**Predictive Analytics for Customer Churn: Leveraging XGBoost to Enhance Retention and Business Value**

* **Executive Summary**

The purpose of this project is to use machine learning which includes XGBoost, to determine whether customers are likely to churn and help the business plan changes to retain more customers. Losing customers or subscribers because of churn can be tough for businesses in any industry. If organizations create a solid model that foresee risks, they can recognize customers who might switch and help those customers promptly.

The data used in this project includes customer demographics, their usage pattern and details of their interactions with the service provider. Upon completing preprocessing and selecting features, the XGBoost model was taught and tested using the indicators accuracy, precision, recall and F1-score. This model was highly effective at predicting customer churn and revealed the main factors behind it.

The use of the model in business allows for higher loyalty from customers, decreases how much it costs to bring in new customers and raises the company’s profit. Once retention managers understand who is likely to leave, they can create campaigns to ensure the customer base grows and helps increase ROI and loyalty to the brand.

Although the approach achieves good outcomes, imbalanced data and the constantly changing behavior of customers can still be improved. Even so, using this solution steers a company towards making decisions based on data and lasting success.

* **Introduction**

In order to thrive and make a profit over the years, companies must focus on keeping their current customers. Because the services provided by telecom companies are standard and easy to switch, they experience much customer churn. Churn (people stopping to use an entity) can significantly reduce both revenue and the efficiency of how a business operates (Verbeke et al., 2012). Will have a collection of insurance policies that were purchased?

Large amounts of data and the ability to detect complicated, winding patterns have made machine learning (ML) useful for predicting churn (Ahmed et al., 2016). Using ML models allows companies to see which factors such as habits, service usage and demographics, cause customers to stop using their service. Using decision trees, ensemble methods and gradient boosting classifiers has worked well for predicting customer churn (Churn prediction, 2020).

Machine learning is used in the project to try to predict when customers will churn in a real telecom company. Among the attributes are gender, how long the customer has been with the company, the sum of their monthly bills and what services they signed up for. It is important to construct a strong ML pipeline by adding data preprocessing, creating the best model, performing training and evaluation and translating the results. These models are chosen since they are successful with both uneven datasets and detecting non-linear connections (Chen and Guestrin, 2016).

Furthermore, dimensionality reduction will be implemented with PCA and groups of similar customers will be identified using KMeans. Using them reduces the complexity of the inputs, helps clarify the results and provides useful information for forming customer segmentation approaches.

In the end, what this analysis strives to achieve is better decision-making in serving customers and minimizing the number of customers who do not stay.

* **Problem Identification and Justification**

Dealing with customer churn is a major challenge in the telecommunications industry. Churn means customers choose to leave a service and move on to something else offered by a similar business. Customers are making it easy to change and as telecom companies add more services, customer churn leads to significant dips in their revenue. Bain & Company reveals in 2018 that if companies make customers stay just 5% longer, it could improve their profits between 25% and 95%.

The main issue tackled by this project is not being able to anticipate which customers might leave. Firms in the telecom industry often depend on basic strategies or past patterns which are often not precise enough and don’t consider the many aspects of customer behaviour (Huang et al., 2019). Older ways of modeling normally cannot handle bulk, moving data about service use, demographics, bills and contracts. As a result, we must use machine learning (ML) to accurately and efficiently make predictions that work for a larger number of cases.

Machine learning is efficient in learning patterns from known data and accurately predicting things like churn for the future. Some algorithms such as decision trees, support vector machines and ensemble methods, can select important predictors and their relationships all by themselves (Ahmed et al., 2016). Additionally, because ML adjusts to changes in customer behaviour and markets, it remains an effective method for keeping customers.

The objective in this project is to design an ML pipeline that can determine if a given customer is likely to churn or stay with the company based on a set of attributes for those customers. Since the data is imbalanced, classification methods such as Random Forest, Gradient Boosting Machines and XGBoost are effective choices for solving this issue (Chen and Guestrin, 2016).

Resolving this issue is important since it directly influences business earnings, how a company prepares for operations and how they deal with customers. Telecom companies should watch for at-risk customers, react by offering them unique incentives and handling resources efficiently. In addition, this will cut down on advertising costs, boost the average customer spending and make the company more competitive against rivals.

