UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

Artificial Intelligence

Assignment 4

Supervised Learning: Decision Tree

Name	ID
Sergio Orlando Alanís De La Rosa	2043932
Diego Castro Galindo	2109304
Sebastian Hernandez Renteria	2109140
Braulio Azael García Treviño	2043046
Miguel Angel Perez Luevano	2052579

Professor Name: Daniel Isaías Lopez Paez

Career: IB   **Group:** 003

Date: 17/10/25

Introduction: In the field of Artificial Intelligence, supervised learning is one of the most widely used approaches to enable computers to learn from labeled data and make predictions or classifications based on patterns found in the information. Within this branch, the decision tree algorithm stands out as one of the simplest yet most powerful models for classification tasks. A decision tree works by dividing data into branches according to specific feature conditions until it reaches a final decision or classification. The main advantage of this algorithm is that its logic is transparent and easy to interpret, making it especially useful in fields such as healthcare, where understanding the reasoning behind predictions is crucial.

This project focuses on the application of a decision tree model to the Breast Cancer Wisconsin (Diagnostic) dataset, which is part of the Scikit-learn library in Python. This dataset contains measurements computed from digitized images of breast tissue, along with a label indicating whether the tumor is benign or malignant. Each record includes several numerical features that describe the cell nuclei characteristics, such as texture, smoothness, compactness, and symmetry.

The purpose of this activity is to train and evaluate a supervised learning model capable of classifying tumors with a high level of accuracy. Through this exercise, we aim to understand the basic steps in developing a machine learning model: data preprocessing, training, testing, and evaluation. Moreover, this practice provides a solid introduction to the implementation of machine learning algorithms using Python and the Scikit-learn library, which are fundamental tools for modern data analysis and biomedical applications. As a biomedical engineering student, this project also reinforces the connection between computational tools and medical diagnosis support systems, contributing to the development of technology that can assist in early cancer detection.

Methodology: The development of this project was carried out using Google Colab as the programming environment, due to its compatibility with Python and its integration with scientific libraries. The methodology followed the typical workflow for supervised learning model implementation and consisted of several steps, as described below:

1. Library Importing:

   The first step was to import the required Python libraries, including pandas, numpy, and several modules from scikit-learn, such as datasets, model_selection, tree, and metrics. These libraries provide the necessary tools for data handling, model creation, and performance evaluation.

2. Data Loading:

   The Breast Cancer Wisconsin (Diagnostic) dataset was loaded directly from Scikit-learn's built-in datasets. The features were stored in a DataFrame for easier visualization and manipulation, while the target variable contained the corresponding diagnosis labels ("malignant" or "benign").

3. Data Preprocessing:

   Before training the model, the dataset was divided into training and testing subsets using the train_test_split() function from Scikit-learn. The split ratio was 80% for training and 20% for testing, ensuring a sufficient amount of data for both model learning and performance validation.

4. Model Training:

   A Decision Tree Classifier was created using the DecisionTreeClassifier class. The model was trained with the training data (train_data and train_labels) using the fit() function. This process allowed the algorithm to learn patterns and relationships between the features and their corresponding labels.

5. Model Testing and Predictions:

   Once the model was trained, predictions were made on the testing data (test_data) using the predict() function. The predicted values were then compared with the actual test labels to evaluate performance.

6. Model Evaluation:

   The accuracy of the model was calculated using the accuracy_score() function, providing a numerical value that reflects how well the model classified the samples. Additionally, the classification_report() function was used to generate a detailed summary of performance metrics, including

precision, recall, and F1-score.

7. Decision Tree Visualization:

    Finally, the decision tree was visualized using Scikit-learn's plot_tree()
    function with a resolution of 300 DPI. This visualization provided a clear
    understanding of the structure of the decision-making process, showing the
    different branches, features, and thresholds used for classification.

Through these steps, the project successfully applied theoretical machine learning
concepts into a practical example, allowing for the interpretation and evaluation of
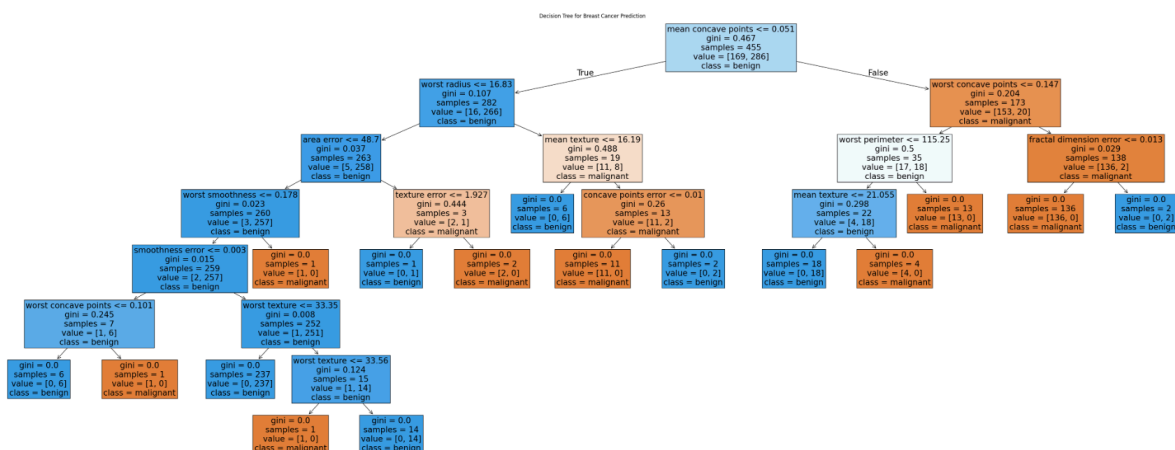real biomedical data in a structured and visual way.

Results:

Figure.1 *Decision Tree*

Google Colab:

https://colab.research.google.com/drive/1SB_S04g73oF0EK_T9vm6lyj3I41EJwJM?usp=sharing

Github:

https://github.com/Sergio-Alan/AI-2025/blob/main/Assignments/AF4_Decision_Tree.ipynb

Conclusion:

The completion of this project provided a comprehensive and practical understanding of the fundamental concepts behind *supervised learning*, with a particular focus on the Decision Tree Classifier algorithm. By applying this model to the Breast Cancer Wisconsin (Diagnostic) dataset, we were able to explore the entire process of developing a machine learning solution, from data preparation and model training to testing, evaluation, and visualization. The results demonstrated that decision trees can effectively classify tumors as benign or malignant, achieving a high level of accuracy and reliability. In conclusion, this project not only strengthened technical programming skills but also highlighted the potential of computational tools in supporting medical diagnosis. Integrating biomedical engineering with artificial intelligence represents a significant step toward developing systems that contribute to early cancer detection and more informed clinical decision-making.

References:

- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.
- Karanam, S. (2023, June 10). *Exploratory data analysis: Breast Cancer Wisconsin (Diagnostic) dataset*. Medium. https://medium.com/@shashmikaranam/exploratory-data-analysis-breast-cancer-wisconsin-diagnostic-dataset-6a3be9525cd
- Matthes, E. (2023). *Python Crash Course* (3rd ed.). No Starch Press.
- W3Schools. (2024). *Python Tutorial*. Retrieved from https://www.w3schools.com/python/
- «1.10. Decision Trees», *Scikit-learn*. https://scikit-learn.org/stable/modules/tree.html