Multivariate Freight Transportation Forecasting

Darsh Chopra, Sriya Chundi, Luke Foltz, Saanidhya Jain, Srikar Lanka, Brian Tam

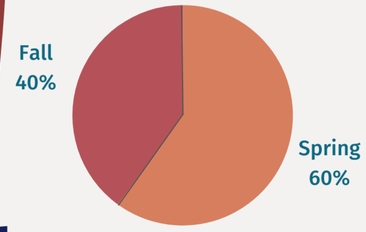
INTRODUCTION

What is Wabash: An American diversified industrial manufacturing firm that is the largest maker of semi-trailers and liquid transportation systems in North America.

Our want: Anticipating demand for their primary products, which includes dry freight vans. The forecasting model is divided into two parts: industry-wide and Wabash-specific. Our Goal: To develop the best possible forecasting models so that Wabash can make informed business decisions.

Fall

Data exploration via groups divided by product types such as Dry vans (Enclosed Semi trailer), Reefer (refrigerated container) and tankers



Spring

Further forecasting via groups divided by models (LSTM for text-based, XGBoost for machine learning, VARMA for variables in ARIMA), transitioning to multivariate focus

Representation of Importance

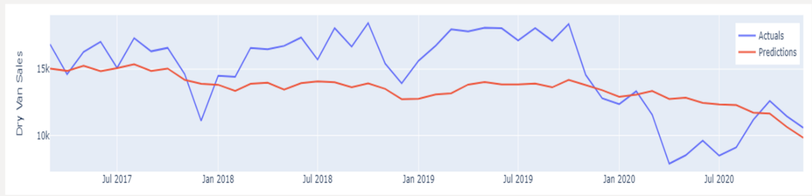
Fall required us to understand data and clean it, while the real work began in Spring to predict the market

Fall Semester:

- Locating and understanding data from ACT & FTR (Private data consultancy firms) and external (databases for potential factors, ex. inflation, gasoline prices, other product sales) sources of data.
- Data cleaning, finding correlations (heat maps), PCA, basic linear regression, producing visualisations
- Initial forecasting via ARIMA and SARIMA (univariate time series analysis via differencing for stationality)

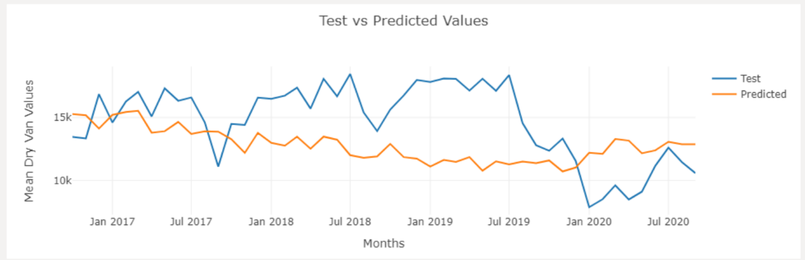
Spring Semester:

- Focusing on dry van data, applying to other product types later



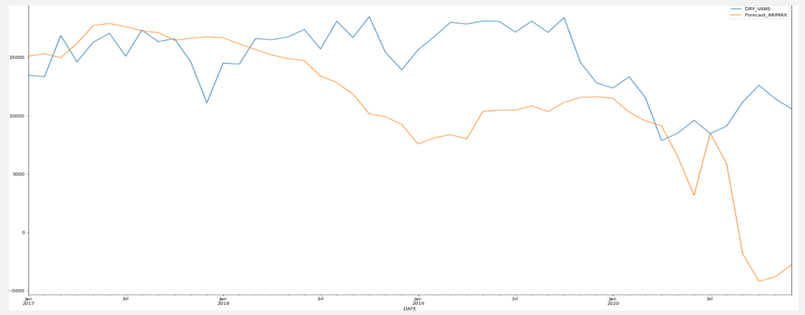
Graph 1

Monthly LSTM Model with a 1 Year Lag



Graph 2

Monthly XGBoost Model



Graph 3

Monthly ARIMAX model

MODEL DESCRIPTION

LSTM – It's a Recurrent Neural network (RNN) with a feedback connection that allows it to process strings of data instead of single points like most RNNs. The results of this model showed even with a full year lag we were able to predict a decline in the 2020 market.

XGBoost – Extreme Gradient Boosting, is the implementation of gradient boosted decision trees designed for speed and performance with great usage in classification and regression problems.

ARIMAX – The multivariate version of ARIMA, a form of dynamic regression model. The model allows for a quicker predictions due to lack of heavy data processing steps required. The prediction were more accurate further out in the lags.

Time	LSTM	ARIMAX	XGBoost	ACT & FTR
1 Month	9.68 %	16.9 %	22.01 %	----
3 Months	13.71 %	19.6 %	14.95 %	----
6 Months	12.97 %	22.9 %	8.92 %	4 %
12 Months	15.39 %	24.8 %	17.73 %	11 %
18 Months	15.30 %	13.9 %	28.41 %	15 %
24 Months	15.32 %	17.2 %	26.54 %	19 %

Inaccuracy Table for Forecasting Models

Conclusions:

- Calculated and used closely related economic factors as weighted inputs
- Implemented a variety of models including LSTM, XGBoost, & ARIMAX
- Tested models with different lags, input variables, and prediction goals
- Achieved 70% + accuracy for all tests
- Helped Wabash gain a better knowledge of their business products and their performance
- Provided reusable code that can be used for predictions in future

Future Goals:

- Forecast for different Wabash products such as reefers, platforms, and tankers
- Forecast for about three years into the future
- Improve and choose forecasting model based on specific time frames like monthly, quarterly, bi-annually, and yearly.
- This is goal is affected by our project findings as seen above. Different forecast models have better accuracy rates.
- Build a UI for business use

ACKNOWLEDGEMENTS

Dr. Mark Ward, Maggie Ann Betz, David Glass, Kali Lacy, Aditya Singh, Anna Brown, Devin Chatterji, John Dill, and Landan Perry, the whole Data Mine team, all members of Wabash.

References

- https://medium.com/analytics-vidhya/combining-time-series-analysis-with-artificial-intelligence-the-future-of-forecasting-52d6f57c8d13
- https://medium.com/@ranjeki/dimensionality-reduction-using-pca-on-multivariate-timeseries-data-b5ec0723bbc4
- https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/
- https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/
- https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/