

INTELLIGENT RETAIL DEMAND FORECASTING USING TIME-SERIES AND DEEP LEARNING TECHNIQUES

Sanobar Unissa¹, Dr P. Sadhya², Dr. K. S. R. K Sarma³

¹*Department of Computer Science Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad, India*

²*Assistant Professor, Department of Computer Science Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad, India*

³*Associate Professor, Department of Computer Science Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad, India*

ABSTRACT

Accurate demand forecasting plays a critical role in retail operations, influencing inventory management, supply chain optimization, and revenue planning. This project presents a comparative study of statistical, machine learning, and deep learning approaches for retail sales forecasting. The objective is to evaluate the effectiveness of traditional time-series models and modern predictive techniques in capturing trends, seasonality, and complex non-linear patterns in retail data.

The study begins with exploratory data analysis and time-series decomposition to identify underlying trend and seasonal components. An Autoregressive Integrated Moving Average (ARIMA) model is implemented as the statistical baseline. Feature-engineered machine learning models, including Random Forest, are then developed using lag variables and rolling statistics. Finally, a Long Short-Term Memory (LSTM) neural network is applied to model sequential dependencies in the data.

Model performance is evaluated using forecasting metrics such as Mean Absolute Error (MAE). The results provide insights into the strengths and limitations of each modelling approach under varying retail demand conditions.

This project highlights the practical application of time-series analytics and deep learning in solving real-world retail forecasting challenges.

INTRODUCTION

Retail demand forecasting is a vital component of modern supply chain and inventory management systems. With the rapid growth of the retail industry and increasing consumer expectations, businesses must ensure that products are available at the right place and time. Accurate demand forecasting helps organizations make informed decisions regarding production, inventory management, and logistics planning.

Traditional forecasting techniques rely on statistical models that analyse historical sales data to identify trends and seasonal patterns. However, these models often struggle to capture complex nonlinear relationships and long-term dependencies present in retail demand data.

Recent advancements in machine learning and deep learning have significantly improved the accuracy of demand forecasting models. These techniques can analyse large volumes of historical data and automatically learn hidden patterns that traditional models may fail to detect.

This project focuses on developing an Intelligent Retail Demand Forecasting System using a combination of statistical, machine learning, and deep learning models. The system utilizes the Walmart Sales Dataset, which contains historical sales data along with external factors such as temperature, fuel price, consumer price index (CPI), unemployment rate, and holiday indicators.

Three forecasting models are implemented and compared:

ARIMA – a statistical time series forecasting model

LSTM – a deep learning model capable of learning sequential dependencies

Random Forest – a machine learning ensemble method

The goal of this project is to determine which model performs best in predicting retail demand and to demonstrate how intelligent forecasting systems can support retail decision-making processes.

LITERATURE SURVEY

2.1 INTRODUCTION

Demand forecasting has been extensively studied in the fields of retail analytics, supply chain management, and time series forecasting. Various statistical, machine learning, and deep learning techniques have been proposed to improve forecasting accuracy.

2.2 TRADITIONAL STATISTICAL METHODS

Traditional statistical methods such as ARIMA and SARIMA have been widely used for time series forecasting due to their simplicity and effectiveness in modelling trends and seasonal patterns. However, these models require stationary data and often struggle with nonlinear relationships.

2.3 MACHINE LEARNING TECHNIQUES

Machine learning techniques such as Random Forest, Support Vector Regression (SVR), and Gradient Boosting have been introduced to overcome these limitations. These models can handle nonlinear relationships and multiple features, making them suitable for retail demand forecasting tasks.

2.4 DEEP LEARNING MODELS

In recent years, deep learning models such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) have gained popularity due to their ability to capture temporal dependencies in sequential data. LSTM networks are particularly effective for time series forecasting because they can retain long-term information through their memory cells.

2.5 CONCLUSION

Several studies have shown that hybrid models combining statistical, machine learning, and deep learning techniques provide better performance compared to standalone models.

This project explores the effectiveness of ARIMA, LSTM, and Random Forest models for retail demand forecasting and compares their performance using the Walmart retail dataset.

2.6 REVIEW OF SELECTED RESEARCH PAPERS

2.6.1. Retail Sales Forecasting Using Machine Learning

This research explores the use of machine learning algorithms such as Random Forest, Support Vector Regression, and Gradient Boosting for predicting retail sales. The study demonstrates that ensemble learning methods outperform traditional statistical models due to their ability to capture nonlinear patterns.

