USIU AFRICA
United States International University-Africa

Samuel Lumumba:664769

Project Topic: Time-Series Machine learning

Student Name: Samuel Lumumba : 664769
Supervisor Name: Mr. Fredrick Ogore

CHAPTER 1

Project Title

Super-Store Time Series Analysis for Enhanced Inventory Management and Sales Forecasting

Problem Statement

Inventory management and sales forecasting are critical challenges faced by superstores, particularly in the furniture, office supplies, and technology categories. Inaccurate sales forecasts and inadequate insights into consumer behavior often lead to understocking or overstocking issues, resulting in lost sales opportunities, increased holding costs, and diminished customer satisfaction. Traditional models like ARIMA and SARIMA, while effective in capturing linear patterns, often struggle with non-linear patterns and complex seasonality prevalent in retail sales data. Recent advancements in machine learning, such as Facebook Prophet and Long Short-Term Memory (LSTM) networks, offer promising solutions but require thorough evaluation and integration with existing data frameworks.

Superstores need a robust forecasting model that can accurately predict sales trends, identify seasonality, and account for external factors such as promotions, holidays, and economic indicators. This project aims to leverage advanced time series analysis techniques to develop a comprehensive solution that addresses these challenges, providing actionable insights for optimizing inventory management and enhancing marketing strategies. By deploying the best-performing models, the project seeks to improve overall business performance, reduce costs, and increase customer satisfaction.

Project Objectives

Main Objective

To develop a robust time series analysis model that accurately forecasts sales and provides actionable insights for inventory management and marketing strategies in the furniture, office supplies, and technology categories of a Superstore.

Specific Objectives

1. **To Integrate** external factors such as promotions, holidays, and economic indicators into the sales forecasting models to improve accuracy.
2. **To Enhance** the performance of traditional forecasting models by combining them with advanced machine learning techniques like Facebook Prophet and LSTM networks.
3. **To Develop** a user-friendly dashboard using Streamlit for real-time visualization of forecasts and insights, facilitating better decision-making.
4. **To Streamline** the model deployment process by utilizing scalable cloud platforms, ensuring continuous refinement and adaptation based on new data and changing consumer behavior.

Justification

Effective inventory management and accurate sales forecasting are essential for the success of any retail business. Implementing advanced time series analysis techniques in a Superstore can significantly improve inventory accuracy, reduce holding costs, and enhance customer satisfaction by ensuring product availability. This project addresses the limitations of traditional forecasting models and integrates external factors that influence sales, providing a more comprehensive and accurate forecasting solution. By deploying the best-performing models and creating an interactive dashboard, stakeholders can make informed decisions in real-time, leading to optimized inventory levels, targeted marketing strategies, and improved overall business performance. The significance of this project lies in its potential to transform inventory management practices, drive sales growth, and maintain a competitive edge in the retail market.

CHAPTER 2: Literature Review

Overview: Time series forecasting has been widely researched, particularly in retail. In Kenya and across Africa, few studies focus on using advanced machine learning for sales forecasting in superstores. Globally, ARIMA and SARIMA models are standard, but they struggle with non-linear data and complex seasonality. Recent advancements like Facebook Prophet and LSTM networks offer improved accuracy, handling irregular data patterns better. This study's uniqueness lies in integrating these models with external factors like holidays and promotions, enhancing prediction accuracy for the retail sector.

Key Studies and Contributions:

1. **Kenya and Africa:**
   - Limited research exists on applying advanced machine learning models like Prophet and LSTM in retail forecasting.
   - Existing studies primarily focus on traditional models or lack integration of external factors affecting sales.
2. **Global Studies:**
   - **Hyndman & Athanasopoulos (2018):** Explored traditional time series models and identified their limitations in retail forecasting.
   - **Taylor & Letham (2018):** Developed Facebook Prophet, emphasizing its ability to handle missing data and outliers in business forecasting.
   - **Zhang (2003):** Introduced hybrid models combining ARIMA with machine learning to improve forecasting accuracy.

Unique Contribution: This project’s distinctiveness lies in its holistic approach, integrating advanced models (Prophet, LSTM) with traditional ones, while incorporating external factors like economic indicators and promotions, specifically tailored for superstores in a developing country context like Kenya.

CHAPTER 3: Methodology

3.0 Methodology Overview: The project uses a mixed-method approach, combining data-driven techniques with advanced machine learning models to forecast sales. The methodology focuses on enhancing traditional forecasting methods by incorporating modern algorithms and integrating external factors that influence sales.

3.1 Preliminary Investigation:

- **Objective:** Understand the existing challenges in inventory management and sales forecasting.
- **Data Collection:** Sales data from a superstore, including external factors like economic indicators, promotions, and holidays.
- **Analysis:** Initial analysis to identify patterns, trends, and seasonality in the data.
- **Outcome:** Identified the need for models that can handle complex seasonality and non-linear patterns.

