Machine Learning;
Worth the Price of Admission?

Burak Gundogdu, Jeffrey Maloney

May 2019

Project Sponsor

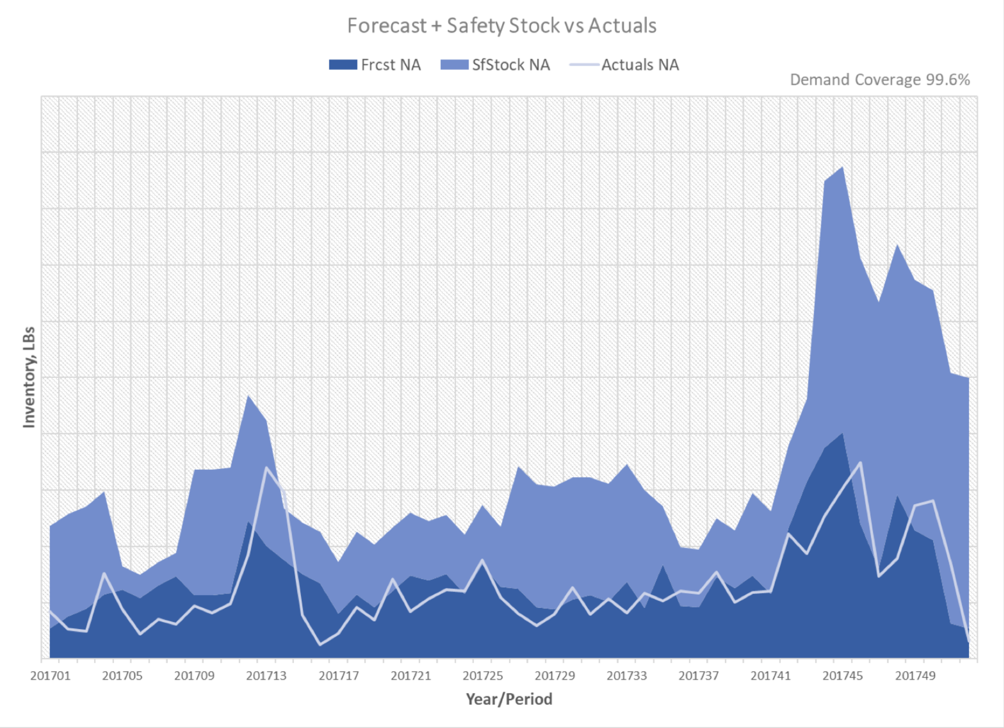
Agenda

- Motivation and research question
- Current state of research
- Our approach
  - Cube search
  - Data flow
  - Machine learning models
- Results
  - Comparison of models
  - Avoiding overfitting
  - Final model performance
- Financial impact

Motivation and Research Question

- A build to stock company, relying on forecasts generated through an S&OP process to meet customer demand
- Highly seasonal demand, peaks occurring during major US holidays
- Holt-Winters is the current forecast methodology
- High interest in applying machine learning but needs to
  1. Identify which model(s) & data to use
  2. Determine whether the potential improvement justifies the costs

Is the improvement (if any) in demand forecast accuracy from a machine learning process over traditional statistical methods significant enough to justify the increased costs?



Current State of Research

The Current State of Literature

Machine Learning in Time Series Forecasting

- ML has significant potential to improve costs over traditional analytical techniques (Chui et al., 2018)
- Some argue that they will not consistently improve over the forecasts using traditional techniques (e.g. Makridakis, Spiliotis and Assimakopoulos, 2018)

Machine Learning Based Demand Forecasting

- A number of studies focusing on a specific ML model and comparing to traditional methods (E.g. Hribar et. al, 2018 & Saloux and Candanedo, 2018)
- No consistent answer to which model performs best overall

FMCG Demand Forecasting with Machine Learning

- Still in its early stages and there is no one-size-fits-all approach
- The selection of the machine learning model, and the hyper-parameters that guide it, play a significant role in the final forecast accuracy

How Our Study Differs?

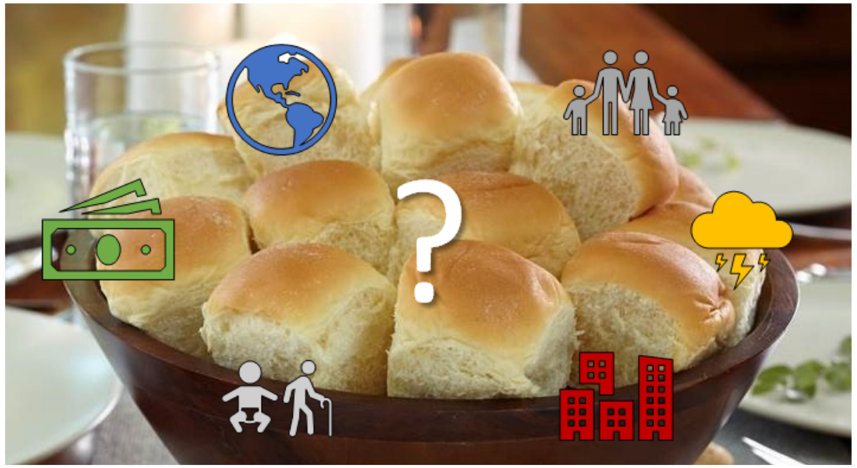
- Most demand forecasting studies focus on single ML model, with few investigating up to 3
- Compared 5 models, ran ~400 iterations in total

- Current studies use standard error metrics
- Converted the error of the model into the custom loss value that the company actually uses in operations

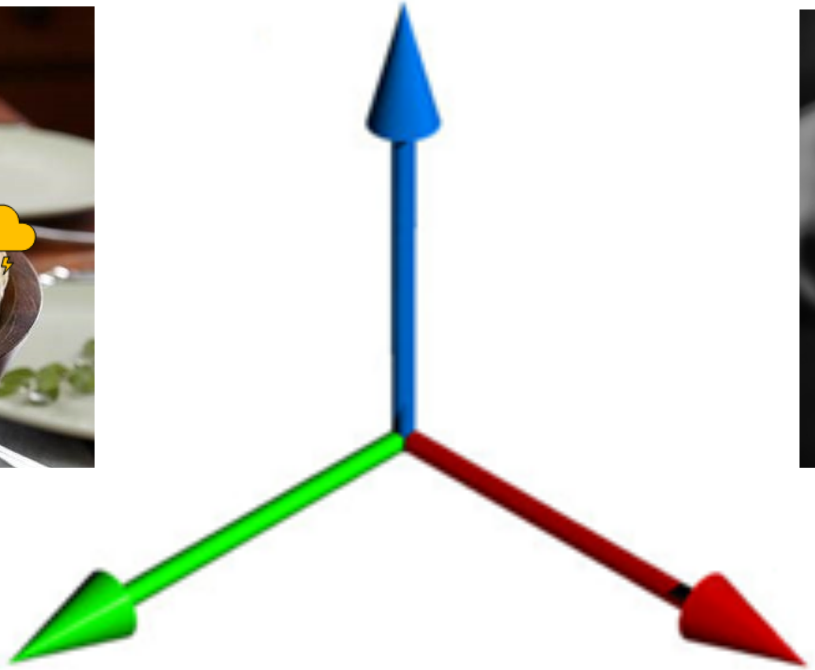
- We did not come across a study that implements a cost-benefit analysis of using ML to improve demand forecasts
- Calculated the expected savings from inventory and compared to the cost of deploying an ML based demand forecast

Our Approach: Cube search

What features to include?



