Identifying Customer Churn in After-market Operations using Machine Learning Algorithms

Vitaly Briker
Southern Methodist University, vbriker@mail.smu.edu

Richard Farrow
Southern Methodist University, farrowr@smu.edu

William Trevino
Southern Methodist University, wtrevino@smu.edu

Brent Allen
Flowserve, BrAllen@flowserve.com

Follow this and additional works at: https://scholar.smu.edu/datasciencereview

Part of the Other Computer Engineering Commons, Other Statistics and Probability Commons, and the Statistical Models Commons

Recommended Citation

Identifying Customer Churn in After-market Operations using Machine Learning Algorithms

Richard Farrow[1], William Trevino[2], Vitaly Briker[3], and Brent Allen[4]

[1] Master of Science in Data Science, Southern Methodist University, Dallas TX 75275 USA {farrowr}@smu.edu
[2] Master of Science in Data Science, Southern Methodist University, Dallas TX 75275 USA {wtrevino}@smu.edu
[3] Master of Science in Data Science, Southern Methodist University, Dallas TX 75275 USA {vbriker}@smu.edu
[4] Flowserve Corporation, Irving, TX 75039 {BrAllen}@flowserve.com

Abstract. This paper presents a comparative study on machine learning methods as they are applied to product associations, future purchase predictions, and predictions of customer churn in aftermarket operations. Association rules are used help to identify patterns across products and find correlations in customer purchase behaviour. Studying customer behaviour as it pertains to Recency, Frequency, and Monetary Value (RFM) helps inform customer segmentation and identifies customers with propensity to churn. Lastly, Flowserve's customer purchase history enables the establishment of churn thresholds for each customer group and assists in constructing a model to predict future churners. The aim of this model is to assist Flowserve in creating a targeted retention strategy to individual customers, with respect to individual customers and priority. Each aspect of the analysis requires a specialized set of tools that will be discussed in detail throughout the paper. The study is based on data from Flowserve's product sales history spanning years from 2014 through 2019. The data includes date of purchase, price, quantity, item description and location of customers.

1 Introduction

To drive profitability in an increasingly competitive marketing landscape, companies have to focus on maintaining and growing their existing customer base. This growth can be achieved through prospecting or solicitation. In prospecting, or the recruitment of new customers, market saturation presents challenges in most industries where the notion of a "brand new" customer is virtually impossible. While the solicitation of customers from a competitor is also difficult and costly.

In general, cost of customer acquisition is significantly higher than customer retention. Therefore, companies generally focus on retaining their existing customers through incentivized loyalty programs, consistency in their quality of

service, and other strategic endeavors. These improvements result in customer satisfaction, leading to customer referrals, and ultimately achieving company growth. Further, longevity is inversely correlated with churn, that is, the longer customers are retained, the easier it is to maintain their loyalty.

In contrast to the benefits customer retention, customer churn can wreak havoc on business operations, including impacts to profitability through revenue loss, greater re-acquisition costs, inconsistent calculation of customer lifetime value, and blurred customer segments [3]. As characteristics of churn vary across industries and organizations, a widely accepted definition of churn is "the propensity of customers to cease doing business with a company in a given time period" [2]. Churn is therefore the point in time at which a customer stops using a product or a service, and these customers are in turn referred to as churners.

"Industries with subscription-based business models traditionally focus on churn more than others. These include banks, telecom companies, insurance firms, energy services companies are among the many types of businesses that use customer attrition analysis and churn rates as one of their key business metrics" [3].

Our churn study focuses on the purchase behaviors within Flowserve's aftermarket operations - specifically for parts, which are used to repair and maintain equipment operating in the field. The scope of this study is (1) to identify associations within and across various part product families and customers, (2) to predict future purchases of various parts for customers, and (3) to understand the propensity of the customer to churn, and when.

The Flowserve Corporation is an International Manufacturing Company with a large and diverse customer base which spans every continent as well as the following industries: Oil and Gas, Chemical, Power, Water, and General Industry (e.g., Agriculture, Mining, and many others). The data in this study is sales volume for Flowserve's Original Equipment Manufacturer (OEM) products.

Using historic customer purchase data, we cluster customers by purchase behavior, identify the probability to churn for each cluster, and suggest optimal response time to reverse churn. Upon the completion of our analysis across Flowserve's customer bookings data we can better understand the propensity of the customer to churn. This analysis enables Flowserve to make proactive decisions intended to improve customer retention.

To effectively manage customer churn events, it is essential to create a predictive model that is robust while also comprehensible. We use statistical and machine learning techniques such as Survival Analysis to create this prediction model. The purpose of the model is to identify initial signals and recognize customers that likely leave voluntarily.

Analysis of individual products or services, and improvements in preventative maintenance was outside the scope of this study. Rather, we use prediction models to identify probability to churn for Flowserve to issue a targeted marketing campaign and incentivize these customers to extend their contract, and continue to procure products or services. Given the high costs in a campaign such as the one described, accuracy in churn prediction is imperative. How-

ever, if high-risk customers are correctly identified, a proactive campaign may strengthen customer loyalty, in turn increase growth, and improve retention.

To accomplish these goals we first used RFM scoring on customers along with K-Means clustering to create customer RFM clusters and then, along with purchase time between the most recent three purchases, we produced customer cohorts to deliver better accuracy on churn prediction. Finally, upon executing survival regression on the entire dataset as well as individual customer groups, we determined that no group had a significant chance of churn before Day 500.

