Predict customer churn in telecom company’s using classification models

1. **Abstract**

Acquisition and the retention of customers are the top most concerns in today's business world. The rapid increase of market in every business is leading to higher subscriber base. Consequently, companies have realized the importance of retaining the on hand customers. It has become mandatory for the service providers to reduce churn rate because the negligence could be resulted as profitability reduction in major perspective. Churn prediction helps in identifying those customers who are likely to leave a company, the classification models that was used in this project are logistic regression, k-nn, random forest, decision tree and xgboost .

1. **Dataset description**

We obtained the dataset from Kaggle (<https://www.kaggle.com/blastchar/telco-customer-churn>) the raw data contains 7034 record (customers) and 21 feature as described in the table below. The “Churn” column is our target.

|  |  |  |
| --- | --- | --- |
| **Features** | **Description** | **Type** |
| customerID | |  | | --- | | A unique ID that identifies each customer | | Object |
| Gender | |  | | --- | | Whether the customer is a male or a female | | Object |
| SeniorCitizen | |  | | --- | | Whether the customer is a senior citizen or not (1, 0) | | int64 |
| Partner | |  | | --- | | Whether the customer has a partner or not (Yes, No) | | Object |
| Dependents | |  | | --- | | Whether the customer has dependents or not (Yes, No) | | Object |
| tenure | |  | | --- | | Number of months the customer has stayed with the company | | int64 |
| PhoneService | |  | | --- | | Whether the customer has a phone service or not (Yes, No) | | Object |
| MultipleLines | |  | | --- | | Whether the customer has multiple lines or not (Yes, No, No phone service) | | Object |
| InternetService | |  | | --- | | Customer’s internet service provider (DSL, Fiber optic, No) | | Object |
| OnlineSecurity | |  | | --- | | Whether the customer has online security or not (Yes, No, No internet service) | | Object |
| OnlineBackup | |  | | --- | | Whether the customer has online backup or not (Yes, No, No internet service) | | Object |
| DeviceProtection | |  | | --- | | Whether the customer has device protection or not (Yes, No, No internet service) | | Object |
| TechSupport | |  | | --- | | Whether the customer has tech support or not (Yes, No, No internet service) | | Object |
| StreamingTV | |  | | --- | | Whether the customer has streaming TV or not (Yes, No, No internet service) | | Object |
| StreamingMovies | |  | | --- | | Whether the customer has streaming movies or not (Yes, No, No internet service) | | Object |
| Contract | |  | | --- | | The contract term of the customer (Month-to-month, One year, Two year) | | Object |
| PaperlessBilling | Whether the customer has paperless billing or not (Yes, No) | Object |
| PaymentMethod | |  | | --- | | The customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)) | | Object |
| MonthlyCharges | |  | | --- | | The amount charged to the customer monthly | | float64 |
| TotalCharges | |  | | --- | | The total amount charged to the customer | | Object |
| Churn | |  | | --- | | Whether the customer churned or not (Yes or No) | | Object |

1. **Feature Engineering:**

* Dummy variable
* Additional column

1. **Feature Importance:**

I extract the feature importance using extra trees

1. **Tools**

**Technologies:**

* Python
* Jupyter Notebook

**Libraries:**

* NumPy
* Pandas
* Matplotlib
* Seaborn
* Sklearn
* Pandas Profiling

1. Chart, box and whisker chart

   Description automatically generated**Communication**

Chart, pie chart

Description automatically generatedFigure 1: Outliers

Figure 2: Solved the imbalanced dataset

Chart, bar chart, histogram

Description automatically generated

Figure 3: Feature importance using extra trees