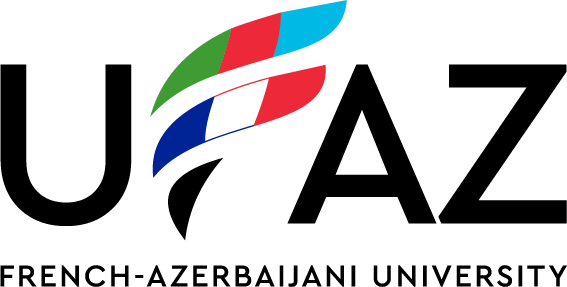
DSAI-22 (2022-2024)

**Project Work:** Introduction to Data Science

**Professor:** Benoit Gaüzère

**Students:** Elnur Maharramov, Oktay Rasulzade, Majid Rahmanov, Ibrahim Aliyev



**Credit card approval prediction**

**Abstract**

Dataset:

A big dataset on Credit Card approval system

Problem description:

Credit cards are a popular risk management strategy in the financial sector. To estimate the likelihood of future defaults and credit card borrowing, it uses the personal information and data provided by credit card applicants. The bank has the authority to choose whether or not to give the applicant a credit card. The degree of risk can be accurately measured by credit scores. Credit score cards are typically based on past data. once experiencing significant economic swings. Previous models might no longer be as accurate. A prominent method for credit scoring is the logistic model. Because Logistic can determine the coefficients of each feature and is appropriate for binary classification applications.

Task:

Create a machine learning model to determine whether a potential client is a "good" or "bad" client. Unlike other assignments, this one does not specify what constitutes a "good" or "bad" client. To build your label, you should employ some technique, such as vintage analysis. Unbalanced data is another significant issue with this task.

