Analysing Internet Service Provider Customer Churn Using Classification Techniques

Sam Tebbet

December 5, 2023

1 Introduction

An Internet Service Provider (ISP) is a service that provides internet connections to its customers, it can also provide subscriptions to mobile and television services, among others [1]. Customer churn refers to the movement of customers from one service to another. It is the goal of a business to minimise the churn for their service and retain as many customers as possible [2].

1.1 Dataset

The dataset contains 72274 rows of customers of the ISP, each with fields indicating TV subscription status, movie package subscription status, subscription age, average monthly bill, remaining contract duration, service failure count, average download speed, average upload speed, and incidents of download over the limit. The data is labelled with the field "churn" with binary values indicating whether or not the customer has churned, with 1 indicating a customer cancelled their service.

1.2 Research Question

To what extent can we accurately predict customer churn of an ISP based on usage patterns and subscription characteristics using machine learning algorithms? The goal of this research is to develop a model that can accurately identify customers at risk of switching services based on their usage patterns and subscription characteristics. Using machine learning algorithms, we aim to uncover patterns and relationships within the dataset that can help the Internet Service Provider proactively address potential issues, optimize customer retention strategies, and reduce churn.

2 Data Analysis

2.1 Feature Selection

Feature selection can be used to reduce training time and can even improve the performance of your model [3].

When first plotting the data, we noticed some data points were very distinct outliers, particularly in the upload and download average fields. To mitigate this we calculated the Z-score for the data and dropped the rows that were more than 3 standard deviations away from the mean. This ensures the model would not be trained on anomalous data.

Using unsupervised learning techniques we can estimate which features contain the most information. We used the K-nearest neighbour method using mutual information classification to select the most important features in the dataset [4]. The results of the feature selection show that the remaining contract of the customer is the most important feature concerning the churn rate, followed by the download and upload average, and subscription age. See Figure 5(b) for diagrams showing the relation of these features against churn.

2.2 Decision Trees

We will be using classification and regression trees (CART), a form of decision tree which uses Gini impurity values to identify the feature to split on [5]. The research question promotes the use of classification techniques and the binary format of the data labels indicates that decision trees would be the best option for making a predictive analysis of customers. One limitation of decision trees is that they are very prone to over-fitting as they tend to fit perfectly into the training dataset. To mitigate this, the tree will need to be pruned to better fit the test data.

2.3 Pruning

Feature selection is one way of pruning the decision tree, however, the model is still prone to over-fitting as the training dataset has only been made smaller. To combat this, you can prune the decision tree after fitting the model.

Post-pruning involves minimising the expected error rate of the model after it has been fit [6]. Multiple methods exist but we will focus on cost-complexity pruning (CCP). The CCP algorithm "penalizes the estimated error based on the subtree size" [6]. A variable α increases the number of nodes pruned, iterating over varying levels of α and fitting decision trees, we can find the optimal α that maximises the accuracy of the test data [6]. Furthermore, pruning the tree makes it significantly smaller, making it far easier for humans to understand the decision process of the model, for a visual comparison see Figure 2 for an un-pruned tree and Figure 1(a) for the pruned decision tree, trained on the data.

2.4 Overview and Analysis of Results

Using Jaccard accuracy, values from the confusion matrix, and F1 score, we calculated the accuracy of the model, see Figure 1(a). After analysing the accuracy vs α to minimise error rate, the accuracy of the model on test data increased by an average of 2.65%. Figure 3(b) shows that an α of approximately 0.00015 pruned the tree optimally. The confusion matrices in Figure 1(c) show the number of true positives increasing and the number of false positives in the test data decreasing significantly after pruning. Although the number of false negatives increased and false positives decreased, the accuracy of the model on the test data improved.