3.2 Design Phase:

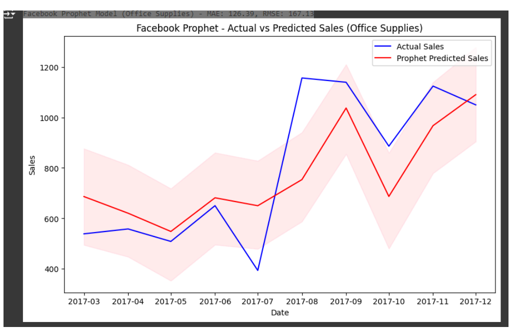
- **Model Selection:** ARIMA, SARIMA for traditional forecasting; Facebook Prophet and LSTM for advanced analysis.
- **Feature Engineering:** Integrated external factors like holidays and economic indicators into the models.
- **Prototype Development:** Designed a user-friendly dashboard using Streamlit for real-time visualization of forecasts.
- **Model Evaluation:** Used Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess model performance.

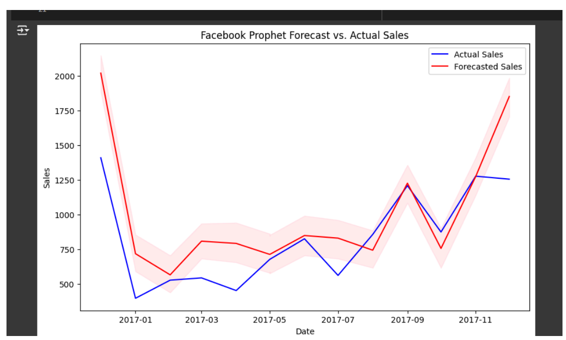
CHAPTER 4: Implementation

Have a look at : https://github.com/Kelta153/Mid-term-Project-

4.0 System Analysis & Design:

- **Objective:** Develop a robust system that integrates various models and external data for accurate sales forecasting.
- **System Components:**
  - Data Preprocessing Module: Cleans and prepares the data for analysis.
  - Forecasting Module: Implements ARIMA, SARIMA, Facebook Prophet, and LSTM models.
  - Visualization Module: Interactive dashboard for real-time insights.
- **Tools & Technologies:** Python, Streamlit, Pandas, Statsmodels, Prophet, Keras (for LSTM).





4.1 Data Flow Diagram (DFD):

- **Level 1: Data Collection and Preprocessing**
  - Sales data is gathered and cleaned.
  - External factors are merged to enhance the forecasting model.
- **Level 2: Model Execution**
  - The data flows into the selected models (ARIMA, SARIMA, Prophet, LSTM).
  - Forecast results are generated and stored.
- **Level 3: Visualization**
  - The dashboard pulls forecast data and displays it in real-time for user analysis.

```
1 technology_data.head()
```

	Order Date	Sales	is_holiday	UMCSENT	GDP	UNRATE
0	2014-01-01	449.041429	0.0	81.2	17197.738	6.6
1	2014-02-01	229.787143	1.0	81.6	17197.738	6.7
2	2014-03-01	2031.948375	0.0	80.0	17197.738	6.7
3	2014-04-01	613.028933	0.0	84.1	17518.508	6.2
4	2014-05-01	564.698588	1.0	81.9	17518.508	6.3

Next steps: Generate code with technology_data | View recommended plots

```
1 office_supplies_data.head()
```

	Order Date	Sales	is_holiday	UMCSENT	GDP	UNRATE
0	2014-01-01	285.357647	0.0	81.2	17197.738	6.6
1	2014-02-01	63.042588	1.0	81.6	17197.738	6.7
2	2014-03-01	391.176318	0.0	80.0	17197.738	6.7
3	2014-04-01	464.794750	0.0	84.1	17518.508	6.2
4	2014-05-01	324.346545	1.0	81.9	17518.508	6.3

Next steps: Generate code with office_supplies_data | View recommended pl

```
1 furniture_data.head()
```

	Order Date	Sales	is_holiday	UMCSENT	GDP	UNRATE
0	2014-01-01	480.194231	0.0	81.2	17197.738	6.6
1	2014-02-01	367.931600	1.0	81.6	17197.738	6.7
2	2014-03-01	857.291529	0.0	80.0	17197.738	6.7

4.2 Entity Relationship (ER) Diagram:

- **Entities:**
  - **Sales Data:** Includes historical sales records by category.
  - **External Factors:** Comprising promotions, holidays, and economic indicators.
  - **Forecast Models:** ARIMA, SARIMA, Prophet, LSTM.
  - **User Interface:** The dashboard where users interact with the data.
- **Relationships:**
  - Sales Data and External Factors are inputs to the Forecast Models.
  - Forecast Models generate predictions that are visualized through the User Interface

```
1 technology_data.head(10)
```