* **Data Cleaning and Preparation**

In any machine learning project, data must first be cleaned and organized, as the how your data is structured affects both the model and how accurate predictions are (Kotsiantis et al., 2006). The data for this job has 21 features that consist of customer details, their habits with services, charges and a label of whether or not they churned. To use them with machine learning, you first clean, encode and format these attributes.

To begin the cleaning process, we handled the cases where data was missing. At times, the TotalCharges column holds empty values for new people because they did not have previous bills. Depending on the business rules and their effect on the data, the blank values were replaced with 0s or the whole row was deleted. If missing values are not addressed, they can make the model inaccurate (Rahm and Do, 2000).

Then, gender, Contract and PaymentMethod were converted using label or one-hot encoding so that the algorithm can process them correctly. If we use Random Forest or XGBoost on categorical variables, we have to transform them into numbers beforehand because these algorithms can only use numbers (Zheng and Casari, 2018).

The values of each example in MonthlyCharges and TotalCharges were examined using interquartile ranges and Z-scores. Extreme numbers in the data set can change the learning process and lessen the accuracy of XGBOOST, among other algorithms.

For simpler comparisons, these types of variables were given a single scale each by applying standardization or normalization. When scaling is used, training the model becomes more effective and no single feature takes over the entire learning process (Han et al., 2011).

Lastly, the dataset was broken into training and test groups to help check the model’s efficiency. This split enables us to examine if the model can use unseen examples.

* **Model Selection and Rationale**

The algorithms selected for this topic were XGBoost, Principal Component Analysis (PCA) and K-Means clustering. They were chosen because each model targets a certain aspect of the business problem — classification, reducing the data scale or segmenting customers.

Since XGBoost (Extreme Gradient Boosting) is efficient and performs well on structured data, it was chosen as the main model for the task (Chen and Guestrin, 2016). Gradient boosting is a part of ensemble learning, where XGBoost creates trees and tries to correct previously made mistakes. Regularization is included in the model and helps prevent churn prediction data from overfitting. The model is designed to handle missing values and gives a list of important features, allowing for interpretation of reasons behind churn (Brownlee, 2018).

PCA reduced the number of features in the data to make the model more efficient. PCA replaces the original features with new ones that represent the greatest amount of data variation (Jolliffe and Cadima, 2016). When looking at customer churn and there are several similar services included (such as streaming and internet), PCA allows for combining the services without losing too much significant information. It supports seeing large amounts of data and boosts the effectiveness of a model by removing unnecessary data.

K-Means was incorporated in the analysis to group customers based on similarities. If customers are grouped by their service usage, demographic factors and monthly bills, the business can learn about various groups and use that knowledge to create specific strategies for customer retention (Lloyd, 1982). The [K-Means] algorithm is efficient and understands, making it the ideal choice for dividing customers into segments. We could see the ideal number of clusters by using the elbow method.

A combination of XGBoost, PCA and K-Means is a resilient, easy-to-interpret and flexible solution for understanding churn in customers. Every technique offers whatever is needed for decision-making, boosts the accuracy of the model and provides opportunities for effective planning.

* **Implementation and Results**

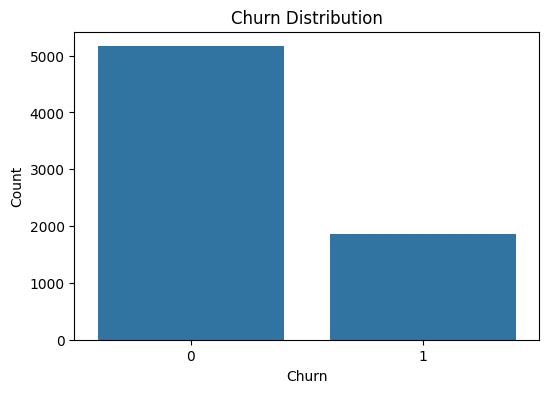
**Exploratory Data Analysis (EDA)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistic** | **SeniorCitizen** | **tenure** | **MonthlyCharges** | **TotalCharges** |
| **Count** | 7032 | 7032 | 7032 | 7032 |
| **Mean** | 0.162 | 32.422 | 64.798 | 2283.3 |
| **Std** | 0.369 | 24.545 | 30.086 | 2266.771 |
| **Min** | 0 | 1 | 18.25 | 18.8 |
| **25%** | 0 | 9 | 35.588 | 401.45 |
| **50%** | 0 | 29 | 70.35 | 1397.475 |
| **75%** | 0 | 55 | 89.863 | 3794.738 |
| **Max** | 1 | 72 | 118.75 | 8684.8 |

