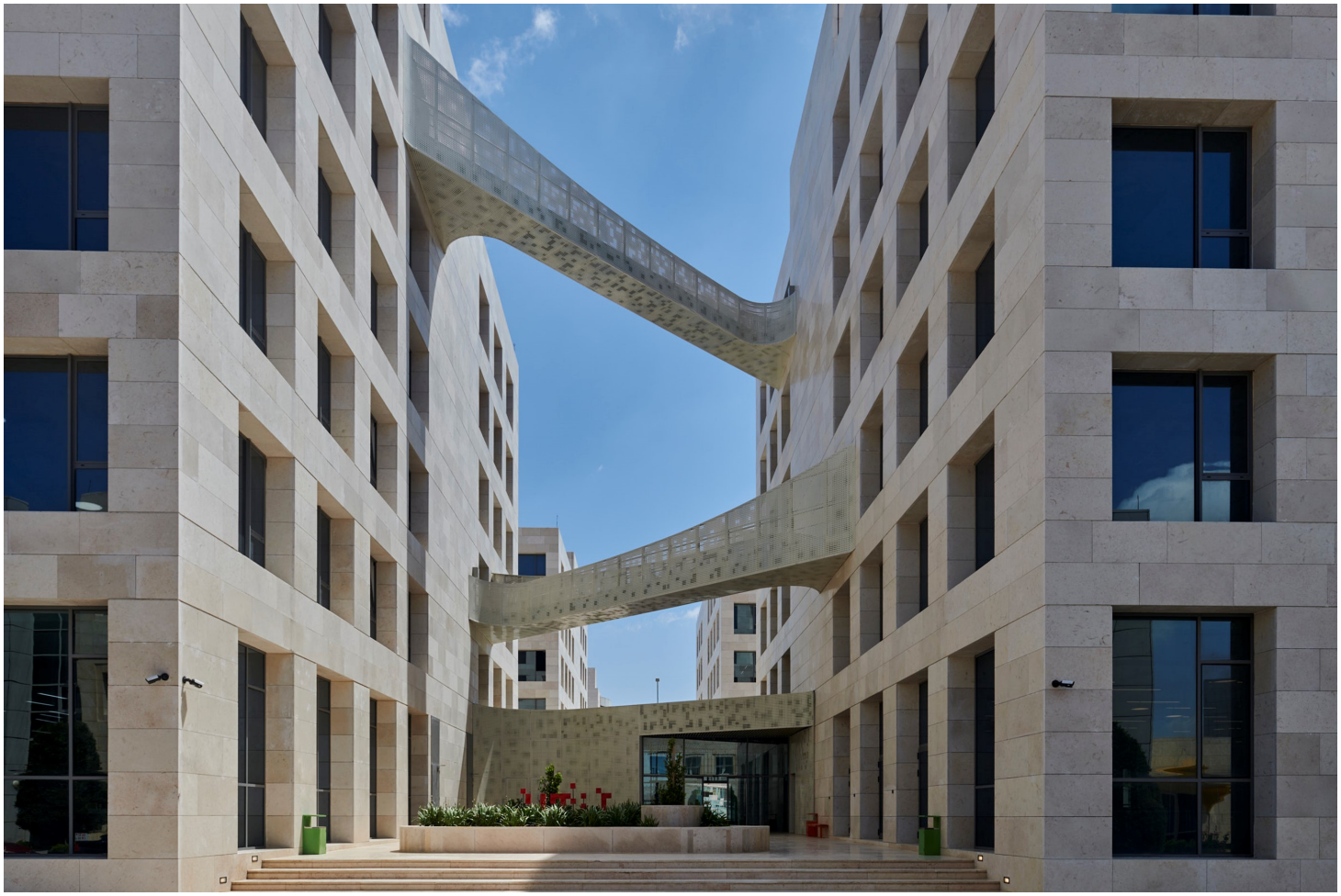
جامعة الحسين التقنية
Al Hussein Technical University

ASSIGNMENT BRIEF

HTU Course No: 10204412	**HTU Course Name:** Applied Analytical Models
BTEC Unit Code:	**BTEC UNIT Name:**



Student Name/ID Number/Section	
HTU Course Number and Title	10204412 Applied Analytical Models
BTEC Unit Code and Title	
Academic Year	2024-2025 1
Assignment Author	Bassam Alkasasbeh
Course Tutor	Bassam Alkasasbeh
Assignment Title	Comprehensive Applied Modeling: Data, Fuzzy Logic, and Evaluation
Assignment Ref No	1
Issue Date	19/12/2024
Formative Assessment dates	From 20/12/2024 to 09/01/2025
Submission Date	28/01/2025
IV Name & Date	Yara Alharahsheh 18/12/2024