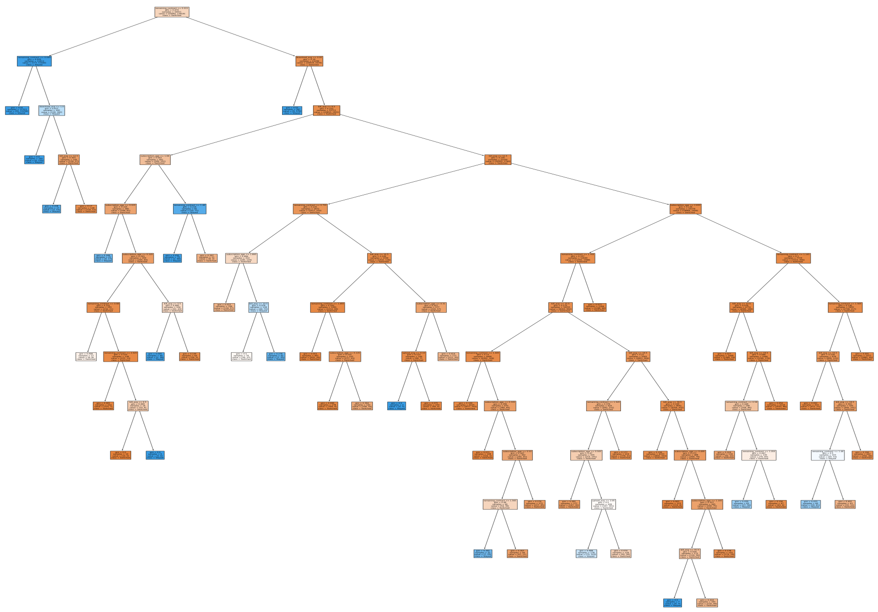
3 Limitations

One key drawback is that the dataset does not have very many features to analyse, especially after the feature selection process where there are only 5 informative features. This makes the model very specialised when trained on this limited data. Furthermore, if that data is not present for some customers, it may be difficult to predict if the customer will churn. This is especially the case with this dataset, with the remaining contract length being the most important feature, there are 21572 missing values for this feature which were dropped to train the model, see Figure 5(a).

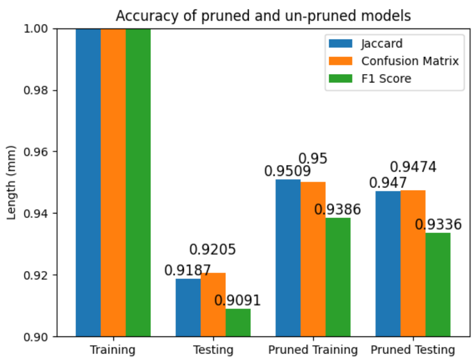
Another key drawback of the dataset is that it only considers the customers' actions. It does not account for any business scandals in the company, or what competitors are doing that might make the customer cancel their service. Other factors including the customer's personal life could also have an effect.

4 Conclusion

We can accurately predict the customer churn of an ISP based on usage patterns and subscription characteristics using machine learning algorithms. The research suggests that the feature with the largest effect on customer churn was the remaining contract length, knowing this the ISP can implement business strategies to prevent customers from churning, especially when closer to their renewal dates. After successfully predicting customer churn the Internet Service Provider can implement targeted interventions for each customer, such as personalized offers to prevent customers from leaving. This research has the potential to enhance customer satisfaction, reduce revenue loss, and contribute to the overall business success of the ISP.



(a) Pruned Decision Tree



(b) Accuracy of the un-pruned and pruned model on the testing and training data

(c) Confusion matrices comparing the un-pruned and pruned trees

Figure 1: Pruning the Decision Tree

References

[1] The Editors Of Encyclopaedia Britannica. "Internet service provider | Benefits, Types & Services". In: *Encyclopedia Britannica* (Nov. 2023). URL: https://www.britannica.com/technology/Internet-service-provider.

[2] Shin-Yuan Hung, David C. Yen, and Hsiu-Yu Wang. "Applying data mining to telecom churn management". In: *Expert Syst. Appl.* 31.3 (Oct. 2006), pp. 515–524. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2005.09.080.