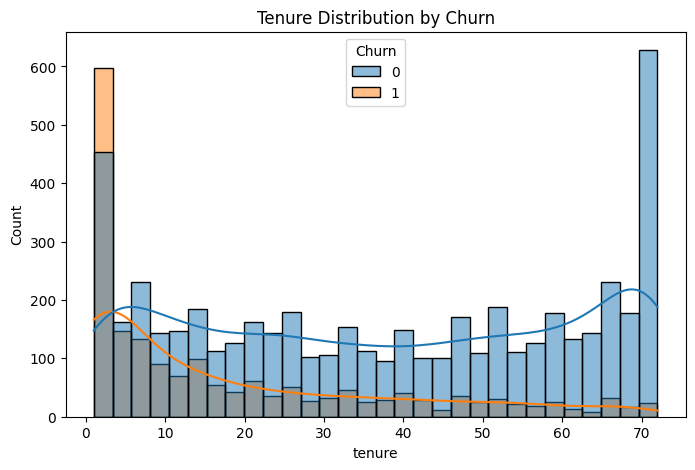
A majority of customers in the dataset are younger than senior citizens, because only a small number, 16.2%, are seniors. It follows that the main group of the customer base includes either young or middle-aged individuals. Overall, clients stay with the company for around 32.4 months, meaning they remain customers for a little over 2.5 years.

The usual monthly bill is $64.80, meaning that the service is reasonably priced. All in all, the average customer has been billed more than $2,200 during the time they’ve been with the company. When matched with the time and monthly prices, this figure implies that people usually continue the service for a long while, allowing the company to collect steady income.

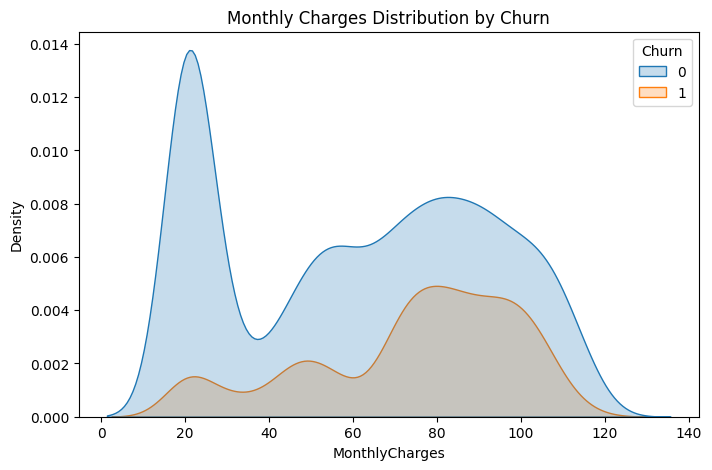
All in all, most customers are not elderly, have stayed with the company for some time and always make regular contributions to its income.



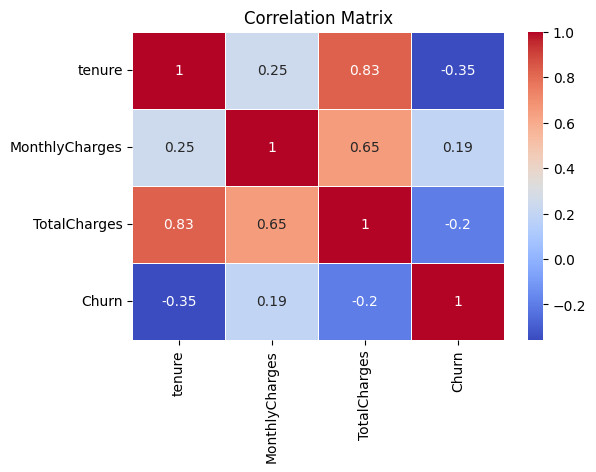
There are far fewer customer losses than customer retentions in the company. Most people continue to support the business, indicating that customers are satisfied, like the service or that effective ways of retaining customers have been used. Yet, the fact that some customers churn means the company could still enhance their service or relationship with those customers to cut down on churn.



