



Florida International University

Capstone II Project (Spring 2022)

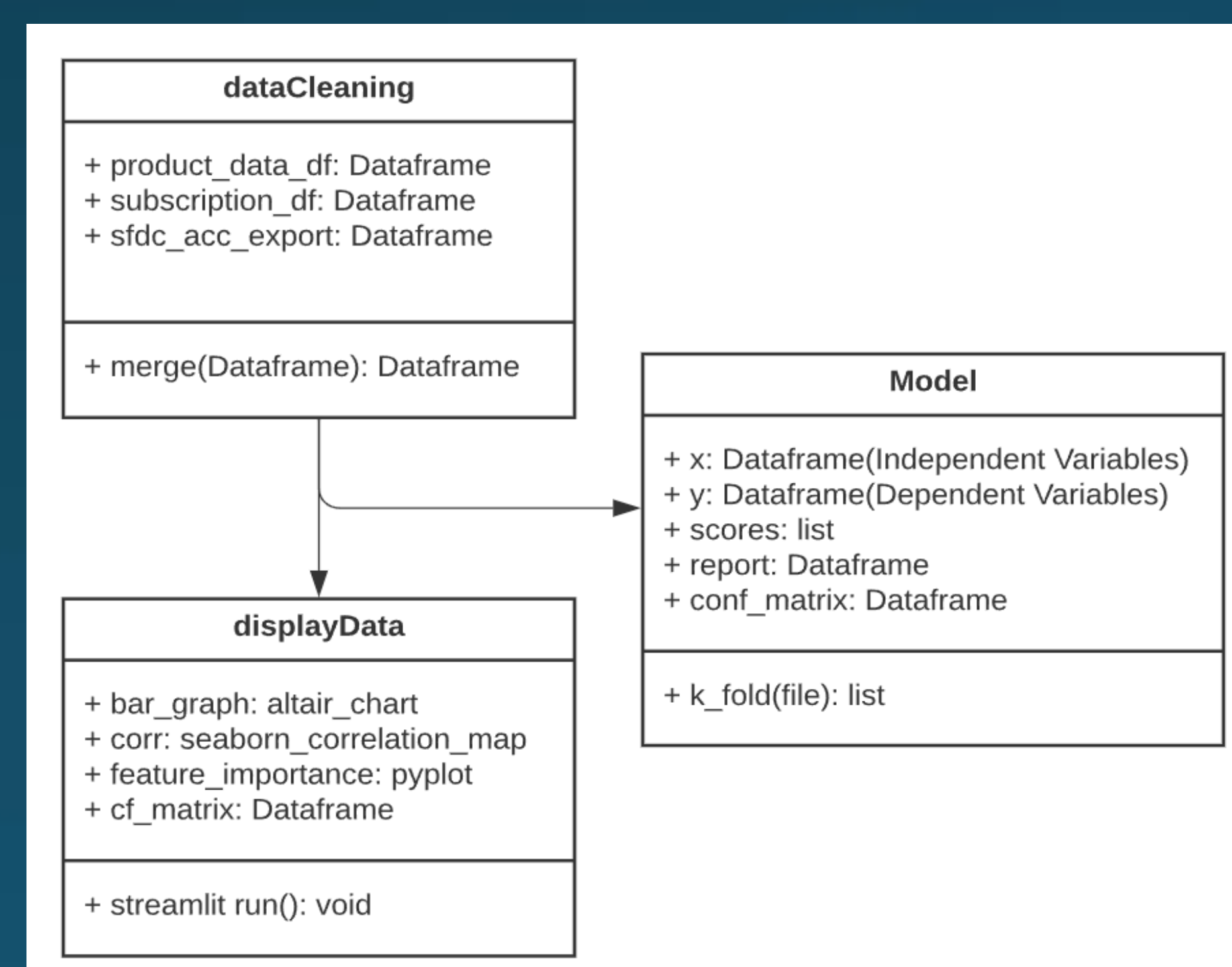
Instructor: Dr. Masoud Sadjadi
Product Owner: Lina Sokol
Student: Josue Daniel Prendes Diaz

Problem

This project had the goal of creating a machine learning model to predict when a customer from the Kaseya company would churn and stop using its services. Detecting customer churn factors would allow Kaseya to identify and try to retain those customers, which have a high probability of cancelling their memberships.