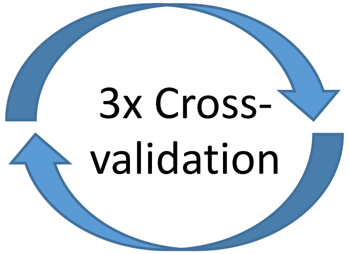
Feature set



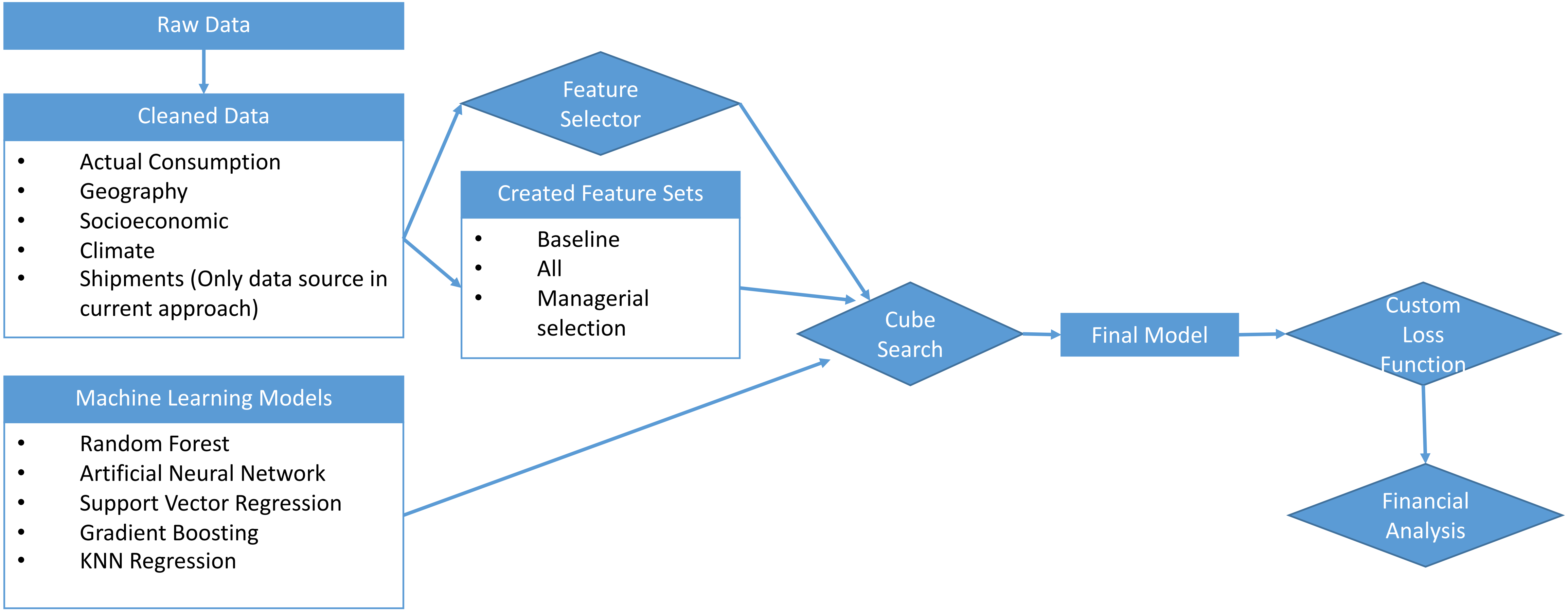
Tuning the model

Hyper-parameter 2

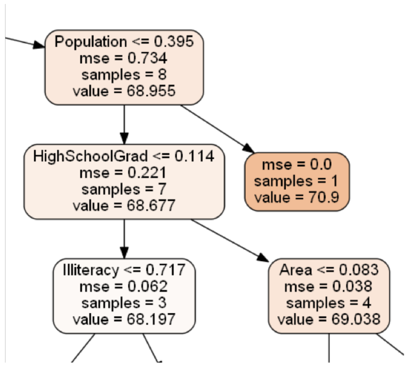
Hyper-parameter 1

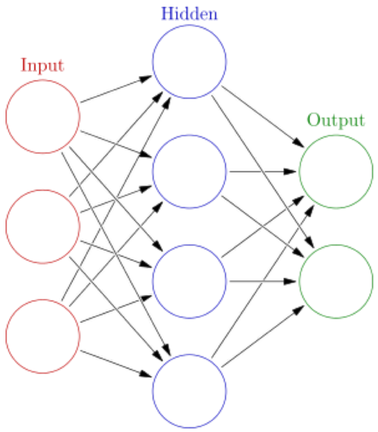


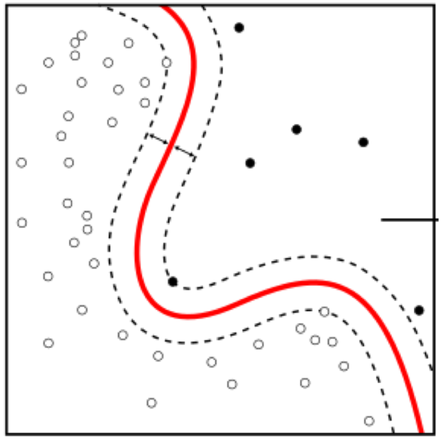
Our Approach: Data Flow

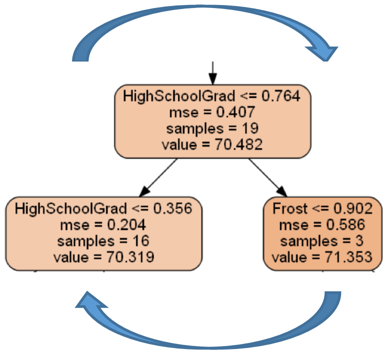


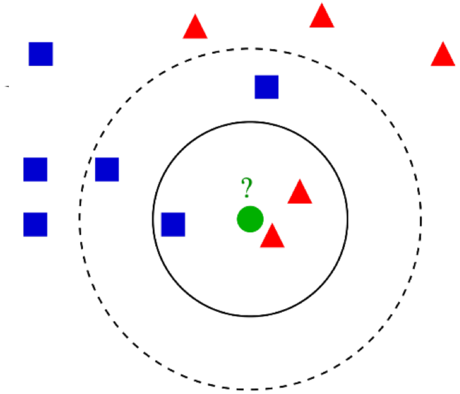
Our Approach: Machine Learning Models











Random Forest

- Ensemble method of decision trees
- Decreases the variance by combining trees

Artificial Neural Network

- Consists of nodes that are used to calculate weights of the features in the model
- Nodes are organized in layers

Support Vector Regression

- Employs a decision boundary called a hyperplane
- Approach: Maximize the minimum margin

