Machine Learning Based Telecom-Customer Churn Prediction

Presented by:

Pushkar Bhuse Aayushi Gandhi Parth Meswani

Riya Muni Neha Katre

Contents

- What is customer churn
- Requirement of Customer Churn in the Telecommunication Industry
- Dataset and Preprocessing
- Algorithms
- Results
- Conclusion

What is Customer Churn?

- Percentage of customers who will discontinue a company’s service.
- Calculated by dividing the number of customers lost over a given time.
- Key to predict success/failure of a business.

Requirement in the Telecom Industry

- A plethora of telecom organisations.
- Heavy competition leads to churners.
- Telecom industries have been battling huge revenue losses.
- Retaining customers is cost effective.

Fig.1 : Churn results from the cumulative experience of many episodes

Preprocessing

- Feature selection.
- Multiple columns are dropped based on correlation to possibility of churn.
- One-Hot Encoding
- Normalization with MinMax Scaler
- Data divided into train and test set (80:20).

Dataset:

No Churn

CustomerID	7590-VHVEG
Gender	Female
SeniorCitizen	0
Partner	Yes
Dependents	No
Tenure	1
PhoneService	No
MultipleLines	No phone service
InternetService	DSL
OnlineSecurity	No
OnlineBackup	Yes
DeviceProtection	No
TechSupport	No
StreamingTV	No
StreamingMovies	No
Contract	Month-to-month
PaperlessBilling	Yes
PaymentMethod	Electronic check
MonthlyCharges	29.85
TotalCharges	29.85
Churn	No

Churn

CustomerID	3668-QPYBK
Gender	Male
SeniorCitizen	0
Partner	No
Dependents	No
Tenure	2
PhoneService	Yes
MultipleLines	No
InternetService	DSL
OnlineSecurity	Yes
OnlineBackup	Yes
DeviceProtection	No
TechSupport	No
StreamingTV	No
StreamingMovies	No
Contract	Month-to-month
PaperlessBilling	Yes
PaymentMethod	Mailed check
MonthlyCharges	53.85
TotalCharges	108.15
Churn	Yes

Algorithms

- Ridge Classifier
- Random Forest
- XGBoost
- K- Nearest Neighbours (KNN)
- Support Vector Classifier (SVC)
- Deep Neural Network

Ridge Classifier

- Converts the label data into [-1,1]
- The highest value in prediction is accepted as a target class.
- Multi-output regression is applied for multiclass data.
- L2 regularization solves the overfitting issue.

Random Forest

- Training is performed on 80% of the dataset with 100 estimators.
- The maximum depth of each decision tree is 10.
- Important features are ranked and predicted based on the voting done on decision trees.

XGBoost

- Combination between gradient descent and boosting.
- Supervised ensemble learning algorithm.
- Good cache optimization.
- High execution speed and great model performance.

K-nearest neighbours

- Non-parametric learning algorithm.
- Uses ‘feature similarity’ to predict the values of new data points.
- Adjusted normalization is used to handle features that have a continuous distribution.

Support Vector Classifier

- Returns a ‘best-fit’ hyperplane that divides or categorizes your data
- Dataset is divided into 2 groups - churn and non churn
- Data points are made linearly separable by mapping them to a higher dimensional plane

Deep Neural Network

- Artificial neural network with multiple layers
- Initially has only one hidden layer, and more layers are added later
- A feature hierarchy of increasing abstraction and learning complex concepts is created by training each layer of neurons on the features/outputs of the previous layer

Results

Algorithms	Training Accuracy	Testing Accuracy
Random Forest	99.69%	**91.26%**
KNN Classifier	77.04%	81.80%
XG Boost	96.42%	88.55%
Ridge Classifier	75.91%	76.28%
SVC	94.85%	87.29%
Neural Network	77.04%	81.84%

Conclusion

- Techniques implemented: Random Forest, SVM, Extreme Gradient Boosting (XGBoost), Ridge classifier, K-nearest neighbors (KNN) and Deep Neural Networks.

- Influence of parameters on churn probability was explored

- A comparison between all the mentioned techniques was made to classify if a customer will churn or not and the efficiency was explored by passing them through a grid search.

- Random Forest model works best particularly for this use case with a prediction accuracy of 90.96% on the testing data.

Future Scope

- Focus on ways to achieve better results such as using different hyper-parameter optimization techniques.

- Different combinations of attributes can be used in the future to determine the customer retention policies.

- Approach newer algorithms to try and improve performance

Thank You!