Outline

Problem Research

Data Pre-Processing

Data Visualization & Business Insight

Machine Learning Implementation

Machine Learning Evaluation

Business Recommendation

Problem Research

- Home Credit is currently using various statistical methods and Machine Learning to make credit score predictions. Based on the submitted credit loan application data, 10 prediction models were made with a classification algorithm and then compared them to obtain the best predictive model in determining whether the loan application was accepted or not.
- The dataset used is application_train.csv
- 10 Machine Learning Models used are Random Forest Classifier, Support Vector Classifier (SVC), Logistic Regression, Decision Tree Classifier, KNeighbors Classifier, Gradient Boosting Classifier, AdaBoost Classifier, XGBoost Classifier, CatBoost Classifier, and ExtraTrees Classifier.

Data Pre-Processing

- Checking for missing values and there are many missing values
- Delete columns containing missing values
- Delete a column whose contents are all zero
- Create an Age column with the formula: (DAYS_REGISTRATION-DAYS_BIRTH)/365
- Check for duplicate data, no duplicate data
- Delete the unnecessary columns: SK_ID_CURR, DAYS_BIRTH, DAYS_EMPLOYED, DAYS_REGISTRATION, DAYS_ID_PUBLISH, WEEKDAY_APPR_PROCESS_START, and HOUR_APPR_PROCESS_START.
- Converting categorical data to binary

Date Pre-Processing

- Overcome columns with outliers, namely: CNT_CHILDREN, AMT_INCOME_TOTAL, AMT_CREDIT, AMT_ANNUITY, AMT_GOODS_PRICE, REGION_POPULATION_RELATIVE, OWN_CAR_AGE, CNT_FAM_MEMBERS, and EXT_SOURCE_2
- Divide the dataset into 80% training data and 20% testing data
- Normalize numeric data with MinMaxScaler()
- Oversampling (SMOTE) for label balancing
- Reduce overfitting by using PCA