[3] Isabelle Guyon and. "An introduction to variable and feature selection". In: *Journal of machine learning research* 3.Mar (2003), pp. 1157–1182.

[4] *Mutual Information between Discrete and Continuous Data Sets - ProQuest.* [Online; accessed 4. Dec. 2023]. Dec. 2023. URL: https://www.proquest.com/docview/1500409249?pq-origsite=primo.

[5] *What is a Decision Tree | IBM.* [Online; accessed 3. Dec. 2023]. Dec. 2023. URL: https://www.ibm.com/topics/decision-trees.

[6] Jeffrey P. Bradford et al. "Pruning decision trees with misclassification costs". In: *Machine Learning: ECML-98.* Berlin, Germany: Springer, 1998, pp. 131–136. ISBN: 978-3-540-69781-7. DOI: 10.1007/BFb0026682.

5 Appendix

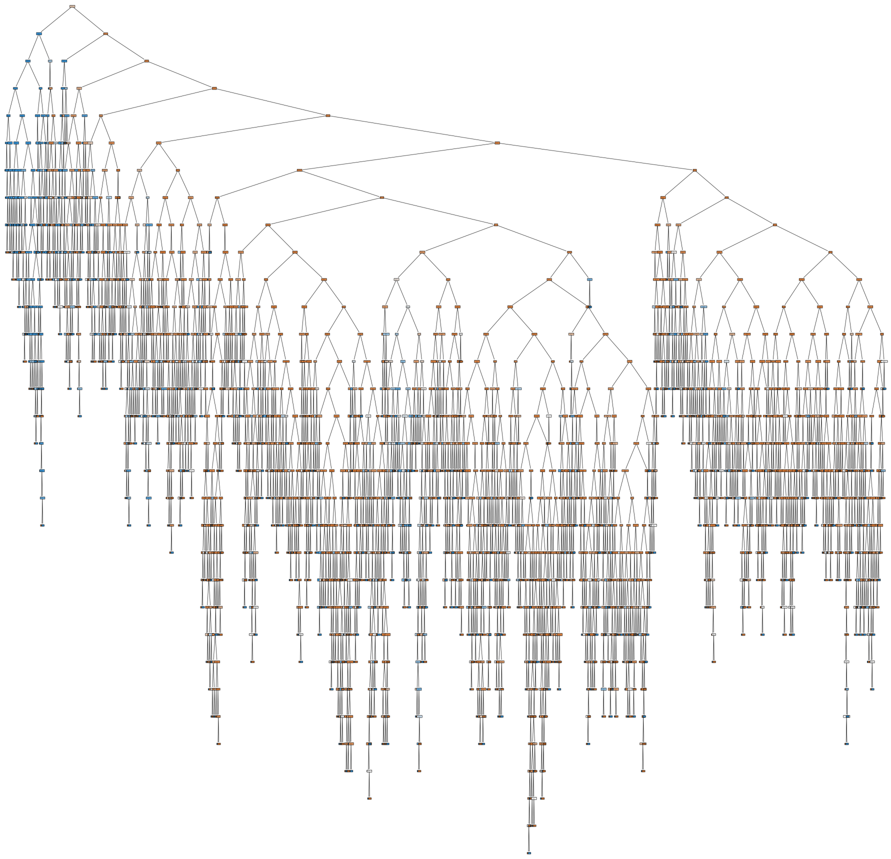
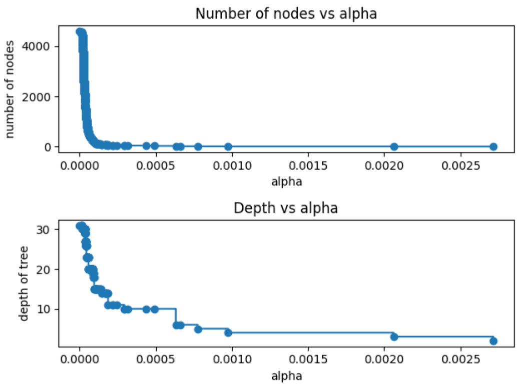
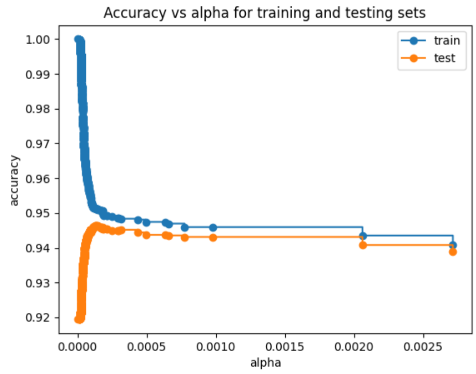