2   Tutorial

2.1   Industry Background

Heavy industry relates to a type of business that typically carries a high capital cost (capital-intensive), high barriers to entry, and low transportability. Flowserve has specific solutions designed to address flow management needs across a variety of heavy industries, such as oil, gas, chemical, nuclear power, waste water and others. High-energy efficiency and maintenance flexibility are one of the major factors that will have a positive impact on company growth in the global market.

While customer details were omitted, it was determined that Flowserve's customer base is comprised of distinct personas observed by their purchase frequency, overall contribution to sales, and longevity of engagement. High value customers, which are frequent buyers and can be attributed to the bulk of sales, make up only 5-10% of the distribution while accounting for 60% of total revenue. Medium value customers, averaging a single order every month, make up half of Flowserve's customer population and are responsible for 30% of the company's revenue. The remaining 40% are low value customers that complete a handful of purchases over a period of five years. The revenue attributed to these customers equals a mere 10% of sales.

2.2   Product and Customer Behavior

Vast portion of the study was focused to better understand the customer and the product using various techniques that are discussed below.

Association or Recommendation Rules

Association rules are used across a wide range of applications including product recommendations as seen in Amazon's e-commerce platform, digital media recommendations that have brought notoriety to companies like Netflix, and other areas such as politics and medical diagnosis to name a few. Retailers are especially drawn to association systems as they inform an array of areas that improve customer satisfaction, drive revenue, and increase profitability. Association rules enable retailers to better understand customer shopping behavior and predict customer purchases as they are moving throughout a store. Retailers tailor

the experience to customers by bringing related products in close proximity and drawing attention to products as customers make their way through the aisles. Store managers pay close attention to store layouts ensuring they are optimized to improve the customer's shopping journey. Marketers target customers using product association rules to prompt relevant discounts or provide notifications on related products. All these predictions are generated using basket analysis, or models built from customer selections using similar cart history. Finally, retailed take advantage of cross-sell opportunities to encourage spend and increase the overall value of the customer's shopping basket.

In this analysis, association rules were applied to identify relationships between products and customers. When reviewing product categories purchased in a single transaction, it was determined that some product categories are often found together and the combination provides evidence of a pattern over time. As an example, pumps and valves are often purchased together while flow meters are bound to isolated transactions. Tangentially, these observations are correlated to customer groups to determine customer behaviour. It may be found that high-value customers, or customers purchasing more frequently, purchase these associated products, which increases their attributed revenue.

Although association rules are often applied in combination with recommendation systems to identify revenue opportunities through up-sell and cross-sell, this was not applied in our scope of work.

Customer Lifetime Value

Customer profitability (CP) is the difference between revenue and cost associated with a customer relationship over a set period. While CP is a monetary measure of the past, customer lifetime value (CLV) is a measure of the future. CLV is the net profit attributed to the anticipated relationship over its future lifespan. Organizations use CLV to cap spending on new customer acquisition and set limits on spending to avoid customer loss.

CLV is defined as the "monetary value of a customer relationship based on the present value and projected future cash flow." Given its uncertainty, future cash flow is discounted over time, and as one would expect, cash flow further into the future is discounted more heavily (e.g., money received 10 years from now is discounted more than money received in 5 years).

$$CLV = margin * [retentionrate_1/(1 + discountrate_2 - retentionrate_1)]$$

1. Retention rate is one minus the churn rate. Churn rate is the percentage of customers who end their relationship with the company in a given period.
2. Discount rate, cost of capital used to discount future revenue from a customer.

Like machines, buildings, and other investments on a company's balance sheet, organizations view a customer relationship as an asset whose monetary representation is CLV. Often used for customer segmentation, companies use CLV to define profitable customer groups and use these cohorts as targets of expensive marketing campaigns. CLV allows organizations to adjust marketing

expenditures and avoid non-profitable customers. This in turn allows companies to maximize resource investments into targeted marketing aimed to reverse churn or restart profitable relationships with former customers.

While the group initially leveraged CLV to infer customer segmentation, the correlations were weak and a different approach was applied to identify customer tiers.

Recency, Frequency, Monetary Value

Another measure of customer value is known as RFM, which is calculated from three distinct components: recency, described by the time span since the previous purchase; frequency, described by how often the purchase is made; and the monetary value, represented by the average spend. RFM is used to analyze customer value in marketing, retail, and professional services industries.

One calculation for RFM can be performed by assigning a score to each dimension on a set scale (e.g., 1 to 10 months since last purchase). Another approach is to define each dimension with subcategories (0 to 90 days, 90 to 365, etc.). Concatenation of each score results in the RFM value for each customer.

RFM was particularly useful in gaining insight from the survival analysis. Originally, modeling an unclustered customer base was too diverse to describe a unique behaviour. RFM delivered a grouping that showcased distinct characteristics of customer classes that corresponded to churn predictions.

Next Purchase Day Prediction

Predictive analysis can be applied to identify a customer's next purchase date, which can be used to lower marketing expenditures and deliver a more targeted marketing campaign. The next purchase date is based on previous purchase intervals and is personalized for each customer. Using this information a company can halt expensive marketing promotions, since the customer is already expected to make a purchase, and incentivize customers with targeted marketing ad, if a purchase was not completed in the predicted time window.

It is important to note that this prediction informs us of the highest probability that the customer will make a purchase, but it does not indicate the probability that the purchase will be completed. To determine the customer's likelihood to complete a purchase, we complement our analysis with Churn Risk Prediction.

Regression for Survival Analysis

Traditionally used in medicine to predict the number of days a person will survive, this type of regression is also applied in other fields to forecast machine failure, calculate drug patients' expected lifetime, and life expectancy of infants. Given its relevancy, survival analysis has been extended to applications in churn predictions. Survival curves show how customer churn changes over time and allows organizations to determine the optimal point for intervention.