Gradient Boosting

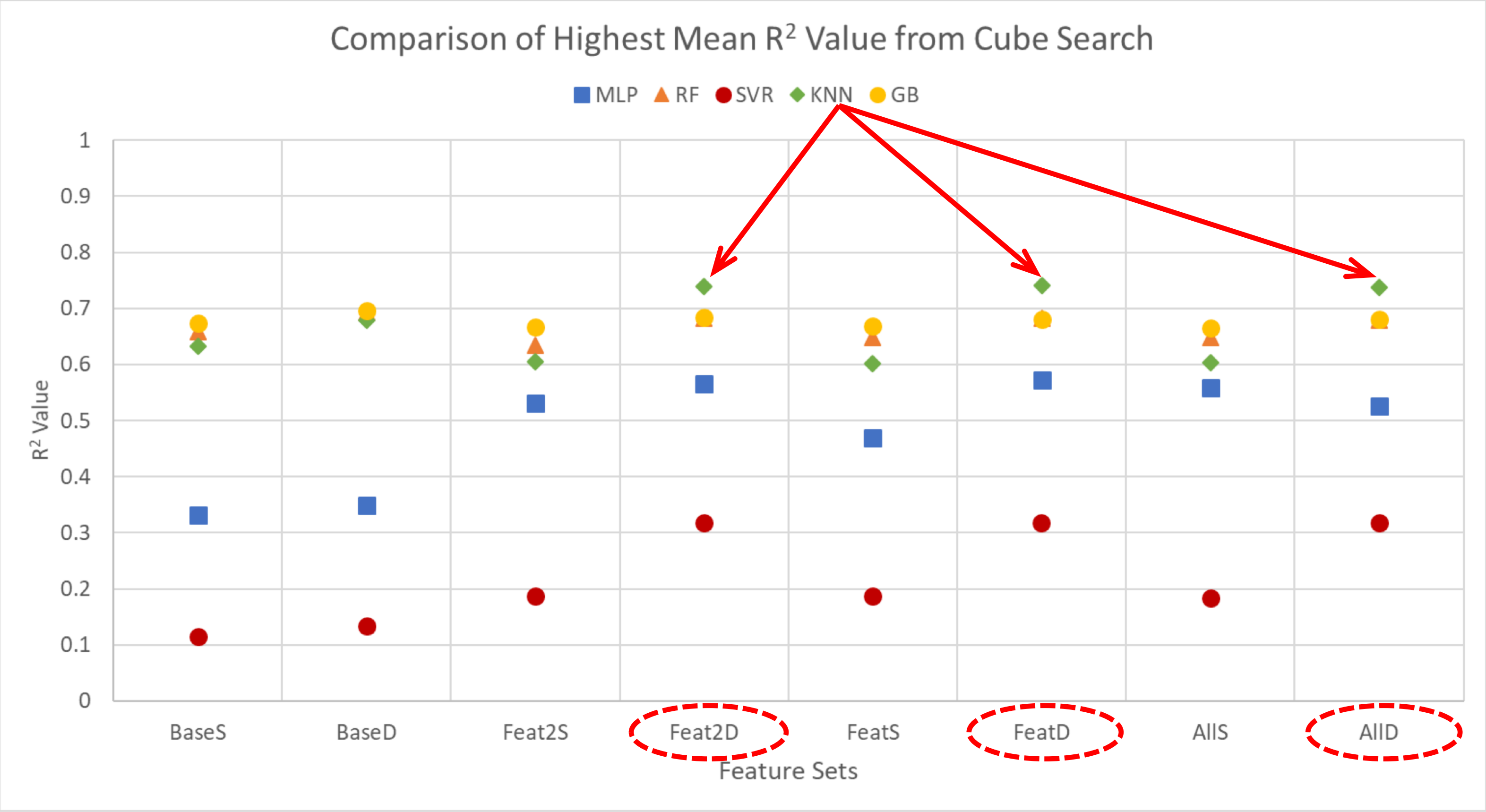
- An ensemble method where predictors are added sequentially
- In each stage, the new predictor is fit into the residual errors

K-nearest Neighbor Regression

- Value of the target: Average of the k closest observations
- Optimal value of k determined by running multiple models

Supply Chain MANAGEMENT

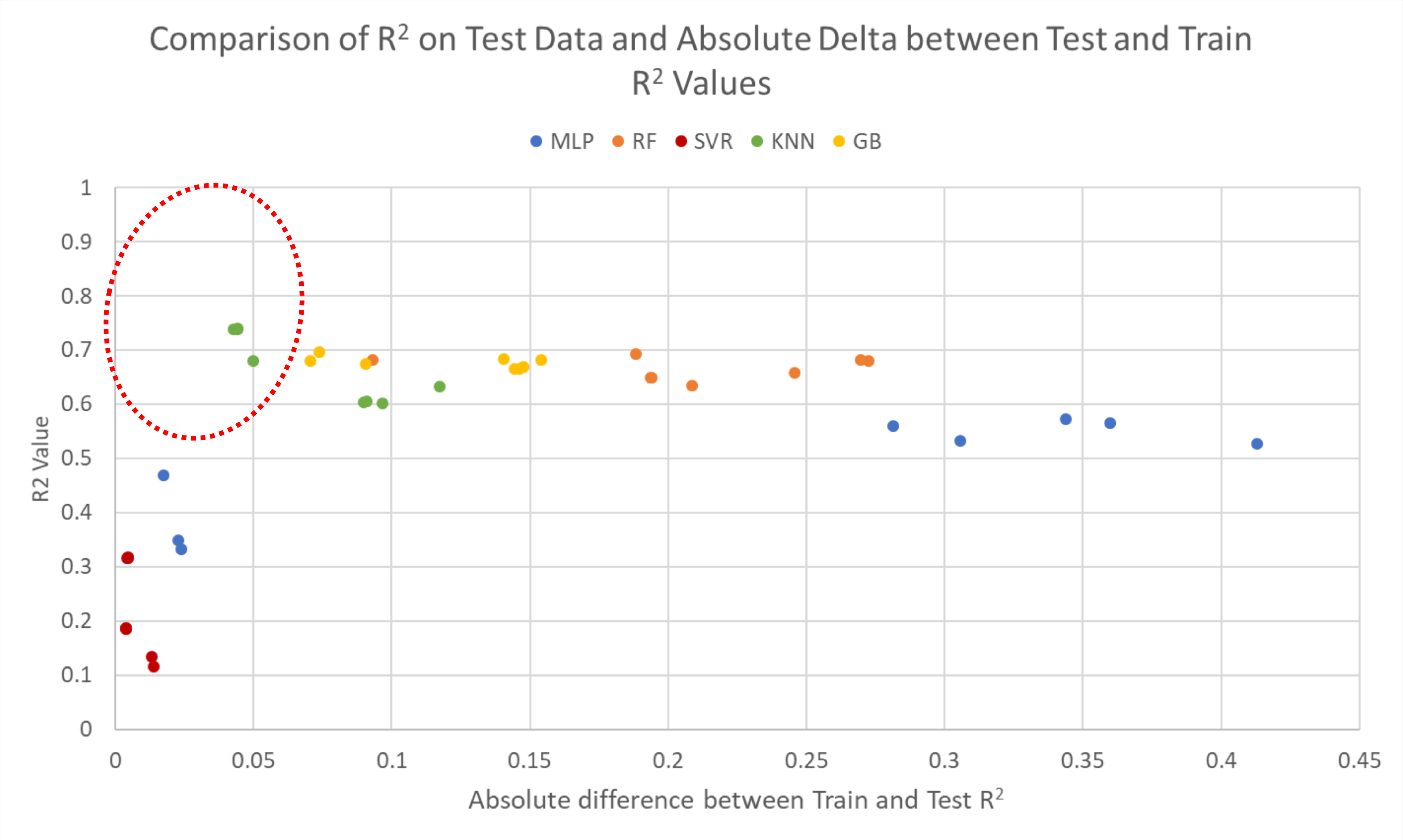
Results: Comparison of Models



- **KNN consistently had higher R^2** when additional features were added and demand was de-seasonalized
- In general the **highest mean R^2** came from **de-seasonalized feature sets**
- The **addition of features tended to improve mean R^2** above baseline
- **SVR's performance was lowest**, and it required the longest run time

