Customer Churn Rate Prediction

Using Deep Learning with Keras

Mukil Raj Kannan

Contents

About the data:

A telecommunications company is concerned about the number of customers leaving their landline business for cable competitors. They need to understand who is leaving. Imagine that you're an analyst at this company and you have to find out who is leaving and why. I am using Telco Customer Churn dataset by IBM scientists to Predict behavior to retain customers: https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2017/06/19/guide-to-ibm-cognos-analytics-sample-data-sets

The dataset includes information about: Customers who left within the last month – the column is called Churn
Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
Demographic info about customers – gender, age range, and if they have partners and dependents

Customer attrition

Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of clients or customers.

Telephone service companies, Internet service providers, pay TV companies, insurance firms, and alarm monitoring services, often use customer attrition analysis and customer attrition rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Companies from these sectors often have customer service branches which attempt to win back defecting clients, because recovered long-term customers can be worth much more to a company than newly recruited clients.

Companies usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another company or service provider, involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care facility, death, or the relocation to a distant location. In most applications, involuntary reasons for churn are excluded from the analytical models. Analysts tend to concentrate on voluntary churn, because it typically occurs due to
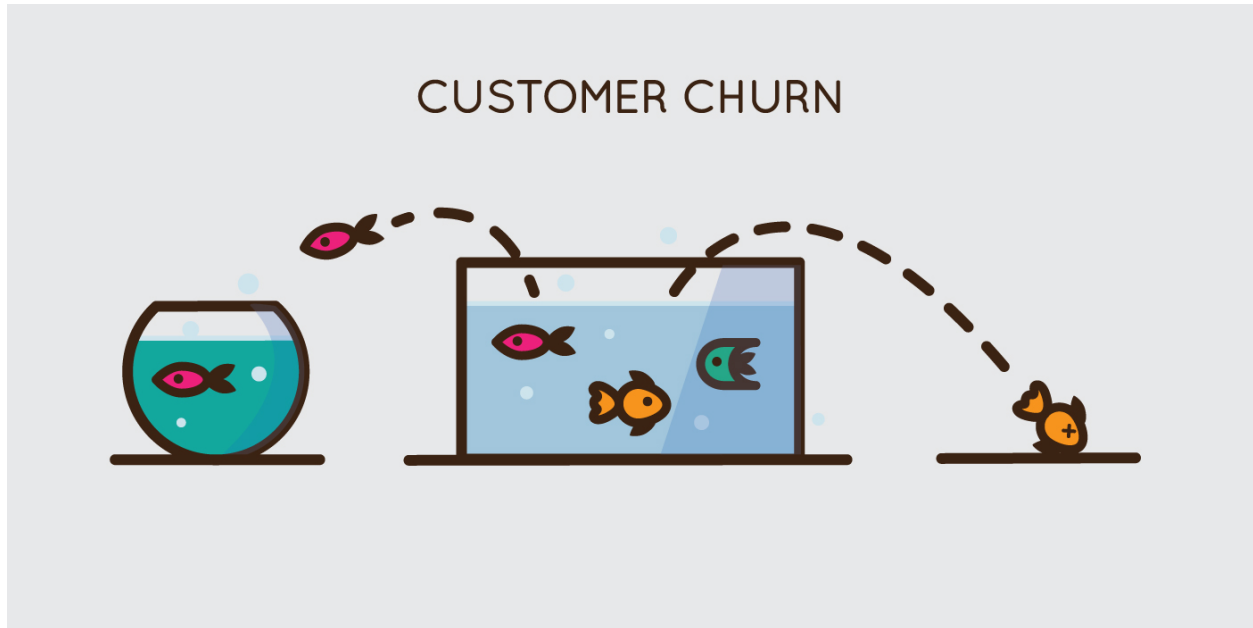
Figure 1: Customer Churn

factors of the company-customer relationship which companies control, such as how billing interactions are handled or how after-sales help is provided.

Predictive analytics use churn prediction models that predict customer churn by assessing their propensity of risk to churn. Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn.