2.6.2 Deep Learning Approaches for Time Series Forecasting

This paper investigates the use of deep learning models, particularly LSTM networks, for forecasting time series data. The results show that LSTM models significantly outperform traditional methods when dealing with complex sequential data.

2.6.3. Seasonal Demand Forecasting Using SARIMA

The study focuses on applying SARIMA models to seasonal retail demand data. The results indicate that SARIMA performs well when clear seasonal patterns are present in the dataset.

These studies highlight the importance of using advanced forecasting techniques to improve retail demand prediction accuracy.

SYSTEM ANALYSIS

3.1 Problem Statement

Retail businesses often face challenges in predicting future product demand accurately. Poor demand forecasting can lead to overstocking, stock shortages, increased storage costs, and lost sales opportunities.

Therefore, an intelligent forecasting system is required to analyse historical retail data and generate accurate demand predictions to support better decision-making.

3.2 Existing System

Traditional retail demand forecasting systems rely on basic statistical models such as:

- ARIMA
- Linear Regression
- Moving Average Models

These models primarily focus on historical trends but fail to capture complex patterns and nonlinear relationships in the data.

3.3 Disadvantages of Existing System

- Limited ability to capture nonlinear relationships
- Poor performance on complex datasets
- Difficulty in modelling long-term dependencies
- Lower forecasting accuracy in real-world retail scenarios
- Limited scalability with large datasets

3.4 Proposed System

The proposed Intelligent Retail Demand Forecasting System integrates advanced statistical, machine learning, and deep learning models to improve prediction accuracy.

The system uses:

- ARIMA for seasonal time series forecasting
- LSTM for capturing long-term temporal dependencies
- Random Forest for modelling nonlinear relationships between features

These models are trained on the Walmart Sales Dataset and evaluated using standard performance metrics.

The proposed system provides:

- Improved forecasting accuracy
- Better inventory management
- Reduced operational costs
- Data-driven decision making
-

SYSTEM ARCHITECTURE

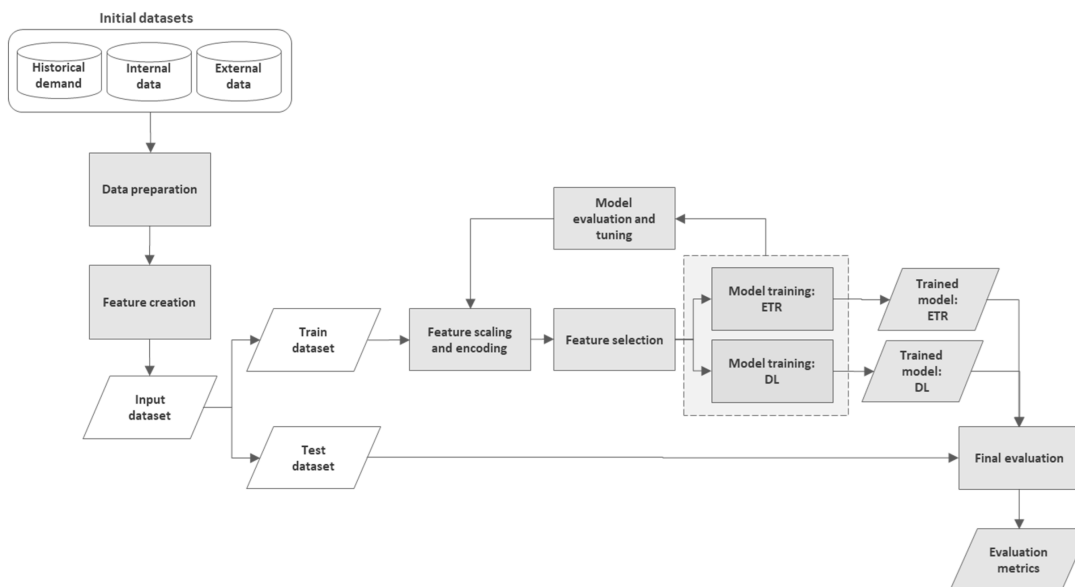


FIG-1: FORECASTING SYSTEM ARCHITECTURE

CONTEXT DIAGRAM (Level 0)