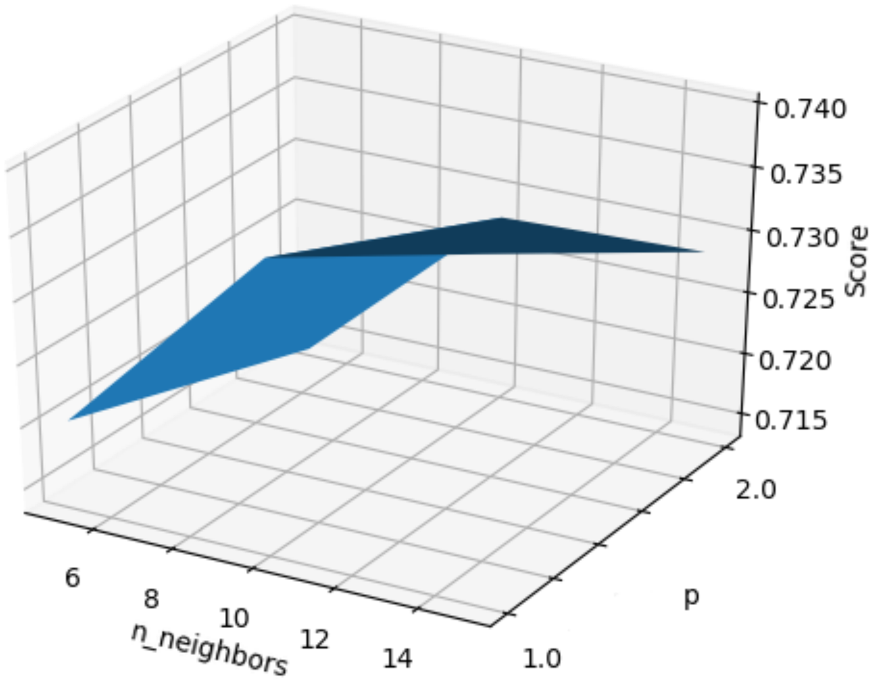
Supply Chain MANAGEMENT

Results: Avoiding Overfitting



- **Compared the absolute variance between** mean R^2 values of the **Training and Test data sets** from the best models
- Comparison helps **resolve differences between models** with similar mean R^2 scores
- **KNN achieved the highest mean R^2, while maintaining a low Train-Test variance** when run on de-seasonalized Feature Select 1 set

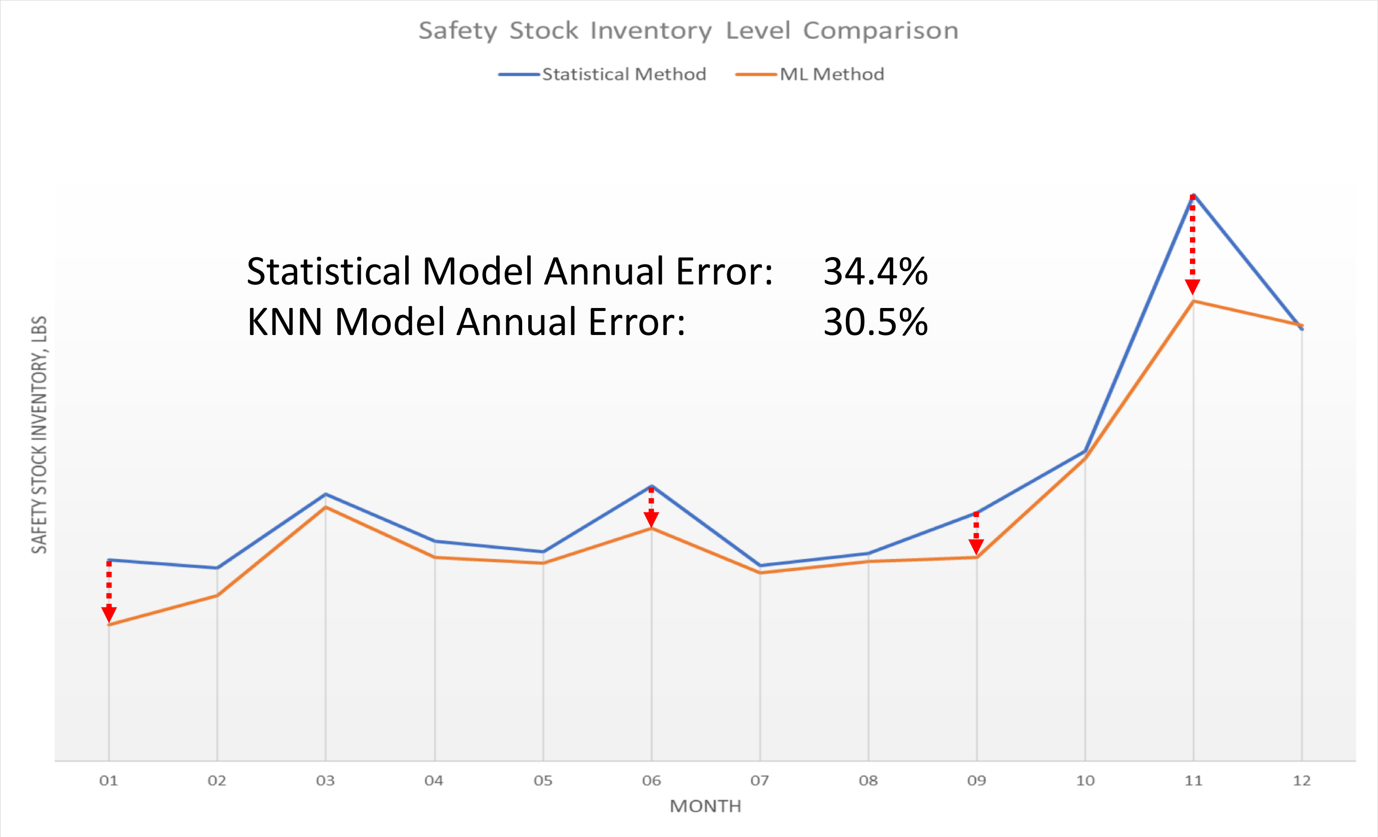
Results: Final Model Performance



Graphical representation of the mean R^2 for each set of hyper parameters evaluated; number of neighbors and distance calculation for KNN on the de-seasonalized Feature Select 1 set

- The hyper-parameter that achieved the best results for KNN were **number of neighbors: 15**, and **distance: Manhattan**
- The best KNN model achieved a **mean R^2 of 0.74 during the cube search**
- The selected model achieved an **R^2 of 0.67 on unseen 2018 test data set,** similar to scores during cross validation
- Final validation of the **selected model resulted in annual forecast error (WAPE) of 30.5%**

Financial Impact



- Selected KNN model **achieved a 3.9% lower forecast error than the current model**
- The ML forecast error driven equation **represents a decrease in the level of safety stock of 10%**
- Lower forecast error **results in a projected reduction in value of safety stock of >$900k**
- **The savings exceeds the annual incremental costs of $500k** (data, software, personnel support), **justifying the use of the advanced methodology**

Definitely worth the price of admission!