**Some important highlights of our code & results**

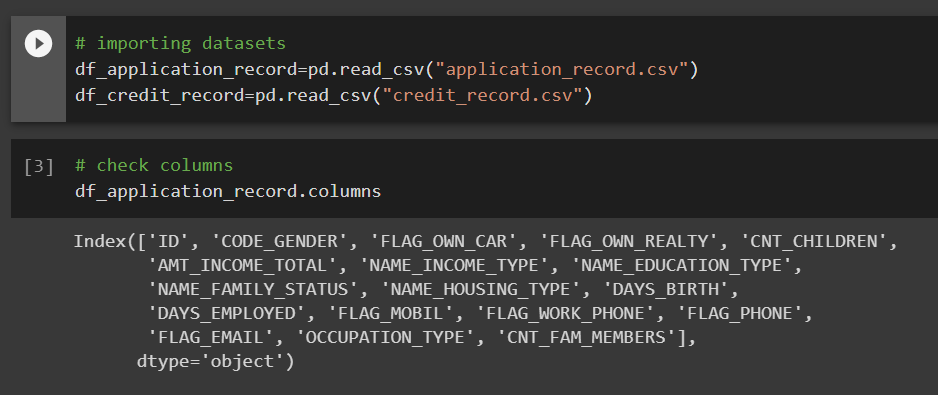
1 – Data acquisition

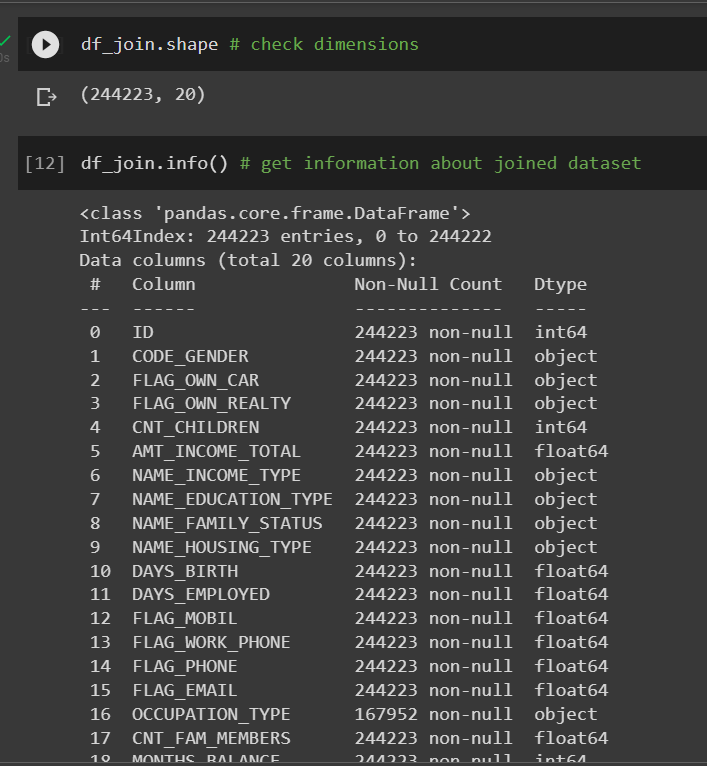
2 – Data cleaning

3 – Feature engineering

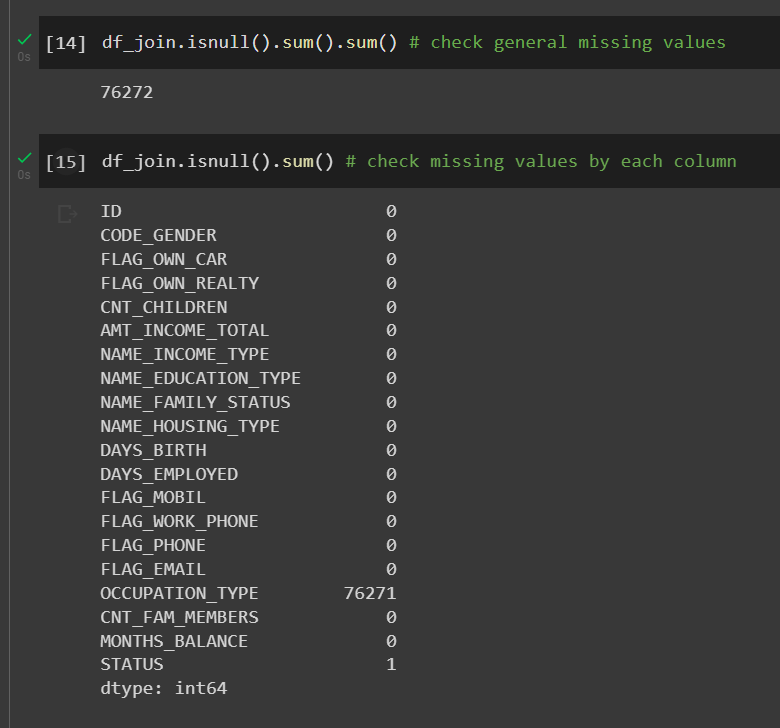
4 – Data Modeling

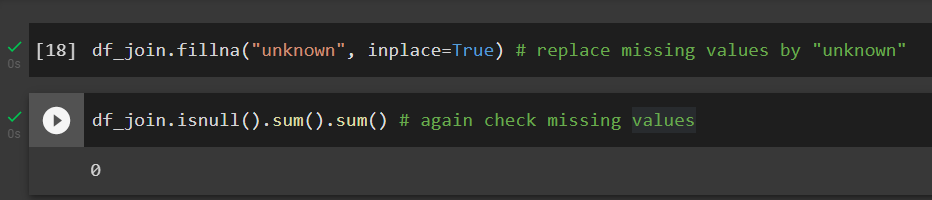
In data acquisition process we import data and trying to understand (types of columns, row count, column count, etc.) this dataset.