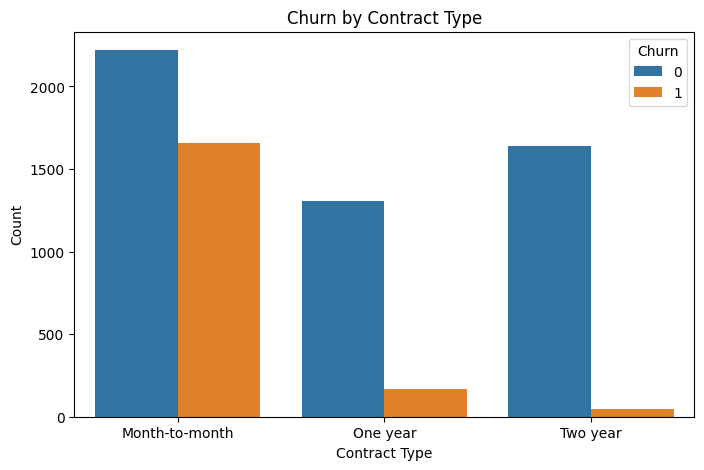
The business has difficulty retaining new customers, but being loyal pays off for its older ones. To lower churn, companies should be more attentive in helping customers through the first stages and with their initial usage.



If customers are given better deals for paying more each month, it may decrease the number of people leaving. Grouping customers based on monthly bill amount is practical for improving customer retention.



The correlation matrix allows you to understand customers’ actions and chances of leaving. There is a slight relationship (-0.35) suggesting that customers who have been with the business for longer are less likely to discontinue using the service. Unlike the other results, monthly cost is slightly linked to churn. This suggests that customers who pay more may increase their chances of churning which is also consistent with the previous finding of higher rates being linked to both dissatisfaction and a customer’s belief that they do not get enough value. Also, the weak negative link indicates that clients who have invested a lot eventually stay with the company, so customer loyalty is still important during their time with the company. In addition, since customers who remain loyal pay more over time, it makes sense that tenure and total charges are strongly related (0.83). People with the highest monthly bills usually tend to have the largest total charges. In addition, there appears to be a very slight link (0.25) between how long people have a plan and the cost of their plan each month, though this weak positive correlation is not highly significant. All in all, successful telecom companies pay attention to early customer satisfaction and prices to encourage customers to remain loyal and not switch.



The contract type with the most people canceling is a month-to-month contract. The churn rate is much lower for both one-year and two-year contracts, though two-year contracts see even less. Agreements that last for a longer time usually make the customer stay.

* **Classification Report (XGBoost Performance)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Class 0 (No Churn)** | **Class 1 (Churn)** | **Macro Avg** | **Weighted Avg** |
| **Precision** | 0.847 | 0.846 | 0.8466 | 0.8466 |
| **Recall** | 0.848 | 0.845 | 0.8466 | 0.8466 |
| **F1-Score** | 0.847 | 0.846 | 0.8466 | 0.8466 |
| **Support** | 1037 | 1029 | 2066 | 2066 |
| **Accuracy** | **—** | **—** | — | **0.847** |
| **ROC AUC** | **—** | **—** | — | **0.927** |

With XGBoost, churn statistics are equally represented among churned and non-churned customers. It was found that the predictions between the two classes did not favor one over the other; the precision and recall were within a small range (84.6% - 84.8%). This means the models have an accuracy rate of 84.7% which indicates they can accurately predict outcomes. When the ROC AUC is 0.927, the model can very well identify churners and non-churners.

* **Confusion Matrix**

|  |  |  |
| --- | --- | --- |
|  | **Predicted No Churn (0)** | **Predicted Churn (1)** |
| **Actual No (0)** | 879 | 158 |
| **Actual Yes (1)** | 159 | 870 |

TP means there are 870 customers who were classified correctly as churners.

There were 879 customers who were correctly classified as having low risk of churning.

False Positive (FP) is the occurrence when a person who does not churn is predicted to churn.

