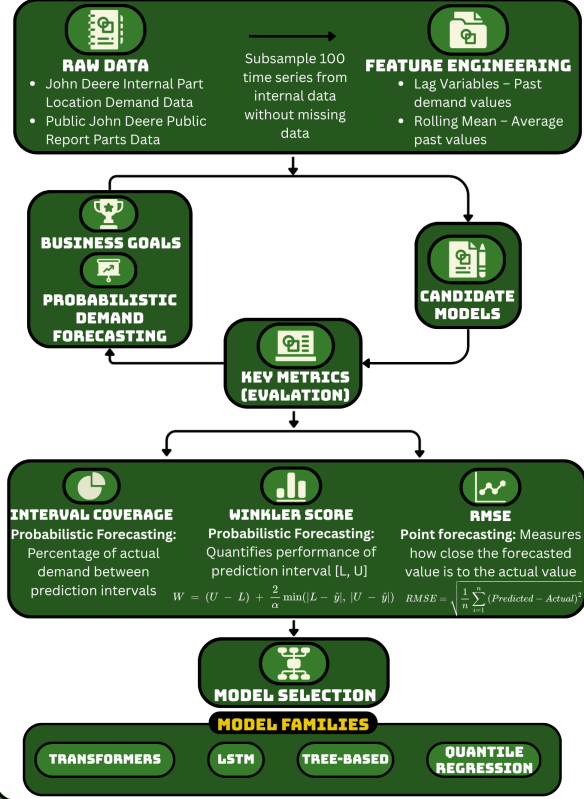


## Introduction

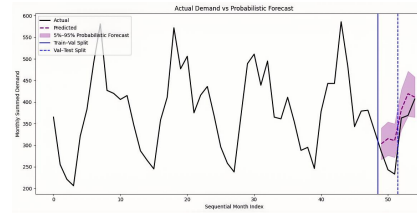
Accurate parts demand forecasting is crucial for companies like John Deere, a global leader in agricultural, construction, and forestry equipment, to balance inventory effectively. With a vast network of dealers and customers relying on timely access to replacement parts, ensuring availability for repairs while minimizing excess stock is essential. Our team developed machine learning models that leverage historical demand data to predict future needs.

## Methodology



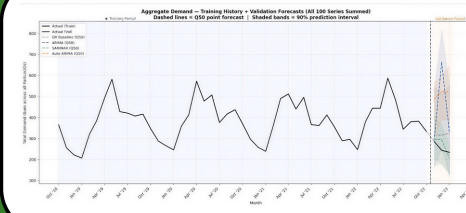
## LSTM

- Built a global Long Short-Term Memory (LSTM) model to forecast monthly demand across part locations.
- External features: Equipment Made, Manufacturers Equipment Inventory, and PPI Manufacturing.



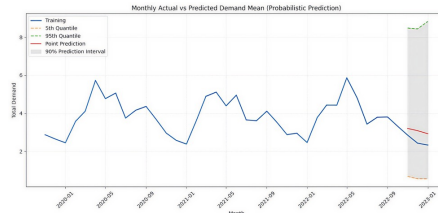
## Quantile Regression

- QuantileRegressor (QR) baseline + ARIMA (Autoregressive Integrated Moving Average)
- SARIMAX: captures seasonal patterns that the original ARIMA model missed and takes in external features.
- Auto-ARIMA: automatically find the best-fitting time series model for each part-location series rather than one-size-fits-all.



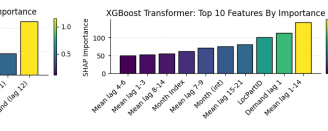
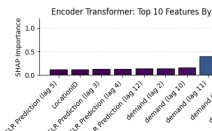
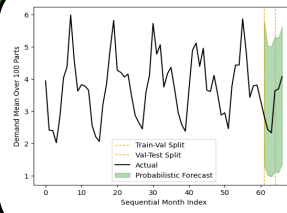
## Trees

We built upon tree-based methods starting with decision trees as a baseline. This was expanded upon with ensemble tree methods like random forest and finally XGBoost. Each iteration of new and more complex tree-based models decreased error metric statistics.



## Transformer

We developed Temporal Fusion Transformer (TFT), Transformer-Based Autoencoder + XGBoost, and Simple Linear Regression (SLR) + Encoder-Only Transformer pipelines.



Metrics Table	Models	RMSE	MWIS	IC
	XGBoost	2.27	8.259	94.00%
Metrics Table: Summarizes the performance of every model	XGBoost TFMR	2.41	11.66	93.00%
	LSTM	3.54	10.88	91.67%
	Encoder TFMR	3.64	6.73	86.33%
	Encoder TFMR SLR	3.92	6.90	88.33%
<b>Key</b>	TFT	4.38	7.29	61.53%
TFMR: Transformer	QR Base	3.04	16.21	64.33%
IC: 90% Interval Coverage (5 – 95)	ARIMA	14.52	27.38	61.67%
MWIS: Mean Winkler Interval Score	SARIMAX	3.09	21.99	50.00%
	Auto-ARIMA	6.40	34.81	56.67%
	Decision Tree	3.09	NA	NA
	Random Forest	2.98	NA	NA
	Seasonal Naive	3.57	NA	NA

## Conclusions

- Both univariate and multivariate models produce successful probabilistic forecasts.
- With a variety of models being developed, our team was able to test features widely. Across the board, features such as demand values from the previous month and from one year earlier are important to model predictions.
- We have also seen improvements by using other models or statistical tools to create engineered features going into a final model.

## Future Directions

- Explore incorporation of different external datasets to improve model performance.
- Further explore data to discover new features.
- Extend to IDs with missing data.
- Implement a full pipeline continuously integrating new data with visual actionable insights.

## Acknowledgements

Special thanks to our John Deere mentors, Yaron Chen and Gonzalo Martinez, for their guidance and insight, along with our TA, Jason Chia and the Data Mine staff for their support throughout the year.