JOHN DEERE PARTS DEMAND

PURDUE UNIVERSITY | The Data Mine | JOHN DEERE

Introduction

John Deere, a leader in agricultural innovation, has **1.6 million parts** in inventory, and are employing various algorithms to facilitate the distribution process, including **statistical models**, **machine learning**, and **deep learning**. This project aims to optimize the algorithms on which these processes are based, in an attempt to further streamline the process and make predictions even more accurate.

Supply chain analytics has transformed the industry by enabling data-driven decisions, enhancing forecast accuracy, optimizing inventory levels, and improving overall responsiveness to market demands.

Project Goal

*Our aim is to create **12-month forecasts** for part demand across locations, helping John Deere make informed **stocking decisions**, Using 3 models- **Deep Learning, Sarima Probabilistic Modelling & Quantile Regression**.*

Business Opportunity

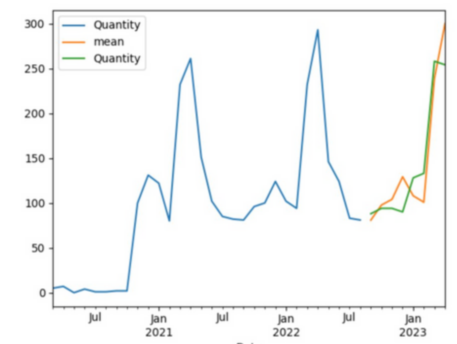
As the industry continues to evolve with **increasing complexity, volatility,** and **customer expectations**, John Deere is proactively taking steps to incorporate artificial intelligence into its **demand forecasting** and **supply chain strategies**. This transition reflects a broader recognition of traditional methods.

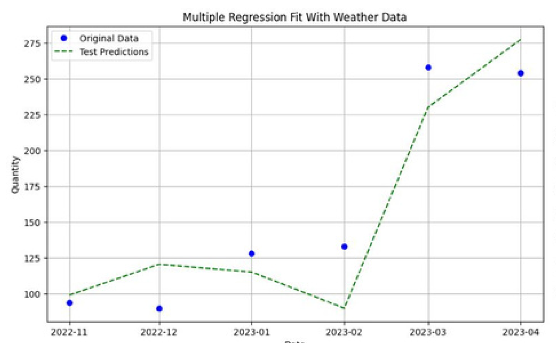
Traditional methods while reliable in the past, are increasingly limited in **handling the complexity and speed required in today's markets**. AI-powered models represent a strategic shift toward more **agile**, **accurate**, and **scalable solutions** that align with John Deere's commitment to **innovation** and **customer-centric excellence**.

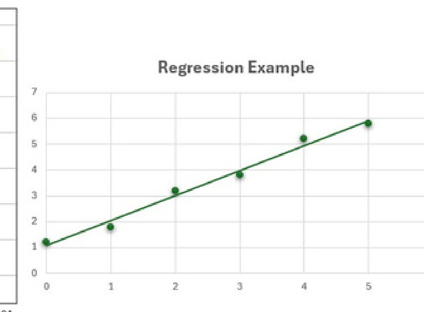
Point Forecasting Models

SARIMA/SARIMAX

- **SARIMA** (Seasonal AutoRegressive Integrated Moving Average) model extends ARIMA by incorporating seasonal components to capture recurring patterns in time series data
- **SARIMAX** (SARIMA with Exogenous Variables) further includes external factors, allowing for more accurate point forecasting when relevant external influences impact the series