Object Design



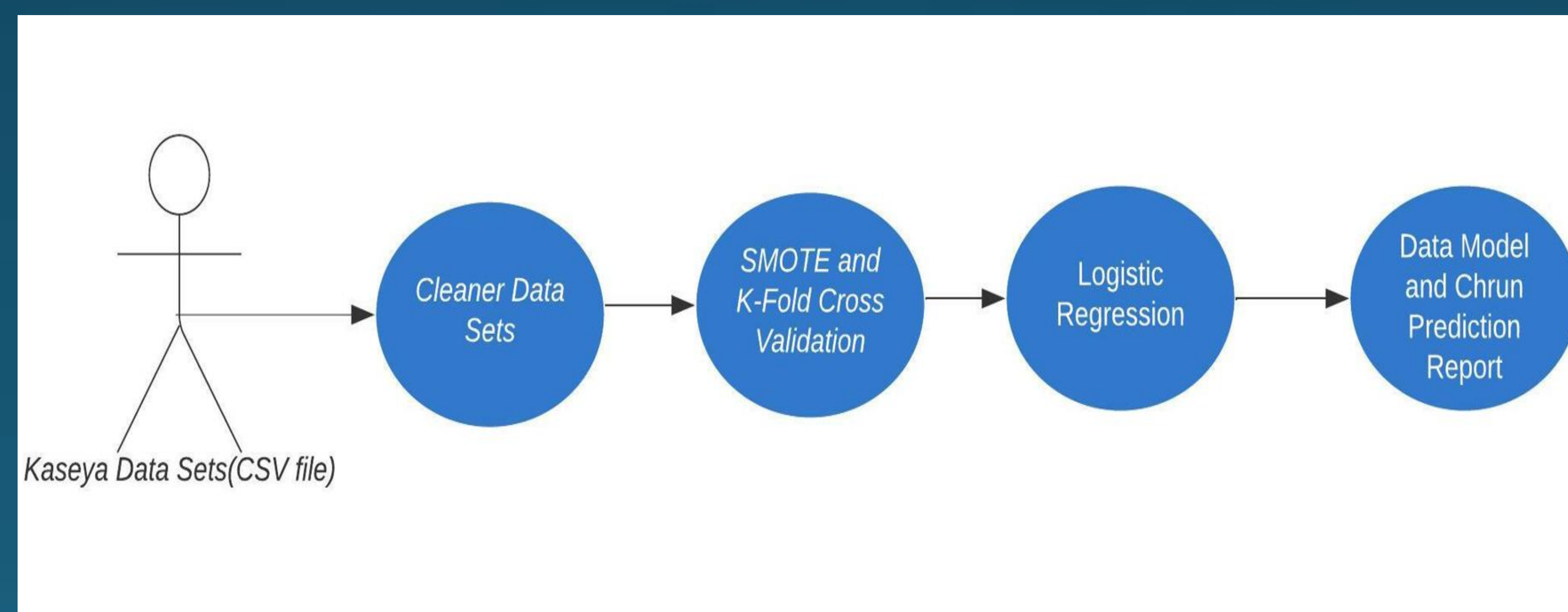
Verification

Logistic regression was used with the initial data sets provided by our product owner, but the data sets needed to be cleaner in order to predict customer churn with a higher accuracy. K-Fold Cross validation was used to improve consistency and accuracy.

Current System

To be able to predict when Kaseya's customers had a high probability of customer churn, our team modified the provided data sets to make them cleaner and better distributed. Then, logistic regression was used to analyze the probably values. Finally, data visualization was implemented using streamlit.

System Design



Summary

- Our team cleaned the data sets provided by Kaseya, and then inputted them into our model as Comma Separated Value (CSV) files.
- The combined CSV file was analyzed using anaconda python (Jupyter Notebook).
- We developed our machine learning model by integrating SMOTE and K-Fold Validation with logistic regression.
- I personally helped in the creation of the User Interface by using the streamlit python library and designed the software product logo.

Requirements



Acknowledgement

The material presented in this poster is based upon the work supported by our product owner Lina Sokol. I want to thanks my Capstone I, Capstone II, and Senior Project team members for their amazing work within our group. Thank you to Dr. Masoud Sadjadi for his help and dedication to this project.