Survival analysis identifies the time between birth and death, while accounting for censorship. While in some case these definitions are quite literal, these

events can be applied to the start and the completion of any occurrence or phenomenon. In cancer patients, the birth event could represent the time of diagnosis. While in a relationship study, the birth event could be the time of a couple's first encounter while the death event would represent the end of their courtship. Censorship refers to the death events not actually observed across the entire sample. Making inferences on the data without seeing the results of the data an observer may be biased and underestimate predictions.

Applying survival analysis to Flowserve churn predictions, the following events are defined in the context of customer churn:

Event	Description
Birth event	Time associated with the beginning of an observation period. A customer makes an initial purchase at Flowserve
Death event	Time associated with the end of an observation period either by an event, end of the study, or withdrawn from it. A customer no longer purchases Flowserve products
Censorship	The death event is not observed.

Table 1. Survival events classification

Survival regression allows us to apply a model to survival analysis and predict when an event is likely to occur. It can be used to model a relationship between customer churn, time, and other customer characteristics. This allows us to determine the probability of churn based on customer attributes and helps determine the significant factors that drive churn.

Given the limited size of our data, we did not observe any "drop offs" or customers that have abruptly stopped doing business with Flowserve. Instead, we applied the RFM clusters, previously discussed, to define distinct customer groups, and then applied the Kaplan Mier estimator that takes a probabilistic approach to the lifespan of each observation.

2.3 Customer Churn

What is churn?

Although churn is one of the most important elements in the Key Performance Indicator (KPI) of a product or a service, its individual qualities vary across industries, organizations, and products.

In the gaming industry, a player's engagement requirements will determine the factors that define churners, and these requirements are specific to individual games. Games with daily engagement will determine churners differently than games with other user requirements[1].

Companies in the travel industry, such as airlines and accommodations, would likely segment their customer base into tiers including business and leisure travelers, which would result in dramatically different churn characteristics. A business

customer who may procure these services weekly would have a different threshold for churn compared to an individual whose purchase history reflects a more seasonal pattern.

While E-commerce engines like Amazon would likely gauge customer churn through customer journey analytics and associated metrics. For instance, the number of times a customer logs in, the total duration the customer spends browsing, and the frequency of product purchases. All these factors in conjunction can serve as the primary elements to determine characteristics of churn.

Churning could be a result of low-level customer satisfaction, aggressive competitive strategies, new products and regulations, to mention a few [4]. These influential factors are referred to as determinants of churn. Companies identify these determinants and work to address or improve in these areas to survive in an increasingly competitive marketplace.

How to address churn?

Study conducted in 1995 surveyed more than 500 service customers producing a list of 800 critical controllable behaviors that were grouped into categories including: pricing, core service failures (e.g., billing errors), service encounter failures, employee responses to service failures, and competition [1]. Determinants such as these can often be controlled or influenced by companies, as opposed to customer relocation and personal circumstances (e.g., death) which cannot be prevented. Companies therefore focus to address controllable determinants of churn in one of two ways: (1) company-focused and (2) customer-focused initiatives.

(1) Company-focused initiatives. A customer may want high quality support when calling a product call center. In telecommunication, one would expect customers demands for exceptional cell phone coverage. Accessibility of locations in fast-food restaurants or gas stations. And in almost every industry, customers expect a fair price. To no surprise, any improvement in these areas could potentially alleviate churn; however, each has limitations, including resource constraints, technology advancements, and cost. Therefore, companies often lean on customer-focused initiatives to preserve product loyalty through improvements in services that tailor to customer's individual needs.

(2) Customer-focused initiatives. Focus on individual customer accounts to improve retention - studying the correlation between churn and customer characteristics, such as an Individual age, income, housing or marital status. These characteristics can often predict certain behaviors, such as how frequently the service is used and the overall spend on a particular service. Therefore, using past churn behavior to predict high-risk customers can help companies initiate a targeted retention campaign, such as marketing vouchers. When applied with focus, these campaigns can result in extended service contracts, which not only create additional revenue but also strengthen a customer's loyalty to the company brand through contract extensions.

Our analysis is focused to assist in company-focused initiatives and target Flowserve's sales and marketing pipeline.

How to address churn through segmentation?

Customer churn is typically segmented into three categories, and each one can have a different approach for reversal.

Short-term churn customers are "piloting" different products to determine which of them, if any, provide value. High rate of short-term churn are signs of inaccurate customer segments and are resolved with shifts in targeted sales and marketing funnel [3].

After this preliminary trial period, customers continue to evaluate products for several months; however, can ultimately drop the product and are referred to as mid-term churners. Retaining mid-term churners requires focus on engagement through the introduction of new features, consistent quality and service, and regular communication [3].

Lastly, long-term churn occurs when an established customer with a long history decides to cancel their service. Retention requires reinforcement of the core value of the product, innovation, and quality. Marketing groups target long-term churners with up-sell and cross-sell opportunities to maintain interest in the product offering [3].

Each of these is discussed and addressed in our analysis.

3   Data Set

Data Attributes

Flowserve's data-set contains 1.8 million records of purchase history from 230 customers in period between 2014-2018. The dataset includes features defining order details; these include the following: item and product description, sell price, date of purchase, and geographical location of a customer. The dataset is clean and has no missing data. Due to the wide array of definitions of churn, attributes used to create a predictive pipeline are highly domain specific. Common attributes capture user behavior with regard to their engagement level (e.g., number of logins, total duration, frequency, intensity). For our analysis we use date of purchase, sell price, and item category. Using this data we develop new features like recency, frequency, monetary value, and purchase distances. Our characteristics for customer churn is the frequency of purchase based on averaged rate of the entire population, and grouping by short, mid-term, and long-term churners.