Submission Format

Practical Assignment Submission

- Submit to the university's eLearning system via https://elearning.htu.edu.jo by the specified deadline.
- The submission must include the following:
  1. A Jupyter source code file (.ipynb) for **Feature Engineering** tasks.
  2. A Jupyter source code file (.ipynb) for **Generative Adversarial Network (GAN)** tasks.
  3. A Jupyter source code file (.ipynb) for **Fuzzy Logic** tasks.
  4. A Jupyter source code file (.ipynb) for **Time Series** tasks.
  5. A Jupyter source code file (.ipynb) for **Explainable AI (XAI)** tasks.

- No compressed files or folders (no .zip or .tar extensions).
- A signed declaration form is required.
- **Plagiarism will result in course failure (U mark). Plagiarism includes using AI tool.**
- The submission deadline is **Tuesday, 28 January 2025 at 8:00 AM.**

Report Submission

- Submit to the university's eLearning system via https://elearning.htu.edu.jo by the specified deadline.
- The submission must be a word file (.docx).
- No compressed files or folders (no .zip or .tar extensions).
- **Plagiarism will result in course failure (U mark). Plagiarism includes using AI tool.**
- A signed declaration form is required.
- The submission deadline is **Tuesday, 28 January 2025 at 8:00 AM.**

Unit Learning Outcomes

LO1 Explore and apply feature engineering and data augmentation techniques to optimize datasets and develop a predictive pipeline for improved analytical modeling.

LO2 Apply fuzzy logic techniques to analyze and solve practical problems involving uncertainty and overlapping datasets.

LO3 Analyze and interpret the results of analytical models while ensuring transparency and fairness.

LO4 Implement and evaluate predictive models using advanced techniques.

Assignment Brief and Guidance

You are a **Data Scientist** tasked with analyzing and predicting trends in the gold futures market. Your project involves transforming raw data into actionable insights using advanced techniques like feature engineering, GANs for data augmentation, fuzzy logic for uncertainty analysis, time series forecasting models, and explainable AI (XAI) for interpreting predictions.

To fulfill the assigned tasks, you must finish the following parts:

- **Dataset Link:** gold_futures.csv

Part 1: Preprocessing, Feature Engineering, and Predictive Pipeline for Gold Futures Dataset

Use Python in Jupyter Notebook and libraries such as Pandas, NumPy, and Scikit-learn to preprocess and enhance the gold futures dataset and develop a predictive pipeline to classify the "Price Movement" column.

1. Create at least two meaningful new features using feature engineering, ensuring they are justified, clearly defined.

2. Use Python in Jupyter Notebook to implement a Generative Adversarial Network (GAN) for augmenting the gold futures dataset. Experiment with training parameters and analyze the impact on the synthetic data quality and GAN performance. Experiment with the following parameters during training:
   - § **Number of epochs:** Train for at least 2000 epochs.
   - § **Batch size:** Experiment with small and large batch sizes.
   - § **Noise vector size (input_dim):** Test different input dimensions for the Generator.
   - § **Learning rate:** Adjust the Adam optimizer learning rate.

   · **Analyze GAN Performance**
   - § Evaluate the loss trends for the Generator and Discriminator during training.

   · **Generate and Evaluate Synthetic Data**
   - § Use the trained Generator to create **10,000 synthetic data samples**.
   - § Compare the synthetic data distributions to real data using:
     - · Visualizations: Histograms, KDE plots, or scatter plots.
     - · Statistical measures: Mean, standard deviation, or advanced measures.

3. Design and implement a machine learning pipeline using Python and Scikit-learn to classify "Price Movement."
   - § Add **at least two advanced technical indicator features** to enhance the dataset by incorporating new meaningful insights.
   - § Justify how these features contribute to improving classification model.
   - § Use feature importance methods (e.g., mutual information, or feature importance) to select the most relevant features.
   - § Evaluate at least **two different machine learning classifiers.**
   - § Choose the best-performing model.
   - § Compare the performance:
     - § The model without advanced feature engineering.
     - § The model with advanced feature engineering.

Part 2: Gold Price Analysis Using Fuzzy Logic and Clustering Techniques

Use Python in **Jupyter Notebook** and scikit-fuzzy to classify gold price data into fuzzy categories.

1. **Implement Membership Functions**
   - **Choose Two Membership Functions:**

     Implement two different membership functions to classify the **High price** column into fuzzy categories: Low, Medium, and High.