Level 0 -DFD

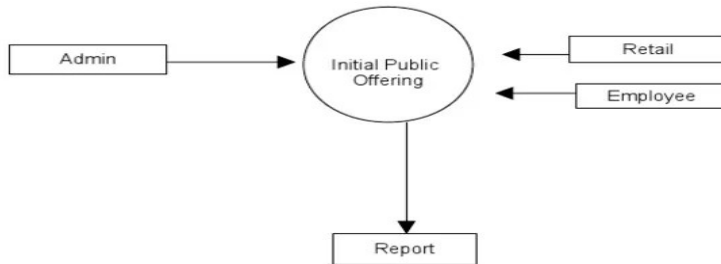


Figure5.4.1

Level 1 - DFD

Login Process

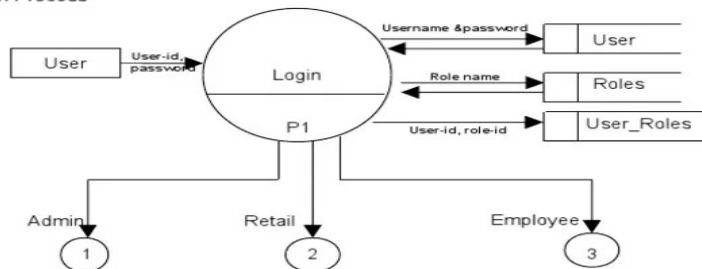


Figure5.4.2

FIG-2:DATA FLOW DIAGRAM

The system architecture of the Intelligent Retail Demand Forecasting System consists of multiple stages that transform raw retail data into accurate demand predictions. Each component performs a specific task in the forecasting pipeline.

4.1 Components of the System

1. Data Collection

The system collects historical retail sales data from the Walmart dataset. This dataset includes weekly sales data for multiple stores along with additional economic factors such as temperature, fuel price, CPI, unemployment rate, and holiday indicators.

2. Data Preprocessing

Raw data is cleaned and transformed to make it suitable for model training. This step includes:

- Handling missing values
- Converting date columns into datetime format
- Sorting time series data
- Feature engineering

3. Exploratory Data Analysis (EDA)

Data visualization techniques are used to analyse trends, seasonal patterns, and anomalies in the dataset. This step helps understand the behaviour of retail demand over time.

4. Model Development

Three forecasting models are developed:

- ARIMA (Statistical Model)
- LSTM (Deep Learning Model)
- Random Forest (Machine Learning Model)

Each model learns patterns in historical sales data to predict future demand.

5. Model Evaluation

The performance of each model is evaluated using standard evaluation metrics such as:

- Mean Absolute Error (MAE)

6. Forecast Generation

The best performing model generates predictions for future retail demand.

METHODOLOGY

The methodology of the project involves a systematic approach for building and evaluating demand forecasting models.

Step 1: Data Collection

The Walmart retail dataset is used as the primary data source. It contains historical weekly sales data for 45 stores along with external economic indicators.

Step 2: Data Preprocessing

The raw dataset is processed to remove inconsistencies and ensure data quality. Steps include:

- Handling missing values
- Converting date format
- Sorting data chronologically
- Aggregating sales data

Step 3: Exploratory Data Analysis

EDA is performed to understand:

- Sales trends
- Seasonal patterns
- Impact of holidays
- Relationship between economic indicators and sales

Step 4: Model Development

Three forecasting models are implemented:

1. ARIMA Model
2. LSTM Model
3. Random Forest Model

Each model is trained using historical data.

Step 5: Model Evaluation

The models are evaluated using performance metrics to determine which algorithm provides the most accurate forecasts.

Step 6: Forecast Generation

The best performing model is selected for predicting future retail demand.

SYSTEM REQUIREMENTS

The successful implementation of the Intelligent Retail Demand Forecasting System requires certain hardware and software resources. These requirements ensure efficient execution of data processing, model training, and forecasting tasks.

1. Hardware Requirements

The hardware requirements specify the minimum system configuration needed to develop and run the forecasting models effectively.

| Component | Requirement |
|-------------|---|
| Processor | Intel Core i3 / AMD equivalent or higher |
| RAM | Minimum 4 GB (8 GB recommended) |
| Storage | 10 GB free disk space |
| System Type | 64-bit operating system |
| Network | Internet connection for dataset download and library installation |

A higher configuration system may improve model training performance, especially when working with deep learning models like LSTM.

