Data-Driven Forecasting of Liquor Sale in Ranchi, Jharkhand:
Insights from ARIMA, SARIMA, and Prophet Models

Indra Mohan Shrivastava
Research Scholar (SBU232593)
Dept. of Computer Science and Engineering
Sarala Birla University, Ranchi, Jharkhand

Dr. Biswarup Samanta
Dean of Students' Welfare; Assoc. Professor, CSE
Faculty of Engineering & Computer Science
Sarala Birla University, Ranchi, Jharkhand

Abstract

The ability to accurately forecast time-dependent data is essential for informed decision-making in various domains, including supply chain management, finance, and public policy. This study presents a detailed comparative analysis of three widely adopted time series forecasting models—Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Prophet—applied to liquor Sale data from Ranchi. The dataset is examined at two distinct temporal resolutions, namely daily and monthly, to evaluate the sensitivity of each model to data granularity. The models are assessed using multiple performance indicators, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE). The results indicate that ARIMA demonstrates superior performance for aggregated monthly data due to its effectiveness in modelling stable trends, whereas Prophet yields better results for daily data owing to its flexibility in capturing irregular fluctuations and multiple seasonal patterns. The study also identifies limitations in the use of MAPE for datasets with zero or near-zero values and emphasizes the importance of selecting appropriate evaluation metrics.

Keywords— Time Series Forecasting, ARIMA, SARIMA, Prophet, Liquor Sale, RMSE, SMAPE, Sale Forecasting

I. INTRODUCTION

The growing reliance on data-driven strategies has significantly increased the importance of time series forecasting in both academic research and industrial applications. Time series analysis enables the identification of underlying patterns in temporal data, facilitating the prediction of future values based on historical observations. In sectors such as retail and beverage distribution, accurate Sale forecasting plays a crucial role in optimizing inventory levels, minimizing losses, and ensuring efficient resource allocation.

Liquor sale presents unique challenges due to its dependency on various dynamic factors such as seasonal Sale, cultural events, and consumer behaviour. The variability inherent in such datasets necessitates the use of robust forecasting models capable of adapting to both stable and volatile patterns.

Traditional models such as ARIMA and SARIMA have been extensively used due to their strong theoretical foundation and interpretability. Their effectiveness is often limited, however, when dealing with non-linear trends and irregular seasonal variations. Newer approaches such as Prophet address these challenges by incorporating flexible components that capture trend changes and multiple seasonal effects.

The primary objective of this study is to perform a systematic comparison of ARIMA, SARIMA, and Prophet models using real-world liquor sale data from Ranchi, providing insights into their applicability under different data conditions.

II. LITERATURE REVIEW

Time series forecasting has evolved significantly over the years. The ARIMA model, introduced by Box and Jenkins, has been a cornerstone in time series analysis due to its ability to model linear dependencies in stationary data. Its reliance on autoregressive and moving average components makes it particularly effective for datasets with consistent patterns.

SARIMA was developed as an extension of ARIMA to handle seasonal variations, incorporating seasonal differencing and seasonal autoregressive terms. While SARIMA improves performance in the presence of strong seasonal trends, it often requires careful parameter tuning and assumes a regular seasonal structure.

The Prophet model has gained attention for its adaptability and ease of use. Unlike traditional models, Prophet decomposes the time series into trend, seasonality, and holiday components. Its robustness to missing values and outliers makes it particularly suitable for high-frequency and noisy datasets.

Existing studies suggest that no single model consistently outperforms others across all scenarios. Model performance is highly dependent on the dataset's frequency, variability, and presence of seasonality — underscoring the need for comparative studies such as this one.

III. METHODOLOGY

The methodology involves a structured sequence of data preparation, model implementation, and performance evaluation. The dataset comprises liquor sale records from Ranchi, filtered to focus exclusively on whisky sale.