Comprehensibility

Due to the high cost of marketing retention campaigns, our goal is to ensure that we are targeting customers that are truly churning as opposed to customers that would have stayed with the company without any additional incentive. The trade-off between model accuracy and comprehension is difficult but worthwhile in the case of customer churn as verification of the model is critical. The assessment from domain expertise allows us to verify our results from predictions to ensure they are logical. These predictions offer insights into our model and

enable the company to take strategic actions prioritizing efforts with highest impact. Lastly, ability to discuss the creation of our model, which is not only predictive but also intuitive, eases adoption with key stakeholders.

Data Imbalance

Accuracy is misleading - for example, a small dataset with 90 customers who are non-churners while the remaining 10 are. Building a model that predicts that all customers are non-churners automatically results in a 90 percent accuracy even though the model itself does nothing to predict churn. Since majority of datasets will be unbalanced by nature, we will need to over sample churners and under sample non-churners, boosting, cost sensitivity (cost sensitivity to misclassifying churners). If organization's dataset contains more churners, there are more immediate issues that need to be addressed. Furthermore non-churners tend to have much higher transactional volume or data points than churners leading to skewness in overall data between these groups and making the necessity for sampling.

3.1 Methods and Experiments

The machine learning algorithms studied in this work consist of Association Rules, K-Means, Random Forest classifiers, and Survival Analysis. All models have been widely implemented successfully in various disciplines and are summarized in the following section.

Association Rules

Rule generation is a two-step process: first, generate an itemset for example Bread, Egg, Milk; second generate a rule from each itemset like {Bread -> Egg, Milk}, {Bread, Egg -> Milk}

Association Terms	Description
Association rule	X->Y is representation of finding Y on the basket which has X
Itemset	X,Y is representation of the list of all times which form the association rule
Support	Fraction of transactions containing itemset
Confidence	Probability of occurrence of Y given X is present
Lift	Ratio of confidence to baseline probability of occurrence of Y

Table 2. Terms and definitions of associations rules

1. Generating itemsets from a list of items. The size of an itemset can vary from one to the total number of items in a sample. The goal is to identify frequent itemsets, or ones that occur at least a minimum number of times in a transaction. These are itemsets for which the support value is above a min threshold. {Bread,Notebook} may not be a frequent itemset if it occurs two times out of a 100 transactions, if 2% is below the value of "minsup". Apriori principle allows us to prune all supersets of an itemset which does not satisfy the min threshold. Apriori algorithm - generate all frequent itemsets (support >= minsup) having only one item. Next generate item sets of length two as possible combinations of above. Then, prune the ones for which support value fell below minsup.

2. Generating all possible rules from the frequent itemsets. Once frequent itemsets are generated, identifying rules out of them is less strenuous. If Bread, Egg, Mile, Butter is a frequent itemset, candidate rules are egg, milk, butter -> bread Bread, milk, butter -> egg Bread, egg -> milk, butter Egg, milk -> bread, butter Butter -> bread, egg, milk From list of all possible candidate rules, we aim to identify those that fall above a minimum confidence level. Confidence of rules generated from the same itemset also follows an anti-monotone property. This means that A,B,C -> D >= B,C -> A,D >= C ->A,B,D.

In our study we use association rules to find the same item-category between customers that belong to different purchase clusters. These associations rules are applied to our survival analysis results in order to create a recommendation system for customers with certain probability of churning.

RFM

RFM analysis into data mining - clustering, classification, and association rules. Clustering based on RFM attributes provides behavioral knowledge of customers' marketing levels [9]. Classification rules using RFM variables help predict future customer behavior [9]. Association rules based on RFM allow analysts to establish patterns and relationships between product properties and customer's contributions to provide recommendations [9].

Next Purchase date

In effort to calculate the next purchase date, we split our dataset into two parts: the first to identify the last purchase date within a period, and second to determine the first purchase. This in turn creates a minimal distance between two purchases for each customer. The feature is our target variable.
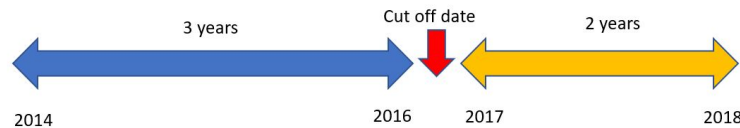

Fig. 1. The schema shows the data split strategy for creating target variable

As a result a feature engineering we calculate and use RFM scores and clusters, days between last three purchases (purchase distances), mean and standard deviation of purchase distances.

Fig. 2. Customer purchase distances (time between purchases) distribution.

In figure two there is skewed distribution of the time between purchases, and data widely distributed, we decided to create multi classification problem and cluster customers into three groups.

Now we can build our classification model to predict next customer purchase. We

Next Purchase Day Range	Customer count	mean	std	min	25%	75%	max
Less than 20 days	21	300.57	278.90	104	125	262	1043
Between 20-100 days	21	50.38	24.01	21	27	69	93
Higher than 100 days	39	9.13	4.40	4	6	13	20

Table 3. Customer classification table based on minimum purchase distance

will build the model based on the below technique and select the best estimator by judging the f1 score.

Logistic Regression

Logistic regression in its specialized form of regression analysis is designed to predict the relationship between one binary variable, in this case churn. To understand logistic regression we must first understand the basics of regression. In regression the predicted or dependent variable is described by the sum of all features or independent variables multiplied by some weight value that is unique per variable and then a final bias term is added to the total. This changes from regression to logistic regression when we add a log-odds function to our output to scale the results to the range of 0 and 1 for the true/false to our binary variable. This function is known as a logit function and thus where we derive the name logistic regression.