	Order Date	Sales	is_holiday	UMCSENT	GDP	UNRATE	is_promotion	lag_1	lag_7	rolling_mean_7	rolling_std_7
0	2014-01-01	449.041429	0.0	81.2	17197.738	6.6	0	0.000000	0.000000	0.000000	0.000000
1	2014-02-01	229.787143	1.0	81.6	17197.738	6.7	0	449.041429	0.000000	0.000000	0.000000
2	2014-03-01	2031.948375	0.0	80.0	17197.738	6.7	0	229.787143	0.000000	0.000000	0.000000
3	2014-04-01	613.028933	0.0	84.1	17518.508	6.2	0	2031.948375	0.000000	0.000000	0.000000
4	2014-05-01	564.698588	1.0	81.9	17518.508	6.3	0	613.028933	0.000000	0.000000	0.000000
5	2014-05-01	564.698588	1.0	81.9	17518.508	6.3	0	564.698588	0.000000	0.000000	0.000000
6	2014-05-01	564.698588	1.0	81.9	17518.508	6.3	0	564.698588	0.000000	716.843092	594.231207
7	2014-05-01	564.698588	1.0	81.9	17518.508	6.3	0	564.698588	449.041429	733.365543	587.109252
8	2014-06-01	766.905909	1.0	82.5	17518.508	6.1	0	564.698588	229.787143	810.096796	543.829569
9	2014-07-01	533.608933	0.0	81.8	17804.228	6.2	0	766.905909	2031.948375	596.048304	78.856998

Next steps: Generate code with technology_data | View recommended plots | New interactive sheet

.

```
1 furniture_data.head(10)
```

	Order Date	Sales	is_holiday	UMCSENT	GDP	UNRATE	is_promotion	lag_1	lag_7	rolling_mean_7	rolling_std_7
0	2014-01-01	480.194231	0.0	81.2	17197.738	6.6	0	0.000000	0.000000	0.000000	0.000000
1	2014-02-01	367.931600	1.0	81.6	17197.738	6.7	0	480.194231	0.000000	0.000000	0.000000
2	2014-03-01	857.291529	0.0	80.0	17197.738	6.7	0	367.931600	0.000000	0.000000	0.000000
3	2014-04-01	567.488357	0.0	84.1	17518.508	6.2	0	857.291529	0.000000	0.000000	0.000000
4	2014-05-01	432.049188	1.0	81.9	17518.508	6.3	0	567.488357	0.000000	0.000000	0.000000
5	2014-05-01	432.049188	1.0	81.9	17518.508	6.3	0	432.049188	0.000000	0.000000	0.000000
6	2014-05-01	432.049188	1.0	81.9	17518.508	6.3	0	432.049188	0.000000	509.864754	164.932810
7	2014-05-01	432.049188	1.0	81.9	17518.508	6.3	0	432.049188	480.194231	502.986891	167.362271
8	2014-06-01	695.059242	1.0	82.5	17518.508	6.1	0	432.049188	367.931600	549.719411	169.029157
9	2014-07-01	601.169500	0.0	81.8	17804.228	6.2	0	695.059242	857.291529	513.130550	108.089798

```
1 office_supplies_data.head(10)
```

	Order Date	Sales	is_holiday	UMCSENT	GDP	UNRATE	is_promotion	lag_1	lag_7	rolling_mean_7	rolling_std_7
0	2014-01-01	285.357647	0.0	81.2	17197.738	6.6	0	0.000000	0.000000	0.000000	0.000000
1	2014-02-01	63.042588	1.0	81.6	17197.738	6.7	0	285.357647	0.000000	0.000000	0.000000
2	2014-03-01	391.176318	0.0	80.0	17197.738	6.7	0	63.042588	0.000000	0.000000	0.000000
3	2014-04-01	464.794750	0.0	84.1	17518.508	6.2	0	391.176318	0.000000	0.000000	0.000000
4	2014-05-01	324.346545	1.0	81.9	17518.508	6.3	0	464.794750	0.000000	0.000000	0.000000
5	2014-05-01	324.346545	1.0	81.9	17518.508	6.3	0	324.346545	0.000000	0.000000	0.000000
6	2014-05-01	324.346545	1.0	81.9	17518.508	6.3	0	324.346545	0.000000	311.058706	124.335462
7	2014-05-01	324.346545	1.0	81.9	17518.508	6.3	0	324.346545	285.357647	316.628548	123.864645
8	2014-06-01	588.774409	1.0	82.5	17518.508	6.1	0	324.346545	63.042588	391.733094	101.921034
9	2014-07-01	756.060400	0.0	81.8	17804.228	6.2	0	588.774409	391.176318	443.859392	171.289912

References

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- Seaman, C. (2018). Inventory management and demand forecasting in retail. *Journal of Business Research, 85*, 239-250.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37-45.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159-175.
- Federal Reserve Economic Data (FRED). (n.d.). *University of Michigan: Consumer Sentiment (UMCSENT)*. Retrieved from https://fred.stlouisfed.org/series/UMCSENT

- U.S. Department of Commerce. (n.d.). *Gross Domestic Product (GDP)*. Retrieved from https://www.commerce.gov/tags/gross-domestic-product-gdp
- WorldPop. (n.d.). *WorldPop Population Data*. Retrieved from https://hub.worldpop.org/geodata/summary?id=50097