



Analysis and Forecasting of the Number of Shipping Bills at Each Destination Station of Nim SeeSeng Transport (1988) Co., Ltd.

Authors: Maneerat Kaeokham, Kodchakorn Aiamsuwan, Nutjira Udsuk, Nattawut Jaitan

Advisor: Assistant Professor Dr. Thanasak Mouktonglang



FACULTY OF SCIENCE
CHIANG MAI UNIVERSITY

ABSTRACT

Nim See Seng Transport (1988) Co., Ltd. continuously collects shipping data; however, the complexity of this data may hinder analysis and forecasting. Efficient data management is therefore crucial. This project aims to enhance data processing through data cleaning, product categorization, forecasting, and data-visualization. The Typhon model achieved 77.20% accuracy in product classification, outperforming traditional dictionary-based keyword-matching methods. For forecasting the number of shipping bills, five models—Prophet, LSTM, ARIMA, SARIMA, and ETS—were evaluated. Prophet demonstrated the most favorable performance, with the lowest error, yielding a Mean Absolute Error of 444.82 and a Root Mean Square Error of 745.45. Therefore, the Prophet model was selected for forecasting. Furthermore, an interactive dashboard was developed using Streamlit, allowing users to filter forecasting results by product category or destination station. This project improves resource allocation and optimizes transportation processes within the organization.

INTRODUCTION

In today's world, proper processing and understanding are crucial for effective utilization. This study analyzes logistics data from Nim See Seng Transport (1988) Co., Ltd., focusing on shipment records by destination stations.



The data was cleaned for accuracy, and product classification was implemented for organization. A forecasting model was then developed to predict shipping bills at each station, improving logistics planning and decision-making efficiency.

OBJECTIVES

- To clean and preprocess data for completeness.
- To categorize data for better organization.
- To develop and compare forecasting models for shipping bill predictions.
- To visualize data to support analysis and decision-making.

THEORIES AND TOOLS

Categorization Model

- TYPHON** Open-source AI for Thai language LLMs, enhancing cultural understanding and empowering the local tech community.

Evaluation Metric for Categorization

- Accuracy:** Measures correct predictions, crucial for classification performance.

$$\text{Accuracy} = \frac{\text{Correctly Classified Samples}}{\text{Total Samples}}$$

$$\text{Overall Accuracy} = \frac{\sum \text{Accuracy of Each Dataset}}{\text{Number of Datasets}}$$

Forecasting Models

- Prophet:** Tool for decomposing trend, seasonality, and holidays in time series.
- LSTM:** RNN for learning long-term dependencies in complex time series.
- ARIMA:** Combines autoregression, differencing, and moving averages for stationary time series.
- SARIMA:** Extends ARIMA with seasonal components for patterns.
- ETS:** Uses exponentially weighted averages for trends and seasonality.

Evaluation Metrics for Forecasting

- MAE (Mean Absolute Error):** Average absolute differences between predicted and actual values.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- RMSE (Root Mean Square Error):** Penalizes larger errors more than MAE by squaring differences.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Tools



CONCLUSION

After data cleaning and preparation, Typhon was used for product categorization, overcoming traditional keyword-matching limitations. The model achieved 77.20% accuracy from 20 randomly sampled datasets. Time series forecasting was then applied using five models: Prophet, LSTM, ARIMA, SARIMA, and ETS. Prophet performed best with the lowest error (MAE = 444.82, RMSE = 745.45), making it the most suitable model for forecasting, handling trends, seasonality, and special events effectively.

METHODOLOGY AND RESULT

1. Importing and Understanding Data

BILL_INDEX	BILL_DATE	CONSIGNEE_TAMBOON	CONSIGNEE_AMPUR	CONSIGNEE_PROVINCE	CONSIGNEE_ZIPCODE	ITEM_CODE	ITEM_DESC	ITEM_NAME	QTY	UNIT_NAME	REG_Code	BEST_CODE
0	20240301	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	50000	10105.005	เอกสารใบแจ้งหนี้	เอกสารใบแจ้งหนี้	1	ซอง	R20	D06
1	20240301	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	50000	10105.005	เอกสารใบแจ้งหนี้	เอกสารใบแจ้งหนี้	1	ซอง	R20	D06
2	20240301	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	50000	10105.005	เอกสารใบแจ้งหนี้	เอกสารใบแจ้งหนี้	1	ซอง	R20	D06
3	20240301	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	50000	10105.005	เอกสารใบแจ้งหนี้	เอกสารใบแจ้งหนี้	1	ซอง	R20	D06
4	20240301	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	51000	10601.229	สินค้าจากต่างประเทศ	สินค้าจากต่างประเทศ	14	ฟอง	R20	D03
1184921	1184921	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	99999.002	1	ฟอง	R22	D06
1184922	1184922	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	99999.002	1	ฟอง	R22	D06
1184923	1184923	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	50300	99999.002	2	ฟอง	R22	D06
1184924	1184924	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	99999.005	2	ฟอง	R05	D06
1184925	1184925	เจ้าชาย	เมืองเชียงใหม่	เชียงใหม่	99999.000	1	ฟอง	R05	D07

2. Data Preprocessing:

"ITEM_DESC" and "ITEM_NAME" were used.

Duplicates were removed, reducing items from 1,184,926 to 99,173. After cleaning in "ITEM_DESC_NEW," 89,804 unique items remained.