Further Investigation

- Expand model evaluation to new model types
- Increase the range and granularity of hyper-parameters
- Evaluate additional features for significance
- Perform comprehensive financial analysis on impact of improved forecast accuracy
- Adapt process to gain additional insights
    - Forecast consumption demand
    - Forecast at the customer account level
    - Expand upon feature selection process to understand key business drivers and customer composition

Questions?

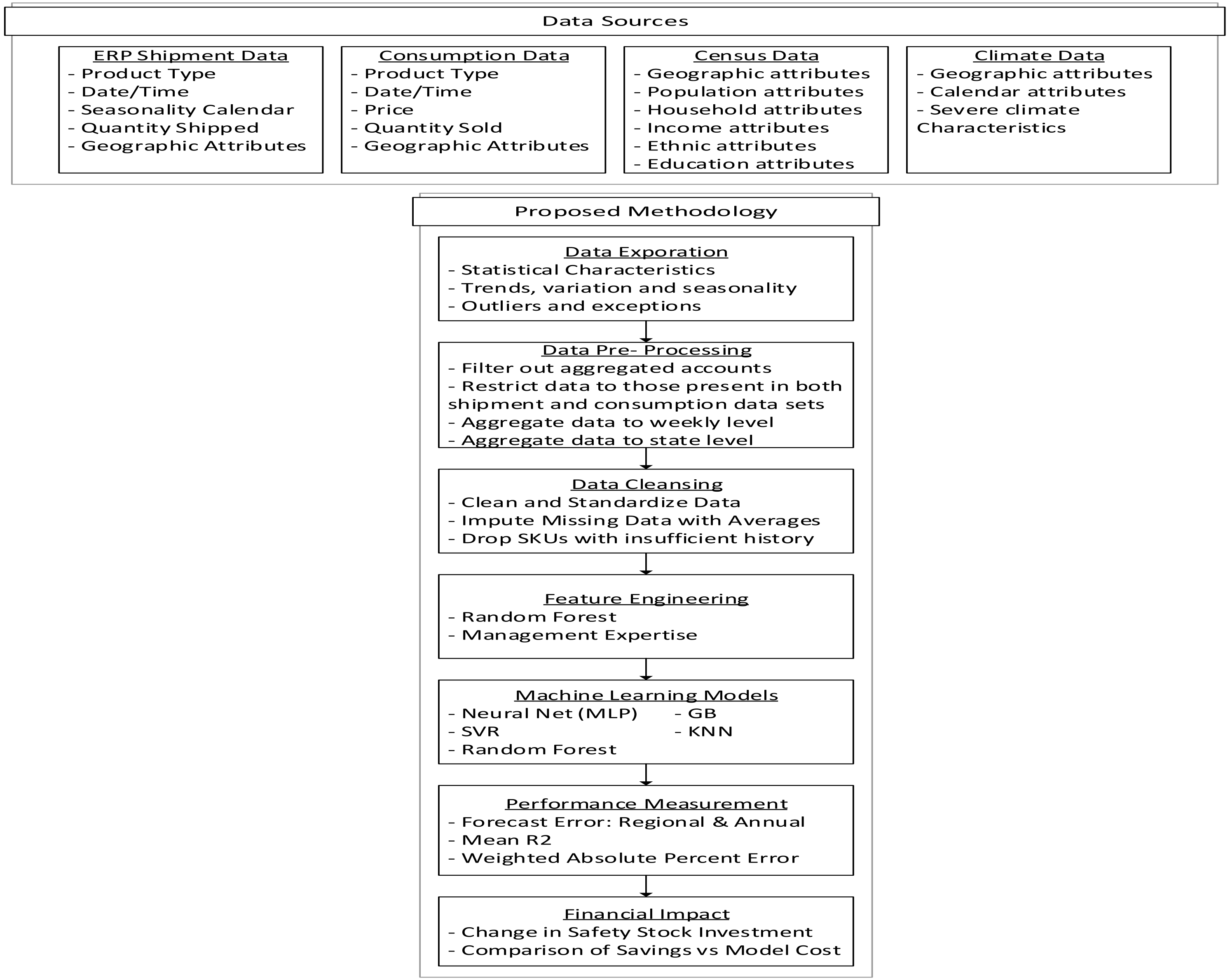
Supply Chain MANAGEMENT

KING'S HAWAIIAN

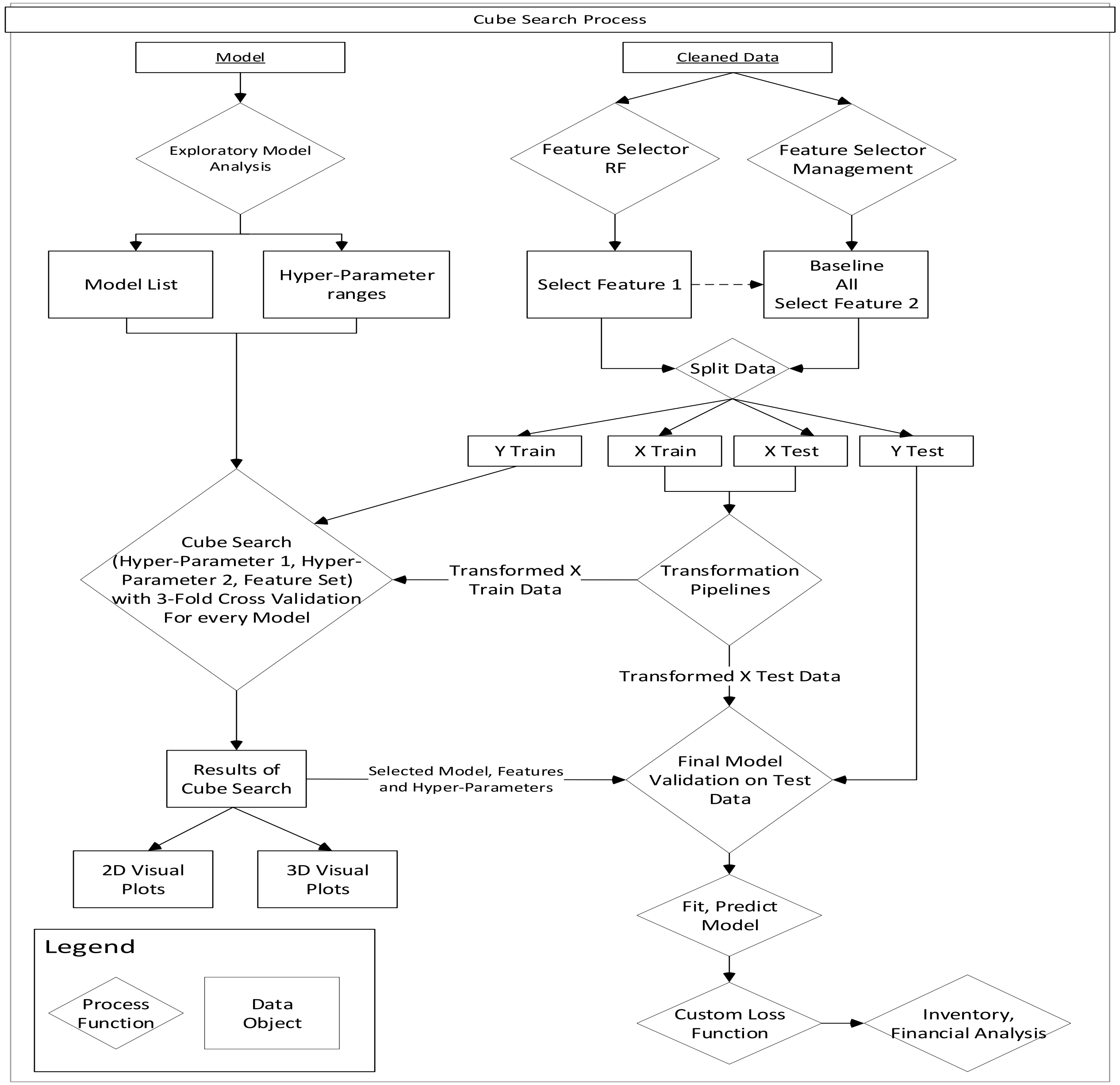
Appendices

- Data sources and overall methodology
- Detailed data flow
- Feature importance curve
- Tuned hyper-parameters
- KNN seasonal vs de-seasonalized performance
- KNN hyper-parameter performance curves
- Comparison of model run time
- Aggregated demand distribution by region

