

# Final Exam – BIA 5402 (30 Marks)

---

## Case Study One (20 Marks)

### Forecasting Weekly Demand for a Soft Drink Product – A Time Series Modeling Challenge

#### Business Context

You are working as a data science intern for a soft drink manufacturer. The company distributes its flagship product nationwide. Recently, the demand for this product has shown fluctuations, resulting in stockouts and overstock situations that cost the company thousands in logistics and spoilage.

You are given historical weekly demand data for this product. Your task is to analyze its time series characteristics and build a forecasting model to predict demand for the next 2 weeks. Your forecast will guide future production planning.

#### Dataset Description (bicup2006.csv)

The dataset contains the following columns:

- DATE: Weekly date index (YYYY-MM-DD)
- TIME: Time of the transaction in 24hr format
- DEMAND: Number of units sold in that week

#### Part A: Data Preparation & Exploration (4 Marks)

1. Preprocess the Data

2. Visualize

- Plot the time series. Comment on any visible trend, seasonality, or irregularity.

#### Technical Checkpoint 1:

Q1. What kind of problems can arise if we do not handle missing time steps in time series data?

Q2. How would differencing help here, and how do you detect the correct level of differencing needed?

#### Part B: Time Series Analysis (4 Marks)

1. Stationarity Check

- Use the Augmented Dickey-Fuller (ADF) test to determine if the series is stationary.
- Plot rolling mean and standard deviation and comment on stability.

2. Decompose the Series

- Use seasonal decomposition to extract trend, seasonality, and residuals.
- Explain what type of seasonality you observe: weekly, quarterly, or yearly?

**Technical Checkpoint 2:**

Q3. Why is stationarity important for ARIMA-type models?

Q4. When would you use additive decomposition over multiplicative decomposition?

**Part C: Forecasting Models (8 Marks)**

1. Implement SARIMA and any other two (02) time series/ Neural Network models you think will fit best.

2. Use training/testing split or cross-validation.

3. Forecast the next 2 weeks of demand and plot predictions.

**Technical Checkpoint 3:**

Q5. How would you decide the values for SARIMA's  $(p, d, q)$   $(P, D, Q)$ s parameters?

**Part D: Model Evaluation & Business Recommendation (4 Marks)**

1. Evaluate your models using at least 2 out of 3 metrics:

- MAE
- RMSE
- MSE

2. Compare the performance of the models and identify which one is more reliable.

3. Write a 1-paragraph recommendation for business stakeholders:

- Should they produce more or less of the product next month?
- How confident are you in your prediction?

**Technical Checkpoint 4:**

Q6. How would you explain residuals and confidence intervals to a non-technical manager?

Q7. What limitations does your forecast have, and how might external factors affect it?

## Case Study Two (10 Marks)

### Predicting Book Purchase Behavior – Classification & Association Rules

#### Business Context

The Charles Book Club is analyzing past purchase behavior to determine which customers are most likely to purchase specialty books, such as a travel guide on Florence. You've been asked to investigate customer profiles and determine how past purchases influence the likelihood of buying this specialty book. You will use ensemble classification methods to build a predictive model and optionally perform association rule mining to uncover book co-purchase trends.

#### Dataset Description (CharlesBookClub.csv)

The dataset contains the following columns:

- Gender, M, R, F: Demographic and transaction recency/frequency/monetary features
- Various book genres: ChildBks, YouthBks, CookBks, etc. (numeric counts of books bought)
- Campaign outcome: Yes\_Florence, No\_Florence – indicates whether the customer bought the Florence travel book

**Note:** Convert Yes\_Florence and No\_Florence into a single binary target column (e.g., FlorenceBuyer = 1 if Yes\_Florence = 1).

#### Part A: Data Exploration and Cleaning (2 Marks)

1. Inspect the dataset and summarize any patterns in book purchases.
2. Remove irrelevant columns (e.g., Seq#, ID#) and handle missing values.
3. Create a binary target column from Yes\_Florence and No\_Florence.

**Checkpoint Q1:** Which features appear to influence specialty travel purchases?

#### Part B: Modeling Florence Book Purchase (6 Marks)

1. Build a classification model using any Ensemble Learning model.
2. Split data into training and testing sets. Evaluate model using accuracy, precision, and recall.
3. Plot feature importance and interpret key drivers of customer behavior.

**Checkpoint Q2:** Why might an ensemble method outperform a simple logistic regression here?

#### Market Basket Insights

- Use association rule mining to identify frequently co-purchased book genres:
  - Convert genre purchase data into binary format
  - Apply Apriori algorithm to extract rules with support, confidence, and lift
  - Interpret the top 3 rules and provide actionable insights

## **Deliverables**

1. Jupyter Notebooks for each case study with code, visualizations, and explanations to technical questions.
2. Summary report (executive summary, visual insights, methodology, and recommendations) for a business manager outlining findings and recommendations. IEEE-style, max three pages for each case study (excluding references and appendix).