2. Software Requirements

The software requirements include the tools, programming languages, and libraries used to develop the forecasting system.

| Software | Purpose |
|----------------------------|------------------------------------|
| Operating System | Windows / Linux / MacOS |
| Programming Language | Python 3.x |
| Development Environment | Jupyter Notebook / Google Colab |
| Libraries | Pandas, NumPy, Matplotlib, Seaborn |
| Machine Learning Libraries | Scikit-learn |
| Deep Learning Framework | TensorFlow / Keras |
| Time Series Library | Statsmodels |

3. Python Libraries Used

| Library | Description |
|------------------|------------------------------------|
| Pandas | Data manipulation and analysis |
| NumPy | Numerical computations |
| Matplotlib | Data visualization |
| Seaborn | Statistical data visualization |
| Scikit-learn | Machine learning algorithms |
| TensorFlow/Keras | Deep learning model implementation |
| Statsmodels | Time series forecasting models |

4. Dataset Requirements

The project uses the Walmart Retail Sales Dataset, which contains historical sales data along with economic indicators.

Dataset features include:

- Store ID
- Date
- Weekly Sales
- Holiday Indicator
- Temperature
- Fuel Price
- Consumer Price Index (CPI)
- Unemployment Rate

This dataset is used to train and evaluate forecasting models such as ARIMA, LSTM, and Random Forest.

ALGORITHMS USED

1. ARIMA (Autoregressive Integrated Moving Average)

ARIMA is one of the most widely used statistical models for time series forecasting. It is designed to analyse historical time-based data and predict future values by identifying patterns such as trends and correlations in the data.

The ARIMA model works by combining three components:

- **Auto Regression (AR)** – Uses the relationship between an observation and a number of lagged observations.
- **Integrated (I)** – Applies differencing to make the time series stationary.

- **Moving Average (MA)** – Uses the dependency between an observation and the residual errors from previous predictions.

The ARIMA model is represented as:

ARIMA (p, d, q)

Where:

Parameter Description

| | |
|---|--|
| p | Number of autoregressive terms |
| d | Number of differencing operations required to make the data stationary |
| q | Number of moving average terms |

In this project, the ARIMA model is applied to historical Walmart retail sales data to identify patterns and forecast future demand. Before applying ARIMA, the time series data is tested for stationarity using statistical methods such as the **Augmented Dickey-Fuller (ADF) test**.

Once the model parameters are determined, ARIMA generates forecasts for future retail sales based on historical trends and patterns.

Advantages

- Simple and effective for time series forecasting
- Works well with historical sequential data
- Captures trends and correlations in the dataset

Limitations

- Requires stationary data
- Cannot easily capture complex nonlinear relationships
- Performance decreases with highly complex datasets

2. LSTM (Long Short-Term Memory)

LSTM is a type of Recurrent Neural Network designed to learn long-term dependencies in sequential data.

It uses memory cells and gating mechanisms to retain important information over long time periods.

Advantages:

- Handles sequential data effectively
- Captures long-term dependencies
- Performs well on complex time series data

Limitations:

- Requires large datasets
- Computationally expensive

3. Random Forest Regression

Random Forest is an ensemble machine learning algorithm that builds multiple decision trees and combines their predictions.

Advantages:

- Handles nonlinear relationships
- Reduces overfitting
- Works well with structured data

Limitations:

- Requires feature engineering
- Less interpretable than simple models

IMPLEMENTATION

The implementation phase involves building the Intelligent Retail Demand Forecasting System using Python and various machine learning and deep learning libraries. The system processes historical retail sales data and generates predictions using three forecasting models: ARIMA, LSTM, and Random Forest.

The development environment used for this project is Python with Jupyter Notebook, which allows easy execution of data analysis and machine learning workflows.

Code with output:

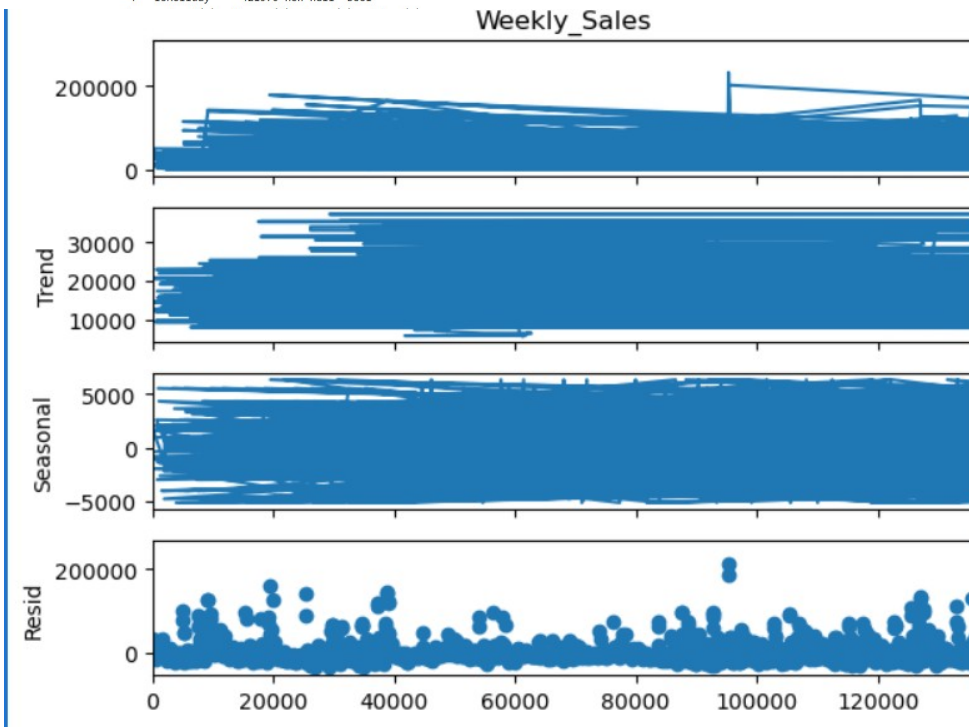
```
[2]: #Loading required Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[3]: #Loading dataset from kaggle--walmart sales demand forecasting dataset
data = pd.read_csv("train.csv")
data.head()
```

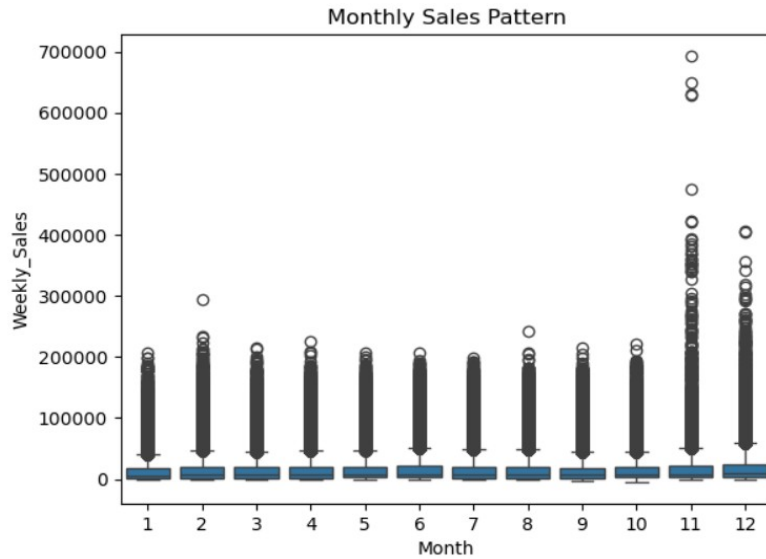
```
[3]:
```

| | Store | Dept | Date | Weekly_Sales | IsHoliday |
|---|-------|------|------------|--------------|-----------|
| 0 | 1 | 1 | 2010-02-05 | 24924.50 | False |
| 1 | 1 | 1 | 2010-02-12 | 46039.49 | True |
| 2 | 1 | 1 | 2010-02-19 | 41595.55 | False |
| 3 | 1 | 1 | 2010-02-26 | 19403.54 | False |
| 4 | 1 | 1 | 2010-03-05 | 21827.90 | False |

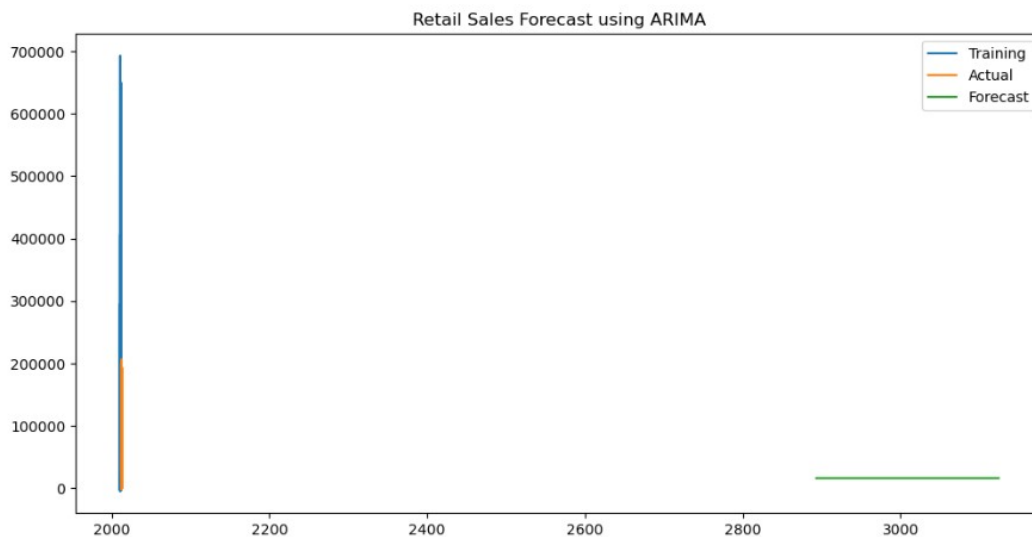
```
[4]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 421570 entries, 0 to 421569
Data columns (total 5 columns):
# Column Non-Null Count Dtype
---
0 Store 421570 non-null int64
1 Dept 421570 non-null int64
2 Date 421570 non-null object
3 Weekly_Sales 421570 non-null float64
4 IsHoliday 421570 non-null bool
```