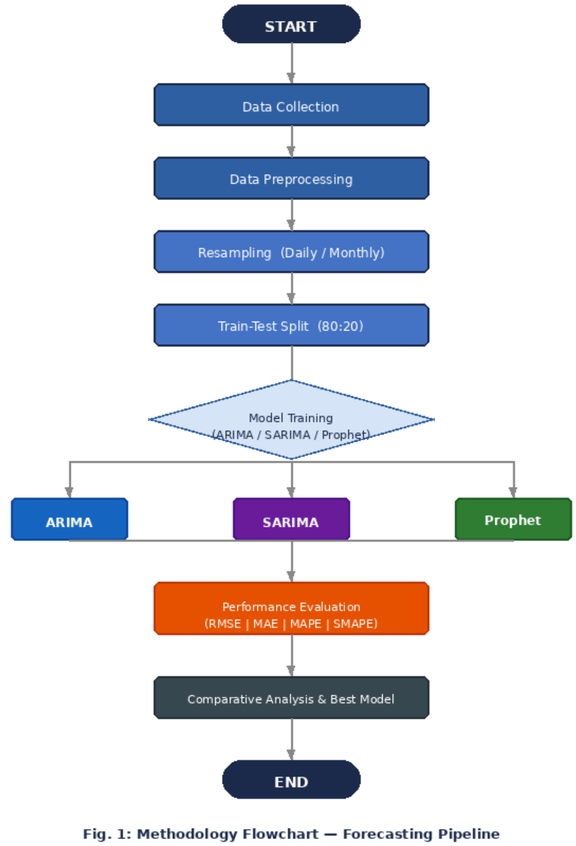

Fig. 1: Methodology Flowchart — Forecasting Pipeline

A. Dataset Description

The dataset attributes used in this study are described in Table I below:

Attribute	Description
District	Geographic location identifier — filtered to Ranchi
Category	Type of liquor — filtered to Whisky
Invoice Date	Date of transaction — converted to datetime format
Quantity in Litres	Volume consumed — converted to numeric format

Table I: Dataset Attributes and Descriptions

B. Data Preprocessing

Data preprocessing ensured accuracy and reliability through the following sequential steps:

- Conversion of date column into standardized datetime format
- Removal of invalid or missing entries to maintain data integrity
- Conversion of quantity values into numeric format
- Chronological sorting of all records

The dataset was subsequently resampled into a daily time series and a monthly aggregate series, enabling comprehensive evaluation across both granularities.

C. Train-Test Split

The dataset was divided using an 80:20 ratio. Training data was used for model fitting; test data was reserved for validation on unseen observations.

Split	Daily Data	Monthly Data
Training (80%)	~80% of daily records	~80% of monthly records
Testing (20%)	~20% of daily records	~20% of monthly records

Table II: Train-Test Split Configuration

D. Model Implementation

1) ARIMA Model

ARIMA is parameterized by (p, d, q), representing autoregression order, degree of differencing, and moving average order. Optimal parameters were identified via grid search using the Akaike Information Criterion (AIC).

Python — ARIMA Implementation

```
from statsmodels.tsa.arima.model
import ARIMA
import numpy as np

def run_arima(train, test):
    best_aic, best_order =
np.inf, (1,1,1)
    for p in range(0, 4):
      for d in range(0, 3):
        for q in range(0, 4):
          try:
           m = ARIMA(train,
order=(p,d,q)).fit()
           if m.aic < best_aic:
               best_aic   = m.aic

               best_order = (p,d,q)
          except Exception: continue
   model = ARIMA(train,
  order=best_order).fit()
   pred  =
  model.forecast(steps=len(test))
   return pred, best_order
```

2) SARIMA Model

SARIMA extends ARIMA with seasonal parameters (P, D, Q)s, where s is the seasonal period (12 for monthly; 7 for daily data), explicitly capturing periodic patterns.

Python — SARIMA Implementation

```
from statsmodels.tsa.statespace
    .sarimax import SARIMAX

def run_sarima(train, test,
```

```
                 seasonal_period):
    model = SARIMAX(
        train,
        order=(1, 1, 1),
        seasonal_order=(1, 1, 1,

seasonal_period),

enforce_stationarity=False,

enforce_invertibility=False
    ).fit(disp=False)
    pred =
model.forecast(steps=len(test))
    return pred
```

3) Prophet Model

Prophet models the time series as a sum of trend, seasonality, and holiday components. It effectively captures irregular fluctuations and multiple seasonal patterns present in daily data.

Python— Prophet Implementation

```
from prophet import Prophet
import pandas as pd

def run_prophet(train, test,
freq,
                data_type):
    df = pd.DataFrame({
        'ds': train.index,'y':
train.values})
    if data_type == 'daily':
        model = Prophet(

daily_seasonality=True,

yearly_seasonality=True)
    else:
        model = Prophet(

yearly_seasonality=True,

weekly_seasonality=False)
    model.fit(df)
    future =
model.make_future_dataframe(
        periods=len(test),
freq=freq)
    fc = model.predict(future)
    return fc['yhat'][-
len(test):].values
```

E. Evaluation Metrics

Models were evaluated using four complementary performance metrics to ensure a balanced assessment:

Python — Evaluation Function

```
from sklearn.metrics import (
     mean_squared_error,
    mean_absolute_error)

def evaluate(actual, forecast):
    a = np.array(actual)
    f = np.array(forecast)
    rmse =
np.sqrt(mean_squared_error(a, f))
    mae  = mean_absolute_error(a,
f)
    mask = a != 0
    mape = (np.mean(np.abs(
        (a[mask]-f[mask])/a[mask]
    ))*100 if np.any(mask) else
np.nan)
    smape = np.mean(
        2*np.abs(f-a) /
        (np.abs(a)+np.abs(f)+1e-
9)
    ) * 100
    return rmse, mae, mape, smape
```