The Importance of Predicting Customer Churn The ability to predict that a customer is at a high risk of churning, while there is still time to do something about it, represents a huge additional potential revenue source for every online business. Besides the direct loss of revenue that results from a customer abandoning the business, the costs of initially acquiring that customer may not have already been covered by the customer's spending to date. (In other words, acquiring that customer may have been a losing investment.) Furthermore, it is always more difficult and expensive to acquire a new customer than it is to retain a current paying customer.

Customer retention helps increase the profitability of your business in a number of ways that you might not realize. Here are five reasons why customer retention is key to your business' success.

1. Save Money on Marketing As we mentioned above, it costs significantly more to acquire a new customer than it does to retain an existing one. So, save your money and reduce your marketing expenses by keeping your old customers who are already familiar with your products and services. Familiarization with your products and services also means your business needs to spend less time on customer support.

2. Repeat Purchases from Repeat Customers Means Repeat Profit Loyal customers will use your business regularly for purchases and tend to spend more money. Existing customers are 3 to 10x more likely to buy than a cold lead. They are 50% more likely to buy new products and spend 33% more than new customers. A valued customer trusts your business and believes that you offer a superior service compared to competitors. This customer believes that your company listens to their needs and requests — so it is important that you do so! Pay attention to which brands, products and purchases this customer prefers, as they are more likely to make additional purchases at your business.

3. Free Word-Of-Mouth Advertising We've said it before, and we'll say it again: word-of-mouth is the most cost-effective advertising you can have and only comes from your loyal, happy customers. Repeat customers are more likely to tell their friends and family about your business and its products, and customers respect the opinion of those close to them. Customers are happy to tell people about excellent service they received or a product that they enjoyed. Just look at some of the numbers:

- 49% of U.S. consumers say friends and family are their top sources of brand awareness.
- People who are referred by a friend are 4x more likely to buy from a business, and one offline word of mouth impression drives sales at least 5x more than a paid impression.
- 92% of people trust recommendations from family and friends more than all other forms of marketing.A successfully retained customer is much more likely to refer other customers. These referees cost less to acquire and have a higher lifetime value than customers gained from other ways.

4. Retained Customers Will Provide Valuable Feedback Customers that you retain provide valuable feedback, and it's important that you listen. 97% of consumers said they are somewhat likely to become more loyal to a company that implements their feedback, while 55% of consumers said they are not likely to continue being a customer of a company that ignores their feedback. Customers who make frequent purchases from your business will know which areas of your business could be improved. How can you improve your business if you are only dealing with new customers? Ask your repeat customers how your business can serve them better. This will lead to new opportunities that you may have overlooked, and lead to increased retention rates and sales.

5. Previous Customers Will Pay Premium Prices Long-time, loyal customers are far less price-conscious than new customers because they value your company already and, thus, are willing to pay the price for your services. Many customers associate higher prices with quality service and retained customers trust that your company can deliver this quality over competitors.

## Data Preprocessing And Exploration

Let's go through the steps to preprocess the data for ML. First, "prune" the data, which is nothing more than removing unnecessary columns and rows. Then I splited the data into training and testing sets. After that, exploration the training set to uncover transformations that will be needed for deep learning.

Remove NULL

The data has 11 NA values all in the "TotalCharges" column. Because it's such a small percentage of the total population (99.8% complete cases), we can drop these observations with the drop_na() function from tidyr. Note that these may be customers that have not yet been charged, and therefore an alternative is to replace with zero or -99 to segregate this population from the rest.
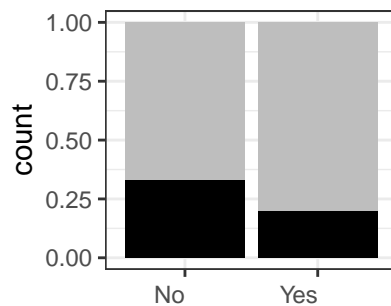
Exploration

Exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.
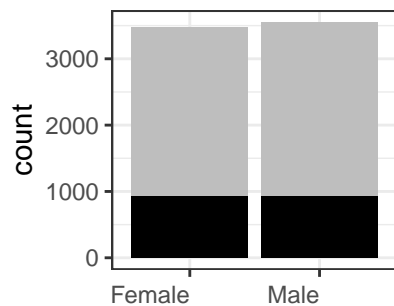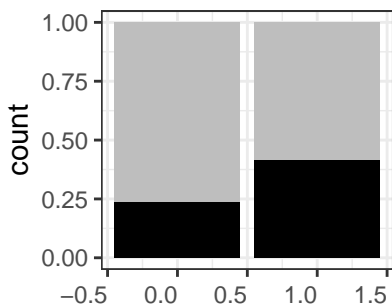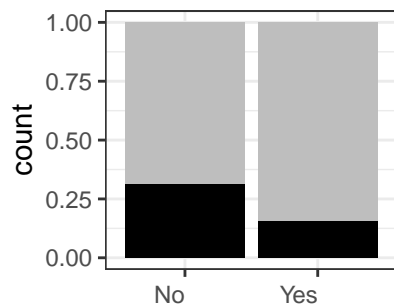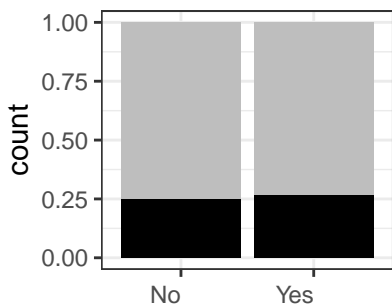
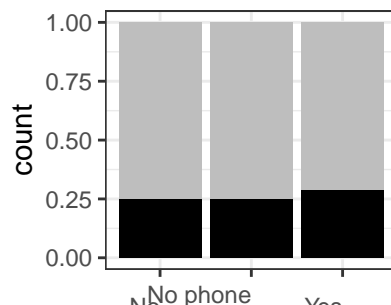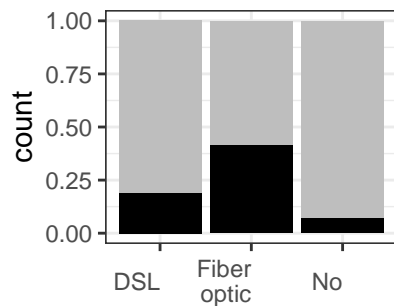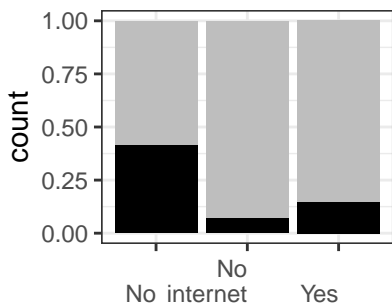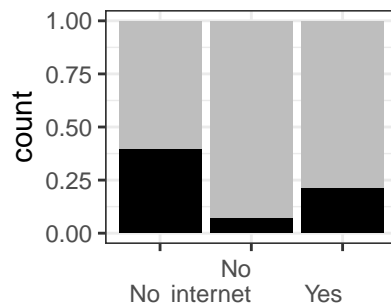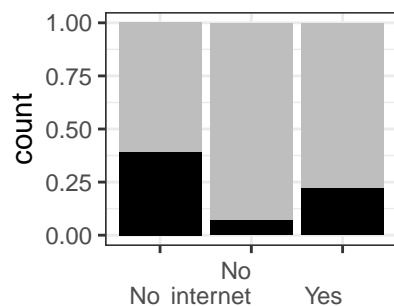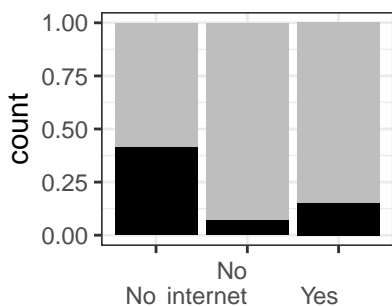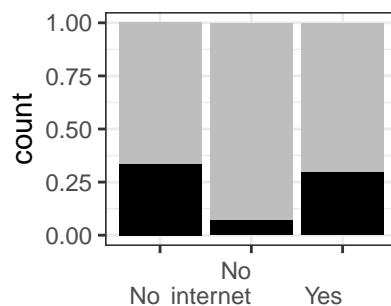- Churn columns tells us about the number of Customers who left within the last month. Around 26% of customers left the platform within the last month.

