John Deere Parts Demand Forecasting

Student Team: Aditya Menon, Alexandra Dorula, Amogh Upadhyaya, Aryan Bhowmik, Jay Sunkara, Saahil Mathur, Saketh Bireddy, Snigdha Mishra, Sriharshita Varanasi

PURDUE UNIVERSITY
The Data Mine

JOHN DEERE

Our Roadmap for Successful Demand Forecasting

1 → Explore Trends: EDA & Seasonality
2 → Single Part-Location Analysis
3 → Model Selection: Baseline & Advanced
4 → Model Testing on 1.6M Part-Locations
5 → Evaluation & Iteration by Metrics
6

Project Premise: Demand Forecasting Essentials

Demand forecasting is the process of predicting future customer demand for a product or service based on historical data and market trends. In the context of the John Deere Dealer Network, it involves **anticipating the demand for parts across various locations (part-location combination)** to ensure optimal inventory management and customer satisfaction.

- Accurate Inventory Control
- Optimization of Production Schedule
- On-Time Delivery
- Increased Dealer & Customer Satisfaction

Project Goal: Our aim is to create **12-month forecasts for part demand across locations**, helping John Deere make informed stocking decisions, improving parts availability and customer service.

Benefits of Demand Forecasting

Consumer, Dealers, and John Deere form the essential triad driving inventory part transactions. Our mission is to **optimize inventory levels** to meet consumer and dealer demand?

Figure 1: Supply Chain Network

Seasonality & Trend Analysis

*The following analysis is based on ONE Part-Location Combination

Seasonality of 12 observed in single part-location combination. Fig. 2 shows a correlation present demands to demands 12 time periods ago.