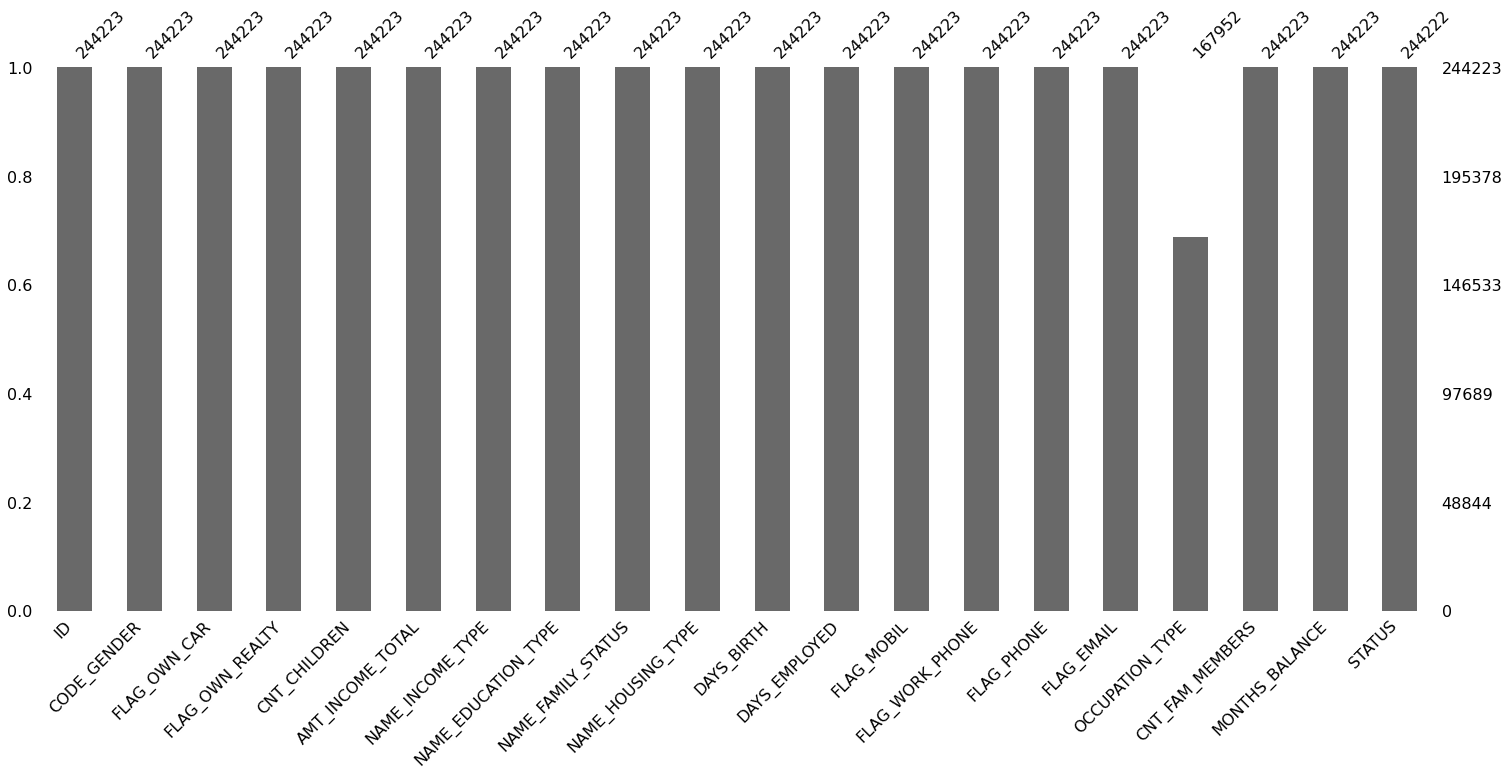


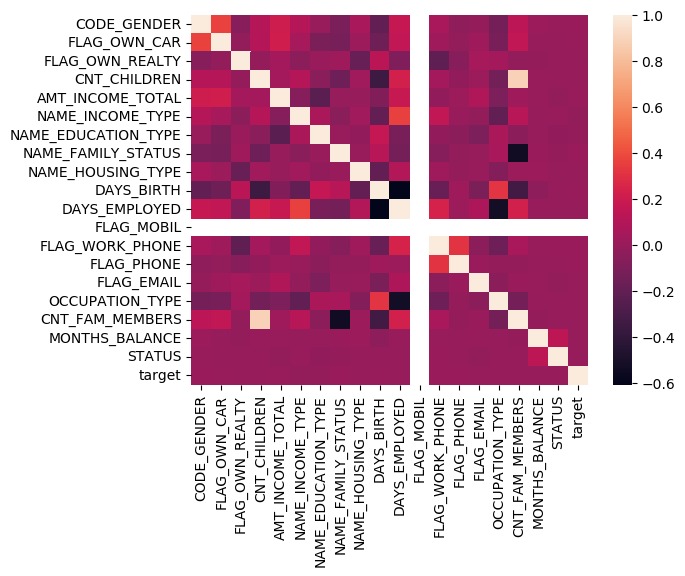


In cleaning process, we remove or replace rows which containes empty cells or NaN data.