- Gender - The churn percent is almost equal in case of Male and Females.

- The percent of churn is higher in case of senior citizens.

- Customers with Partners and Dependents have lower churn rate as compared to those who don't have partners & Dependents.
- Churn rate is much higher in case of Fiber Optic InternetServices.

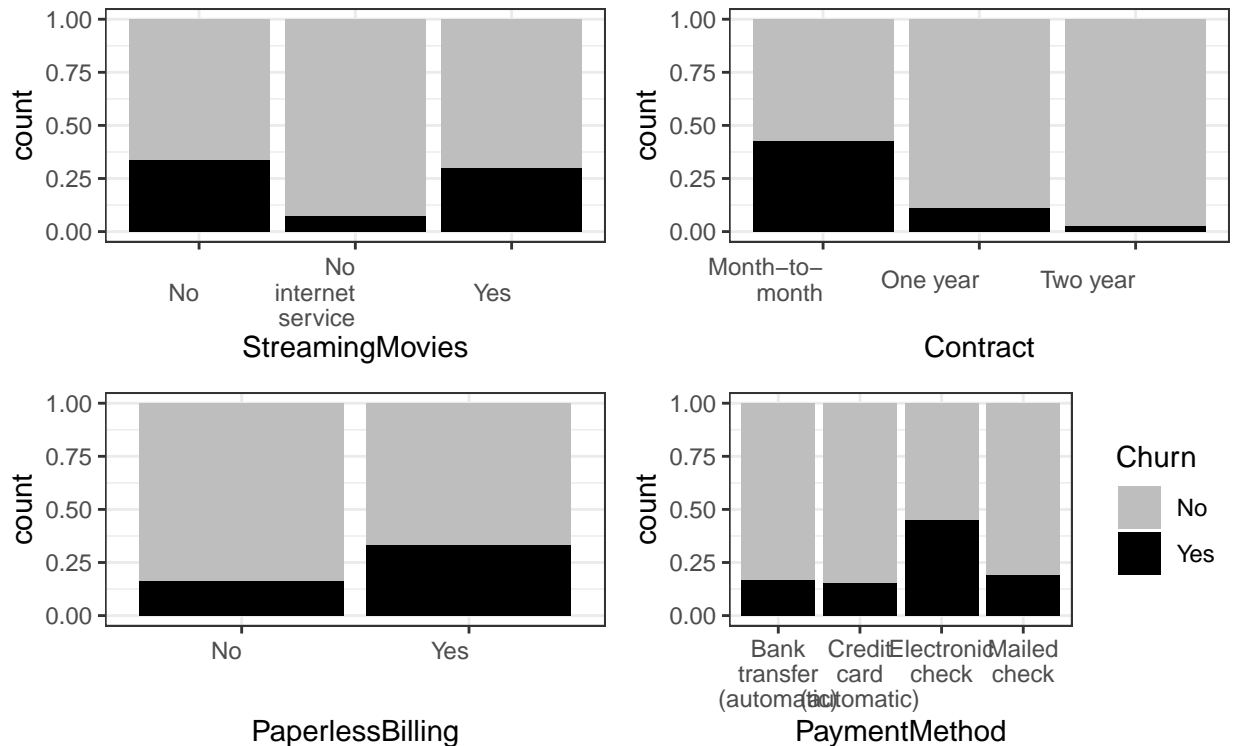- Customers who do not have services like No OnlineSecurity , OnlineBackup and TechSupport have left the platform in the past month.

- A larger percent of Customers with monthly subscription have left when compared to Customers with one or two year contract.

- Churn percent is higher in case of cutsomers having paperless billing option.

- Customers who have ElectronicCheck PaymentMethod tend to leave the platform more when compared to other options.

Data_Partition

Split the data with training and testing sets using sample.split. Training set will have 80% of dataset and testing set will have 20%.

