CHURN PREDICTION TELCOM USER

‘Azzam Naufal Mu’adz

MIND MAP

01 USE CASE

02 CRISP-DM

03 SUMMARY

01

USE CASE

PROBLEM AND SOLUTION

PROBLEM

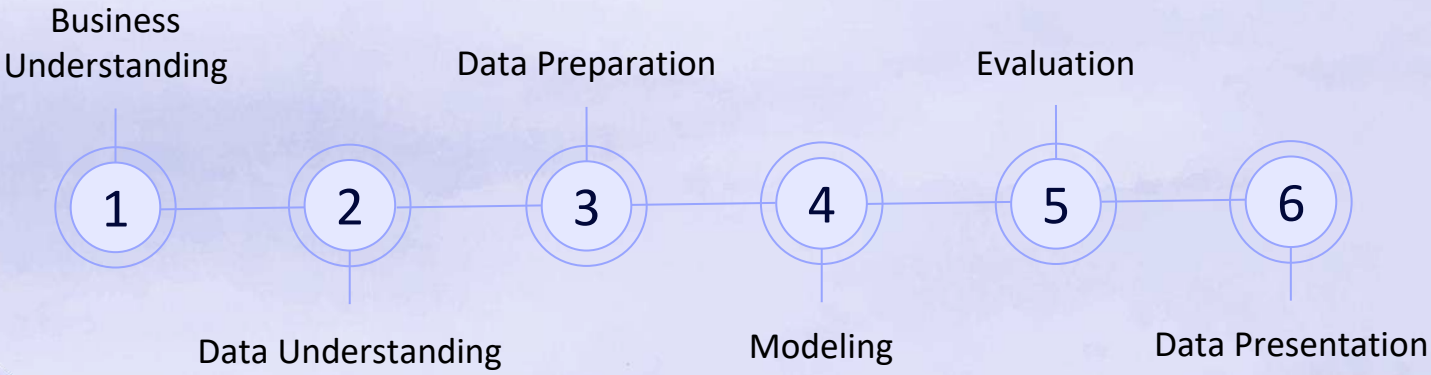
How to analyze and predict telecom user churn through demographics, services that each customer used, and customer account information?

SOLUTION

Machine learning enables to predict customer churn using data such as user demographics, services used and customer account information

02 CRISP-DM

CRISP-DM PROCESS



BUSINESS UNDERSTANDING

Telecommunications company focus providing phone and internet services wants to maximize the number of customers. Therefore, is important not only to try to attract new ones, but also to retain existing ones.

Predicting the churn can react in time and try to keep the client who wants to leave. This will make the task of retention easier to implement than the task of attracting new users, about which we do not know anything yet.

DATA UNDERSTANDING

Column Name		Definition
customerID	:	Customer id
gender	:	Client gender (male / female)
Age	:	Client Age
Partner	:	Whether the customer has a partner or not (Yes, No)
Dependents	:	Whether the customer has dependents or not (Yes, No)
tenure	:	How many months a person has been a client of the company?
PhoneService	:	Whether the customer has a phone service or not (Yes, No)
MultipleLines	:	Whether the customer has a multiplelines or not (Yes, No, No phone service)
InternetService	:	Client's Internet service provider (DSL, Fiber optic, No)
OnlineSecurity	:	Whether the customer has a online security or no t(Yes, No, No internet service)
OnlineBackup	:	Whether the customer has a online backup or not (Yes, No, No internet service)
DeviceProtection	:	Whether the customer has a device protection or not (Yes, No, No internet service)
TechSupport	:	Whether the customer has a tech support or not (Yes, No, No internet service)
StreamingTV	:	Whether the customer has a streaming tv or not (Yes, No, No internet service)
StreamingMovies	:	Whether the customer has a streaming movies or not (Yes, No, No internet service)
Contract	:	Type of customer contract (Month to month, One year, Two year)
PaperlessBilling	:	Whether the client uses paperless billing (Yes, No)
CashBilling	:	Whether the client uses cash billing (Yes, No)
PaymentMethod	:	Payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
MonthlyCharges	:	Current monthly payment
TotalCharges	:	The total amount that the client paid for the services for the entire time
Churn	:	Whether there was a churn (Yes or No)

