UGANDA CHRISTIAN UNIVERSITY
A Centre of Excellence in the Heart of Africa

FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY
DEPARTMENT OF COMPUTING AND TECHNOLOGY
EASTER 2025 SEMESTER **EXAMINATION**

PROGRAM: BACHELOR OF SCIENCE IN DATA SCIENCE AND ANALYTICS

YEAR: 2 SEMESTER: 2

COURSE CODE: DSC 2206

COURSE NAME: TIME SERIES ANALYSIS AND FORECASTING

EXAMINATION TYPE: PROJECT-BASED EXAM

PROJECT DURATION: MAR-APR 2025

TIME ALLOWED: Two Weeks

Examination Instructions

1. The general Uganda Christian University examination guidelines and academic & financial policies apply to this examination. Violating any of the policies by the student automatically makes this examination attempt void, even if you have completed and submitted the answer booklet.
2. This exam consists of a project to be executed in Two weeks.
    i. Assessment of the project shall be based on five milestones, evaluated during the duration of the project. Each milestone shall be evaluated out of 20 marks.
3. Every student has a responsibility to prove their contribution towards every milestone, and marks may be awarded to every student individually.

PART A:
PROJECT DESCRIPTION

Submission Format (to moodle): Project presentation PPT + Jupyter Notebook (code + answers to theoretical analysis questions) + Report/documentation + Web App
Total Points: 100

Exam Structure

Each Question in this project exam consists of **two parts**:

1. **Practical Forecasting Tasks**
   - ✓ Students work with two Kaggle datasets, applying **statistical, machine learning, and deep learning forecasting techniques**.
   - ✓ A **functional web app deployment** is required.
   - ✓ Students must provide clear explanations for each step.
2. **Short Answer Analysis & Critical Thinking Questions**
   - ✓ These appear **after each practical stage**, asking students to **analyze, justify, or critique** their work.
3. Some question parts are optional.

1. Datasets

1. **New York City Energy Consumption Dataset**
   Link: https://www.kaggle.com/datasets/jeanmidev/smart-meters-energy-usage-in-london
   - ✓ **Goal:** Forecast energy demand trends.
2. **S&P 500 Stock Prices Dataset**
   Link: https://www.kaggle.com/datasets/camnugent/sandp500
   - ✓ **Goal:** Predict future stock prices for selected companies.

2. Practical Exam Tasks & Analysis Questions

Section 1: Data Exploration & Preprocessing (15 points)

Tasks:

1. Load both datasets and **inspect missing/null values**.
2. Perform **Exploratory Data Analysis (EDA)**:
   - ✓ Identify **trends, seasonality, and outliers**.
   - ✓ Compute and plot **ACF/PACF** to analyze autocorrelation.
3. Perform **stationarity tests (ADF/KPSS)** and apply **time series decomposition**.
4. Engineer time-based features (**lagged values, rolling mean, Fourier terms (optional)**).
5. Split data into **training (80%) and testing (20%)**.

Deliverable:

- **Jupyter Notebook with Data Preprocessing + EDA.**

Analysis & Critical Thinking Questions (5 points)

1. **Why is stationarity important in time series forecasting? Explain the impact of non-stationary data on ARIMA models.**

2. **Based on your visualizations, does seasonality appear to be a major factor in either dataset? Justify your answer with observations from your analysis.**

Section 2: Statistical Forecasting (15 points)

Tasks:

1. Implement **Moving Averages, Exponential Smoothing (SES or Holt-Winters)**.
2. Build **ARIMA models**, selecting optimal **p, d, q** using **AIC/BIC**.
3. Extend to **SARIMA/VAR** if relevant.
4. Evaluate models using **RMSE, MAE, and MAPE**.
5. Forecast **next 30 days** for energy and stock price data.

Deliverable:

- **Notebook section on Statistical Forecasting.**

Analysis & Critical Thinking Questions (5 points)

1. **Compare and contrast ARIMA and Exponential Smoothing. When would you prefer one over the other? Provide a real-world example.**
2. **Look at your ARIMA predictions—how well do they match observed trends? If there is a large error, what do you think caused it, and how could you improve the model?**

Section 3: Machine Learning Forecasting (15 points)

Tasks:

1. Convert time series data into **ML-ready format**:
   - ✓ Create **date-based features** (e.g., day of week, month, holiday).
   - ✓ Engineer **lagged variables & rolling window statistics**.
2. Train **Linear Regression, Random Forest, and XGBoost models (optional)**.
3. Perform **hyperparameter tuning** for optimization.
4. Compare ML models to **ARIMA/SARIMA**.