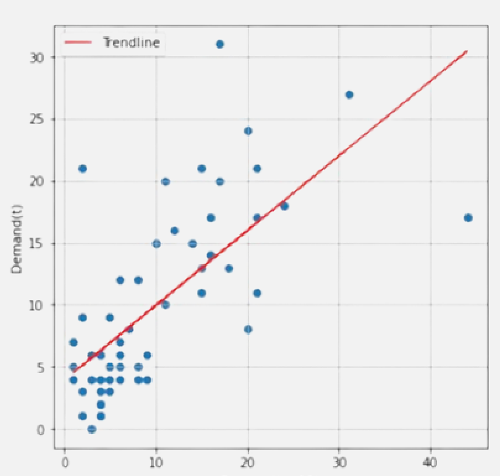


Figure 2: Seasonality Plot

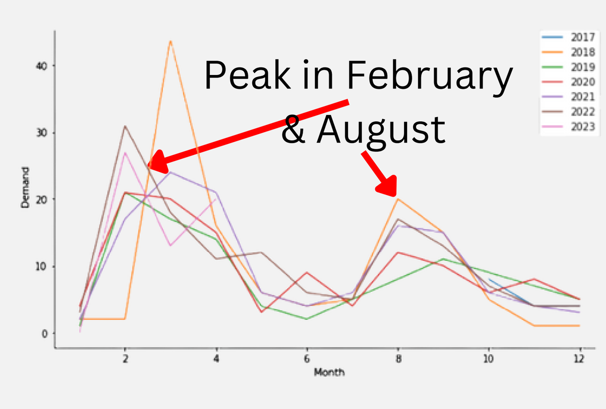
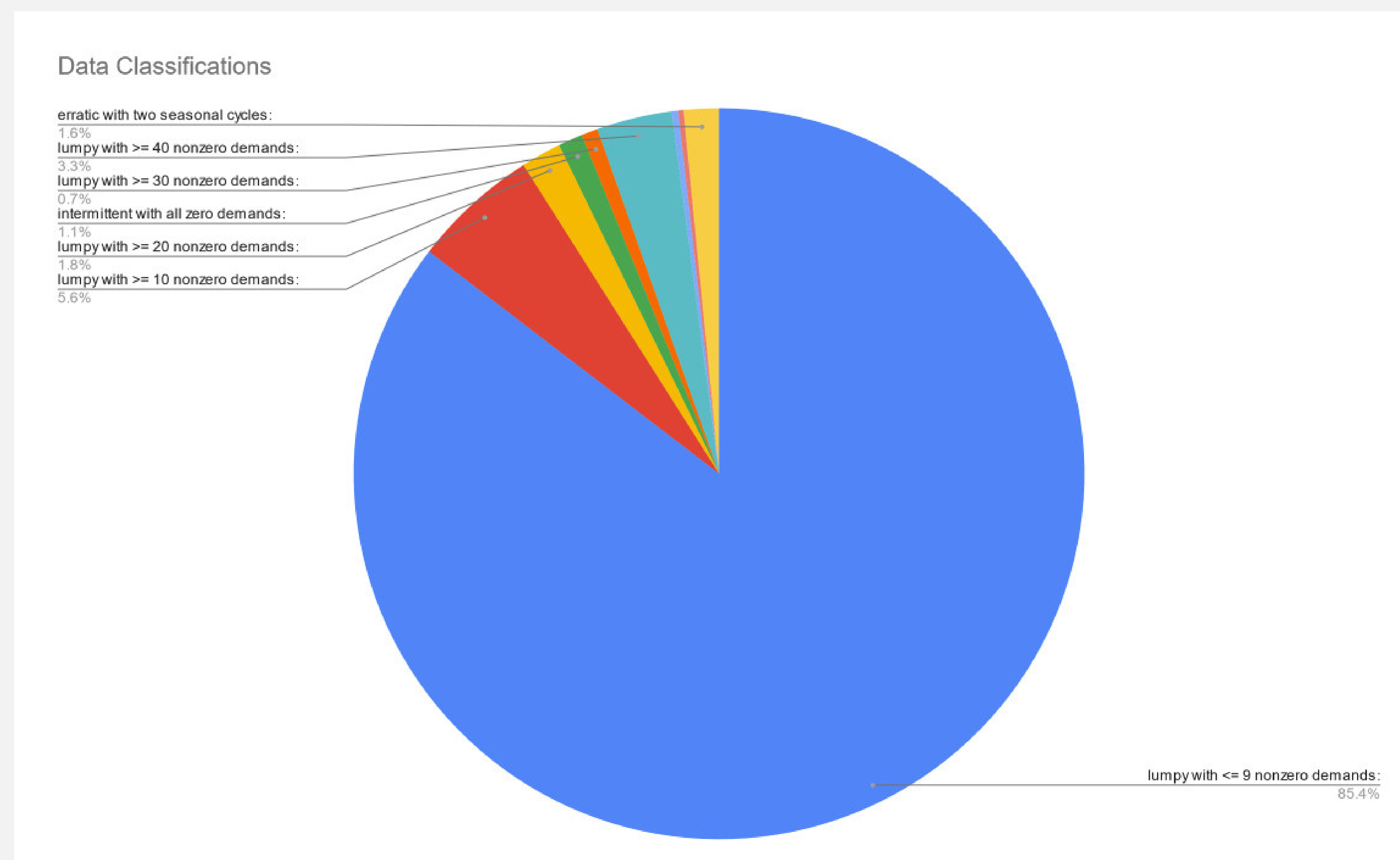


Figure 3: Demands Per Month by Year

Time Series Clustering & Classification

Why is it important to **cluster the data**?

Given **1.6 million unique part-location combinations**, iterating 1.6 million times to determine the best model is not a feasible approach. Time series clustering categorizes data into patterns like Lumpy, Smooth, Erratic, and Intermittent. This classification (bucketing technique) helps select appropriate forecasting models, allocate resources effectively, and manage risks associated with different patterns.



Bucket Part-Location Combinations → Test Appropriate Models for each Bucket → Evaluate Metrics & Compare Models → Iterate & Bucket Further

Figure 4: Bucket Labels and Classification of Part-Location Combinations

The classifications represent **bucketing similar data** by **variation in demand interval and variations in demand quantity** such that different models can be applied for different "buckets".