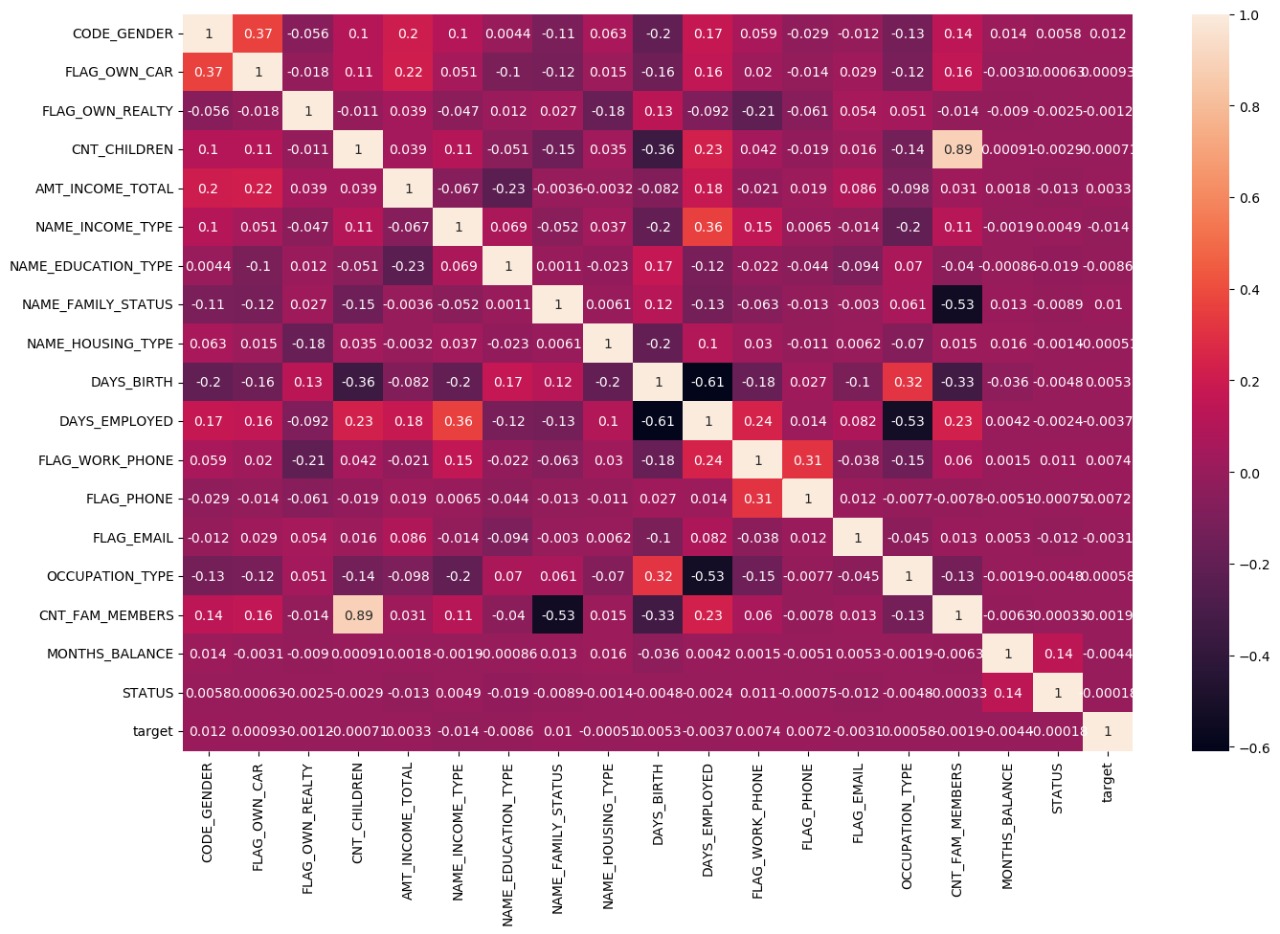




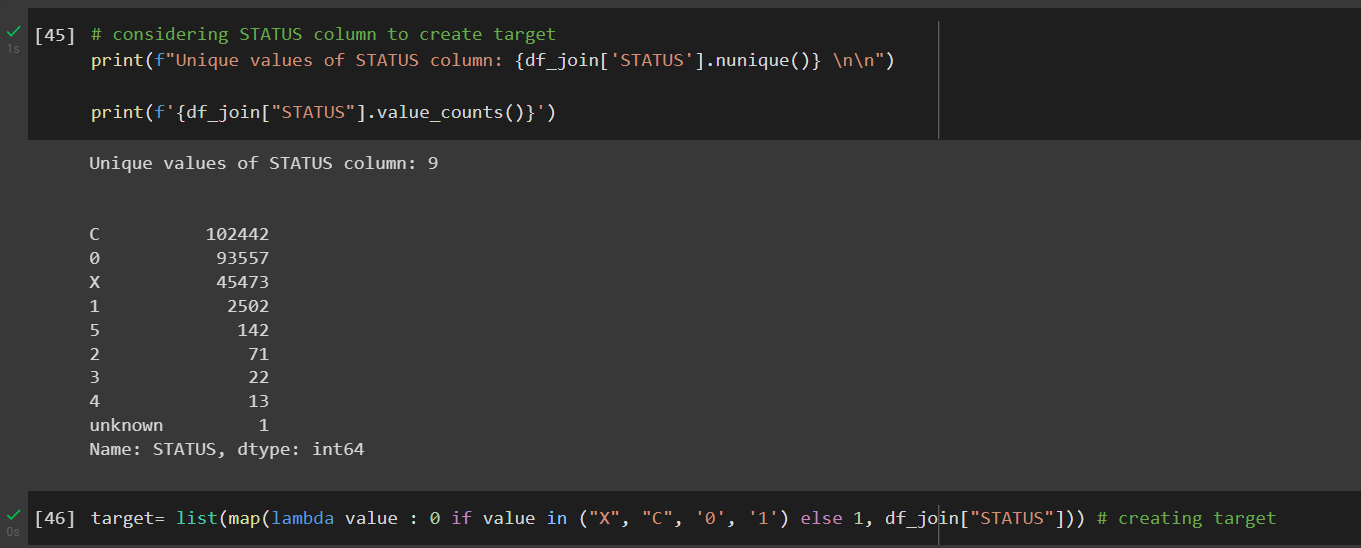




As column Flat mobile column doesn't affect to our