Novel Application of Bootstrap for New Store Sales Forecasting

Snehalatha Doddigarla, Srishti Bisen, Sukhsagar Jaiswal, Vikas Das, Matthew A. Lanham
Purdue University Krannert School of Management
sdoddiga@purdue.edu; sbisen@purdue.edu; jaiswal5@purdue.edu; das104@purdue.edu; lanhamm@purdue.edu

ABSTRACT

We provide a bootstrapped-inspired design to both cluster and predict the performance of newly opened stores for an automotive aftermarket retailer. Opening a new store location is a significant investment and can often take years to recuperate the initial financial investment. Forecasting sales for new stores is challenging since new stores do not have actual sales.

We first identified similar stores based on the new store's demographic and store characteristic profiles using multiple clustering techniques and store these sets. Using these different store sets, we develop store forecasts for each store using several different prediction approaches. These models showed varying statistical performance and varying store sales forecasts. For each cluster-group/predictive-model-group, we average the prediction. Ensembling results from predictions together obtains a more reliable and accurate result that is used for better decision-making.

INTRODUCTION

Dependable sales forecasts are essential for strategic planning for any successful business, especially for a market-oriented industry like automotive aftermarket. Forecasting goes beyond the financial analysis, and it assists in setting the goals for the future. It helps in answering questions about footfall, operations, supply and customer spending for a given period. It also serves as a guide for internal expenses to maintain functioning, manufacturing, marketing, and administration.

Our objective is to bootstrap forecasts using historic sales from similar stores. We use PAM clustering to group stores similar to the new store. ETS, ARIMA and Holt-Winters models are used to forecast sales.

Research Questions

- How well does PAM clustering work in grouping similar stores?
- How does bootstrapping improve the overall time-series forecast?

LITERATURE REVIEW

Forecasting has been used extensively in many industries. Research suggests the use of Time Series Forecast models such as ARIMA, Holt-Winters, Error Trend Seasonality (ETS) models to forecast sales.

Motivation	Algorithms used
Forecast Sales for Automobile markets	OLS, QR, SVM, KNN, DT, RF
Aggregated Retail Sales Forecasting at 3 levels: market, chain and store	ARIMA, NN, Winters exponential smoothing, Spatial interaction model,
Convenience store sales forecast based on high dimensionality data	ARIMA