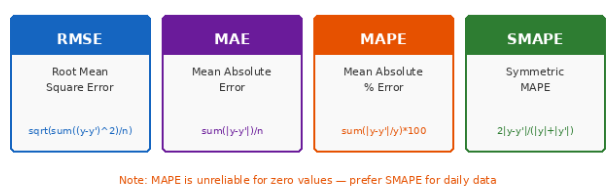

Fig. 3: Evaluation Metrics Summary

Fig. 3: Evaluation Metrics Summary

IV. RESULTS

The evaluation results reveal distinct patterns in model performance across the two temporal granularities. For the monthly dataset, ARIMA consistently achieves the lowest error values, indicating that monthly data exhibits relatively stable trends well-suited to its linear modelling approach. SARIMA shows no significant improvement, suggesting a weak or inconsistent seasonal component, while Prophet underperforms due to the absence of pronounced seasonal effects.

The analysis of daily data presents a contrasting scenario. The daily time series exhibits considerable variability with irregular fluctuations and potential zero-sale days. In this context, Prophet outperforms both ARIMA and SARIMA in terms of RMSE and MAE, demonstrating superior ability to capture complex non-linear patterns.

A. Monthly Data Analysis

Model	RMSE	MAE	MAPE (%)	SMAPE (%)
ARIMA	75,727.71	51,292.16	20.28	15.84
SARIMA	83,551.74	59,964.90	21.95	18.16
Prophet	84,115.94	63,082.65	22.25	19.85

Table III: Model Performance on Monthly Data

Observations: ARIMA outperforms all models across all metrics for monthly data. SARIMA provides no additional benefit due to weak seasonality. Prophet underperforms owing to the absence of strong seasonal signals.

B. Daily Data Analysis

Model	RMSE	MAE	MAPE (%)	SMAPE (%)
Prophet	12,386.23	9,296.42	415.98	98.63
SARIMA	12,471.98	9,488.21	403.34	98.70
ARIMA	12,843.68	9,898.33	316.34	99.83

Table IV: Model Performance on Daily Data

Observations: Prophet achieves the lowest RMSE and MAE for daily data. ARIMA shows the lowest MAPE but performs poorly in absolute terms. SMAPE values are high due to data volatility and zero-sale days.

C. Interpretation of Metrics

MAPE values are significantly elevated for daily data due to: (i) the presence of zero or near-zero sale values causing division instability, and (ii) high variability in daily Sale. RMSE and MAE are therefore more reliable indicators, while SMAPE provides a balanced relative assessment.

D. Visualization Insights

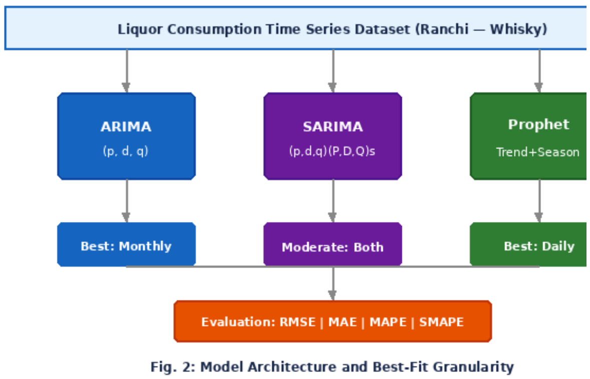

Fig. 2: Model Architecture and Best-Fit Granularity

Graphical analysis confirms that monthly data exhibits smooth trends while daily data contains spikes and irregular fluctuations. Residual analysis shows that ARIMA residuals are well-distributed for monthly data, while Prophet residuals are more stable for daily data.

V. DISCUSSION

The findings emphasize the importance of aligning model selection with dataset characteristics. ARIMA proves highly effective for aggregated monthly data where trends are stable and seasonality is minimal. Its performance diminishes when applied to highly volatile daily data. Conversely, Prophet demonstrates strong performance in handling irregular patterns but provides limited advantage in stable datasets.

These observations reinforce the notion that no single model can universally address all forecasting challenges. Practitioners must consider data frequency, variability, and seasonality when selecting an appropriate model. The critical role of metric selection is also highlighted — MAPE is unsuitable for datasets containing zero values, and SMAPE should be preferred.