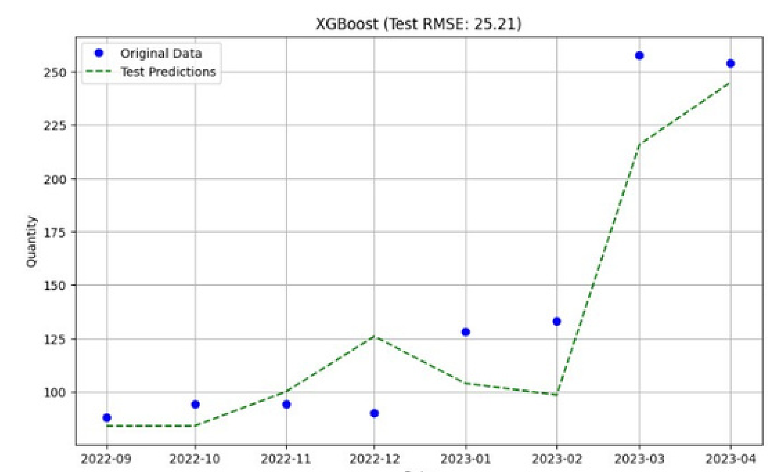


REGRESSION

- **Regression** works by minimizing the least squared error between points to try and fit

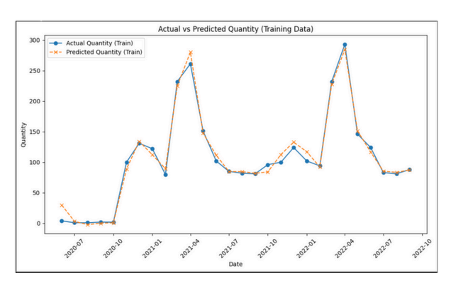
XGBOOST

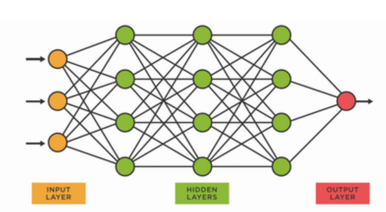
- **XGBoost** works by using a random forest and minimizing the error in each tree then it will iterate again **creating a new forest**

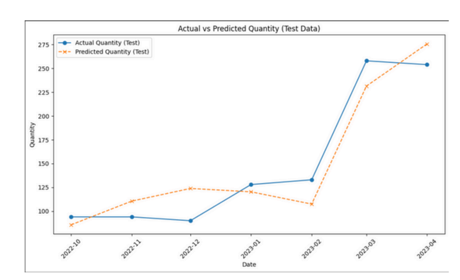


NEURAL NETWORK MODEL

- **A Neural Network** is a machine learning model that processes data through layers of interconnected nodes, adjusting weights to minimize error.



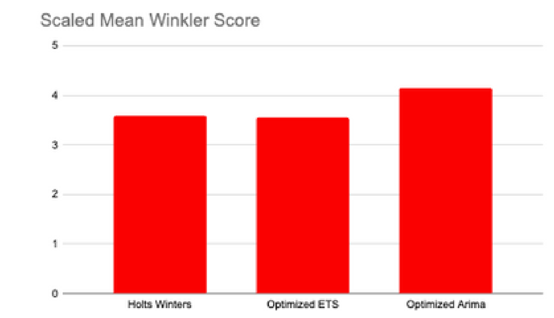


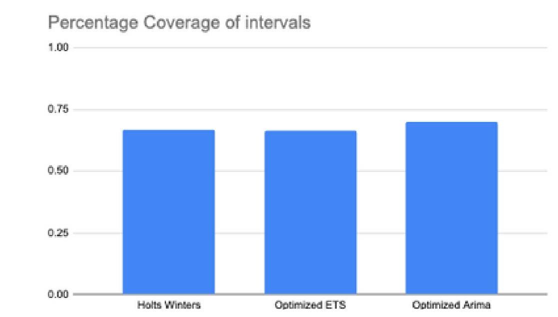


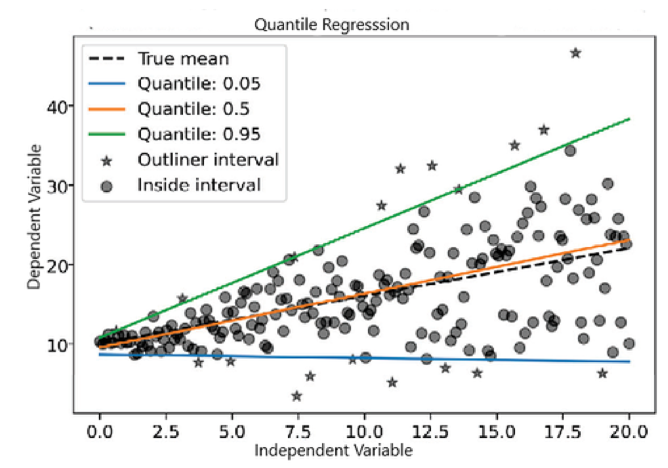
Probabilistic Forecasting Models

SARIMA

- The statistical forecasting models **SARIMA, ETS** are commonly used in industry as an **effective and relatively lower cost means of forecasting**.
- They can be applied to both point, and probabilistic forecasting directions, and allow for an effective forecast and benchmark against other models.