ITEM_DESC	ITEM_DESC_NEW
อุปกรณ์ D (ลำปาง) ขากร...	อุปกรณ์ ขากร...

ITEM_DESC	ITEM_NAME
ประจักษ์ ใต้ดิน ...	ประจักษ์ ใต้ดิน ...
ยูทีบี ...	ยูทีบี ...
ไฟฉาย ...	ไฟฉาย ...
ผ้า ...	ผ้า ...
พวงมาลัย ...	พวงมาลัย ...
...	...
ถุงผ้า ...	ถุงผ้า ...
...	...
มะเขือเทศ 28 kg (ทองหยอด) ...	มะเขือเทศ 28 kg (ทองหยอด) ...
...	...
รองเท้า ...	รองเท้า ...
เกลือ (เม็ด) 42*42*31 ...	เกลือ (เม็ด) 42*42*31 ...
...	...
ไม้ทรง ...	ไม้ทรง ...

3. Analysis and Categorization:

89,804 unique entries from "ITEM_DESC_NEW" were labeled using Typhon into 17 categories, with unclassified items in "Others".

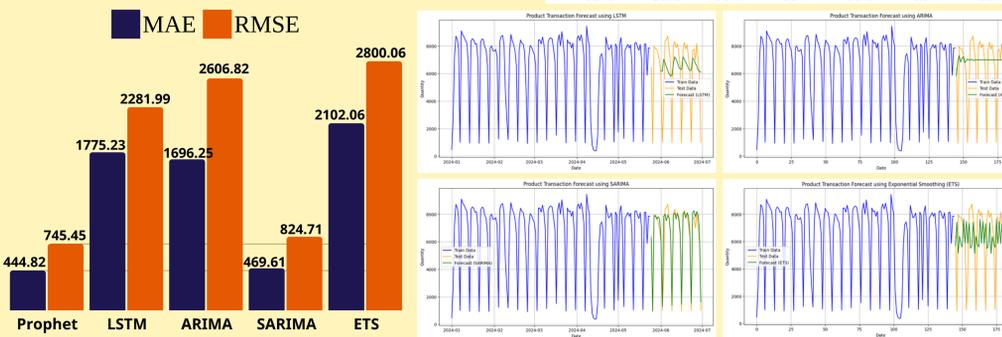
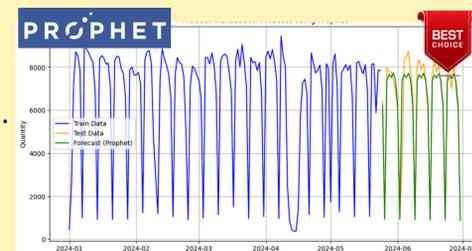
- Food and Beverages
- Fabrics and Apparel
- Electrical Appliances
- Household Items
- Office Supplies
- Tools and Equipment
- Personal Care Items
- Petrochemical and Chemical Products
- Construction Materials and Equipment
- Industrial Materials and Machinery
- Automotive Materials and Equipment
- Packaging
- Medical and Pharmaceutical Supplies
- Additional Shipping Costs
- Return Bills
- In-house Products
- Others

A random sampling approach with 20 datasets and manual validation resulted in an average **accuracy of 77.20%**.

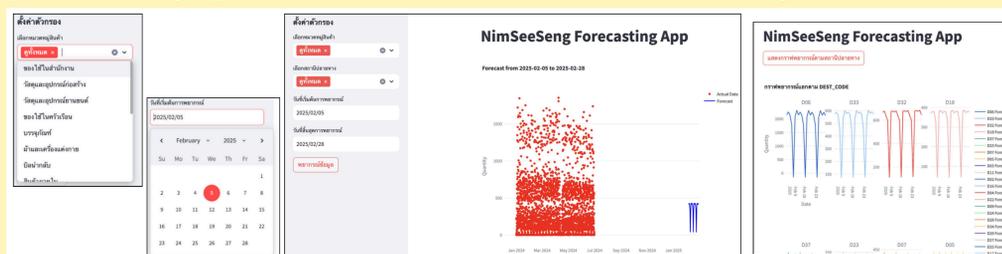
Product	Predicted	Actual
ชุดปะเก็น	ผ้าและเครื่องแต่งกาย	เครื่องมือช่าง

4. Data Forecasting:

The dataset (Jan 1–Jun 30, 2024) was split 80-20% for training and testing. Five models were tested, with Prophet proving most accurate for shipping bill prediction.



5. Data Visualization: A Streamlit dashboard, developed on VS Code, visualizes shipping bill forecasts using Prophet. Users can filter results by product category or destination station for better accessibility.



REFERENCES:

- ALGOADDICT. (2019). หลักการทำงานของ ARIMA ในการทำนายค่าของ Time Series. AlgoAddict Blog. เข้าถึงได้จาก <https://shorturl.at/ZvFtl>
- NUTHDANAI WANGPRATHAM. (2021). ทำนายข้อมูล Time Series ด้วย Prophet Python. Medium. เข้าถึงได้จาก <https://shorturl.at/UybSc>
- Sirinart Tangruamsub. (2017). Long Short-Term Memory (LSTM). Medium. เข้าถึงได้จาก <https://medium.com/@sinart.t/long-short-term-memory-lstm-e6cb23b494c6>
- Santra, R. (u.u.u.). Introduction to SARIMA Model. Medium. เข้าถึงได้จาก <https://medium.com/@ritusantra/introduction-to-sarima-model-cbb885ceabe8>