Data Transformation

Artificial Neural Networks are best when the data is one-hot encoded, scaled and centered.Transformations that have been done for the data set are as follows:

DISCRETIZE THE "TENURE" FEATURE: The "tenure" feature falls into this category of numeric features that can be discretized into groups.

TRANSFORM THE "TOTALCHARGES" FEATURE: The "Total Charges" variable is skewed, We can use a log transformation to even out the data into more of a normal distribution.

ONE-HOT ENCODING: One-hot encoding is the process of converting categorical data to sparse data, which has columns of only zeros and ones (this is also called creating "dummy variables" or a "design matrix"). All non-numeric data will need to be converted to dummy variables. This is simple for binary Yes/No data because we can simply convert to 1's and 0's. It becomes slightly more complicated with multiple categories, which requires creating new columns of 1's and 0's for each category (actually one less). We have four features that are multi-category: Contract, Internet Service, Multiple Lines, and Payment Method.

FEATURE SCALING: ANN's typically perform faster and often times with higher accuracy when the features are scaled and/or normalized (aka centered and scaled, also known as standardizing).

Preprocessing with Recipes

A "recipe" is nothing more than a series of steps you would like to perform on the training, testing and/or validation sets.

For our model, we use:

1. step_discretize() with the option = list(cuts = 6) to cut the continuous variable for "tenure" (number of years as a customer) to group customers into cohorts.
2. step_log() to log transform "TotalCharges".
3. step_dummy() to one-hot encode the categorical data. Note that this adds columns of one/zero for categorical data with three or more categories.
4. step_center() to mean-center the data.
5. step_scale() to scale the data.

Building A Deep Learning Model

Build a three-layer MLP with keras.

Initialize a sequential model: The first step is to initialize a sequential model with keras_model_sequential(), which is the beginning of our Keras model. The sequential model is composed of a linear stack of layers. Apply layers to the sequential model: Layers consist of the input layer, hidden layers and an output layer. The input layer is the data and provided it's formatted correctly there's nothing more to discuss. The hidden layers and output layers are what controls the ANN inner workings.

