2017-2018 Academic Year

PREDICTION OF CORPORATE CUSTOMER CHURN IN TELECOMMUNICATION SECTOR

Muhammet BOYRAZ // Sami KAYA // Süleyman DAŞ

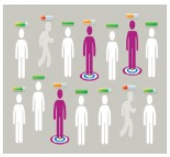
Academic Advisor: Assoc. Prof. Selçuk GÖREN // Company Advisor: Murat KÜLCÜ

ABDULLAH GÜL UNIVERSITY

DEPARTMENT OF INDUSTRIAL ENGINEERING

INTRODUCTION

Churn, *also known as customer attrition*, could be defined as the state of a customer's abandonment of doing business with a company or service.



Product-Focused

Increase in Competition

Customer-Focused

Customer Relations Management

Increases Revenue — *Increases Customer Satisfaction*

Decreases Churn

2008 - Legal Regulation

Moving Mobile Number

Increased the Competition

2008 → 2018

112 M

Mobile Number Movement

3rd Quarter of 2017

Competitors	TURKCELL
INCREASED	DECREASED
	LOSS
# of Customers	42 K CUSTOMERS

# of Mobile Subscribers

78 M

11% of them are Corporate Customers

PROBLEM DEFINITION

ARPA < ARPC

Revenue per New Customer — ***Loss per Customer Churn***

Since attracting new customers require sacrifices in the company side, such as relatively low price than existing customers, the ARPC value is expected to be greater than the ARPA value.

TURKCELL does not have automated system to indicate the risk of churn for corporate customers, the managers use their judgment and experience for prediction.

CUSTOMER CHURN *Causes* ***%2 REVENUE LOSS***

DATA PRE-PROCESSING

1. ***Data pre-processing consumes***

60% Total Effort

Outliers + Redundant+Missing Values

Removed from the system

2. ***Jan, 2016 - Sep, 2017***

The raw data contains:

- ***Last Visit Day of Customer***
- ***Associated Account Manager***
- ***Sector***
- ***Monthly Roaming Bill***
- ***Monthly Total Bill***
- ***Monthly IT Service Bill***
- ***Monthly Mobile Number Activated***
- ***Monthly Mobile Number Cancelled***
- ***Existance of Contract***

for each 21 month period.

3. ***758 Customer → 666 Customer***

142 Churn → 50 Churn

4. ***Time period: Last 6 Month***

Calculated by considering last 6 month:

- ***Average Amount of Bills***
- ***Standard Deviation of Bill Types***
- ***Amount of Mobile Subscribers for each of the 6 month***

37 Feature: ***3 Categoric*** — ***34 Numeric***

5. ***Multicollinearity Test Results:***

High

Variance Inflation Factor

Principal Component Analysis

10 Principal Component — ***%90 of the variability***

Low Performance

6. ***Multicollinearity Test Results:***

High

Variance Inflation Factor

Removed the features with HIGH VIF

37 Feature → 13 Feature: ***3 Categoric*** — ***10 Numeric***

Acceptable VIF

LEARNING MODELS

Artificial Neural Network (ANN)

- The objective of the neural network is to transform the inputs into meaningful outputs.
- Each input value is multiplied by the corresponding weighting layers, then summed and a scalar parameter called bias is added.
- There are three types of transfer functions used commonly in literature such as sigmoid, ***rectifier****, hyperbolic tangent and linear.

** transfer function which is used in our model*

Support Vector Machine (SVM)

- A learning model that is based on structural risk minimization, controlled by associated learning algorithms that analyze data and define patterns.
- A support vector machine attempts to find the line that "best" separates two classes of points. By "best", it means the line that results in the largest margin between the two classes. The points that lie on this margin are the support vectors.

Naive Bayes (NB)

- Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of independence between every pair of features.
- The result of Naïve Bayes modeling technique became acceptable when the high dimensional data was transformed into low dimension (Huang, Kechadi, & Buckley 2012).

Random Forest (RF)

- It is a model of multiple decision trees created using more than one decision tree. It has been improved by adding randomness feature to bagging method. (Breiman 2001)
- This model breaks each node into branches using the best among randomly selected features in each node, instead of dividing each node by using the best parameter among all features.

EVALUATION

Cunfusion Matrix

		Predicted -	Predicted +
Real	-	TN	FP
Real	+	FN	TP

TN: Predicted true but do not leave the system
FP: Predicted the customer will the system bu it does not leave
FN: The customer leave the system but model predict it wrong
TP: Model precited correct and customer leaves the system

Accuracy

$$\text{Accuracy} = \frac{TN + TP}{TN+TP+FP+FN}$$

Sensitivity

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Recall

$$\text{Recall} = \frac{TP}{TP+FP}$$

F-measure

$$\text{F-measure} = 2 \times \frac{\text{Recall} \times \text{Sensitivity}}{\text{Recall} + \text{Sensitivity}}$$

Model/Measure	*Accuracy*	*Recall*	*Sensitivity*	*F-measure*
ANN	94%	75%	25%	37,5%
SVM	95,2%	100%	33%	50%
NB	91%	30%	25%	27%
RF	96,3%	80%	66%	72,3%

RESULTS

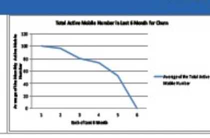
- The Random Forest model predicted customer loss with an accuracy of 96.3%.
 In addition, the features with the highest effect on the prediction are identified by using feature importance.
- The Company uses this information to revise customer relationship management strategies.

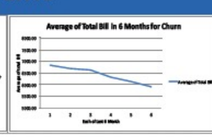
FEATURE IMPORTANCE

Feature importances are measured for learning models which have highest accuracies. SVM and RF models approximately gives the same results. The most important 5 Features are:

- Average # of Roaming Bill
- Total # of Cancelled Mobile Numbers
- Std. Deviation of Roaming Bill
- Average of the Total Bill
- Number of port-out mobile number in the last 6 months

INSIGHTS





REFERENCES

Breiman, 2001; Breiman,http://oz.berkeley.edu/users/breiman/randomforests.html

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), 1414-1425.

INDUSTRIAL ENGINEERING

TURKCELL