Deliverable:

- **Notebook section on Machine Learning models.**

Analysis & Critical Thinking Questions (5 points)

1. **What advantages does machine learning have over ARIMA in time series forecasting? What are some limitations?**
2. **Based on your results, which model (ML vs ARIMA) performed better? Why do you think that is?**

Section 4: Deep Learning for Time Series (15 points)

Tasks:

1. **Transform data for deep learning models**:
   - ✓ Normalize values using **MinMaxScaler**.
   - ✓ Convert data into **supervised learning format** (X_train, y_train).
2. Train a **Basic LSTM model** using TensorFlow/Keras/Pytorch.
3. Experiment with **CNN-LSTM (optional).**
4. Compare **LSTM results with ML & ARIMA models**.

Deliverable:

- **Notebook section on Deep Learning models.**

Analysis & Critical Thinking Questions (5 points)

1. **What is the key advantage of using LSTMs for time series forecasting? How does it handle long-term dependencies compared to traditional models?**
2. **Examine your deep learning model's performance—how does it compare to ML and ARIMA? Did it generalize well, or does it show signs of overfitting?**

Section 5: Model Deployment & Business Analysis (20 points)

Tasks:

1. Deploy the best-performing model as a **Streamlit or Flask Web App**.
2. Users should be able to **input a date range** and receive predictions.
3. Write a **3-5 page report (PDF)** covering:
   - ✓ Best-performing model & evaluation metrics.
   - ✓ Forecast insights & implications.
   - ✓ Challenges faced & future recommendations.

Deliverables:

- **Functional App (8 points).**
- **Report with Insights (8 points).**

Analysis & Critical Thinking Questions (4 points)

1. **Imagine your energy demand forecast is used by a power company. What business decisions could be influenced by your predictions?**
2. **If you were to deploy your forecasting model for real-time use, what additional steps would you take to ensure robustness and reliability?**

PART B: Project-based assessment guidelines

S/N	Milestone Description	Maximum Marks
1	**MILESTONE ONE** **Problem Understanding and Dataset Analysis** Students must analyze and explain the **importance of time series forecasting** in the selected datasets (**New York Energy Demand and S&P 500 Stock Prices**). Clearly describe **business problems** these forecasts can help solve. Conduct an **initial exploratory data analysis (EDA)**, including visualization of trends and seasonality. **Key Focus:** Clear problem definition, dataset exploration, relevance of forecasting.	**20%**
2	**MILESTONE TWO** **Data Preprocessing and Baseline Models** Students must perform **data cleaning, transformation, and feature engineering** (e.g., lag features, rolling means, holiday effects). Apply **basic statistical forecasting models** (Moving Averages, ARIMA/SARIMA) and **compare initial results**. Explain choices made in **handling stationarity, missing values, and model tuning**. **Key Focus:** Data preparation, feature selection, baseline statistical models.	**20%**
3	**MILESTONE THREE** **Machine Learning and Deep Learning Implementation** Train **at least one machine learning model (e.g., Random Forest, XGBoost, or Linear Regression)**. Train **one deep learning model (e.g., LSTM or CNNLSTM)** and compare performance with statistical models. Tune **hyperparameters** and explain tradeoffs. **Key Focus:** Proper model implementation, tuning, comparative analysis between ML, DL, and statistical methods.	**20%**
4	**MILESTONE FOUR** **Model Evaluation, Interpretation, and Business Impact** Evaluate models using **MSE, RMSE, MAE, MAPE**. Interpret and compare results from **statistical, ML, and DL models**. Discuss **business implications** of forecasts and limitations of the models. **Key Focus:** Model evaluation, real-world impact, comparative analysis.	**20%**
5	**MILESTONE FIVE** **Deployment, Documentation, and Presentation** Deploy the **best-performing model** in a **Streamlit or Flask web app**. Submit a **well-documented report** (3-5 pages). Deliver a **structured final presentation PPT, Notebook and** a **live demonstration** of the web app. **Key Focus:** Functional deployment, clarity of documentation, and presentation effectiveness.	**20%**
	TOTAL MARKS	100%

~END OF EXAM GUIDELINES~