Naive Bayes classifier

The Naïve Bayes classifier algorithm is based on Bayes' theorem which is better defined here: "Bayes' theorem is a method for calculating the validity of beliefs (hypotheses, claims, propositions) based on the best available evidence (observations, data, information). Here's the most simplified description: Initial belief plus new evidence = new and improved belief." [8] In essence it is a likelihood of an event given an already known relationship between a condition and that event. This algorithm further gets its name from the naïve assumption than independence between all features exists. This assumption is made to simplify the calculations and make them traceable. "This is a very strong assumption that is most unlikely in real data, i.e. that the attributes do not interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold." [7]

Support Vector Classification

The objective of the Support Vector Machine or SVM is to separate two groups of data points with boundary and the maximum distance between the two data sets and that boundary. This boundary is a hyperplane since in order to split any N dimensional object into two complete sets you must use an N-1 dimensional plane. "...the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3." [6]

Unfortunately, a SVM is very difficult to visualize beyond 3 features represented on a 3D graph and even harder to attempt to explain. Another major disadvantage is the training time for an SVM quickly grows with the addition of any new feature. That is the more features you are working with the slower an SVM performs on train time. However, SVM models tend to have good accuracy when they are trained but usually not stand out enough to justify the train time on large feature data sets.

K-Neighbors Classifier

K- Nearest Neighbor or KNN is at its core one of the simplest machine learning algorithms and in fact is considered an instance-based learning since no training is required before prediction can occur. The user first defines how many neighbors to consider (K) and then a distance formula calculates the distance to every neighbor and the lowest K are considered. The new data point is considered to be most related these nearest K neighbors and is labeled accordingly. This is chosen for classification problems the the mode and on regression problems by the mean of all K data points.

For KNN to work best a correct value of K should be defined. One pitfall that should be avoided is picking an even number of K and opening the algorithm up to a tie between two labels. To avoid this simply pick an odd number of K.

Decision tree and random forest

Decision tree learning is performed via the use of a decision tree predictive model which takes an item as its features and makes a conclusion on some target value for that item. A decision tree, like the name implies, is arranged in a tree-like structure that can best be described as a flowchart where each node represents a "decision" on a feature with some condition. This True or False decision node does a binary branch for both outcomes to the next level of decision nodes. The process continues a leaf node is reached that contains an outcome or prediction.

Due to the flowchart like nature of decision trees they have a few key advantages. The main advantages being they are very straightforward to visualize and thus to explain and they can handle any categorical information without the requirement of data manipulation. For example "the person is male?". They also benefit from the simple conditional statements of the tree being very light computationally and thus have quick execution time for prediction.

However decision trees also carry a few notable cons the biggest being that they can be very prone to overfitting. This most commonly occurs when a tree has many layers of nodes. The more node layers we have the less data each layer must consider for each split and when we are only considering a few possible matches in the training data set we make incorrect conclusion that are not generalizable.

To solve this problem we must expand upon decision trees with random forest.

"Ideally, we would like to minimize both error due to bias and error due to variance. Enter random forests. Random forests mitigate this problem well. A random forest is simply a collection of decision trees whose results are aggregated into one final result. Their ability to limit over-fitting without substantially increasing error due to bias is why they are such powerful models." [5]

Random forest further protects against the over-fitting mentioned above by creating the collection of decision trees that only use a subset of all features provided. The technique ensures that any one feature would not be over weighted in the final result.

While other techniques including Logistic Regression, SVM, and Naive Bayes were applied, RF resulted in 84% accuracy.

Evaluation Techniques

To evaluate of our results, we applied the following techniques.

To measure model optimization we used the f1 score. It provides a better measure of incorrectly classified cases while accuracy only measures cases that are correctly identified. In addition, the f1 score is a better metric for an imbalanced dataset like the one in our case study.

For survival analysis we created a censored feature ("dead" or "alive") based on the mean value of the rate. The rate is the frequency of purchases during a customer's lifetime.

For the apriori algorithm we used the support of 10 percent, a value selected to obtain enough useful combinations. Lastly, in the application of rules, we selected all rules with lift value greater than one.

4 Results

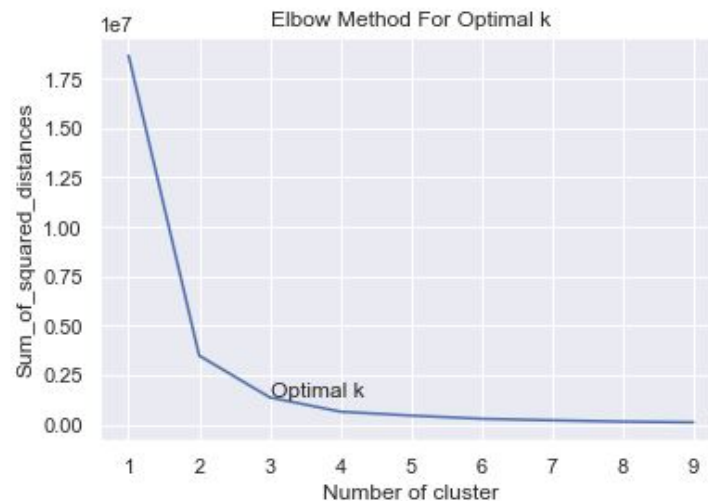Clustering RFM scores based on recency calculation using K-Means algorithm.