   - **Visualize Membership Functions:**

     Create clear visualizations for both membership functions (e.g., Triangular and Gaussian).

2. **Apply C-Fuzzy Means (FCM) Clustering**

   Apply the C-Fuzzy Means (FCM) clustering technique to classify overlapping data points and analyze the effect of hyperparameter tuning.

   - **Select Features**: Use **Open**, **High**, **Low**, **Close**, and **Volume**.
   - **Cluster the Data**: Implement FCM clustering with **three clusters**.
     - Vary the **fuzziness parameter (m)** (e.g., 1.5, 2.0, 2.5).
   - **Evaluate Performance**:
     - Use **Fuzzy Partition Coefficient (FPC)** and **Silhouette Score** for each **fuzziness parameter(m)**.
   - **Visualize Results**:
     - Plot scatter plots for each **fuzziness parameter(m)** showing cluster memberships and centres.
     - Summarize results in a table comparing FPC and Silhouette Scores.

3. **Construct a Fuzzy Inference System (FIS)**

   Build and evaluate a Fuzzy Inference System (FIS) to analyze market conditions using fuzzy rules.

   - **Define Inputs and Outputs:**
     - **Inputs:**
       - **Volume (Low, Medium, High)**
       - **Price Movement (Up, Down)**
     - **Output:**
       - **"Market Condition" (Bearish, Neutral, Bullish)**
   - **Create Fuzzy Rules:**

   Write at least three rules to infer market conditions. Example:

     - IF Volume is High AND Price Movement is Up, THEN Market Condition is Bullish.
   - **Evaluate the System:**

   Test the system using sample inputs from the dataset and display defuzzified outputs.

   - **Visualize Results:**
     - Plot input/output membership functions.
     - Visualize the fuzzy rules and surfaces.

4. **Design an Advanced Fuzzy Logic System**

Develop a fully functional fuzzy logic system integrating multiple variables to analyze risk levels.

- **Inputs:** High Price, Low Price, Volume
- **Output:** "Risk Level" (Low, Medium, High)

  Risk Level: Represents the likelihood of significant price fluctuation or market instability based on the input variables.
- **Fuzzy Rules:** Construct a comprehensive fuzzy system without limiting the number of rules, ensuring the system fully captures the relationships between inputs and outputs.
- Use scikit-fuzzy to build the system and validate it with real data from the dataset.
- **Visualize Results:**
  - § Fuzzy membership function plots
  - § Fuzzy rule visualizations
  - § Fuzzy surface plots for the inputs and output

Part 3: Gold Futures Closing Price Forecasting Using Advanced Time Series Models

Use Python in Jupyter Notebook to forecast gold futures closing prices by implementing advanced time series models. Apply two predictive models—**Prophet** and **AutoTS**—to analyze the dataset, evaluate their performance, and compare their effectiveness in forecasting future prices.

1. **Implement Predictive Models**
   - **Model Selection:**
     - § Use **Prophet** and **AutoTS** for forecasting.

2. **Document Parameters:**
   - List the key parameters used in each model and explain their significance.

3. **Evaluate and compare** model performance using performance metrics such as Mean Absolute Error (MAE) and R^2 (Coefficient of Determination).

4. **Visualize and Discuss Findings:**
   - Visualize the actual vs. predicted closing prices for each model using line charts.
   - Create bar charts or tables to summarize the performance metrics.

Part 4: Explainable AI (XAI) for Predicting Gold Price Movement

Leverage Explainable AI (XAI) techniques to understand and interpret the predictions of machine learning models for predicting "Price Movement" (Up/Down) using the provided gold futures dataset.

1. **Build the Predictive Models:**
   - Use the provided gold futures dataset to build a predictive model to predict the "Price Movement" (Up/Down) using any **two** machine learning models.
   - Use performance metrics like accuracy to assess the initial performance of the models.

2. **Use SHAP for Explanation:**
   - Use SHAP to explain the predictions made by your models.
   - Generate a summary plot to visualize the overall feature importance.
   - Select a few individual predictions and explain how each feature contributes to the output.

3. **Use LIME for Explanation:**
   - Use LIME to explain the predictions for specific data points.
   - Identify why the model predicted "Up" or "Down" for selected cases.

4. **Compare SHAP and LIME:**
   - Compare the explanations provided by SHAP and LIME for the same predictions.

