**Report on Customer Churn Predication**

**using ML**

|  |
| --- |
| 1. Business Problem |
| 2. Introduction |
| 3. Problem Definition |
| 4. Solution |
| 5. Data overview |
| 6. Installation and metrics |
| 7. File Description |
| 8. Result |
| 9. Limitation |
| **List of References** |

1. **Business Problem**

Identify customers likely to churn and develop focused intervention plans to retain as many customers as possible.

1. **Introduction:**

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

Telephone service companies, ISPs, 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 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.

1. **Problem Definition**

400 million subscribers in the US telecommunication industry. There are 54 telecommunication companies, according to Forbes 2018.

Monthly loss from Churn is $65M

1. **Solution**

Predict customers’ churn decisions, Develop retention plans

1. **Data Overview**

Telco-Customer-Churn Data

Link to data: <https://www.kaggle.com/blastchar/telco-customer-churn>

● Each row represents a customer; each column contains customer’s attributes described on the column Metadata

● The raw data contains 7043 rows (customers) and 21 columns (features)

● After cleaning the data for later analysis, the final data set contains 7043 rows (customers) and 41 columns (features).

● Data reports: Customers who left within the last month: Churn Customer Services: Phone Service, Multiple Lines, Internet Service, etc. Customer Account information: Tenure, Contract, Payment Method, Charges, etc. Customer Demographic Information: Gender, Partner, Dependents, etc.

● Source: Telco Customer Churn programs from Kaggle

**6. Installation**

The code in this project is written in Python 3.6.6.

Anaconda custom (64-bit). The following additional libraries have been used:

Pandas numpy matplotlib

Seaborn plotly sklearn

imblearn (Synthetic Minority Oversampling TEchnique (SMOTE))

**6.1 Metric**

The following metrics have been used to compare the model performances:

Accuracy Recall AUC ROC

Precision F1 score

1. **File Descriptions**

The Jupyter notebooks included in this project are:

Telecom Customer Churn Prediction.ipynb.ipynb

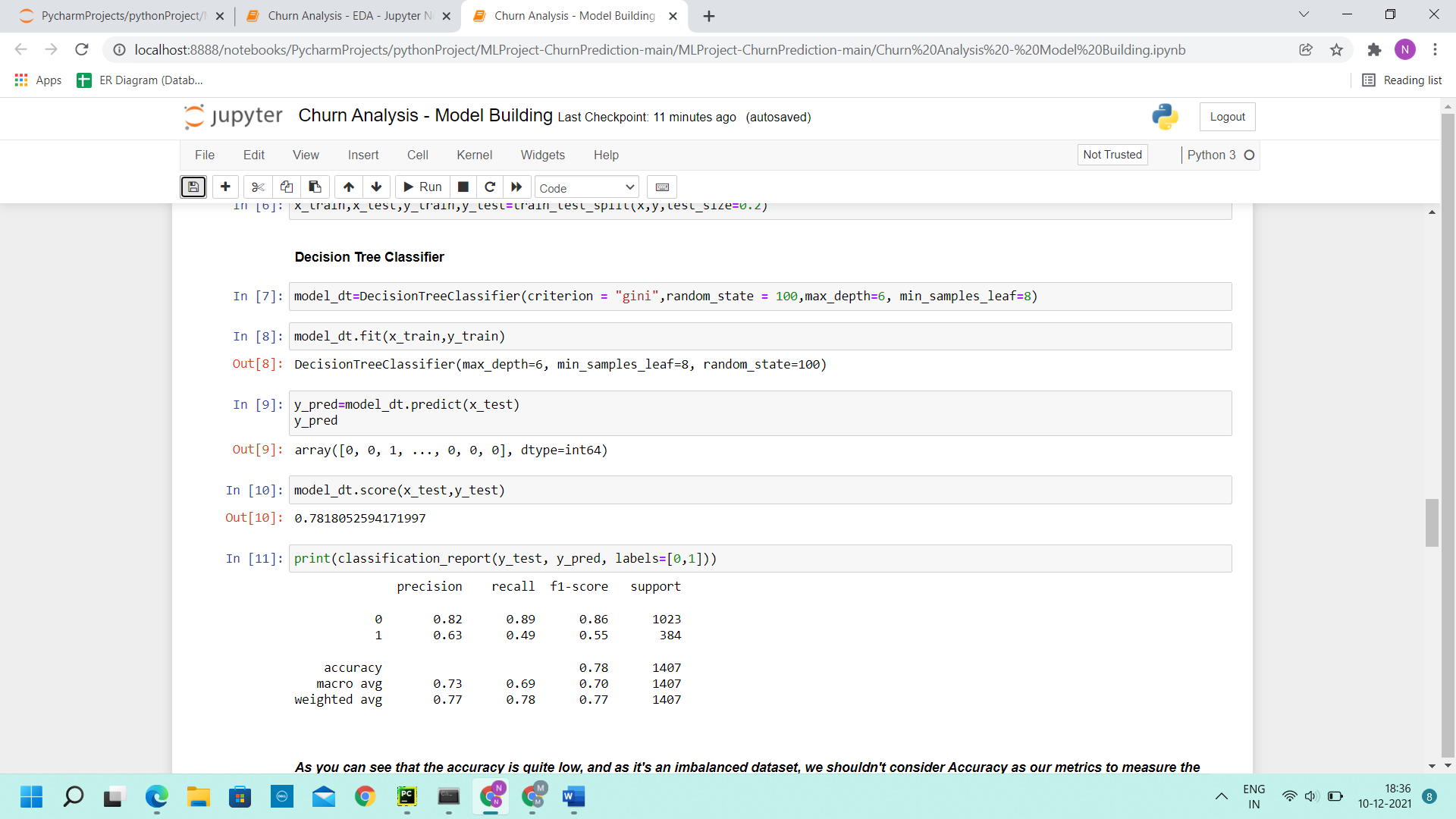
Data files (under data directory):

