Labs Guidelines

Machine Learning and Artificial Intelligence in Finance

FINC 5322

Fall 2025

The American University in Cairo

Onsi Sawiris School of Business

Heikal Department of Management

Interpretability in Machine Learning – Assignment 3

In this assignment, you will focus on model interpretability, understanding how machine learning models make predictions and how these can be explained to non-technical stakeholders.

Using Logistic Regression and Decision Trees, you will apply interpretability techniques like Weight of Evidence (WoE), Information Value (IV), Scorecards, and SHAP values to reveal how features influence outcomes. The goal is to translate complex model behavior into clear, human-understandable insights.

You will visualize feature impacts, compare explainability methods, and reflect on their role in responsible model deployment. By the end, you should be able to explain what a model predicts and why, bridging the gap between predictive power and interpretability

---

Accessing the Assignment

- All assignments are hosted on our course Code Repository (GitHub): https://github.com/kelkess43/AUC-Material
- The assignment will also be shared weekly through Canvas modules for easy access.
- The dataset to be used is provided along with the assignment files.
- A detailed data documentation sheet (Excel format) is also included to describe each variable and its meaning.

Structure of the Assignment

Your notebook is divided into seven sections.
You must complete the tasks and reflections directly within the. ipynb file.
All questions, instructions, and grades are provided inside the notebook.

Section 1 – Dataset Understanding

- **Q1.1 (2 pts)**: What is the shape of the dataset (number of rows and columns)?
- **Q1.2 (2 pts)**: What are the primary data types of the features (numeric, categorical, etc.)?
- **Q1.3 (3 pts)**: Are there any missing values in the dataset? If yes, mention which columns contain them and their approximate percentage.

Section 2 – Exploratory Data Analysis (EDA)

- **Q2.1 (3 pts)**: Plot the distribution of the target variable. What can you infer about class balance?
- **Q2.2 (5 pts)**: Plot box plots of at least two numeric variables against the target variable.

Section 3 – Feature Engineering (Weight of Evidence and Information Value)

- **Q3.1 (4 pts)**: Perform *fine classing* on all continuous variables to generate detailed initial bins.
- **Q3.2 (18 pts)**: Apply *coarse classing* by merging adjacent fine classes and categorical feature levels that share similar WoE values across **all features**.
- **Q3.3 (5 pts)** – Create two DataFrames: one with original features and one with WoE-transformed features.

Section 4 – Model Comparison

- **Q4.1 (10 pts)**: Compare both models in terms of precision, recall, and accuracy. Which performs better overall? What trade-offs do you observe between the two?

Section 5 – Logistic Regression Interpretability

- **Q5.1 (16 pts)**: Build a scorecard from the logistic regression model and transform coefficients into business scores within the 350–800 range.
- **Q5.2 (6 pts)**: Interpret three model coefficients. What does a positive versus negative sign indicate about credit risk?

Section 6 – Decision Tree Interpretability

- **Q6.1 (10 pts)**: Visualize the decision tree. Which features appear at the top, and why do you think they are important?
- **Q6.2 (5 pts)**: Choose one leaf node (final decision). Write out the full rule path leading to that node and describe what kind of borrower it represents.
- **Q6.3 (3 pts)**: Compare the simplicity of the decision tree explanation to the numeric interpretability of the logistic model. Which is easier to justify to regulators and why?

Section 7 – Model Explainability with SHAP

- **Q7.1 (4 pts)**: Generate and display a SHAP summary plot for the decision tree. Which three variables contribute most globally, and why might that be?
- **Q7.2 (4 pts)**: For one test instance, display SHAP force plots for both models. Which model's explanation aligns better with business intuition?
- **Q7.3 (2 pts)**: In your own words, explain the difference between *global* and *local* interpretability in the context of SHAP.

Deliverables and Submission

- Submit one Jupyter Notebook (.ipynb) file with all your work.
- Make sure all cells are executed, and outputs are visible.

- Add short comments/markdown where required
- Deadline: Monday, 10th November at 11:59 PM
- Submission: Upload your completed. ipynb notebook on Canvas