**Group 2 : Summary**

**Demand Forecasting for Commercial Truck Assembly: A Data-Driven Approach**

Chris Cooke, Brian Enns, Martin Guzman, Kuljinder Kaur, Sigao Li

**Objectives:**

Our project was designed to create a demand forecasting model for a top-tier global commercial truck and bus manufacturer that employs flexible manufacturing systems (FMS). FMS inherently blends Job-Shop and Mass Production methodologies, often leading to inventory inconsistencies and efficiency concerns. By analyzing historical data and tapping into previously overlooked variables, our chief aim was to anticipate future demand for crucial parts requirements for the truck & bus assembly process. Ultimately, we sought to offer data-driven recommendations for forward-thinking inventory management, streamlining costs and boosting operational efficiency.

**Data Preparation:**

Datasets for this project were sourced from the company's Hadoop data lake, replicating their ERP system tables. Data was accessed through SQL queries and exported as .csv. Initial data preparation involved cleaning and standardizing date formats. The three primary datasets, namely 'tbl\_znmps3400', 'ldos\_archive', and 'tbl\_nav830\_hist', provided data on forecast percentages, item attributes, and purchase schedules respectively. However, due to challenges in mapping items to individual truck models, 'tbl\_znmps3400' was omitted in the final analysis.

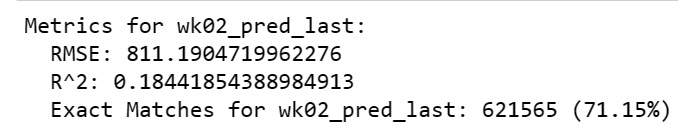
**Model Design:**

Initially, we utilized unsupervised machine learning, specifically clustering using k-means after dimensionality reduction via PCA, to better understand the data's structure. Following this, our supervised approach began with a naïve model where we shifted past data to predict current outcomes. Further models deployed include BayesianRidge, RandomForest, Ridge, Decision Tree, and many more. Complex feature engineering and data transformations were fundamental to our modeling phase, as well as customized Scalers and Cross-Validation and implementing parallel processing.

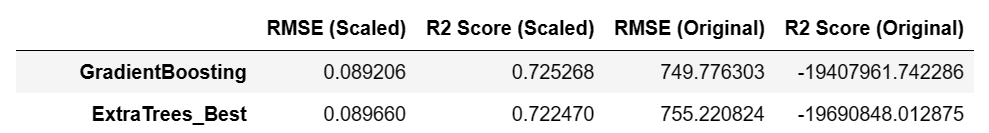
**Model Evaluation:**

Based on the results for future-predicted ‘wk02’ values, we have two approaches related to normalized data vs non-normalized data. The GradientBoostRegressor for “**Model: Predict shifted ‘wk02’ (Future, Row Normalized Data)**” may hold information that can be useful for further analysis of predictive factors for ‘wk02’, but would not be useful for direct prediction (although it did have an RSME marginally better than the **Naïve Model**). Preliminary results from “**Model: Predict shifted ‘wk02’ (Future, Non-Normalized Data)**” suggest that kNN with default values may be best for direct prediction, with score results trending better than the **Naïve Model.**

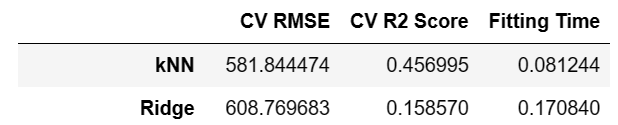
**Model Results: Naive Model**

****

**Model Results: Predict shifted ‘wk02’ (Future, Row Normalized Data)**

****

**Model Results: Model: Predict shifted ‘wk02’ (Future, Non-Normalized Data)**

****

**Conclusions:**

Our analysis provided a robust comparison of various regression models tailored to this dataset. The insights drawn highlighted the importance of considering both performance metrics and computational efficiency. The project underscores the potential of data-driven approaches in demand forecasting for optimizing operations in complex industries like commercial truck assembly.

**Source of Data:**

Datasets were derived from digital replicas of the company's ERP system stored on the company's Hadoop data lake. Specific links or direct data access methods are proprietary and therefore not shared here.