DATA PREPARATION

DATA PREPARATION

Prepare the required packages such as pandas, seaborn, sklearn and dataset

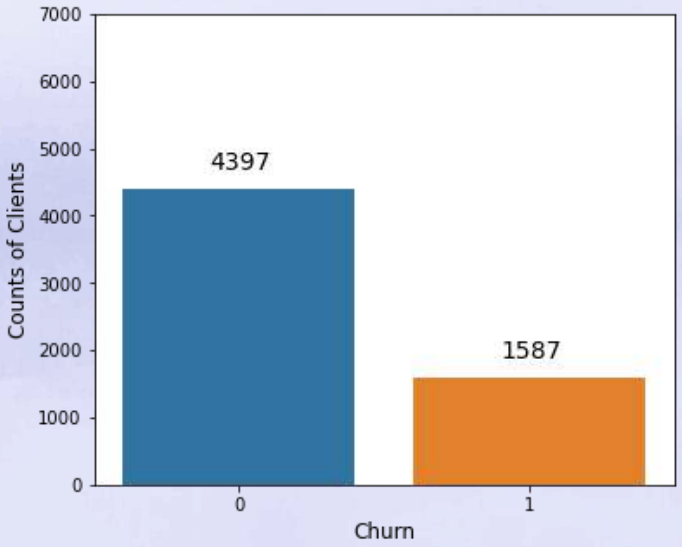
DATA CLEANSING

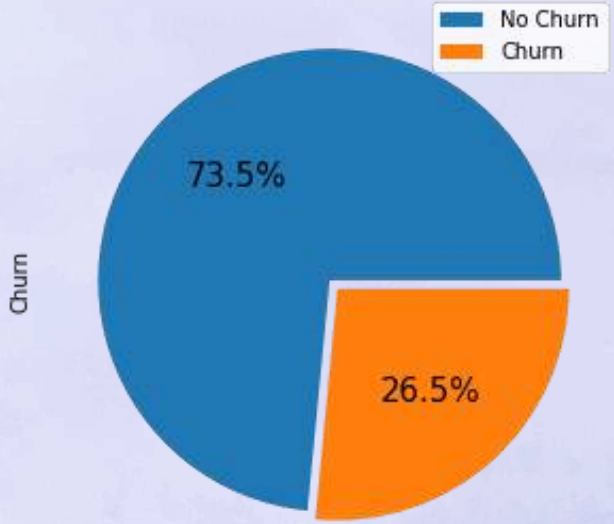
Solve missing value, correct wrongly typed value.

EXPLORATORY DATA ANALYSIS

Looked up the distribution of data and visualize to get insights

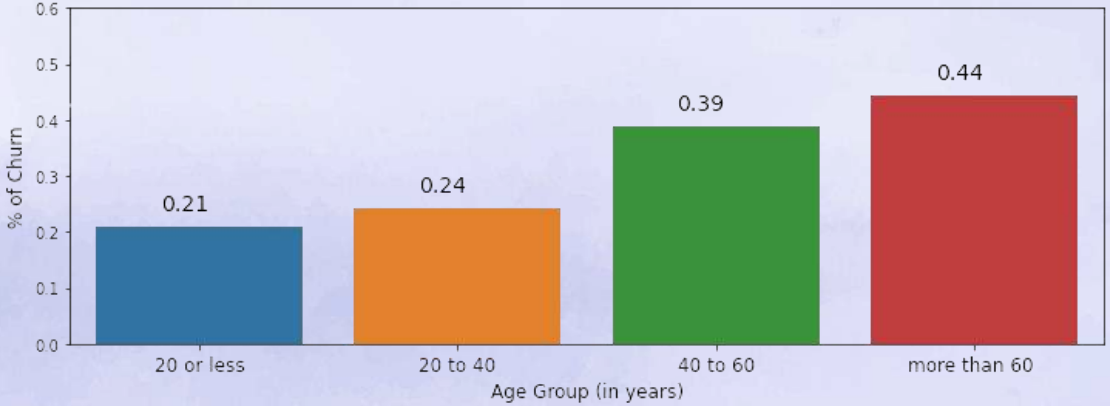
CUSTOMER CHURN





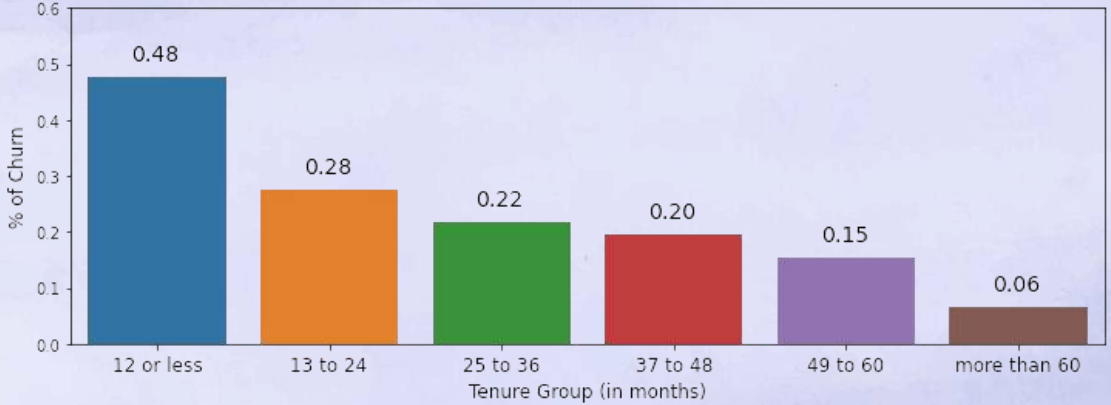
At the period represented in this dataset, there is a 26,5% of customer churn.

