

DATA ANALYSIS AND DATA SCIENCE WITH PYTHON - Task 6 Report

Part 1: Time Series Analysis – Sales Forecasting

Objective

To analyze historical sales data and forecast future sales using statistical time series modeling, specifically the ARIMA model.

1. Dataset Overview

- Columns Required:
 - Date: Timestamp of each recorded sale.
 - Sales: Number of units sold or revenue generated on that date.

2. Steps Performed

a. Visualization of Sales Trends

- Created line plots of sales over time.
- Identified patterns such as:
 - Trends: Long-term growth or decline in sales.
 - Seasonality: Regular patterns occurring at specific intervals.
 - Irregularities: Spikes or drops due to external events.
- Applied moving averages to smooth short-term fluctuations.

b. Forecasting Using ARIMA

- Preprocessed the data (e.g., converting dates, ensuring stationarity).
- Trained the ARIMA (p,d,q) model on historical sales data.
- Forecasted sales for future time periods.
- Evaluated the model using:
 - RMSE (Root Mean Square Error)
 - MAPE (Mean Absolute Percentage Error)

3. Deliverables

- Forecast Table: Predicted sales values for upcoming periods.

- Plots Included:
 - Historical trend visualization.
 - Forecast plot comparing predicted and actual sales.

Part 2: Predicting Heart Disease Using Logistic Regression

Objective

To build a predictive model using logistic regression to determine the likelihood of heart disease based on patient medical data.

1. Dataset Overview

- Dataset: heart_disease.csv
- Features:
 - Age
 - Gender
 - Cholesterol levels
 - Blood Pressure (Systolic/Diastolic)
 - Heart Disease (Target: 1 = Disease, 0 = No Disease)

2. Steps Performed

a. Data Cleaning

- Checked for missing values and duplicates.
- Performed appropriate imputation or removal.

b. Feature Engineering

- Normalized numerical variables for better model convergence.
- Encoded categorical variables (e.g., Gender).

c. Model Training

- Trained a Logistic Regression model to classify heart disease presence.

d. Model Evaluation

- Evaluated performance using:
 - Confusion Matrix

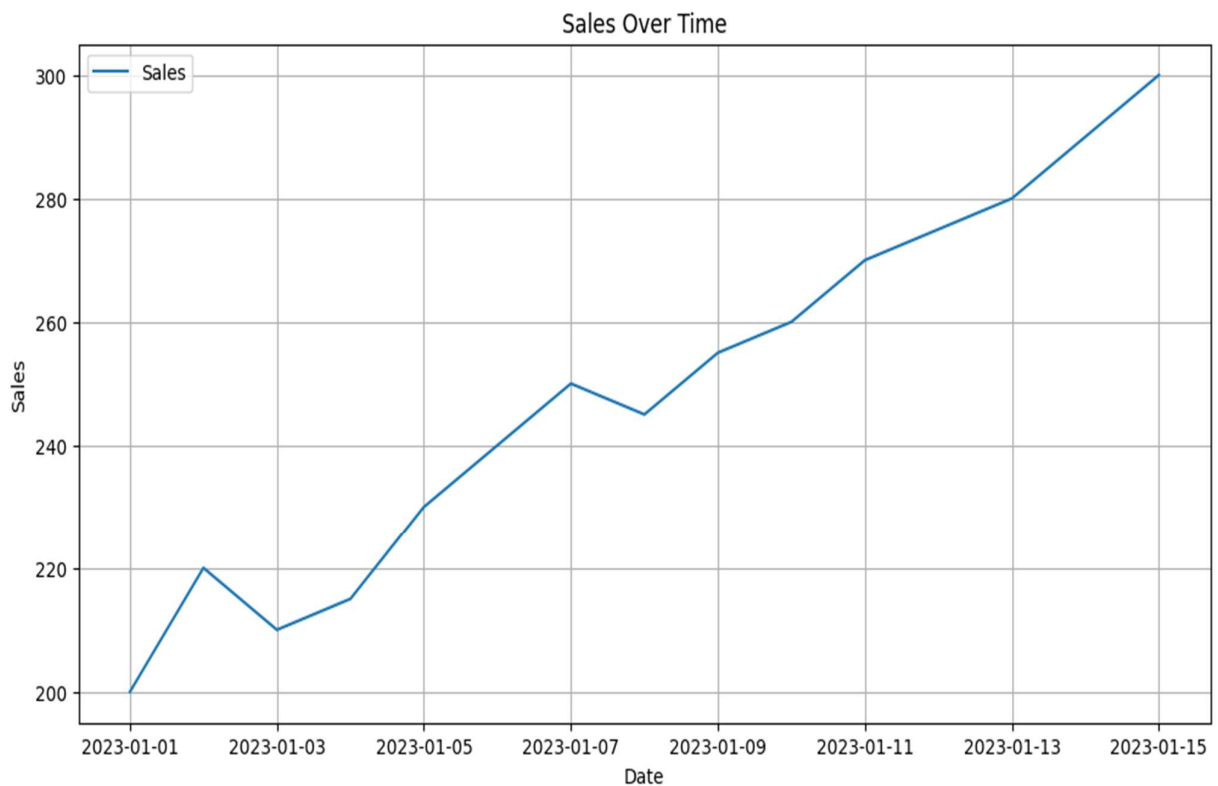
- Accuracy
- Precision
- Recall
- F1-Score