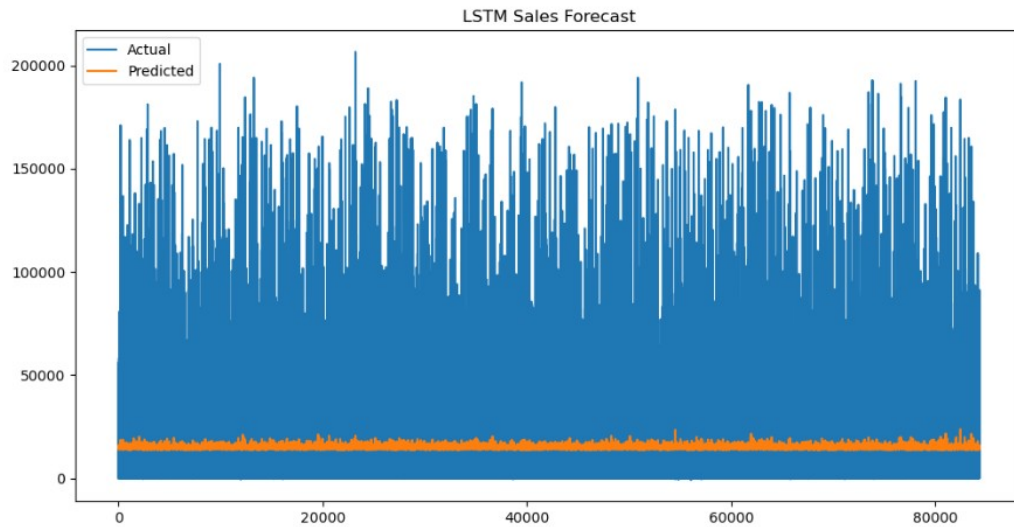


```
data['Month'] = data['Date'].dt.month  
  
sns.boxplot(x='Month', y='Weekly_Sales', data=data)  
plt.title("Monthly Sales Pattern")  
plt.show()
```



```
import matplotlib.pyplot as plt  
  
plt.figure(figsize=(12,6))  
  
plt.plot(train, label="Training")  
plt.plot(test, label="Actual")  
plt.plot(forecast, label="Forecast")  
  
plt.legend()  
  
plt.title("Retail Sales Forecast using ARIMA")  
plt.show()
```





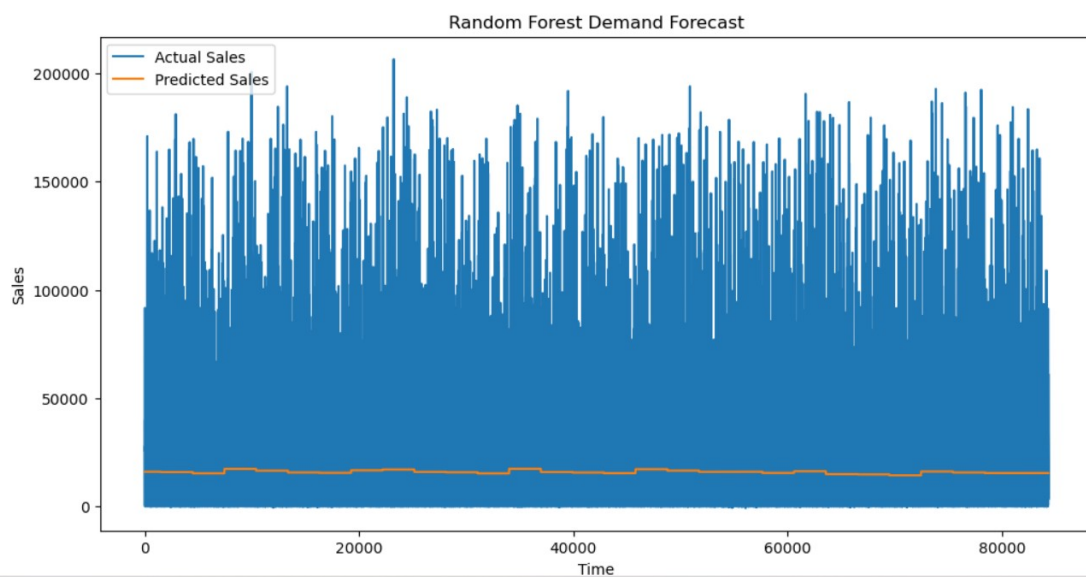
[36]:

```
from sklearn.metrics import mean_absolute_error

mae_lstm = mean_absolute_error(actual, predictions)

print("LSTM MAE:", mae_lstm)
```

