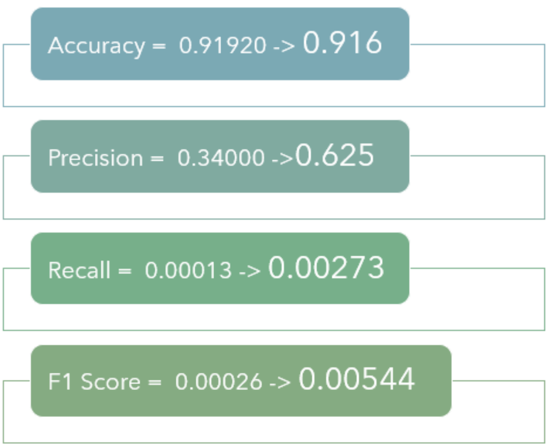
HOME CREDIT RISK ANALYSIS

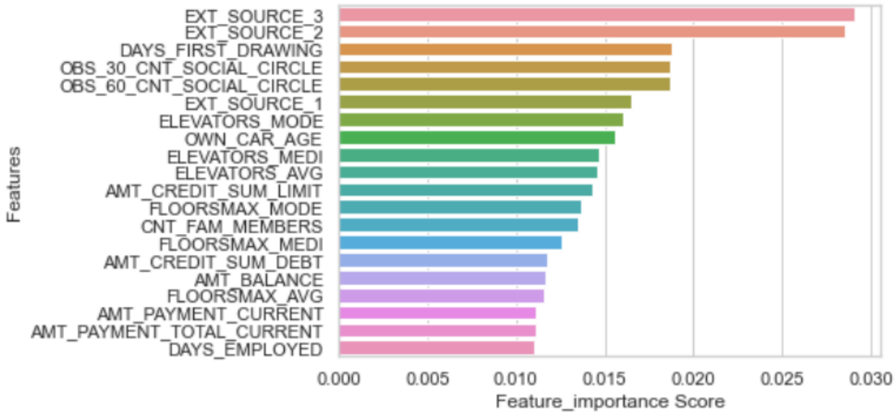
Home Credit B.V. is an international non-bank financial lending institution founded in 1997 in the Czech Republic and headquartered in Netherlands. Home Credit is already using various statistical and machine learning methods to make the best predictions as to who will most likely be a defaulter. It is important to ask the right question, get the correct information from possible loan takers, and find patterns among defaulters and non-defaulters to feed into statistical methods. The project focuses on finding the most important features that characterizes a defaulter and a non-defaulter.

The data consisted of five individual tables namely – User Installment Payments, User Credit Card Balance, User POS Cash, Previous Application Data and Bureau Information. User installment payments consisted of repayment history for the previously disbursed credits in Home Credit related to the loans in the sample data. User Credit Card Balance has monthly balance snapshots of previous credit cards usage that the applicant has with Home Credit. User POS cash comprises of monthly balance details of previous POS (point of sales) and cash loans that the applicant had with Home Credit. Previous application data consisted of all previous applications for Home Credit loans of user. Lastly, Bureau information referred to all client's previous credits provided by other financial institutions that were reported to Credit Bureau. Data Cleaning and data manipulation tasks such as missing value treatment, type casting and data transformation were performed on each table. The final master table was then prepared by merging each table on the primary key (SK_ID_CURR). The resultant master table consisted of 307511 entries and 251 features.

After extensive data cleansing and data wrangling, the next aim was to identify the most important features that classify a defaulter and a non-defaulter. In order to implement classification algorithms, the data was split into training (80%) and testing (20%) subsets. Before the split, to handle the class imbalance, SMOTE technique was implemented on the Defaulter class as it was the minority. To understand the importance of all features, two classification models were built. The first model was designed where we considered all users, including users that did not have any credit history. On the contrary, the second model comprised of all users who has credit history. On comparison, it is understood that the second model performed much better than the first model (as seen in the figure below). This proves that that prior historical credit history makes a difference in classifying a customer as a defaulter or a non-defaulter.



Finally, the Random Forest classifier was used to further understand the top features that described the target group (defaulter vs non-defaulters). Based on importance, the top features recognized are as follows:



"EXT_SOURCE_3", "EXT_SOURCE_2" and "EXT_SOURCE_1" were recognized as some of the most important features. This makes sense as these features provide credit scores of the client's prior transactions as maintained by third party sources. Therefore, Home Credit would ideally want to utilize these scores in determining whether to give out a loan or not. The social circle feature provides details into the group that is responsible for paying in case the client is unable to pay back the loan. The features describing an individual's house played an important role in determining a defaulter and a non-defaulter too. Lastly, the number of days an individual is employed also helped Home Credit in its classification.

In conclusion, it can be inferred that along with the customer's credit history, personal attributes such as the value of one's assets and sources of income play a vital role in determining whether an individual would be categorized as a defaulter. Home Credit can implement these features into their existing statistical and machine learning models to better understand their customer portfolio.

References:

https://www.kaggle.com/c/home-credit-default-risk

https://www.homecredit.net/about-us.aspx

https://github.com/Chaitanya2593/home_credit_risk_analysis

https://public.tableau.com/app/profile/mvs.chaitanya/viz/CreditRiskAnalysis_16362852398140/External_sources_1