CUSTOMERS TRENDS



AGE

Almost 50 percent of those who become customers are older people



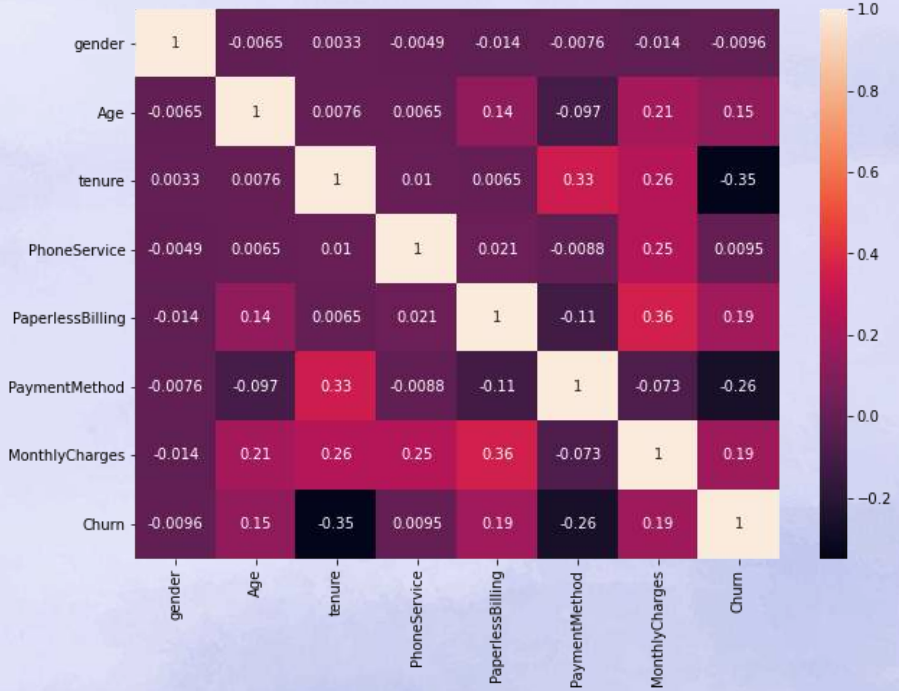
TENURE

Almost 50 percent of those who became a customers for a year or less ended up leaving.

HEATMAP CORRELATION

Features:

- Gender
- Age
- Tenure
- Phone Service
- Paperless Billing
- Payment Method
- Monthly Charges
- Churn



MODELING

Split the data into train and tests sets with a test size of 20%. Use 3 different model:

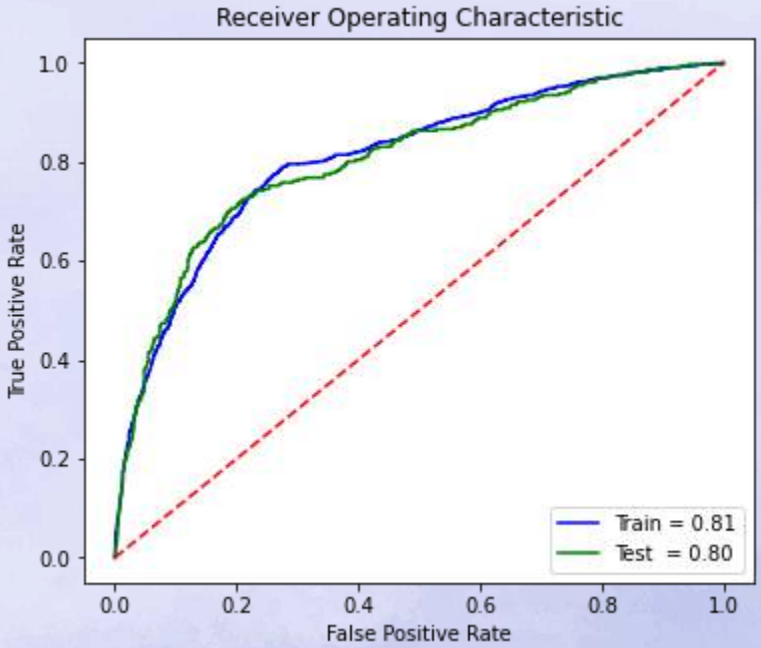
- Support Vector Machine
- Decision Tree Model
- Random Forest Model

Evaluate the best one of the three using the AUC (Area Under ROC Curve). AUC is relatively easy to interpret model performance.

EVALUATION

The evaluation metric that be used is AUC (Area Under ROC Curve). Despite the slight discrepancy, the Support Vector Machine Model performs better than other approaches on test and validation sets.

AUC train & test	: 80.79% & 80.39%
Confusion Matrix Evaluation	
Accuracy train & test	: 79.40% & 80.79%
Recall train & test	: 40.44% & 41.83%
Specificity train & test	: 93.64% & 94.16%
Precision train & test	: 69.91% & 71.11%
F1 Score train & test	: 51.24% & 52.67%
Log Loss train & test	: 7.1141 & 6.6366



SUMMARY 03

SUMMARY

- Machine learning enables decision makers to predict customer churn using company-owned historical data.
- The evaluation metric that be used is AUC (Area Under ROC Curve).
- The Support vector machine model has better performance than the other approaches on the test and validation sets.

RECOMMENDATION

We recommend adding more products for young people considering the low churn rate among young people. As well as to improve services so that users do not turn away when they enter their first year.

THANKS

Do you have any questions?