Fig. 3. Result of K-Mean clustering.The elbow graph with optimal k=3.

The results of our Elbow test for best cluster size in our RFM analysis can be found here on Figure 3.

With each separate RFM cluster of size three combined we computed an overall value score for customers in Figure 5, with zero being the lowest value score and six being the highest.

Segmentation	Customer Count
High-Value	19
Low-Value	74
Mid-Value	77

Table 4. Customer count segmentation based on recency K-Means result with k=3

The results of the overall value scores were subsequently clustered to create a High, Mid, and Low value customer clusters as shown in Figure 6.
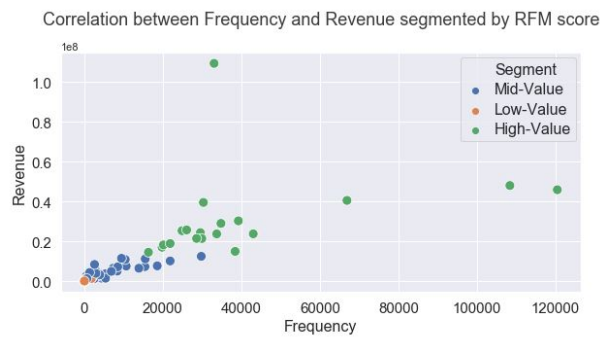
Fig. 4. Correlation plot for each segmented group High, Medium, Low. Higher frequency yields higher revenue
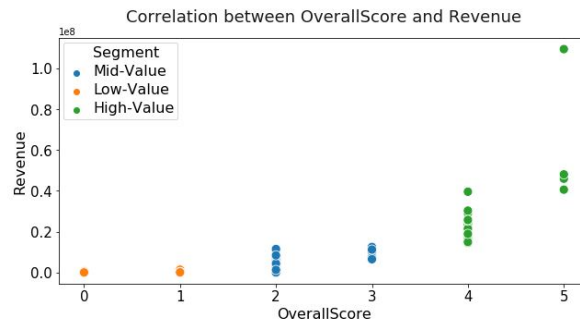
Fig. 5. Overall score as function of revenue. Higher score yields higher revenue

Figure 4 displays the scatter plot of High, Mid, Low value clusters plotted against Frequency and Revenue. While Figure 4 represents a scatter plot of the same group compared to the overall score verse revenue.

Next chart represents next purchase date for each class verses the overall score. We define our classes in Table 3.

Fig. 6. Minimal purchase distance as a function of Overall RFM score. Higher score yields more frequent purchases.

The last step before modeling we observed correlations between our features. There is a correlation above 0.7 between the next purchase date and RFM score parameters.
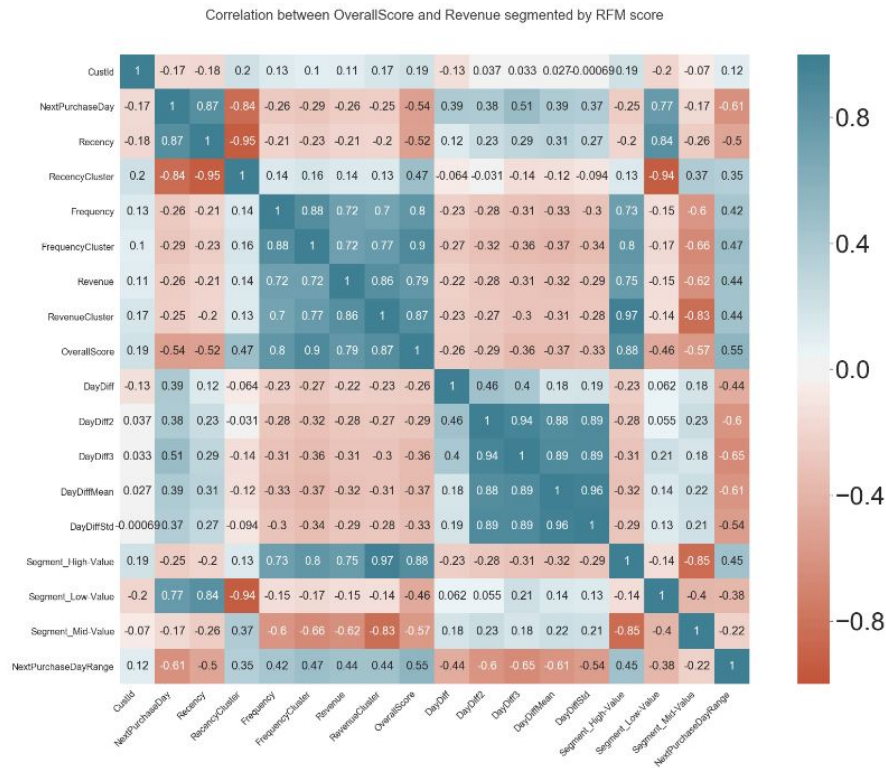
Fig. 7. Correlation Matrix of Features

Using different classifiers we fit a model to our dataset and select the best performer based on the f1 score.

Classifier	f1 Score (%)
Random Forest	87
KNN	63.2
Logistic Regression	60.3
Naïve Bayes	57.1
Support Vecto Classification	31.4

Table 5. Baseline f1 score for each classifier.

Based on the result of the best classifier, in our case Random Forrest, the classification report in Table 5 represents results of fitting the classifier. f1 score

is around 0.7 for Class 0 and 1 and 0.9 for Class 2. Class 2 is a customer with most purchased activity.

Using grid search we fine-tune the model optimizing Random Forest hyper parameters such as bootstrap, 'max depth', 'min samples leaf', 'min samples split' and 'n estimators'. Table 6 represents the difference before and after optimization.