1. Hidden Layers: Hidden layers form the neural network nodes that enable non-linear activation using weights. The hidden layers are created using layer_dense(). We'll add two hidden layers. We'll apply units = 16, which is the number of nodes. We'll select kernel_initializer = "uniform" and activation = "relu" for both layers. The first layer needs to have the input_shape = 35, which is the number of columns in the training set. Key Point: While we are arbitrarily selecting the number of hidden layers, units, kernel initializers and activation functions, these parameters can be optimized through a process called hyperparameter tuning that is discussed in Next Steps.
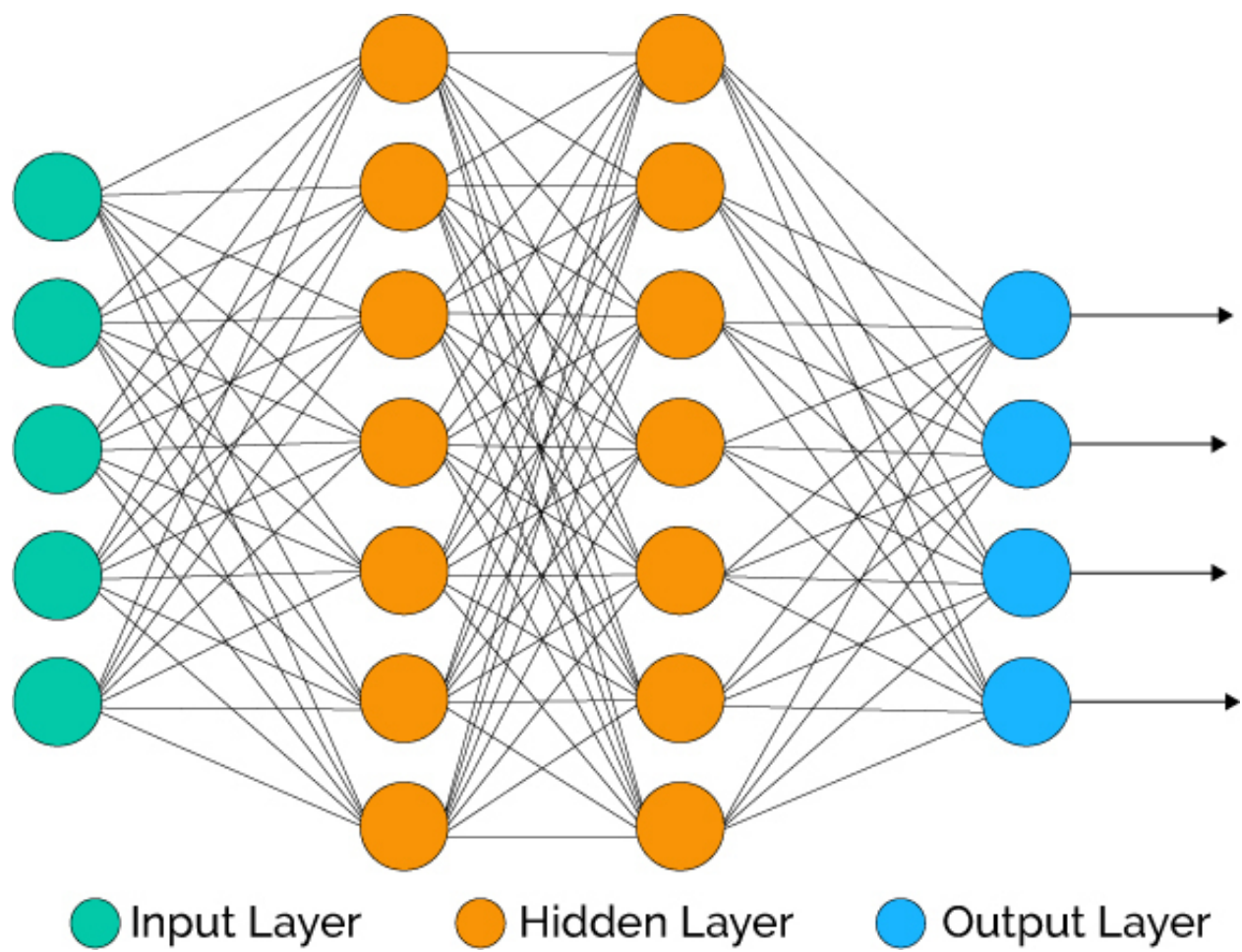
Figure 2: Alt Text

2. Dropout Layers: Dropout layers are used to control overfitting. This eliminates weights below a cutoff threshold to prevent low weights from overfitting the layers. We use the layer_dropout() function add two drop out layers with rate = 0.10 to remove weights below 10%. Output Layer: The output layer specifies the shape of the output and the method of assimilating the learned information. The output layer is applied using the layer_dense(). For binary values, the shape should be units = 1. For multi-classification, the units should correspond to the number of classes. We set the kernel_initializer = "uniform" and the activation = "sigmoid" (common for binary classification).

3. Compile the model: The last step is to compile the model with compile(). We'll use optimizer = "adam", which is one of the most popular optimization algorithms. We select loss = "binary_crossentropy" since this is a binary classification problem. We'll select metrics = c("accuracy") to be evaluated during training and testing. Key Point: The optimizer is often included in the tuning process.

