustrates how model performance depends on the underlying data structure and the objectives of the analysis. Logistic Regression is a linear and interpretable model, while Random Forest is non-linear and more flexible. The fact that both models performed well suggests that the dataset could be separated in more than one way, and it also reflects the importance of technical strategies such as feature scaling, normalization, regularization, hyperparameter tuning, dimensionality reduction, and optimizing the number of training iterations. These insights indicate that model choice is not absolute but conditional, depending on sample size, type of variables, and the balance between accuracy and interpretability.

Another important aspect was the engagement with technical concepts and their application to real-world scenarios. Reflections not only defined the types of ML but also connected them to practical domains. Healthcare was a recurrent example, with suggestions to use supervised and unsupervised learning for cancer diagnosis or supervised classification for psychiatric risk prediction by combining neuroimaging with personality traits. These applications highlight the potential of ML in high-stakes fields where integration of heterogeneous data and model explainability are essential. Other reflections proposed applications in customer service, where unsupervised learning could analyze feedback to uncover hidden needs, and in architectural design, where ML could predict land development outcomes using zoning and property characteristics. These examples show both the flexibility of ML across industries and the critical role of domain knowledge in shaping model selection and data requirements.

Despite these strengths, a common limitation emerged in the reflections: the challenge of deciding which model is the most appropriate in a given context and why. While Logistic Regression and Random Forest were central to the discussion, more advanced algorithms such as Gradient Boosting Machines, Support Vector Machines, or Neural Networks could potentially achieve better results depending on the nature of the dataset. The decision requires careful attention to several factors: the size of the sample, the type of variables, the degree of linearity in the data, the objective of the study, and the need for interpretability. Moreover, as some reflections pointed out, a crucial issue is not only achieving accuracy but also explaining the model's predictions to non-technical stakeholders. This highlights the growing importance of explainable artificial intelligence, especially in domains such as medicine, neuroscience, or finance, where trust and transparency are as essential as performance.

In conclusion, the group demonstrated a solid grasp of machine learning types and workflows, showing the ability to move beyond definitions into applied analysis and real-world application. The next step in this learning journey is to develop a deeper understanding of advanced model selection, improve skills in preprocessing complex and unstructured data, and strengthen awareness of explainability as a necessary condition for the responsible use of ML. By addressing these challenges, machine learning can be applied more effectively, ensuring not only technical success but also ethical and meaningful impact in society.