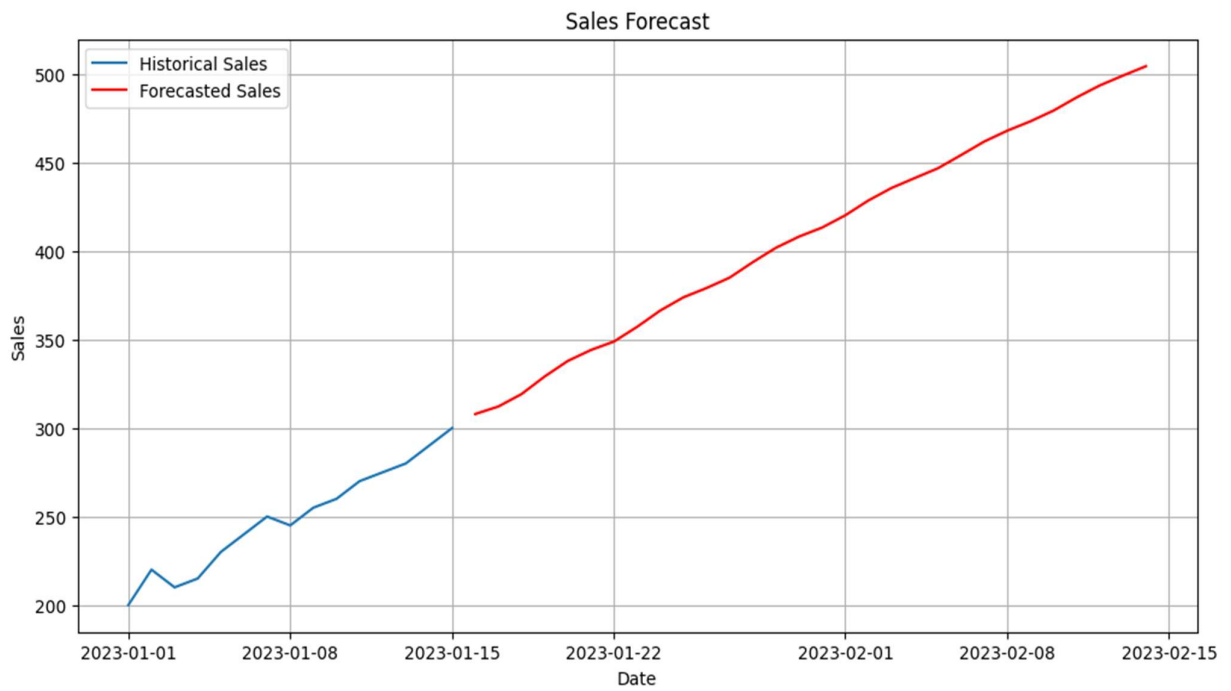
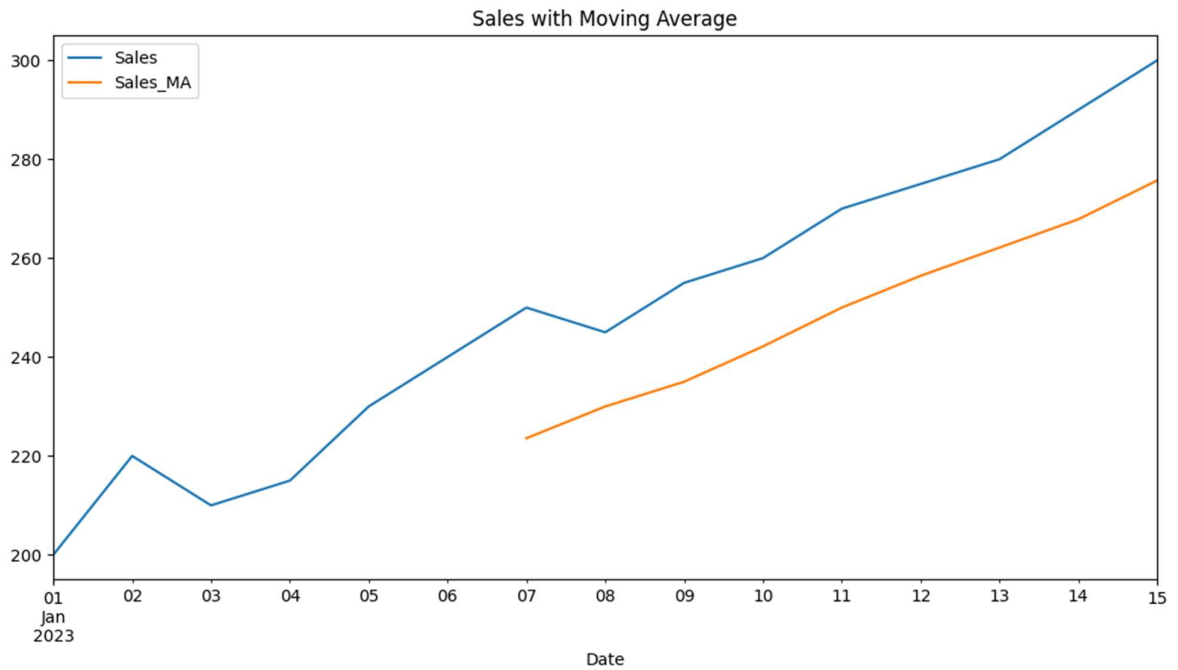
3. Deliverables

- Model: Trained logistic regression classifier.
- Evaluation Report:
 - Confusion matrix visualization.
 - Metric values with interpretation.

RESULTS:

Part 1: Time Series Analysis – Sales Forecasting





RMSE: 52.28, MAPE: 8.92%

Part 2: Predicting Heart Disease Using Logistic Regression

```
➦ Confusion Matrix:
[[1 0]
 [0 1]]

Classification Report:
              precision    recall  f1-score   support

      0       1.00      1.00      1.00         1
      1       1.00      1.00      1.00         1

   accuracy       1.00
  macro avg       1.00
 weighted avg       1.00
```