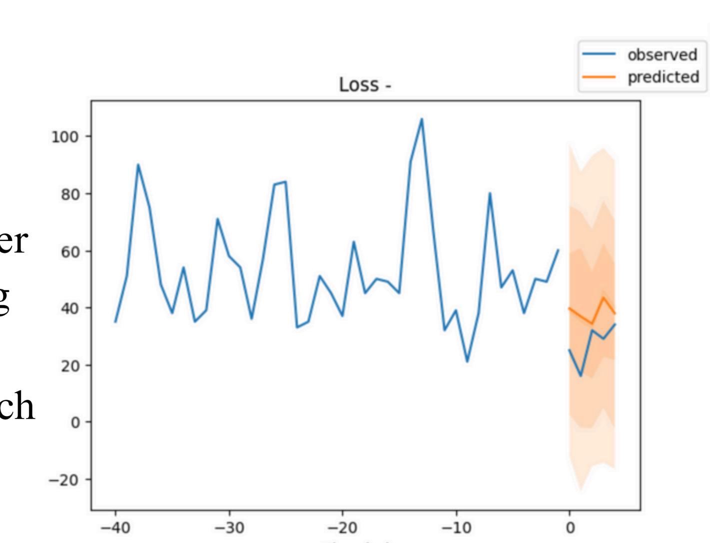


QUANTILE REGRESSION

- This model works similarly to a **Linear Regression Model**.
- The main difference is that it uses **Pinball Loss** as the error it's trying to minimize for fitting the line for a **specific quantile**.
- This model has worked fairly well on the data it has been given.

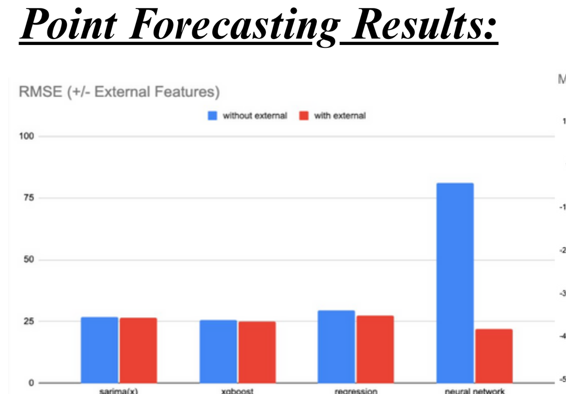
DEEPAR

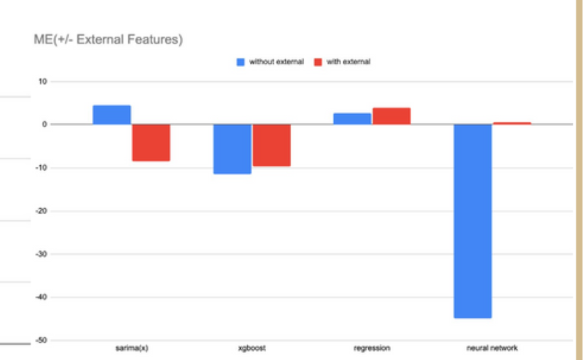
- **DeepAR** is an **LSTM** based **RNN**. Meaning it **incorporates a long-short term memory mechanism.**
- First, after a model sends the input through a layer that output of that layer is sent through the layer again allowing the **model to remember past data.**
- Next the model has a memory cell which can store data for extended periods of time. The model is then **trained on multiple time series data.**



Model Results

Point Forecasting Results:

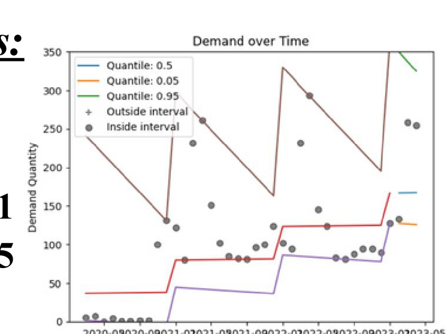




The Neural Network model had the best performance when using external features with an **RMSE score of 21.9795** and an **ME score of 0.6184**. However, **it performed the worst without external features**, highlighting the major impact of external features on model accuracy.

Probabilistic Forecasting Results:

- The Winkler score of quantile regression was **211**
- The **95th** percentile Pinball Loss is **61**
- The **50th** percentile Pinball Loss is **35**
- The **5th** percentile Pinball Loss is **44**



Point Error Metrics

Root Mean Squared Error (aka RMSE) was used in our models to compare the average accuracy of the model. This method also was resistant to outliers since it took the square root

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

Mean Error (aka ME) was used to calculate the average bias in our models. If the ME was positive, we knew that the model was overpredicting, while if the value was negative then the model must be underpredicting.

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)$$

Probabilistic Error Metrics

Pinball Loss, or **Quantile Loss**, is used in quantile regression to **predict specific percentiles of a target variable** (e.g., median, 90th percentile). It's useful in applications like risk management, weather forecasting, and demand prediction.

$$\text{pinball}(y, \hat{y}) = \frac{1}{n_{\text{samples}}}\sum_{i=0}^{n_{\text{samples}}-1} \alpha \max(y_i - \hat{y}_i, 0) + (1-\alpha)\max(\hat{y}_i - y_i, 0)$$

y= actual value *ŷ= predictive value* *α = alpha (target quantile)*

Winkler Score evaluates interval forecasts by considering both the width of the prediction interval and whether the actual value falls within it. **It penalizes intervals that are too wide or miss the true value.**

$$W_{\alpha,t} = \begin{cases} (u_{\alpha,t} - \ell_{\alpha,t}) + \frac{2}{\alpha}(\ell_{\alpha,t} - y_t) & \text{if } y_t < \ell_{\alpha,t} \\ (u_{\alpha,t} - \ell_{\alpha,t}) & \text{if } \ell_{\alpha,t} \le y_t \le u_{\alpha,t} \\ (u_{\alpha,t} - \ell_{\alpha,t}) + \frac{2}{\alpha}(y_t - u_{\alpha,t}) & \text{if } y_t > u_{\alpha,t}. \end{cases}$$

Point Forecasting Conclusion

- After using benchmark models, Exponential Smoothing and Seasonal Naive methods, we compared them with more **advanced/robust methods** such as XGBoost, Regression, SARIMA, and Neural Networks.
- After we reached **strong performances** with our advanced models, especially the **Dense-Layer Neural Network**, we learned that external features such as weather data and federal funds rate were crucial for accurate predictions.
- We began to move toward accurately accounting for how demand may vary from our predictions by adapting our **Point forecasting models** into **Probabilistic models.**

Probabilistic Forecasting Conclusion

- We compared each model through error metrics like **Pinball Loss** and **Winkler Score**.
- The small quantity of **outliers** from the machine learning model indicates that it was **effective in its representation of the data**.
- **Error regions** in the deep learning graph represent the range of possible values, showing **high potential variability** (from -67% to +300%)
- General **strong performance** from traditional statistical models, with a need for **further research and development** for probabilistic machine learning models.

Future Goals

- **Enhance Model Accuracy**: Continue refining the forecasting models by experimenting with advanced deep learning architectures and feature engineering techniques.
- **Integrate External Factors**: Incorporate additional variables like economic indicators, backorder rates and geopolitical events to improve model robustness.

Acknowledgement

We would like to acknowledge **The Data Mine** and our corporate partners from **John Deere**, **Yarong Chen** and **Gonzalo Martinez**, also our TAs, **Harsh Srivastava** and **Sankaran Iyer** for their helpful insights during our research.