Class	Precision		Recall		f1-score		support
	Before	After	Before	After	Before	After	
0	0.75	1	0.6	0.53	0.67	0.7	15
1	0.59	0.6	0.83	1	0.69	0.75	12
2	0.95	1	0.86	0.95	0.9	0.98	22

Table 6. Classification report comparison before and after hyper-parameters optimization
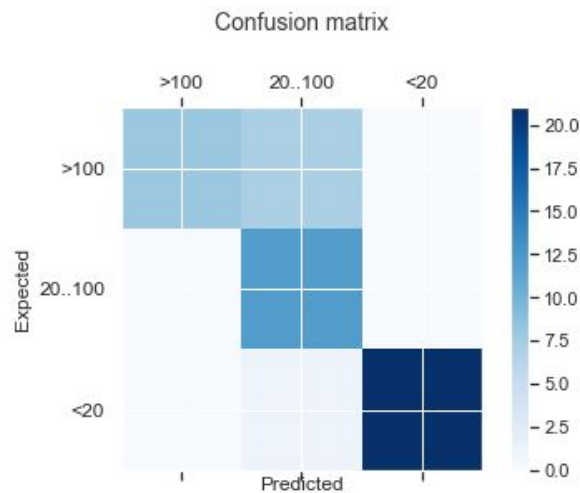

Fig. 8. Confusion Matrix after fine tuning

The next step represents the results of our survival analysis with and without implementing the clustering algorithm that we was described above. Each of the charts show the confidence interval as well.
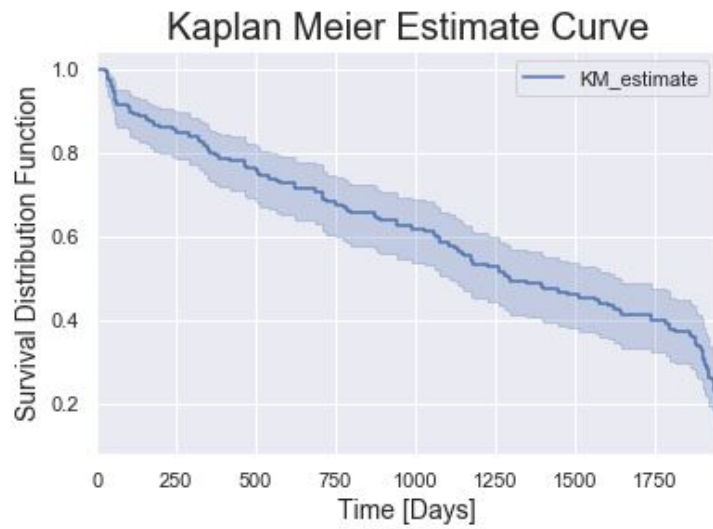
Fig. 9. Survival Analysis KM Estimator

Figure 9 represents the Kaplan Meier survival regression of entire dataset unclustered by customer.
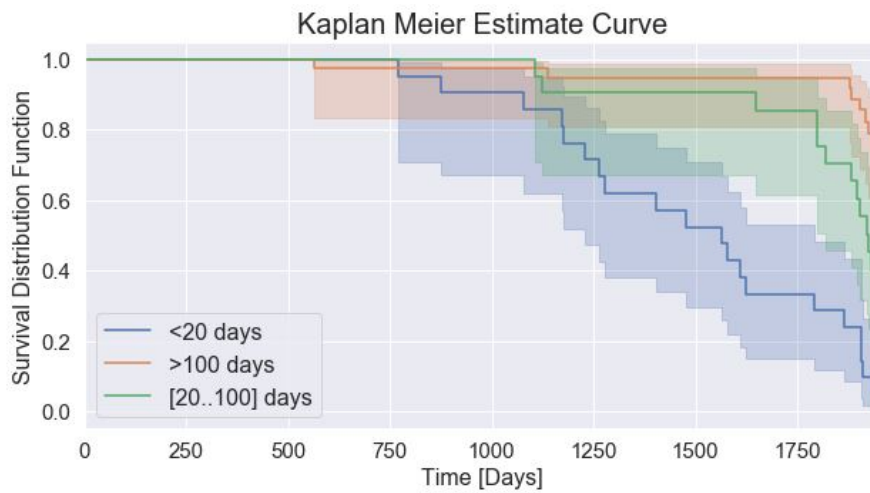

Fig. 10. Survival Analysis across Clusters

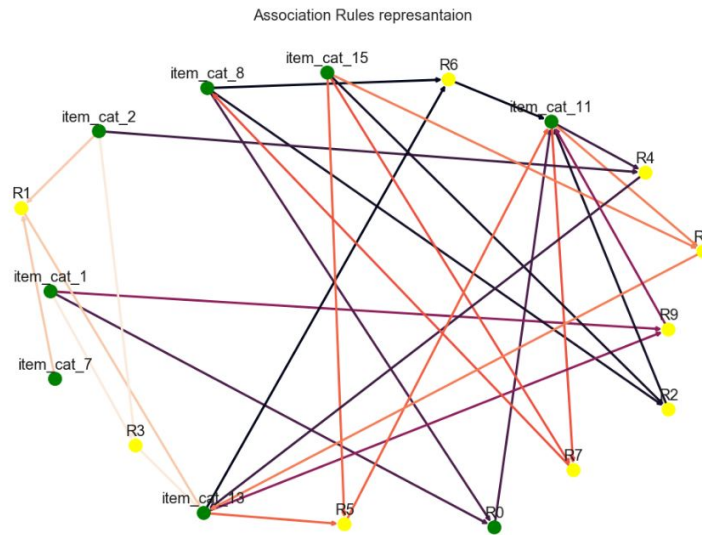Figure 10 shows the results of the Kapaln Mier survival regression while applying the customer cohorts.

