Project Presentation: Predict Customers Churn

Slide 1: Topic

Have you ever had bad days of internet or phone service? You must have felt need to leave such Telcom Company and go for the better one!

But company do not want that!
And it even be the costliest decision for them!

Slide 2: Introduction

Predicting *customer churn* is an important piece of information across all industries because businesses are not only concerned with attracting new customers but also with retaining current customers. If a business can better understand what factors may cause a person to be more likely to leave, then the business can adjust its marketing and corporate strategy in hopes of lowering their customer churn. The telecommunications industry in particular is especially concerned with decreasing customer churn since, in most countries, a customer can choose from several different companies and the process of switching telecom providers is fairly easy.

We will apply two data analytic models of Decision Tree and Logistic Regression on telecom data to predict whether or not a customer will leave Telco, a fictional telecom company.

Slide 3: Data Exploration

Our dataset contains customer information for a fictional telecom company and we found it on Kaggle. We are interested in predicting the churn value (either “Yes” or “No”) for a particular customer. The churn value indicates whether a particular customer has left the company within the last month (resulting in a “Yes”) or remains a customer at the company (resulting in a “No”). The raw data contains 7,043 rows with one row per customer and 21 features.

To get a better understanding of the data set, we first looked at how the output, i.e. whether an employee leaves the company or not, is distributed. As you can see in the pie chart, the output is very unevenly distributed (74% No, 26% Yes).

Slide 4: Data Exploration – numeric features

Next we looked at the numerical features. These are tenure, total charges and monthly charges. When we look at the correlation matrix, we notice that there is a high correlation between total charges and tenure, a medium correlation between total and monthly charges and hardly any low correlation between tenure and monthly charges. If we look at the graphs for the individual numerical features divided into churn = yes and churn = no, it becomes clear that customers usually cancel the service in the first few months.

Slide 5: *Data Exploration – numeric features*

Then the categorical features were examined. There are 19 of these, such as gender, various extra services, contract length and so on. We came to the following observations:

1. Customers with Online Security, Device Protection, Online Backup and Tech Support service are less likely to leave
2. Increase in contact length leads to decrease in churn rate
3. Demographic effects churn rate
4. Customers in relationship have lower churn rates than singles

Slide 6: Clean and Prepare Data

Now we want to clean and prepare the data. After the exploration we knew that the data is almost ready for further processing. Only minor changes had to be made. Rows in which no TotalCharges were specified have been deleted. We also decided that it would make more sense to convert the column "tenure", which was a numerical feature, into a categorical feature, since this will later affect a better model. Here we have chosen the following categories: 0-12; 12-24; 24-48; 48-60; >60. In the graph you can see how the column churn behaves for each category.

Slide 7: Modelling – Preparation

Before we can start building the model, we have to do some things to make sure that the model can process the data correctly. We plan to create a model using Logistic Regression and a model using Decision Tree. First we split the data into a training set and a testing set. Here we have decided to use 70% for training and 30% for testing, as is usual. We set the class-weight to balanced, because as we have seen in exploration, the values for churn are distributed differently and equally. This ensures that in the training and testing set there are equally distributed people with churn = yes and churn = no. After constructing our training and test datasets, we transformed numerical and category features respectively. Standard scaling is used on numerical features to rescale the and One Hot Encoding is used on the categorical features to convert them into a form that could be provided to ML algorithms. The reason why we didn't use K-Nearest Neighbor and Naive Bayes as a model is that we have too many features and have these dependencies on each other, as we noticed before. We use GridSearchSV to get the best hyper parameters for our models.

Slide 8: Modelling – Decision Tree (1)

Now we can start creating a Decision Tree model. Decision Tree algorithm is a supervised learning algorithm. The main aim of this algorithm is to predict the class or the target feature by learning simple decision rules inferred from prior data. A basic structure of a decision tree has root nodes, decision nodes and leaves nodes. Nodes are split based on providing the highest information gain for a given criteria. Our model uses gini Index as the

criteria. Information gain can be loosely defined as the difference between Information before and after the parent node split. Gini Index is the number of samples that belong to a specific class for a particular node. The aim is to get Gini to zero for the leaf nodes so they are pure leading to no more information gain.

Through GridSearchSV we knew for which parameters we should not use the default values.

Slide 9: Modelling – Decision Tree (2)

On the left side you can see the finished tree. This is the post-pruned variant, so all the children of a parent node were deleted if all children nodes have the same classification.

In order to read the decision tree:

- Left Child means the condition is True and vice versa.
- For categorical features, less than and equal to 0.5 means 0 (No) and vice versa.
- Numerical features are standardized and hence they can be from 0 to 1.
- Class refers to whether the leaf is going towards the Churn value of Yes or No.

After plotting the feature importance, we found out that the five most important features are Contract, Monthly Charges, Total Charges, Tenure and Tech Support. Hence, these features are very important for Decision Tree to split the observation into Churn Yes or No.

On the bottom right you can see the performance of this model. It has an accuracy of 0.755 and a F1 score of 0.715. This is what we want to surpass with the next model

Slide 10: Modelling – Logistic Regression

In the Logistic Regression model, the dependent variable assumes two values: zero or one. Thus, this model is used for predictions in which there are only two possible outcomes, which makes the model suitable for our task of predicting whether a customer has churned. Logistic Regression uses the logistic function with the features we are using. With the function, one should note that the result will be a number between 1 and 0; every time we obtain a feature greater than 0.5 we classify it as 1 and if it is lower we classify it as 0.

We used again GridSearchSV to hypertune the parameter.

Based on the coefficients, the four most important features that push the model towards Churn value of Yes are Monthly Charges (0.87), Tenure of less than year (0.83) ,Monthly Contract (0.79) and Electronic Check Payment Method (0.32). Similarly, the four most important features that push the model towards Churn value of No are Two Years Contract (-0.78), Tenure of greater than 60 months (-0.38) and having TechSupport (-0.24).

You can see the Performance on of the Logistic Regression Model at the table. It has an accuracy of 0.752 and a F1 score of 0.719 for the testing set.

Slide 11: Compare Results

If we compare the two results of the models, we see that Logistic Regression has a slightly better F1 score. Another measure that we employed is the Receiver Operating Characteristics (ROC). The ROC curve is obtained by changing the threshold value and evaluating the combination of false positive rate and true positive rate for each threshold.

One measure that we can use from the ROC curve is the area under the curve (AUC). While the ROC compares different thresholds, the AUC gives us an average of performance for various thresholds. Logistic Regression does a better job of predicting Churn status based on the ROC curve as you can see on the diagram.

We find Logistic Regression to be the best model with its high Macro f1-score, which serves as a proxy for the actual performance of the classifier, as well as a high AUC value, which tells how well the two kinds of responses can be separated. As our data contains more overlapping regimes then Logistic Regression should perform better as this leads to overfitting of Decision Tree**.**

Slide 12: Result

Based on our best model, we made different conclusions: We found out that price is the biggest reason to switch and higher prices pushes customer to find competitor with the cheapest service. The more customers have spent on the company, the more likely they are not churn. Customers are also learning to use the product correctly, so they will churn at the beginning if they don't like it. But if they use it long enough and get familiar with it, they are more likely to commit. Customers with a one-month contract very often have the possibility to terminate and more often ask themselves the question whether it is worthwhile and generally deal with it more often, so they are more likely to churn.

Slide 13: Conclusion

With the help of data science processes, new knowledge has been gained that can be used in the future to improve the service and to target and convince high-risk customers to remain loyal to the provider, thereby saving a considerable amount of money.