Models & Evaluations

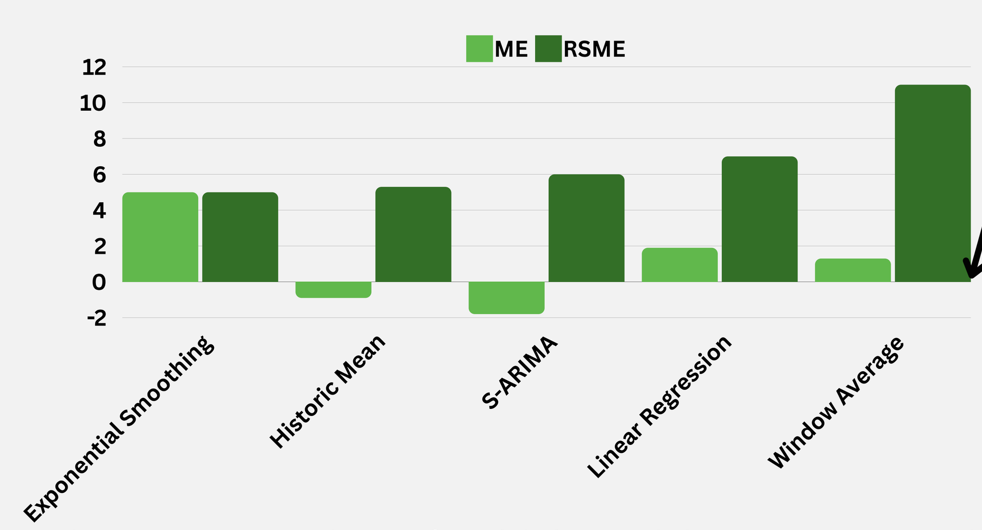


Figure 6: Comparison of Multiple Models on 1 Part-Location

Identify best models based on accuracy and bias metrics

Compare Advanced Model to Benchmark to determine accuracy

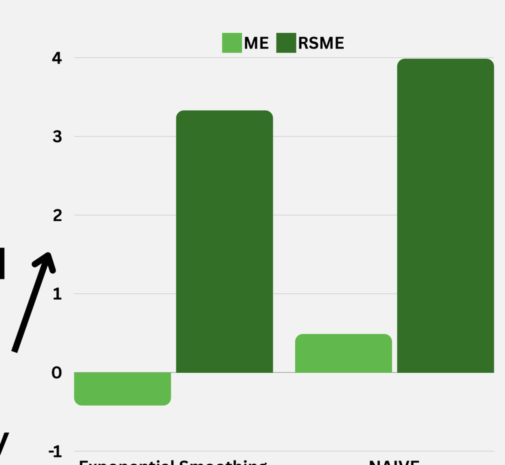


Figure 7: Advanced vs. Benchmark Metrics

Forecasting Result & Metrics

- Forecasts generated on erratic classified data using the Exponential Smoothing Model
- Low but negative Mean Error indicates under forecasting slightly on average
- Low Root Mean Squared Error indicates relatively high accuracy

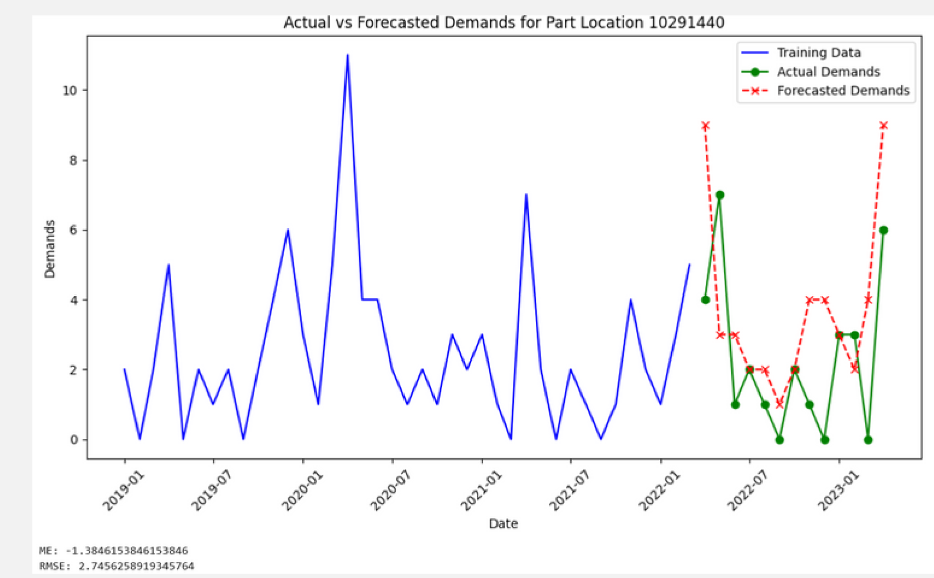


Figure 8: The difference between forecasted and actual demands is used to calculate error metrics.

Future Goals

Explore More Models → Identify Relevant Error Metrics → Perform Further Clustering, Repeat

- Implement ML models such as LSTM, XGBoost, and Random Forest to discover if they have better prediction accuracy.
- Add additional features such weather and geographic data, along with multiple seasonalities to improve model performance.

Conclusion

1) **Lumpy data (irregular and sporadic demand)** constituted to around **1.5M data points**. Our current methods struggle with this, resulting in forecast inaccuracies.

2) Portions of data classified as Intermittent and Lumpy had a **trivial forecast of 0 demand** due to data consisting of 0 demand throughout. Low number of **non-zero data points** presents a challenge in forecasting.

3) Leveraging **Exponential Smoothing** and **Seasonal Naive** methods, we aim to forecast Smooth, Erratic, or Intermittent data more effectively, enhancing inventory management and operational efficiency.

Acknowledgements

We would like to thank our TAs Sankaran Iyer and Wenyin Jiang, our John Deere mentors Michael Kaminski and Yarong Chen, and the Data Mine staff for guiding us throughout this project.