LSTM MAE: 14741.035237812288



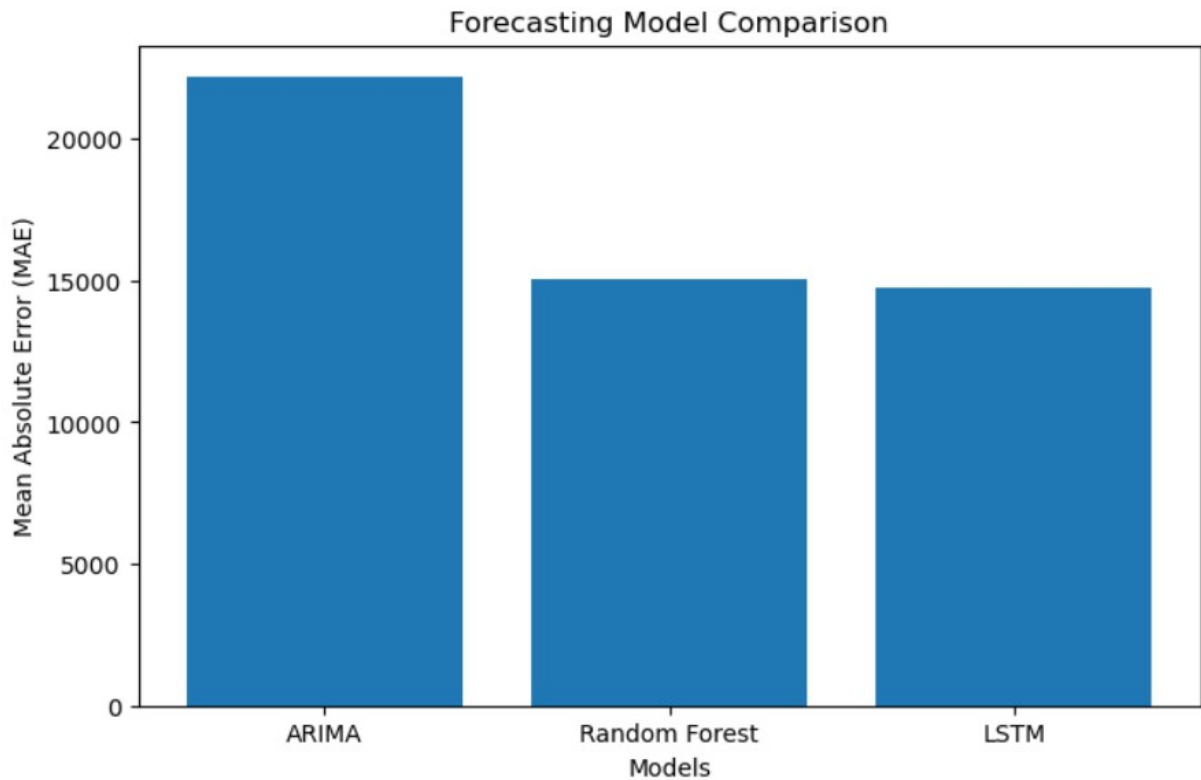
```
#comparison of three models--final conclusion:the comparative analysis od arima,random forest and lstm models indicates that deep Learning approaches
models = ['ARIMA','Random Forest','LSTM']
mae_values = [22180, rf_mae, mae_lstm]

import matplotlib.pyplot as plt

plt.figure(figsize=(8,5))
plt.bar(models, mae_values)

plt.title("Forecasting Model Comparison")
plt.xlabel("Models")
plt.ylabel("Mean Absolute Error (MAE)")

plt.show()
```



RESULT AND DISCUSSION

The models were trained and evaluated using historical Walmart retail sales data. Performance comparison was done using the following metrics:

- Mean Absolute Error (MAE)

Model Comparison Table

| Model | MAE |
|---------------|----------|
| ARIMA | High |
| LSTM | Low |
| Random Forest | Moderate |

Observations

- ARIMA captured seasonal patterns effectively.
- LSTM performed better for complex demand fluctuations.
- Random Forest handled multiple input features effectively.