Fig. 11. Association 1 (product nodes diagram)

Figure 11 is the visual result of the top ten product association rules that were grouped by each customer e.g., product A is purchased with product B.

5   Analysis

RFM next purchase prediction

To compute a customer's next purchase, we split out data into two datasets: purchases in years 2014-2016 and purchases in years 2017-2018. We used the first set to create new features using customer purchase history and time when the purchases were made and second group to create a target variable for prediction. First we calculate RFM value for each customer and cluster them using K-Means algorithm based on the best recency score. Elbow shows a major jump at two clusters but still a sufficient improvement at the three cluster mark to justify proceeding. After clustering each component we combine them into one overall score. Group 0, 1 and 2 with average of half a year to years since last purchase and total revenue under approximately 1 million average. Groups 3-5 with a purchase a few times a months. Finally with group 6 that appear to purchase multiple times a week. Subsequently, customers were segmented into three (3) distinct categories: Low, Medium, and High based on their corresponding RFM scores. Next we calculate the days between last three purchases, mean and standard

deviation, and use it as explanatory variables in our dataset. Next we create a classifier. We use last purchase date in from the first group and the first purchase in the second group. Cluster the result base on purchase distribution into three classes, Table 3, customer with purchase rate higher than 100 days, between 20 and 100 days and less than 20 days.

The end result was a multi-classification problem where the proposed model predicted next purchase date range for each customer. For our final model we used Random Forrest classifier to predict a customer group. The predicted accuracy after optimization on the test data set is 84 percent. The f1 score for each category are as follows: more than 100 days group is 70 percent, between 20 to 100 days is 75 percent, and less than 20 days is 98 percent, see Table 6.

Survival Regression

For Survival Regression we trained with both the full unclustered dataset and a grouping based on last three-day purchases. From Figure 9 we can see a fairly consistent decrease in likelihood not to churn and a confidence band of about 10% throughout the entire 2000 day period ending with a 20% chance. The slit out of data into three groups gives a lower confidence band (Figure 10) that can probably be attributed to the decrease in sample size when compared to the whole but also each section seems much more meaningful and with these distinctions it manages to last much longer before the first decrease in likelihood below 100% happening at approximately day 500. These numbers are much more actionable when compared to the results of the unclustered data survival analysis results.

Application of Association Rules for Item Categories

When association rules for product categories were computed across all order we ended up with 10 rules with high confidence, Figure 11. Those associations rules applies to survival analysis result in order to create a recommendation system for customers with certain probability of churning across clustering groups.

6 Ethics

It is important to note that while clustering customers to obtain more general customer behavior, the grouping may have lost some customer characteristics that could have pertained to a protected class. Although not used in this study, protected classes are often overlooked and due to this are protected by law. Ensuring correct representation of these classes in any future analysis of this work may be required.

Further, any application of this model should be ran with the knowledge that attributes pertaining to customer uniqueness were not taken into consideration and that these would need to be applied on a per customer bases to ensure no future legal or ethical violation.

7 Conclusions and Future Work

As mentioned previously the product association rules lead to mixed results that were not useful as features for our analysis. This issue seems to stem from our small customer base and skewness in product categories and customer purchases. Association rules tend to work best in situations with high transitions between a large set of diverse customers and neither condition was met in our dataset.

The RFM overall score clustering method presented here shows great promise for predicting which customers fall in which of the three: High, Mid, and Low value sections and is in a high enough accuracy where it should be considered for predicting customers that may transition between groups.

We have shown that with properly clustered customer cohorts that we can generate distinct and reliable churn rates with survival analysis. This analysis can provide confidence that a customer will not churn and deliver relative likelihood of churn at any given time for the business to take actions to reverse attrition. This will provide Flowserve the ability to focus marketing and enriching the Mid value customers in hopes to turn them into future high–value customers and to focus less energy on top High value customers who have a very low probability of seeking business elsewhere.

Further work on this subject should be expanded to the entire customer base of Flowserve and not just one industry. This could improve performance in area of survival analysis and, with the larger customer base, allow for smaller confidence bands around such models. This larger customer pool could also allow sub-sampling of customers based on possibly more meaningful clusters such as the overall RFM score. This expanded customer grouping could show High value customers who are at such a low risk for churn that marketing can ignore and the High value customers who are beginning to trend towards being Mid value so these groups can be approached in a similar manner as the current Mid value customers.

References

1. Braun, Michael; Schweidel, David A.; Modeling Customer Lifetimes with Multiple Causes of Churn
2. Xie, Yaya; Li, Xiu; Ying, Weiyun; Customer churn prediction using balanced random forests.
3. Pointillist; The Data-Driven, Journey-Based Guide to: Predicting and Reducing Customer Churn
4. Vafeiadis, T.; Diamantaras, K.I.; Sarigiannidis, G.; Chatzisavvas, K.Ch.; A comparison of machine learning techniques for customer churn prediction
5. Liberman, Neil; Decision Trees and Random Forests
6. Gandhi, Rohith; Support Vector Machine — Introduction to Machine Learning Algorithms
7. Brownlee, Jason; Naive Bayes for Machine Learning
8. Horgan, Jason; Bayes's Theorem: What's the Big Deal?
9. [D. Birant]- http://cdn.intechweb.org/pdfs/13162.pdf
10. [S. Prem] - https://medium.com/@shubhampr31/how-to-calculate-the-customer-life-time-value-whe