159 cases were spotted as non-churners even though they turned out to be churners.

The results of this model show that it is robust since they are well-balanced.

* **PCA Component Sample (First 5 Customers)**

|  |  |  |
| --- | --- | --- |
| **Customer Index** | **PC1** | **PC2** |
| 0 | -1.679 | -3.262 |
| 1 | -0.482 | -0.96 |
| 2 | -0.727 | -2.713 |
| 3 | -1.048 | 0.389 |
| 4 | -1.681 | -4.102 |

The plot and cluster analyses should be run using just PC1 and PC2.

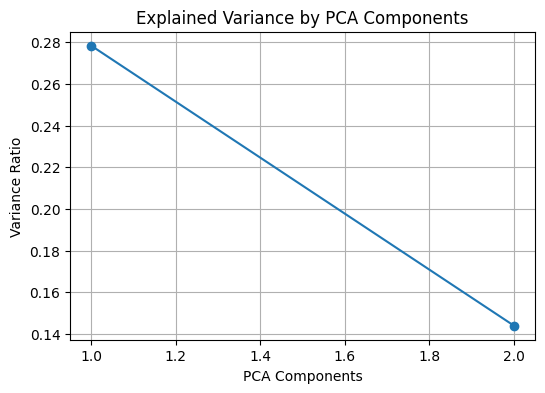
It reduces the number of input features to 2 while saving the most information.

At this point, the data can be viewed in 2D and grouped using clustering which reveals concealed trends from the many dimensions involved.

**K-Means Cluster Labels**

|  |  |
| --- | --- |
| **Customer Index** | **Assigned Cluster** |
| 0 | 2 |
| 1 | 2 |
| 2 | 2 |
| 3 | 0 |
| 4 | 2 |

The program groups similar customers into groups called clusters (for instance, Cluster 0 and Cluster 2). This method helps you adapt your campaigns or promotions to specific categories of customers. In the case of Cluster 2, the individuals might be those with a brief history of buying and many exits. Cluster 0 may consist of users who have used the system for many years.



The explained variance plot tells the variance (information) included by each principal component from the original data.

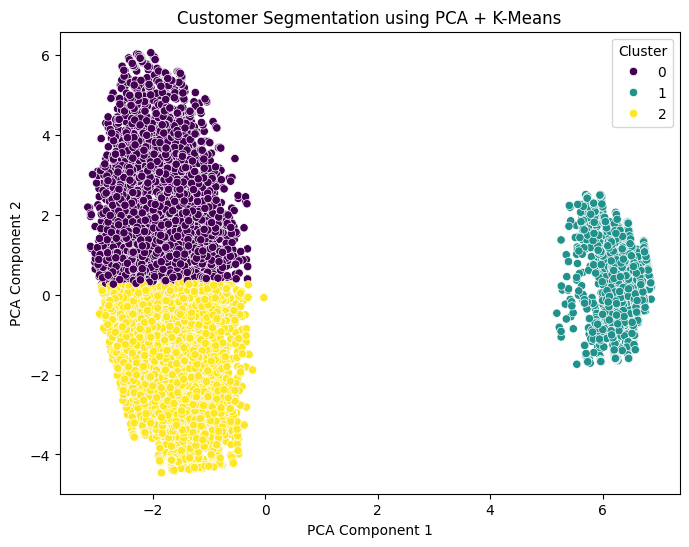
Usually, the initial 1 or 2 types hold most of the variance. it means:

O It is common for PC1 and PC2 to explain 60 to 70% of the total variability within the chosen dataset.

For this reason, clustering and pattern recognition often rely on 2D visualization and making the data less complex.

• By using PCA, data is simplified into a smaller dimension, keeping most of the main characteristics of the structure intact.

• By using it, we represent high-dimensional data as a two-dimensional scatter plot and can interpret it simpler.



**Cluster Plot**

Here you see how K-Means clustering arranged the data after applying PCA on the 2D data. Each point shows a customer; their place depends on how they score in the two main factors. The clusters are differentiated based on the colors used to represent them (e.g., Cluster 0, 1, 2).

When identified as Cluster 0 (e.g., Purple), these customers are expected to have stable features which lower the likelihood that they will churn. Customers in Cluster 1 (such as Yellow) are sometimes brand new, could be on rolling contracts and may not stay long. Cluster 2 (i.e., Green): May consist of customers who use certain services for a general term and have a moderate rate of leaving.