Figure 2: Unpruned Decision Tree

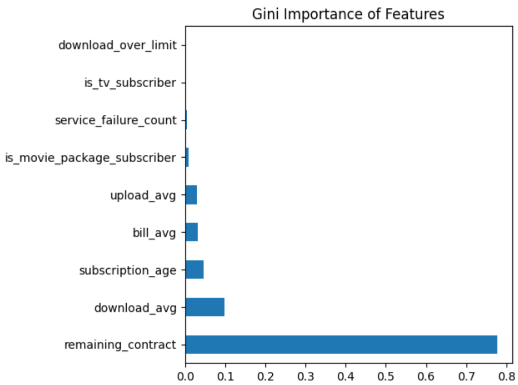


(a) Alpha vs depth and number of nodes in the tree

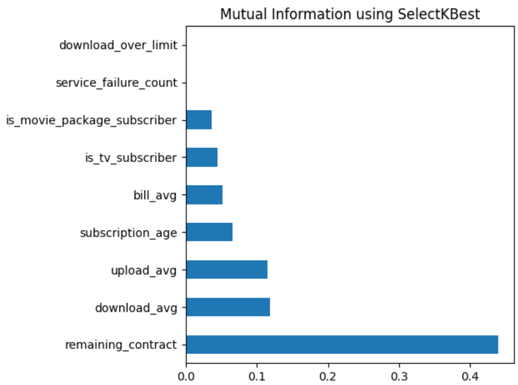


(b) Alpha vs Accuracy of the decision trees

Figure 3: Analysing the alpha variable

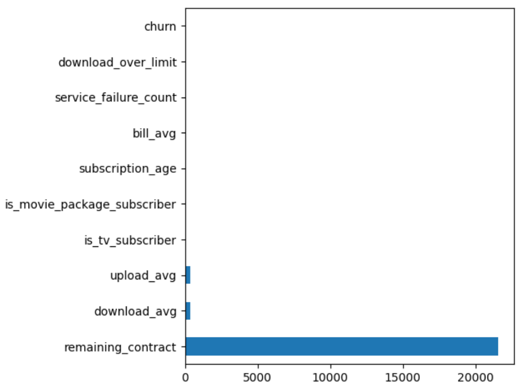


(a) Gini importance of features relative to churn

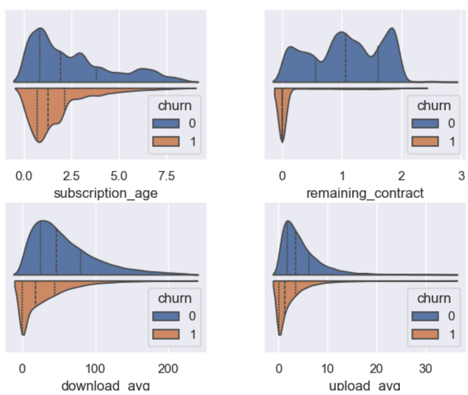


(b) Mutual information classification of features

Figure 4: Analysing Features using Machine Learning Techniques



(a) Number of null values for each field in the dataset



(b) EDA of features in relation to churn

Figure 5: Analysing the Dataset