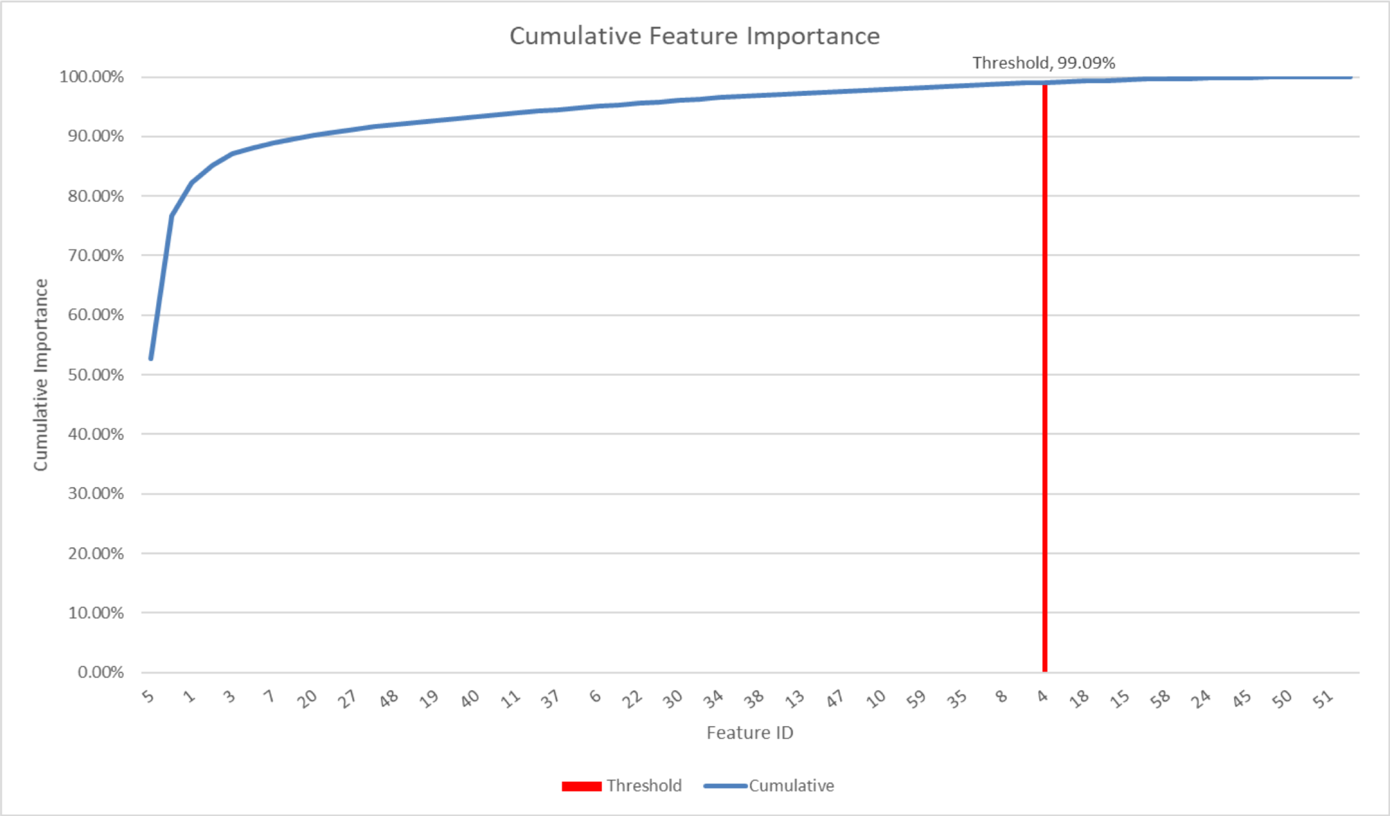
Data Sources and Overall Methodology



Detailed Data Flow



Feature Importance Curve

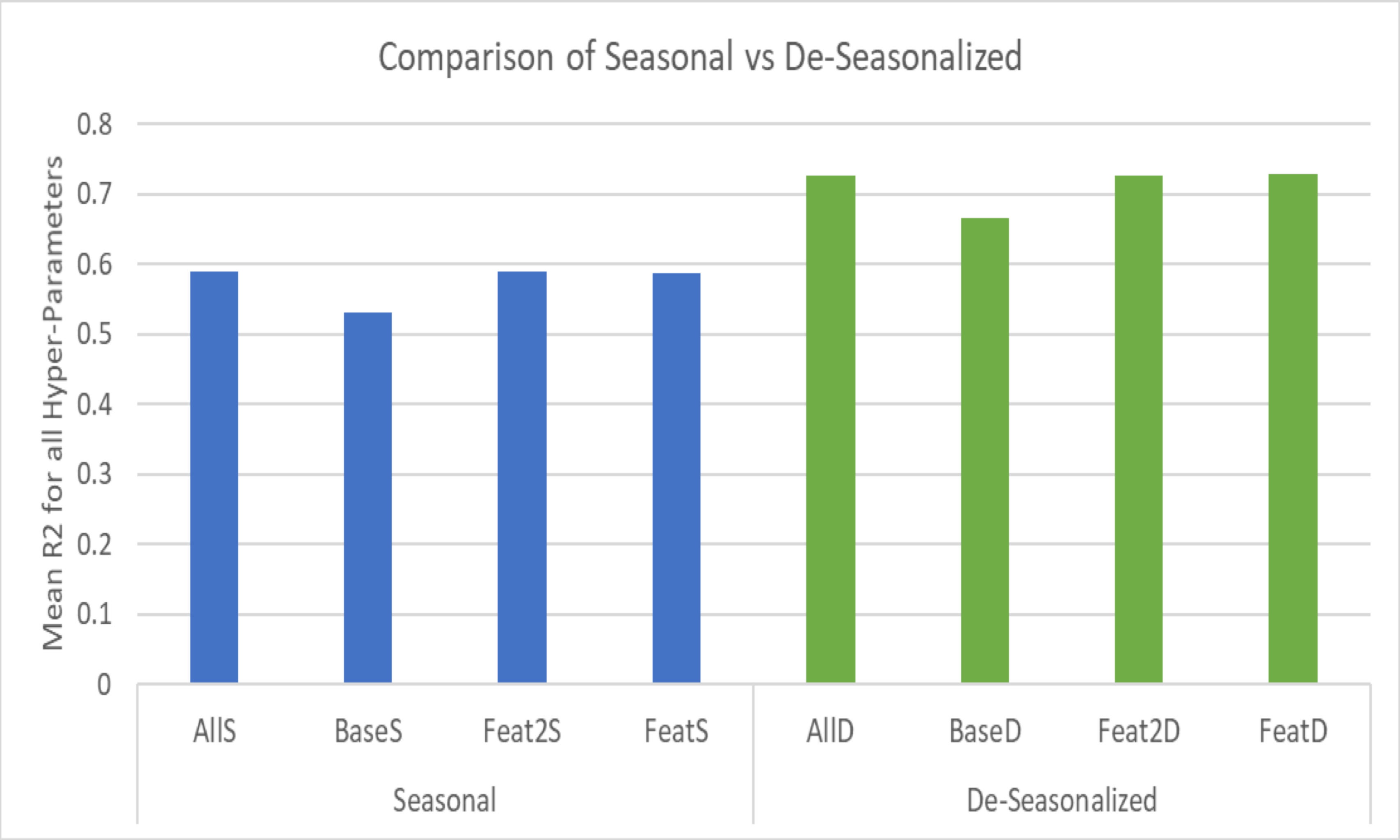


Feature Importance Curve. Number of features plotted against the cumulative weighted value for each feature according to the Random Forest classifier, with the threshold indicating cumulative importance captured in features selected

Tuned Hyper-parameters

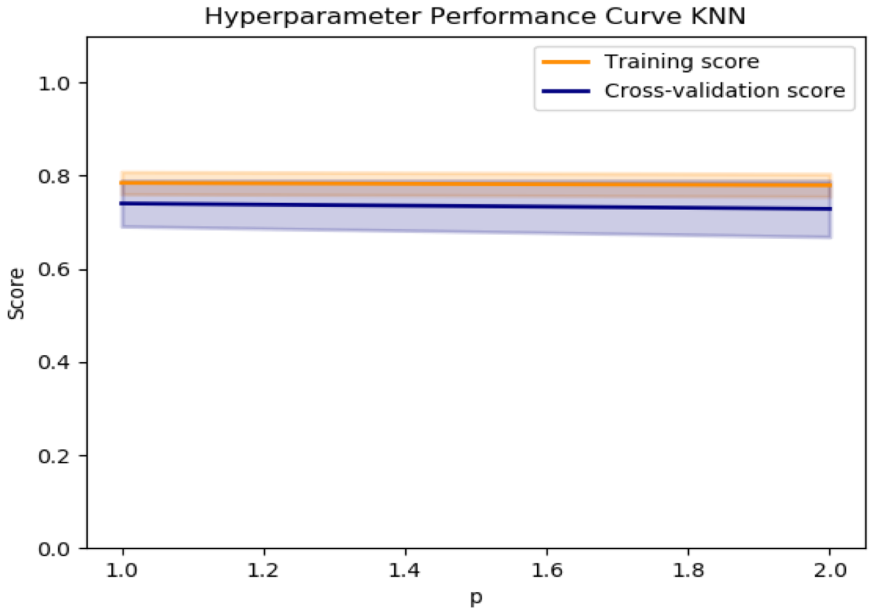
Model	Hyperparameter 1	Hyperparameter 2
Support Vector Regression	C: The penalty for the error	Epsilon: The margin of tolerance for errors
Artificial Neural Network	Number of layers	150 neurons per layer