Visualize the Keras training history using the plot() function.Validation accuracy and loss leveling off, which means the model has completed training. There is some divergence between training loss/accuracy and validation loss/accuracy. The model indicates that training can be stopped at an earlier epoch. Use enough epochs to get a high validation accuracy. Once validation accuracy curve begins to flatten or decrease, training should be stopped.
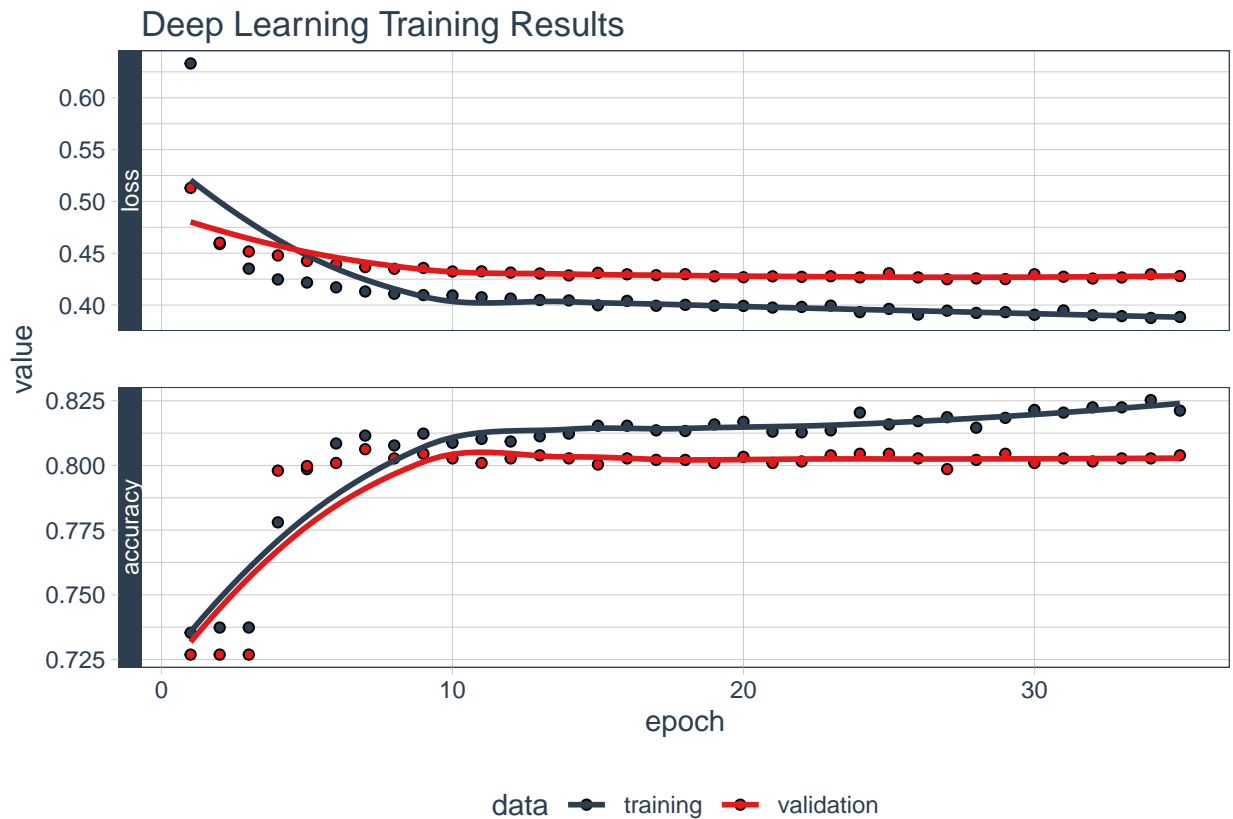
Prediction

Predictions can be made from keras model on the test data set, which was unseen during modeling. Two functions to generate predictions:

predict_classes(): Generates class values as a matrix of ones and zeros. Converted the output to a vector as I am dealing with binary classification. predict_proba(): Generates the class probabilities as a numeric

matrix indicating the probability of being a class. Converted to a numeric vector because there is only one column output.

Model Metrics

- CONFUSION TABLE: Used the conf_mat() function to get the confusion table.
- ACCURACY: Used the metrics() function to get an accuracy from the test set.
- AUC: Calculate the ROC Area Under the Curve (AUC) measurement.
- PRECISION AND RECALL: Precision is when the model predicts "yes", how often is it actually "yes". Recall (also true positive rate or specificity) is when the actual value is "yes" how often is the model correct.
- F1 SCORE: Calculate the F1-score, which is a weighted average between the precision and recall.

Recommendation for Customer Retention Using Shiny

It's critical to communicate data science insights to decision makers in the organization. Most decision makers in organizations are not data scientists, but these individuals make important decisions on a day-to-day basis. The Shiny application below includes a Customer Scorecard to monitor customer health (risk of churn).