result, we can remove it

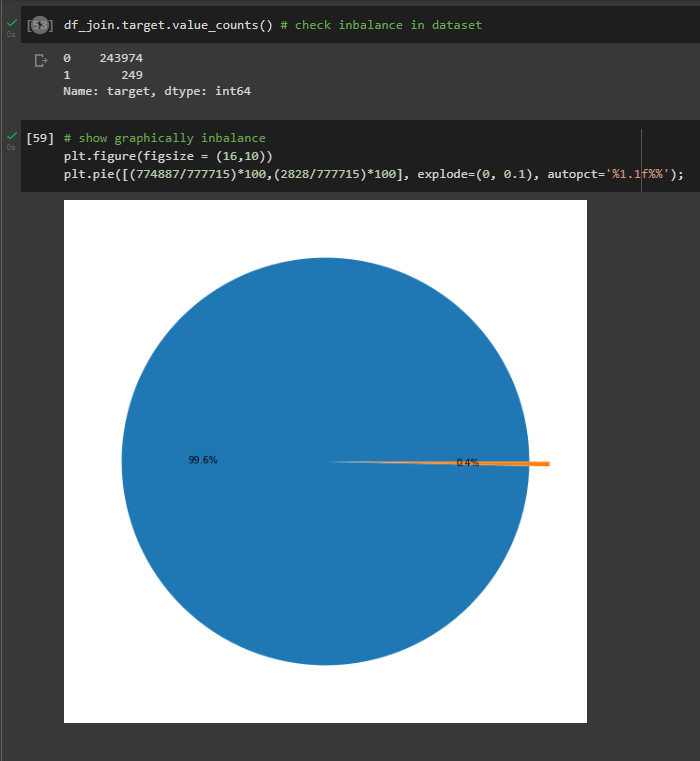


In feature engineering step, we extract characteristics of our data.

