Orange Fiber Customer Churn Prediction

Data science graduation project

Meet our Team

Mohammed Alkarmi

Thamer Smadi

Mohammed Sharap

Work flow

1. Problem Definition
2. Data Cleaning
3. Modeling
4. User Interface
5. Conclusion

Problem Definition

Our dataset is for Orange fiber customers , according this data we build a ML model that predicts customer churn or not depends on some features.

The dataset contains 21 columns (features) and 94473 rows

Data Cleaning

- A Drop columns
- B Remove outliers
- C Filling null values
- D Encoding

Feature Selection

For feature selection, we used SelectKBest to select the top 8 features

Selected Features

- GOVERNORATE.
- Customer with Orange Months.
- COMMITMENT_FG.
- OF_SPEED.
- OF_PREV_SPEED.
- MIGRATION_FLAG.
- GB_TOTAL_CONSUMPTION_Month1.

Building Model

Basically, we are dealing with a binary classification problem.

01 **LogisticRegression**

```
LogisiticRegression
accuracy for test : 0.99
accuracy for train : 0.99
```

02 **RandomForestClassifier**

```
RandomForestClassifier
accuracy for test : 0.99
accuracy for train : 1.0
```

Building Model

- **Imbalanced data**

  In our data, the ratio of 1 over 0 was 0.005, indicating an imbalance.

Solution :
Giving a higher weight to the lower class by using compute_class_weight

Building Model

01

LogisticRegression

```
LogisiticRegression
accuracy for test : 0.82
accuracy for train : 0.82
```

02

RandomForestClassifier

```
RandomForestClassifier
accuracy for test : 0.99
accuracy for train : 1.0
```

introducing overfitting problem

Evaluating and Comparing

According Grid search method

01 LogisticRegression

Parsmeters :

solver='newton-cg'

Results:

```
LogisiticRegression
accuracy for test : 0.82
accuracy for train : 0.82
```

02 RandomForestClassifier

Parsmeters :

max_depth: 6, max_features: 'auto,' and n_estimators: 100

Results:

```
RandomForestClassifier
accuracy for test : 0.83
accuracy for train : 0.83
```

User Interface

Using (Tkinter) we built a Graphical User Interface, to help the company to predict the churn customers.

Conclusions

Customer churn prediction is crucial to the long-term financial stability of a company. In this project, we successfully created a machine learning model that's able to predict fiber customer churn with an accuracy of 83%.

Thank you for your attention

Q & A