5. **Analyze and evaluate Model Bias:**
   - Check if the model's performance varies across the data, such as analyzing whether the model is biased toward certain price ranges (e.g., "High" vs. "Low").
   - Critically evaluate if the model treats "Up" and "Down" predictions equitably in performance measures.
   - Use SHAP and LIME to identify if the model relies too heavily on certain features.
   - Adjust the model to ensure that it makes decisions based on meaningful and interpretable features.
   - Experiment with hyperparameter tuning **or** alternative machine learning algorithms to improve predictive accuracy while maintaining interpretability.
   - Use performance metrics like accuracy to compare models before and after optimization.

Part 5: Report

Prepare a report outlining the analysis of your tasks. The report should include the following sections:

1. **Feature Engineering**

   Explain the advanced technical indicator features added, their relevance to the analysis, and justify how these features contribute to improving the classification model.

2. **GAN Analysis**
   - Provide an overview of the GAN architecture and training parameters.
   - Include visualizations and analysis of Generator and Discriminator loss trends.
   - Compare synthetic and real data distributions using visualizations and statistical measures.

3. **Fuzzy Inference System (FIS)**
   - Describe the inputs, outputs, and fuzzy rules created.
   - Provide defuzzified outputs for sample inputs with a brief analysis.
   - Support this section with visualizations of input/output membership functions and fuzzy rule surfaces.

4. **Advanced Fuzzy Logic System**
   - Summarize the system design, including inputs, output, and fuzzy rules.
   - Validate the system using dataset examples.

- Include visualizations of membership functions, rules, and surface plots illustrating the relationship between inputs and the "Risk Level" output.

5. **Time Series Models**
   - List key parameters for each model and their significance.
   - Summarize performance metrics (e.g., MAE, R^2) for Prophet and AutoTS using a table or bar chart.
   - Include line charts showing actual vs. predicted closing prices.
   - Discuss and compare model performance based on insights from the visualizations.
6. **XAI Framework for Price Movement Prediction**
   - Present a summary table of initial model performance metrics (e.g., accuracy) for predicting "Price Movement" (Up/Down) using two machine learning models.
   - Include a SHAP summary plot for overall feature importance.
   - Explain specific predictions using LIME (e.g., why the model predicted "Up" or "Down").
7. **SHAP vs. LIME**
   - Compare the insights provided by SHAP and LIME for the same predictions.
   - Highlight strengths and weaknesses of each method.
8. **Bias and Equity Evaluation**
   - Evaluate model performance across data (e.g., "High" vs. "Low" days).
   - Discuss whether predictions (e.g., "Up" and "Down") are treated equitably.
   - Support this section with insights from SHAP and LIME on feature reliance and potential biases.
9. **Model Optimization**
   - Summarize adjustments made to improve model performance or mitigate biases.
   - Include a table comparing performance metrics before and after optimization.
   - Provide a comprehensive set of SHAP and LIME plots, bias analysis charts, and tables summarizing feature importance and performance metrics.

Learning Outcomes and Assessment Criteria			
Learning Outcome	**Pass**	**Merit**	**Distinction**
LO1 Explore and apply feature engineering and data augmentation techniques to optimize datasets and develop a predictive pipeline for improved analytical modeling.	**P1** Apply basic feature engineering techniques to create new features from existing data.	**M1** Apply advanced data augmentation techniques to generate synthetic data and explain how it enhances the dataset for improved model training in cases of limited data.	**D1** Develop a custom predictive pipeline using advanced feature engineering to improve classification accuracy.
LO2 Apply fuzzy logic techniques to analyze and solve practical problems involving uncertainty and overlapping datasets.	**P2** Implement using at least two membership functions to classify input data into fuzzy categories. **P3** Apply C-Fuzzy Means (FCM) clustering to classify overlapping data points into clusters from the given dataset.	**M2** Construct and evaluate a fuzzy inference system (FIS) using the given dataset to process fuzzy inputs and produce defuzzified outputs.	**D2** Design an advanced fuzzy logic system with multiple membership functions and rules, integrating all relevant variables from the given dataset to solve a real-world problem.
LO3 Analyze and interpret the results of analytical models while ensuring transparency and fairness.	**P4** Apply techniques to explain the predictions of a machine learning model and provide insights into how the model makes decisions.	**M3** Compare Explanation Techniques and Evaluate Model Fairness.	**D3** Critically evaluate and optimize the model for fairness, transparency, and performance, ensuring equitable predictions and providing evidence-based justifications.
LO4 Implement and evaluate predictive models using advanced techniques.	**P5** Implement and apply different time series forecasting models to a basic dataset, and explain the key parameters used in each model.	**M4** Compare the results of different time series forecasting models using key performance metrics. Discuss which model performs better based on these metrics and explain the reasons behind the observed differences.	