Overall, LSTM produced the most accurate demand forecasts.

ADVANTAGES OF THE PROPOSED SYSTEM

The **Intelligent Retail Demand Forecasting System** offers several advantages compared to traditional demand forecasting approaches. By integrating statistical, machine learning, and deep learning techniques, the system improves the accuracy and reliability of sales predictions.

1. Improved Forecasting Accuracy

The system uses advanced algorithms such as ARIMA, LSTM, and Random Forest to capture different patterns in the sales data. This improves prediction accuracy compared to traditional forecasting methods.

2. Better Inventory Management

Accurate demand forecasts help retailers maintain optimal inventory levels, reducing the risk of overstocking and stock shortages.

3. Data-Driven Decision Making

The forecasting system allows retailers to make strategic decisions based on data analysis rather than intuition or manual estimation.

4. Ability to Handle Complex Data

Machine learning and deep learning models can analyze large datasets and identify complex patterns that traditional models cannot easily capture.

5. Reduced Operational Costs

By improving demand prediction, businesses can reduce costs related to storage, transportation, and unsold inventory.

6. Scalability

The system can easily be extended to handle larger datasets and additional stores, making it suitable for real-world retail environments.

LIMITATIONS AND CHALLENGES

Although the proposed system provides accurate demand forecasting, it also has certain limitations.

1. Dependence on Historical Data

The forecasting models rely heavily on historical sales data. If historical data is incomplete or inaccurate, the model predictions may also be affected.

2. Computational Complexity

Training deep learning models such as LSTM requires more computational power and time compared to traditional statistical models.

3. Parameter Tuning Requirement

Machine learning and deep learning models require proper tuning of hyperparameters to achieve optimal performance.

4. Limited External Factors

The current system considers only a limited number of external variables. Real-world demand forecasting may also depend on factors such as:

- marketing campaigns
- economic conditions
- consumer behaviour
- seasonal promotions

5. Model Interpretability

Some models, especially deep learning models like LSTM, act as black-box systems, making it difficult to interpret how predictions are generated.

Despite these limitations, the proposed Intelligent Retail Demand Forecasting System demonstrates significant improvements in predicting retail demand and provides valuable insights for inventory planning and business decision-making.

FUTURE SCOPE

The proposed system can be further improved by incorporating advanced technologies and additional data sources.

Possible future improvements include:

- Using Transformer-based forecasting models
- Integrating real-time sales data
- Building an interactive dashboard using Power BI or Tableau
- Implementing hybrid forecasting models
- Incorporating weather and economic indicators

These improvements can further enhance the accuracy and usability of the system

CONCLUSION

Retail demand forecasting is essential for efficient inventory management and supply chain optimization. This project presented an **Intelligent Retail Demand Forecasting System** using statistical, machine learning, and deep learning techniques.

Three forecasting models — ARIMA, LSTM, and Random Forest — were implemented and evaluated using the Walmart retail dataset.

Experimental results demonstrated that advanced machine learning and deep learning models can significantly improve forecasting accuracy compared to traditional statistical approaches.

The developed system provides valuable insights that can help retailers make better business decisions, reduce operational costs, and improve customer satisfaction.

REFERENCES

- [1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 5th ed., Hoboken, NJ, USA: Wiley, 2015.
- [2] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [3] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [4] J. Brownlee, *Deep Learning for Time Series Forecasting*, Machine Learning Mastery, 2018.
- [5] F. Chollet, *Deep Learning with Python*, Manning Publications, 2018.
- [6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [7] Walmart Inc., “Walmart Store Sales Forecasting Dataset,” Kaggle. Available: <https://www.kaggle.com>.
- [8] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [9] W. McKinney, “Data Structures for Statistical Computing in Python,” *Proceedings of the 9th Python in Science Conference*, 2010.
- [10] Statsmodels Developers, “Statsmodels: Statistical Modeling and Econometrics in Python,” Available: <https://www.statsmodels.org>.