Random Forest	Max depth: Limits the depth and fit of the tree	Max features: Defines the limit of features considered for each split
Gradient Boosting	n estimators: Sets the number of boosting stages	Min sample split: Defines the samples required to split a node
K-Nearest Neighbor Regressor	P: The Minkowski distance parameter	n neighbors: Determines the number of neighbors evaluated for each observation

KNN Seasonal vs De-seasonalized Performance

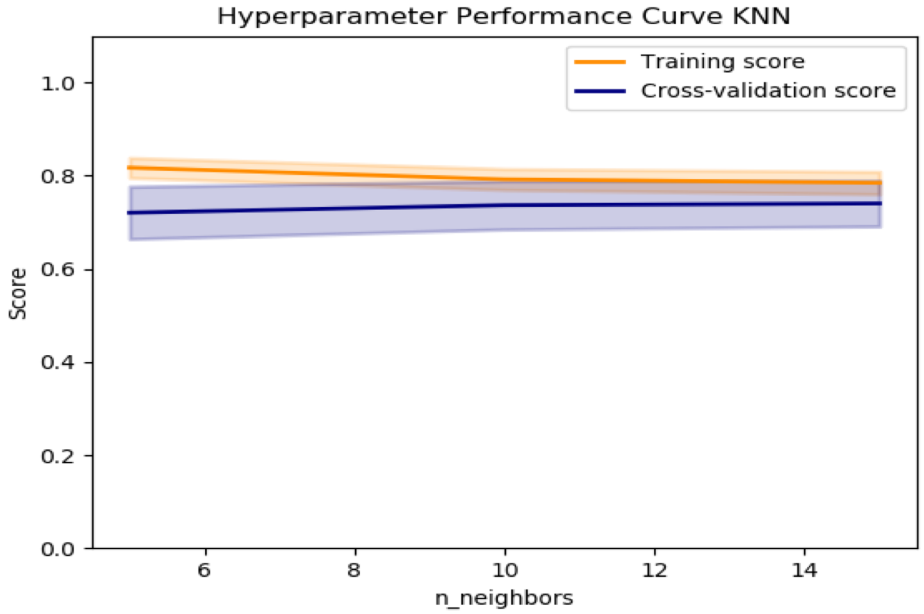


KNN Seasonal vs De-Seasonalized Performance. Comparison of *mean R^2 values for all KNN hyper-parameters tested on seasonal vs de-seasonalized feature sets*

