InspironLabs
inspiration to innovation

Shipment Delivery Prediction and Real time Dashboard System

CASE STUDY

Client: **AI/ML Company**

Background

- E-commerce businesses can significantly increase their profits by choosing non-guaranteed (or local) shipment services:
- Since these non-guaranteed (or local) shippers give an unreasonably long range of delivery time, thus it drastically degrades the customer experience, and ultimately affects sales
- However, it has been observed that these non-guaranteed (or local) shippers often deliver shipments on time and at a lucrative price but they don't guarantee it
- Due to the absence of a guarantee, it becomes hard for e-commerce websites to rely on these non-guaranteed (or local) shippers

Problem Statement

- Build a system that can predict the delivery time of various non-guaranteed (or local) shipment services with high accuracy
- Along with this, the system should be capable of considering a variety of other situational parameters, such as weather, traffic, etc.

Required Solution

- It is desired to have a system that can predict delivery times with high accuracy after considering other situational parameters, such as weather, traffic, etc.
- The system should be self-learning so that it keeps improving its accuracy for delivery time with the changing scenario
- Along with this, a real-time dashboard displaying all the delivery time predictions and their accuracy is expected
- The solution needs to be implemented on AWS

Importance of Logistics

Logistics is the backbone of our economy due to the following reasons:

1. Knowing the time that it takes to transport an item from one place to another maximizes operational efficiency
2. It gives support to supply chains to
3. It reduces costs
4. It provides the flexibility to match the demand fluctuations

Importance of Shipment Prediction

By reliably knowing when a delivery will happen benefits:

E-commerce enterprises

Local stores & Customers

Trucking

Logistics and Warehousing businesses

Solution Architecture

Machine Learning Solution Stack

Matplotlib & Plotly: To visualize data and performance of trained models

Tensorflow, ScikitLearn, & Keras: Model Training packages for Shallow & DL

ScikitLearn: Data preparation for model training, such as train-test split, etc.

Numpy & Pandas: Data preprocessing, such as cleaning data, filling missing values, etc.

Map, AccuWeather, GoogleNews, Holidays: Gather data from external sources

Boto3: AWS SDK for Python. To access data in AWS S3

Predicting Shipping Time with ML (2019) LINK

- Researchers applied ML algorithms to predict the shipping time of containers between South East Asia and North America, from factory to port of destination.
- They used Random Forest algorithm and four models to produce estimates at each step of the shipping process. This functional tool yielded better results than traditional approaches.
- The freight forwarding arm of Maersk has developed a tool, Harmony, to help the decision-making process when it comes to organizing transportation for a shipment. It uses historical data to provide the use with an average and its associated deviation of the elapsed time between transportation booking and delivery at destination. This tool can be classified as a Descriptive Tool as it does not make any recommendations, but provides statistics regarding past data.
- Their model delivers predictions with a 90% accuracy (measured on the Mean Absolute Percentage Error (MAPE) of days), as early as the transport booked.
- The model relies on historical data & also on data from external sources (such as port congestion, holidays, etc.).
- This algorithm, coupled with Harmony, introduces a prediction component to the output of the tool by giving the
- user an estimated date of unloading at the destination port for a shipment.
- The used dataset had a great number of route-carrier combiantions. After consulting with Maersk, we decided to
- focus our study to shipments loaded in South East Asia (China, Hong Kong, and Vietnam) and unloaded in USA.
- Observations without any unload date were dropped.
- After this, the dataset had 1,744,278 instances across 74 different routes. The shipments are moved by 31 different carriers and 2,997 unique shippers.

Predicting Shipping Time with ML (2019) LINK

- Overall 14 different features were chosen and extracted from the historical data and used for modelling.
- Data was split as training, validation, and test data
- The training set was consists of first two years of data and used for training.
- Test set was consist of data from the most recent year.
- Python programming language was used.
- Three scripts were written:
  - First Script: It was used to clean the data, remove missing values, and transform all columns to the right format.
  - Second Script: It automatically reads in all the relevant data, trains (on new data) and saves the final model.
  - Third Script: It is used for making predictions in a production environment.
- Three different ML algorithms were used:
  - Random Forest
  - Neural Network
  - Linear Regression
- Since Random Forest was performing better than the other two, thus Random Forest was used as the final model.

