

## Title: Forecasting Trends Using Predictive Models

### Objective:

To forecast trends and evaluate the predictive accuracy of models using real-world-like data. This study compares Linear Regression, Logistic Regression, and ARIMA (AutoRegressive Integrated Moving Average) to understand their strengths in trend forecasting and classification.

### Dataset Overview:

A synthetic monthly retail **sales dataset (Jan 2022 – Dec 2023)** with 24 entries was used to simulate time-dependent sales behavior. The dataset helps demonstrate practical forecasting models for business decisions.

### Columns:

- Month: Categorical (converted to datetime)
- Sales: Numerical (target variable)

### Sample:

Month	Sales
Jan-2022	120
Jun-2023	260
Dec-2023	300

### Applied Forecasting Techniques:

#### 1. Linear Regression

- Models the linear relationship between time (MonthIndex) and sales.
- Captures consistent upward/downward trends.

#### 2. Logistic Regression

- Converts the regression task to classification:
  - 1: High sales ( $> 200$  units)
  - 0: Low sales ( $\leq 200$  units)
- Helps identify high-performing months.

#### 3. ARIMA (1,1,1)

- Captures autoregressive and moving average patterns.
- Good for stationary time series with trend differencing.

### Evaluation Metrics:

Metric	Description
RMSE	Root Mean Squared Error – sensitive to large errors.
MAE	Mean Absolute Error – average prediction error.
R <sup>2</sup> Score	Measures variance explained by the model (closer to 1 = better fit).

### Results & Performance Analysis:

Model	RMSE	MAE	R <sup>2</sup> Score	Accuracy
Linear Regression	6.60	5.50	0.98	—
ARIMA (1,1,1)	7.91	6.43	0.98	—
Logistic Regression	—	—	—	95.83%

- Linear Regression** performed well for capturing steady growth trends.
- ARIMA** closely matched actual values but was slightly less accurate in short intervals.
- Logistic Regression** effectively flagged “High-Sales” months with 95.83% classification accuracy.

### Visualization Summary:

- A **line graph** compared actual sales with Linear and ARIMA forecasts.
  - High correlation observed between predicted and actual values in both models.
  - The classification chart showed precise month-wise identification of "High" vs. "Low" sales.
- 

### Insights:

- **Linear Regression** is excellent for trend estimation with minimal computation.
  - **ARIMA** is suitable when past values significantly influence future ones.
  - **Logistic Regression** can aid in binary forecasting, such as campaign planning or threshold-based decisions.
- 

### Tools Used:

- **Python** (Jupyter Notebook)
  - **Libraries:** pandas, numpy, matplotlib, sklearn, statsmodels
- 

### Conclusion:

Forecasting is essential for strategic planning. This comparative study shows that:

- **Linear Regression** is simple yet powerful for linear trend forecasting.
  - **ARIMA** provides temporal smoothing and precision in noisy time series.
  - **Logistic Regression** helps classify sales outcomes effectively for actionable insights.
- 

### Future Enhancements:

- Include external variables like seasonality, holiday effects, or ad campaigns.
- Explore deep learning models like **LSTM** for long-term forecasting.
- Automate rolling forecasts with model retraining for dynamic data updates.