WA\_Fn-UseC\_-Telco-Customer-Churn.csv

1. **Results**

The following classifiers have been compared:

**Decision Tree**

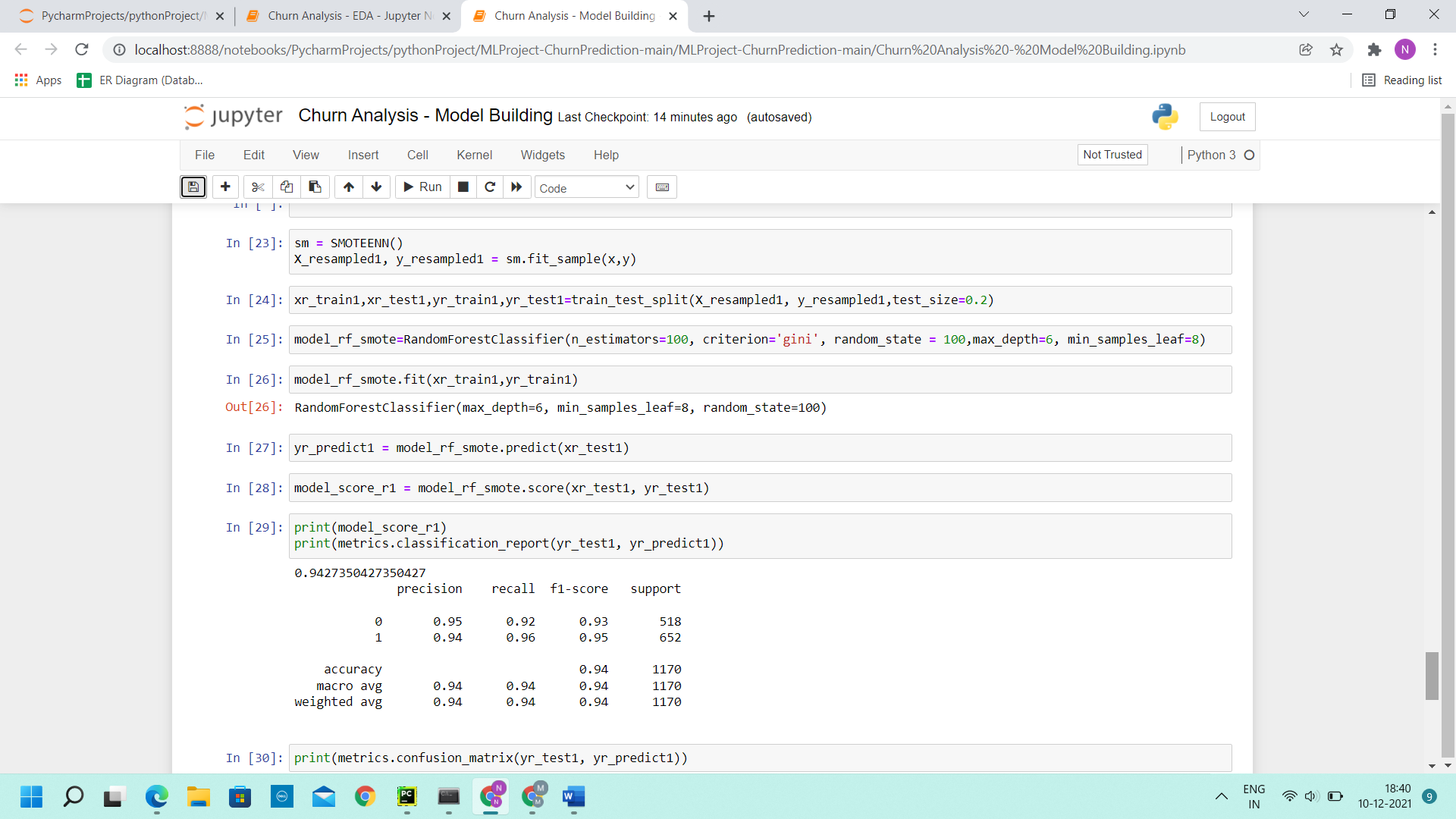


**Random Forest**

As we can see that the accuracy is quite low, and as it's an imbalanced dataset, we shouldn't consider Accuracy as our metrics to measure the model, as Accuracy is cursed in imbalanced datasets.

we need to check recall, precision & f1 score for the minority class, and it's quite evident that the precision, recall & f1 score is too low for Class 1, i.e. churned customers.

Hence, moving ahead to call SMOTEENN (UpSampling + ENN ) achieved the best precision of 0.94



1. **Limitations**

Limited Data (7,043 observations with 26 variables) and Imbalanced Data (26.54% of churned customers)

**List of References**

<https://blog.aidatabases.in/churn-analytics/>

<https://towardsdatascience.com/predict-customer-churn-with-r-9e62357d47b4>

<https://www.kaggle.com/blastchar/telco-customer-churn>