KNN Hyper-parameter Performance Curves

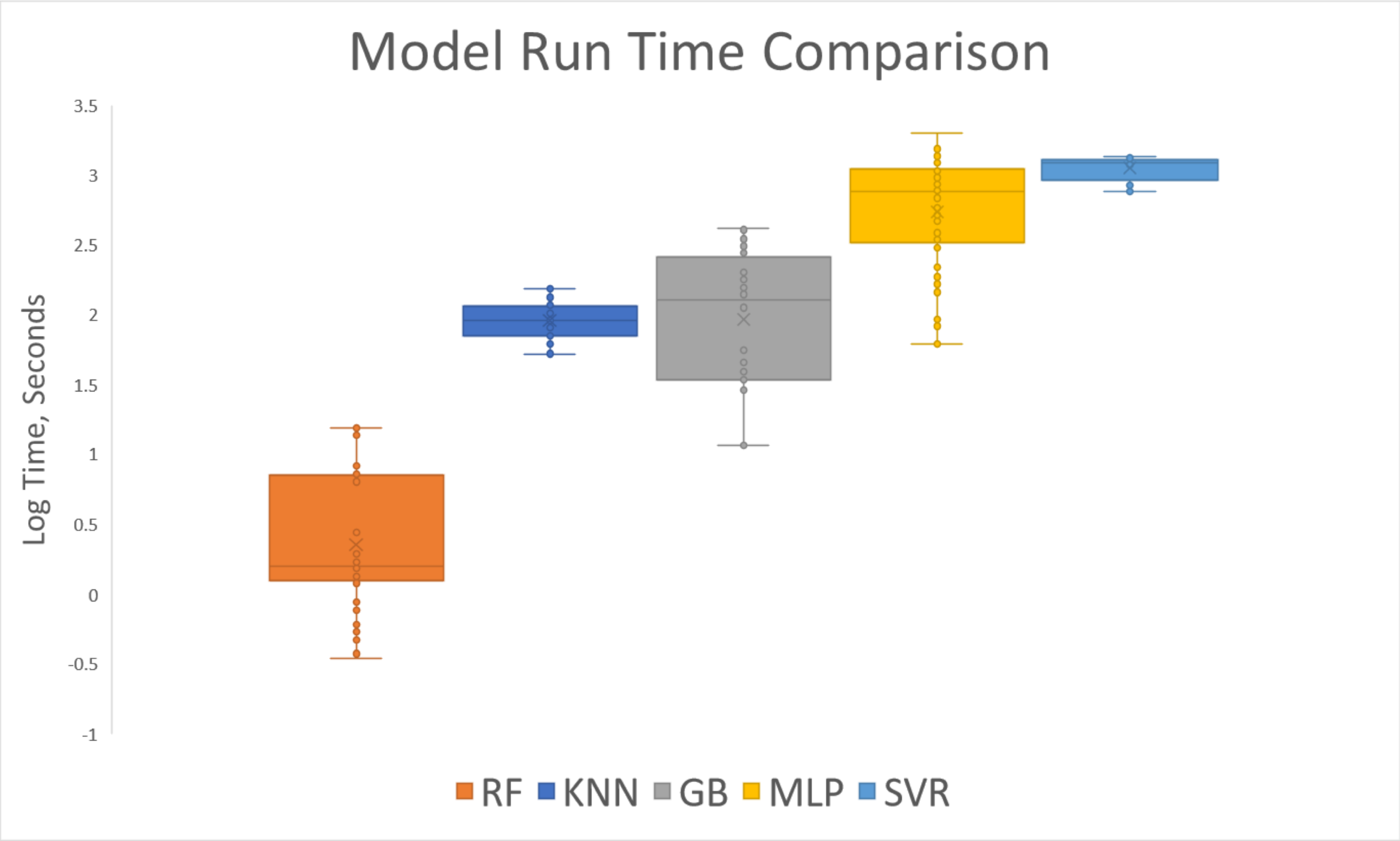


KNN P hyper-parameter Performance Curve. Comparison of Train-Test mean R^2 for the two different values of the p- parameter which determine the distance calculation on the de-seasonalized Feature Select 1 set



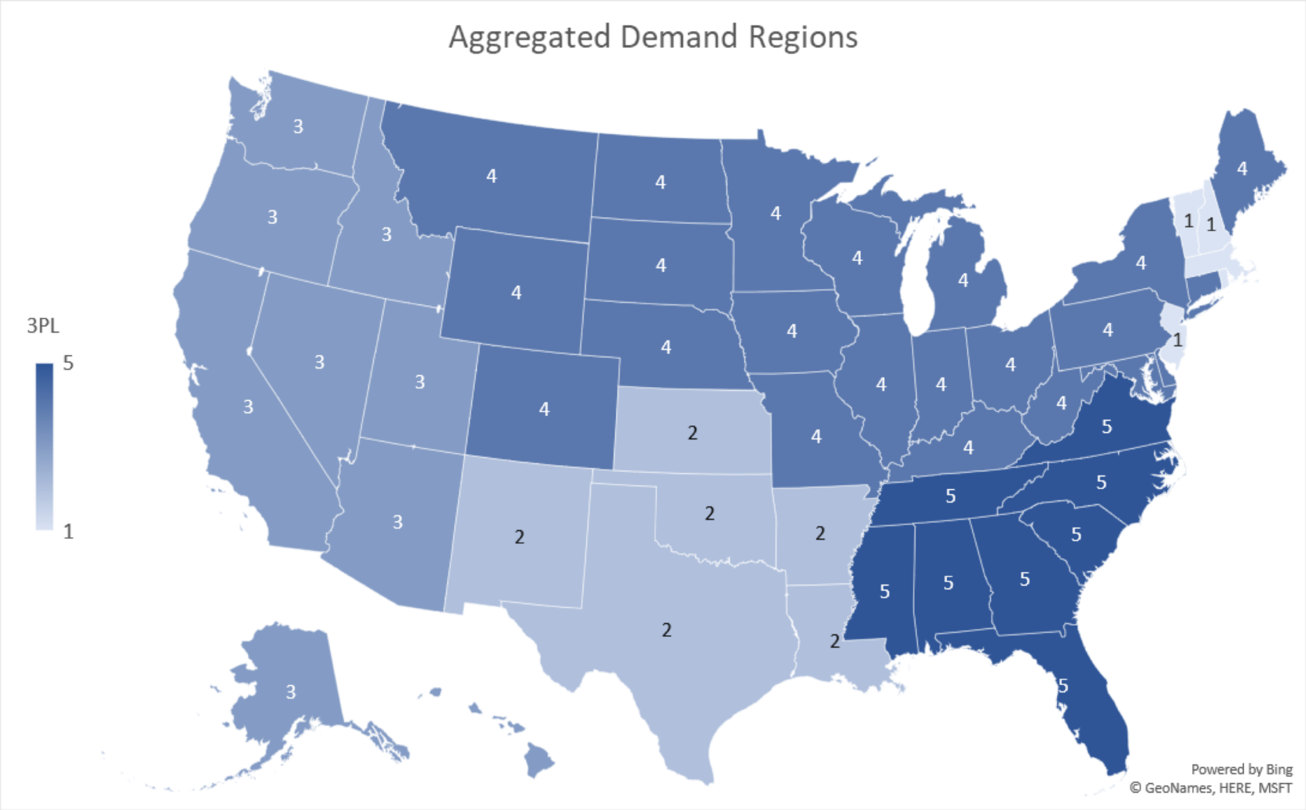
KNN N-Neighbors hyper-parameter Performance Curve. Comparison of Train-Test mean R^2 for varying values of the number of neighbors on the de-seasonalized Feature Select 1 set

Comparison of Model Run Time



Comparison of Model Run Time. Log time in seconds for each feature set run during cube search

Aggregated Demand Distribution by Region



Geographic Distribution Regions. The states included in each aggregated region for forecast evaluation and comparison