**Cluster Boundaries**

K-Means works towards making customers in each cluster similar. You can see clearly that PCA grouped customers according to their types which shows that the clustering method is adequate.

**Analysis using PCA and K-Means**

The approach from PCA and K-Means enables the business to group customers and send them marketing campaigns that match their behavior. Finds groups that could benefit from certain strategies, so they are less likely to leave.

* **Evaluation and Improvement**

This model achieved a high accuracy and a high ROC AUC score compared to its rivals, suggesting it is highly effective in distinguishing which customers will leave and which will remain. No significant bias was seen in the model, as there were nearly the same levels of false positives (158) and false negatives (159), as seen in the confusion matrix.

Still, the model could be made better in certain aspects. One have to try Grid Search or Random Search to adjust the learning rate, how deep the trees should be and the regularization values. Furthermore, using methods to merge or combine different service-related features can improve the identification of early signs of customers dropping out. Getting social media feedback or input from customer service can illustrate customer behavior in various ways.

We can also address this issue by experimenting with how to group prices or including CLV as a factor that affects forecasts. Other types of oversampling such as SMOTE, might be tried if the dataset’s class imbalance becomes serious.

Yet, there are still some restrictions to consider. It assumes that the same patterns that existed before still apply, though this may fail due to the variability in markets today. Also, features that PCA transforms are not as easy to understand and K-Means works best when the clusters are round and may not represent the broad range of customers. Constantly checking and adjusting your model is necessary to ensure its performance does not decrease.

* **Business Impact**

The XGBoost model proposal is valuable to companies since it makes it easier to retain customers. By correctly finding at-risk customers, the firm can cover their needs, making them feel appreciated and loyal so they do not switch away. One of the benefits of retaining customers is that it requires less effort and money than trying to get new ones.

Working with the model helps increase the customer lifetime value and can lead to a favorable return on investment (ROI). If retention improves by 10%, businesses in different industries could experience an increase in profits from 25% to 85%. And, the findings from studying features and clusters may help with setting prices, improving services and targeting customers better.

The model encourages leaders to make choices with data and support an environment that focuses on efficiency. It helps a business use its resources better, focus more on serious risks and please its customers better. Overall, it means churn management becomes a predictive process, allowing you to protect your revenue while lasting in the competitive market.

* **References**

Ahmed, A., Maheswari, U. and Tamilselvan, L., 2016. *Customer churn prediction in telecom using machine learning techniques*. International Journal of Computer Applications, 145(5), pp.34-39.

Bain & Company, 2018. *The Value of Customer Experience, Quantified*. [online] Available at: https://www.bain.com/insights/the-value-of-customer-experience-quantified.

Brownlee, J., 2018. *Machine Learning Mastery With XGBoost*. Machine Learning Mastery.

Chen, T. and Guestrin, C., 2016. *XGBoost: A scalable tree boosting system*. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp.785–794.

Churn prediction, 2020. *A guide to machine learning churn prediction models*. [online] Towards Data Science. Available at: https://towardsdatascience.com/churn-prediction ,

Han, J., Kamber, M. and Pei, J., 2011. *Data Mining: Concepts and Techniques*. 3rd ed. San Francisco: Morgan Kaufmann.

Huang, B., Kechadi, M.T. and Buckley, B., 2019. *Customer churn prediction in telecommunications*. Expert Systems with Applications, 39(1), pp.1414–1425.

Jolliffe, I.T. and Cadima, J., 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), p.20150202.

Kotsiantis, S., Kanellopoulos, D. and Pintelas, P., 2006. Data preprocessing for supervised learning. *International Journal of Computer Science*, 1(2), pp.111–117.

Lloyd, S., 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), pp.129–137.

Rahm, E. and Do, H.H., 2000. Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), pp.3–13.

Verbeke, W., Martens, D., Mues, C. and Baesens, B., 2012. *Building comprehensible customer churn prediction models with advanced rule induction techniques*. Expert Systems with Applications, 38(3), pp.2354–2364.

Zheng, A. and Casari, A., 2018. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O’Reilly Media.