VI. CONCLUSION

This research presents a comprehensive comparison of ARIMA, SARIMA, and Prophet models for forecasting liquor sale in Ranchi. ARIMA is most suitable for monthly data due to its ability to capture stable trends, while Prophet is better suited for daily data characterized by irregular fluctuations. Key findings include:

- ARIMA is most suitable for monthly aggregated data
- Prophet performs best for daily high-frequency data
- SARIMA provides limited improvement in the absence of strong seasonality
- MAPE is unreliable for datasets with zero or near-zero values
- SMAPE is recommended as the preferred balanced evaluation metric

VII. FUTURE WORK

Future research may explore integration of external variables such as weather conditions, cultural festivals, and economic indicators to enhance forecasting accuracy. Specific future directions include:

- Integration of exogenous variables — weather, festivals, economic indices
- Application of deep learning models: LSTM, GRU, Temporal Fusion Transformer
- Development of hybrid ARIMA + deep learning forecasting models
- Deployment as real-time inventory and Sale management systems
- Extension to other beverage categories and geographic regions

VII. DECLARATIONS

A. Ethics Approval and Consent to Participate

This study did not involve human participants, human tissue, personal data, or animal subjects requiring approval from an Institutional Ethics Committee. Therefore, ethical approval and consent to participate were not applicable for this research.

Where applicable, all research activities were conducted in accordance with the academic and ethical guidelines prescribed by the affiliated university and relevant regulatory authorities.

B. Consent for Publication

Not applicable.

The manuscript does not contain any individual person's data, images, videos, or personal information requiring consent for publication. All authors have reviewed and approved the final version of the manuscript and consent to its publication.

C. Availability of Data and Materials

The research data are maintained in CSV format and will be made available to editors, reviewers, or interested researchers following acceptance of the manuscript, subject to applicable institutional policies and research requirements.

D. Competing Interests

The authors declare that they have no known competing financial interests, personal relationships, or non-financial interests that could have appeared to influence the work reported in this paper.

E. Funding

The authors received no specific grant, financial support, sponsorship, or funding from any public, commercial, or not-for-profit funding agency for the conduct of this research.

F. Authors' Contributions

The Research Scholar conceived and developed the research problem, conducted the literature review, collected and processed the data, performed the analysis and experimentation, interpreted the results, and prepared the initial draft of the manuscript.

The Supervisor provided overall academic guidance and research supervision throughout the study. He contributed to the conceptualization of the research framework, reviewed the methodology and findings, provided critical feedback for improving the quality and rigor of the work, and revised the manuscript for important intellectual content. Both authors reviewed, approved, and agreed to the final version of the manuscript for submission and publication.

G. Acknowledgements

The authors express their sincere gratitude to their affiliated University for providing the academic environment, infrastructure, and support necessary for conducting this research.

The authors also acknowledge the encouragement, understanding, and moral support received from their family members and friends throughout the course of this work.

Special thanks are extended to all individuals who directly or indirectly contributed to the successful completion of this research.

REFERENCES

[1] Mahajan, H. Comparative Study of ARIMA, LSTM and Prophet Models for Time Series Forecasting: A Comprehensive Review. Preprints 2026, 2026011377. https://doi.org/10.20944/preprints202601.1377.v1

[2] Dogan, A. (2025). Time Series-Based Sale Forecasting: A Comparative Analysis of Holt-Winters, SARIMA, and Prophet Models on Retail Inventory Data. Istanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, 24(53), 641–668. https://doi.org/10.46928/iticusbe.1725584

[3] Chango, W., et al. (2026). SARIMA vs. Prophet: Comparative Efficacy in Forecasting Traffic Accidents Across Ecuadorian Provinces. Computation, 14(1), 5.

https://doi.org/10.3390/computation14010005

[4] Adebayo, P., Ahmed, I., & Oyeleke, K. T. (2025). Comparative Analysis of Prophet and ARIMA Models for Forecasting Economic Indicators. SSRN. https://ssrn.com/abstract=5206738

[5] Khanal, I., Dhakal, O. P., Guragain, M. K., Karki, P., & Pandey, B. (2023). Comparative analysis of forecasting models for PM2.5 in Kathmandu. Proc. 14th IOE Graduate Conf., 14, 1182–1189.

[6] Yadav, S., & Shukla, S. (2022). A comparative study of ARIMA, Prophet and LSTM for time series prediction. J. Artif. Intell. Mach. Learn. Data Sci, 1(1), 1813–1816.

[7] Samal, K. K. R., Babu, K. S., Das, S. K., & Acharya, A. (2019). Time series based air pollution forecasting using SARIMA and Prophet model. Proc. Int. Conf. on IT and Computer Communications, 80–85.

[8] Veikkolainen, J. E. (2024). Comparison of Predictive Models for Time Series Forecasting. Information Technology.

[9] Sunki, A., et al. (2024). Time series forecasting of stock market using ARIMA, LSTM and FB Prophet. MATEC Web of Conferences, 392, 01163. EDP Sciences.

[10] Duarte, D., & Faerman, J. (2019). Comparison of time series prediction of healthcare emergency department indicators with ARIMA and Prophet. CS & IT Computer Science Conference, 123, 33.

[11] Henderi, H., & Sofiana, S. (2025). Comparative study of traditional and modern models in time series forecasting for inflation prediction. Int. J. for Applied Information Management, 5(3), 155–167.