Predicting Shipping Time with ML (2019) LINK

Conclusion

- An insight from researcher's study is that their approach could be utilised in any shipping industry.
- As long as the necessary data is available and the impactful factors can be identified, the method can be adapted for any shipping route in the world.

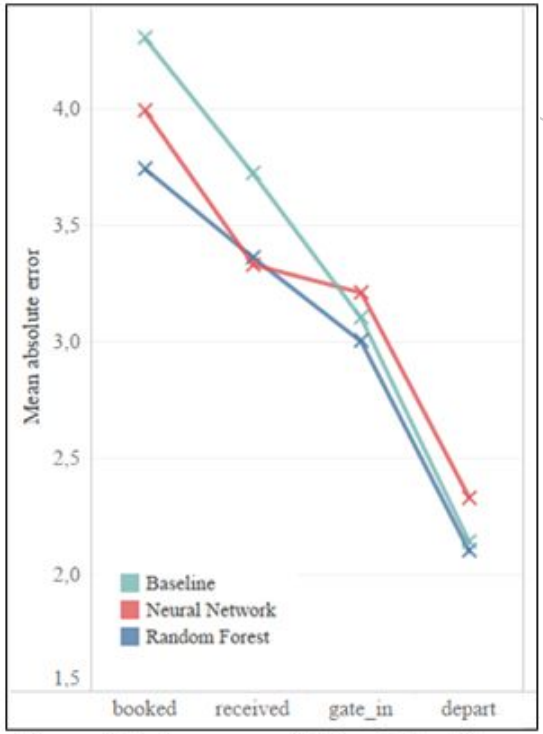


Figure 3: Performance of different algorithms tested for the models

Reliable Shipping Time Prediction

- Dataset available on Kaggle was used. It had 5,114 records containing the following features: (Rejected, Selected)
  - Shipment_id: Removed because unique ID are of no use in ML
  - Send_timestamp: Removed as timestamps are not relevant to ML unless converted to something more meaningful such as treating the data as a time-series
  - Pick_up_point: Removed as it was the same for all orders and therefore made no difference
  - Drop_off_point: Considered. Converted drop off locations from text to numerical format – 0 & 1; as there are only two unique values in the dataset
  - Sourcre_country: Removed as it was the same as pick_up_point.
  - Destination_country: Removed as it is the same as drop_off_point
  - Freight_cost: Considered
  - Gross_weight: Considered
  - Shipment_charges: Considered
  - Shipment_mode: Considered. Converted from text to numerical – 0 & 1, as there are only two unique values
  - Shipping_company: Considered. Three categories of shipping companies were observed. These values were converted into numeric format – 0 & 1; as there are only two unique values
  - Selected: Removed because every candidate company was marked as selected and was hence redundant to this experiment.
- The target feature was shipping_time.
- Following are the most important features:

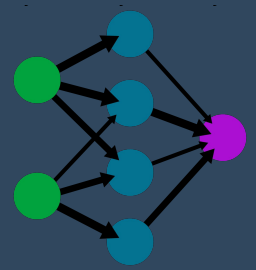
Target (Output)	Problem Type
shipping_time	Regression(~7.62)

Feature (Input)	Score
shipment_mode	0.58
freight_cost	0.22
gross_weight	0.2

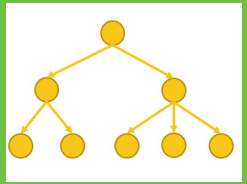
Reliable Shipping Time Prediction

Modelling: Three ML algorithms were used -

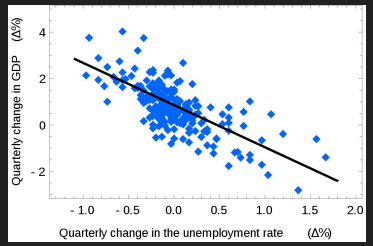
Neural Network



Random Forest



Linear Regression



THANK YOU