Churn Risk

1.Success:-If the probalitity of churn for a particular customer is between 0 to 0.33, then the guageSectors will be Green(customer will not churn)

2.Warning:If the probalitity of churn for a particular customer is between 0.33 to 0.66, then the guageSectors will be Yellow(chance to be churned)

3.Danger:If the probality of churn for a particular customer is above 0.66, then the guageSectors will be Red(customer will Churn)

Churn probability is provided based on various feature. Developed a Recommendation system to reduce the churn probability

- Main Strategy

    - Strategy 1: If tenure less than a year ,then **Retain until one year**
    - Strategy 2: If tenure greater than 9 and with month-to-month contract ,then **Upsell to annual contract**
    - Strategy 3: If tenure greater than 12 and with No internet Service , then **Offer internet service**
    - Strategy 4: If tenure greater than 18 with monthly charges greater than 50 , then **Offer discount in monthly rate**
    - strategy 5: If tenure greater than 12 and has no additional service , then **Offer Additional Services**
    - Strategy 6: Customers not falling in above categories are less likely to churn, then **Retain and Maintain**

- Commercial Strategy

    - Startegy 1: If customers doesn't have any additional services, then **Offer Additional Services**
    - STrategy 2: If customers has Fiber optic services, then **Offer tech support and services**
    - Strategy 3: If customers doen't have Internet Services, then **Upsell to internet service**
    - Strategy 4: Customers not falling in above categories are less likely to churn, then **Retain and Maintain**
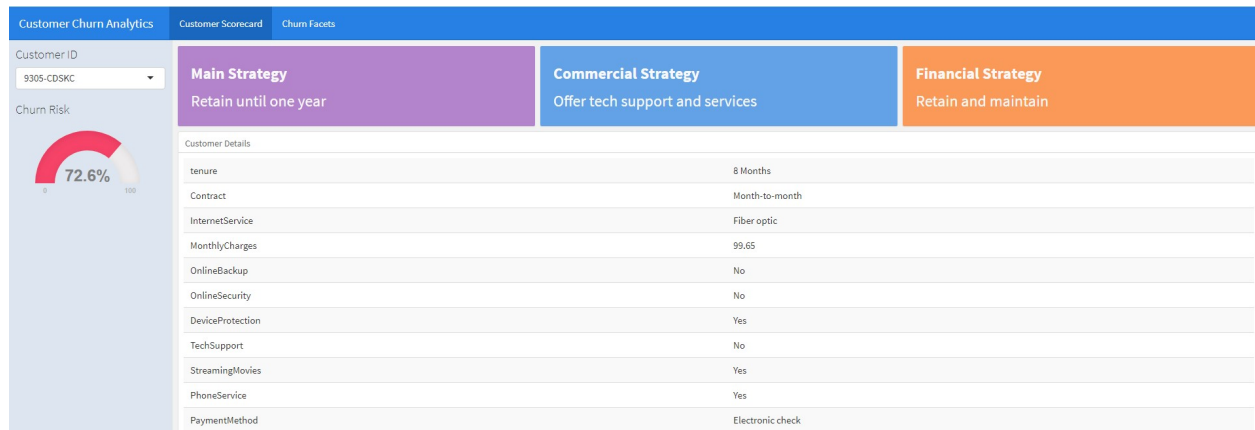
- Financial STrategy

Figure 3: Alt Text

- Startegy 1: If customers are using Payment Methods like Mailed Check and Electronic Check, then **Move to credit card or bank transfer**
- Strategy 2: Customers not falling in above categories are less likely to churn, then **Retain and Maintain**

Conclusion

Throughout the analysis, I have learned several important things: 1. Features such as tenure_group, Contract, PaperlessBilling, MonthlyCharges and InternetService appear to play a role in customer churn. 2. There does not seem to be a relationship between gender and churn. 3. Customers in a month-to-month contract, with PaperlessBilling and are within 12 months tenure, are more likely to churn; On the other hand, customers with one or two year contract, with longer than 12 months tenure, that are not using PaperlessBilling, are less likely to churn.

Customer churn is a costly problem. The good news is that machine learning can solve churn problems, making the organization more profitable in the process.I achieved 82% predictive accuracy (without tuning) by building an ANN model. using the keras package.