METHODOLOGY

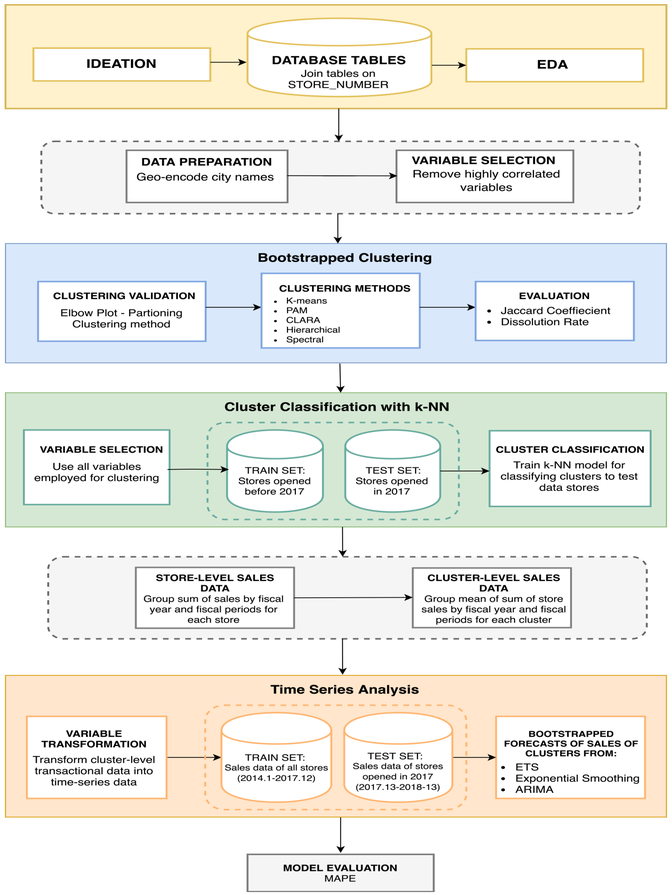

Fig 2. Process Flow

PAM CLUSTERS

PAM (Partitioning around Medoids) is a more robust form of K-means which is intended to find a sequence of objects called medoids that are centrally located in clusters. Therefore this clustering is more capable of handling outliers. The goal of the algorithm is to minimize the average dissimilarity of objects to their closest selected object or medoid. Equivalently, we can minimize the sum of the dissimilarities between object and their nearest selected object.

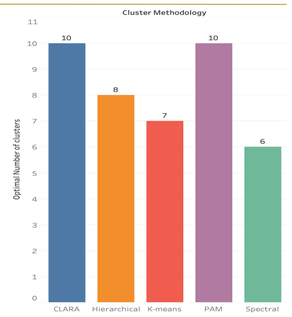

Fig 3. Optimal number of clusters by methodology

RESULTS

Step 1: We performed bootstrap evaluation of clusters built from K-means, PAM, CLARA, Hierarchical and Spectral Clustering to develop the best clusters

Step 2: We classify the new store to one of the clusters created and attribute the mean gross sales of the cluster as historical sales of the new store

Step 3: We bootstrapped time series data for the new store and split it into train and test data. We then build time series forecasting models like ARIMA, HW and ETS on the train data and used MAPE as the evaluation measure on the test data

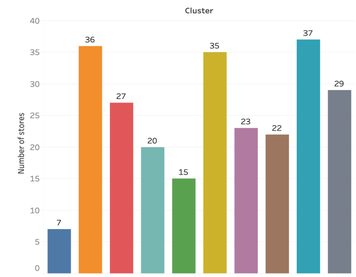

Fig 4. Store distribution across clusters

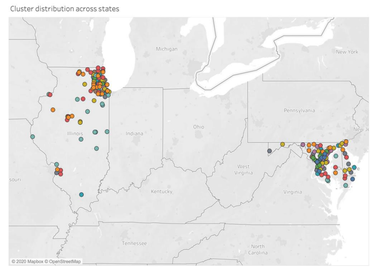

Fig 5. Cluster distribution across geography

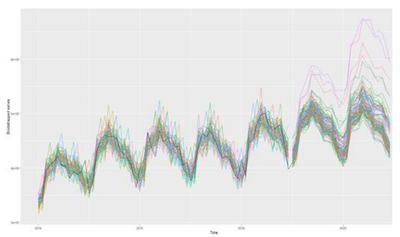

Fig 6. Block Bootstrapping time series data

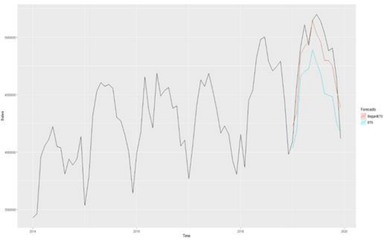

Fig 7. Bagged ETS gives most accurate forecasts

CONCLUSIONS

Thus, Bootstrapped Forecast of similar store from Clustering are found to be **reliable at predicting sales for a new store with no sales history**

If we have demographic details of a new store and find a cluster of similar stores, we can use the average sales of the cluster as a proxy for historical data. Using a **single time series forecast model results in an accuracy of around 94%** whereas **bootstrapping forecasts from multiple simulations increases the accuracy to around 98%**. This improvement in prediction helps the business save around **$1M in revenue** by helping their planning and budgeting for the fiscal year.

This bootstrapped design allowed the retailer to 1) estimate performance (point estimate) and risk (forecast distribution), and 2) provide a means to estimate when the new store would likely break-even from their initial investment.