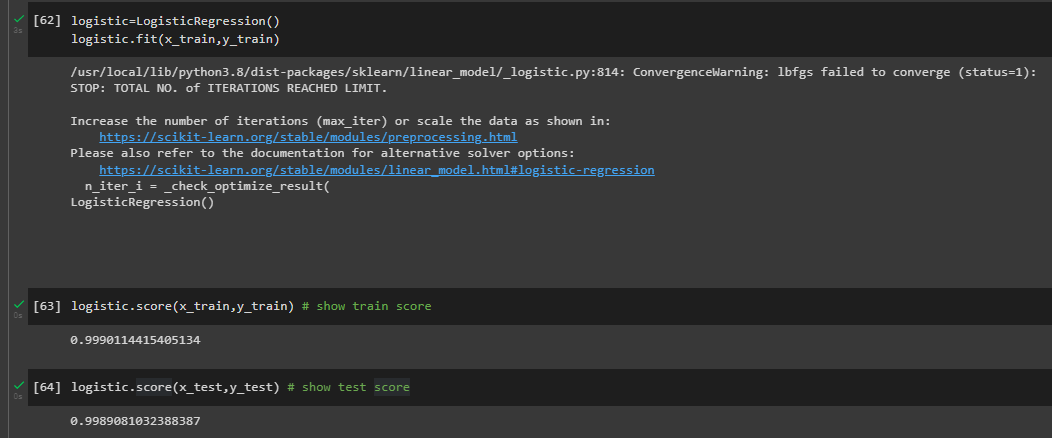


Text

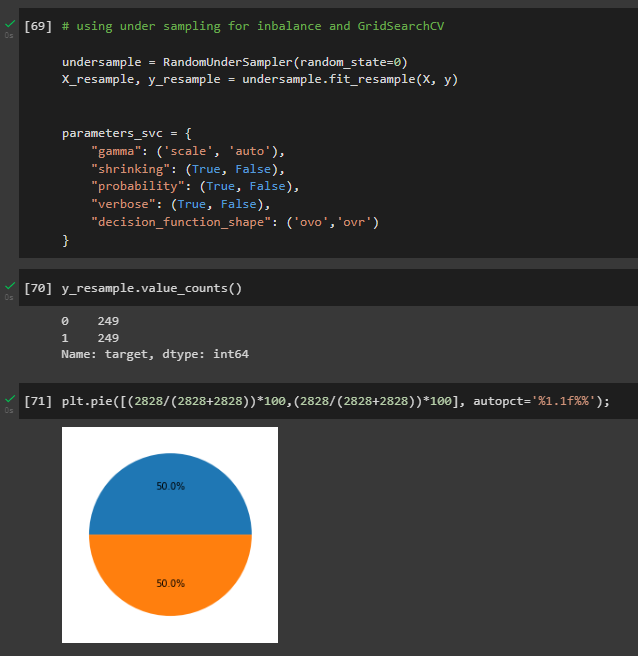
Description automatically generated

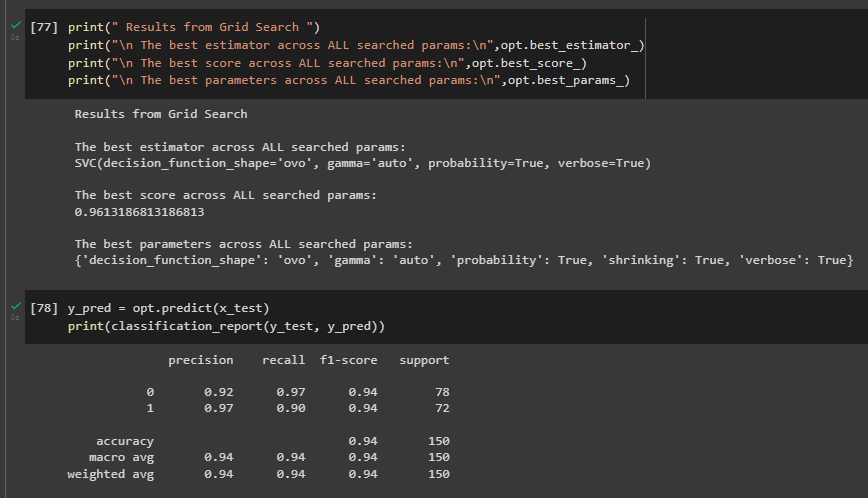


Before solving inbalance problem, we did logistic regression model and we get result below:

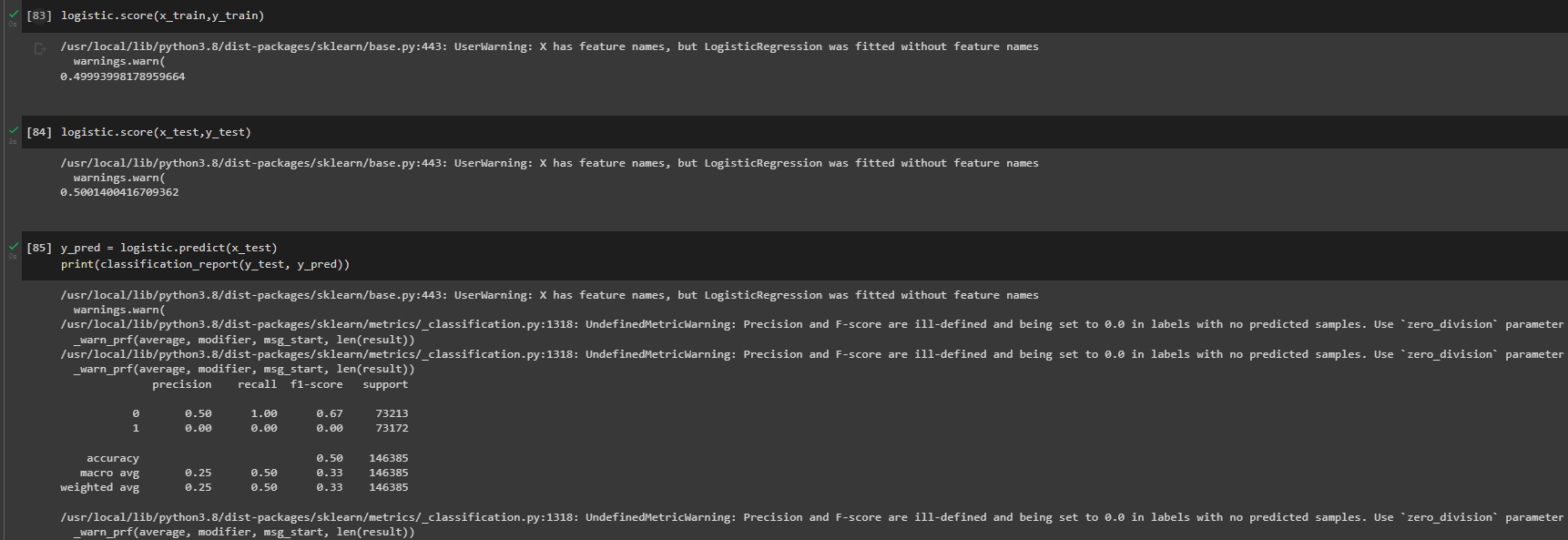


After we have solved inbalance problem by using under sampling method, we have done support vector machine classifier model by using GridsearchCV and get get results below:





We have also used over sampling method to solve inbalance problem, in this case we have used ligistic regression model and we got this result:



**Conclusion**

We import data and we cleaned it then we used some methods for feature engineering and data cleaning then we create some models like Kmeans, SVC (classifier variant of SVM), Logisticregression. in this case we also used pipeline and also we used GridsearchCV for hiper parameters tuning.

Our dataset has inbalance problem and we solved it, after solving we trained and we got results for comparing after solving this inbalance problem.