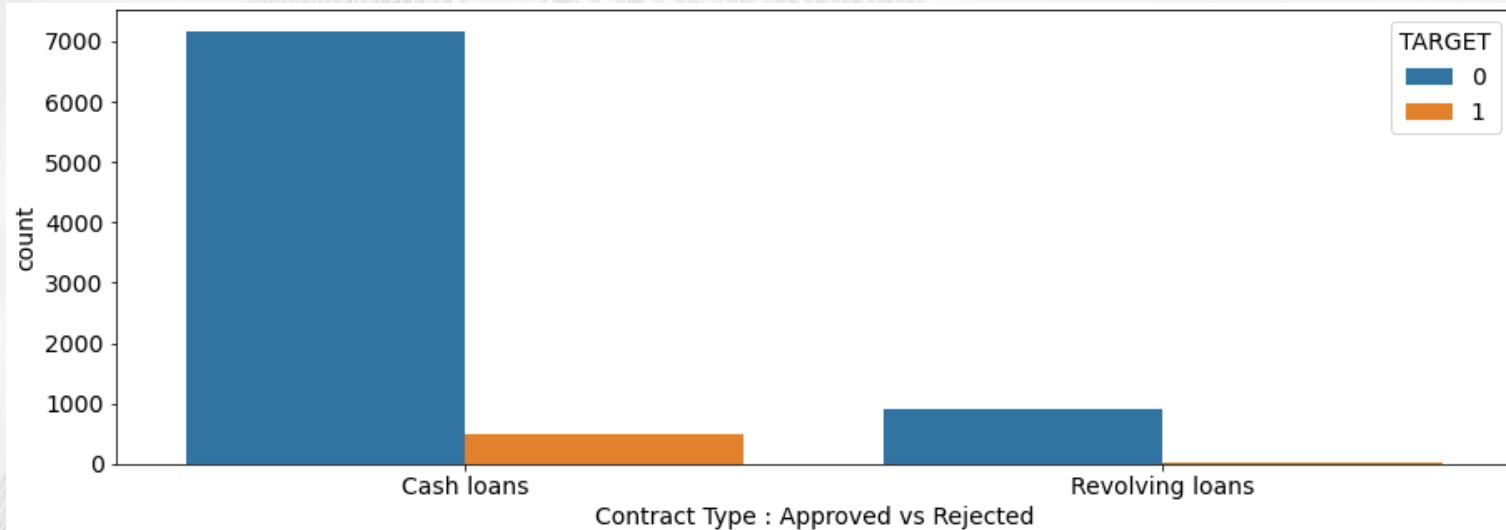# Data Visualization & Business Insight

Most of the customers who received and rejected loans were male.
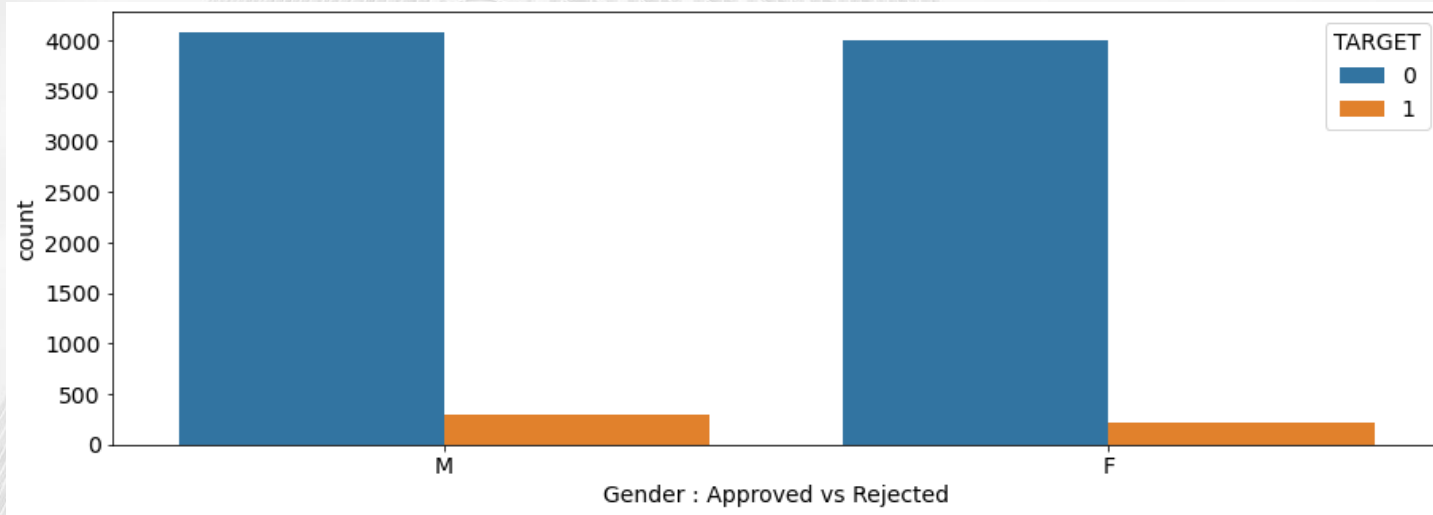
Data Visualization & Business Insight

The type of loan contract that is more widely accepted and rejected is Cash Loans.

Machine Learning Implementation

There are 12 Machine Learning models using hyperparameter tuning, namely:
1. Random Forest Classifier
2. Support Vector Classifier (SVC)
3. Logistic Regression
4. Decision Tree Classifier
5. KNeighbors Classifier
6. Gradient Boosting Classifier
7. AdaBoost Classifier
8. XGBoost Classifier
9. CatBoost Classifier
10. ExtraTrees Classifier

	Model	Score
0	Random Forest Classifier	0.944218
1	Support Vector Classifier (SVC)	0.944218
4	KNeighbors Classifier	0.944218
6	AdaBoost Classifier	0.944218
9	ExtraTrees Classifier	0.944218
7	XGBoost Classifier	0.943637
8	CatBoost Classifier	0.941894
5	Gradient Boosting Classifier	0.920395
2	Logistic Regression	0.777455
3	Decision Tree Classifier	0.763510

Machine Learning Evaluation

The model with the best accuracy is the Random Forest Classifier with an accuracy of 94.42%.
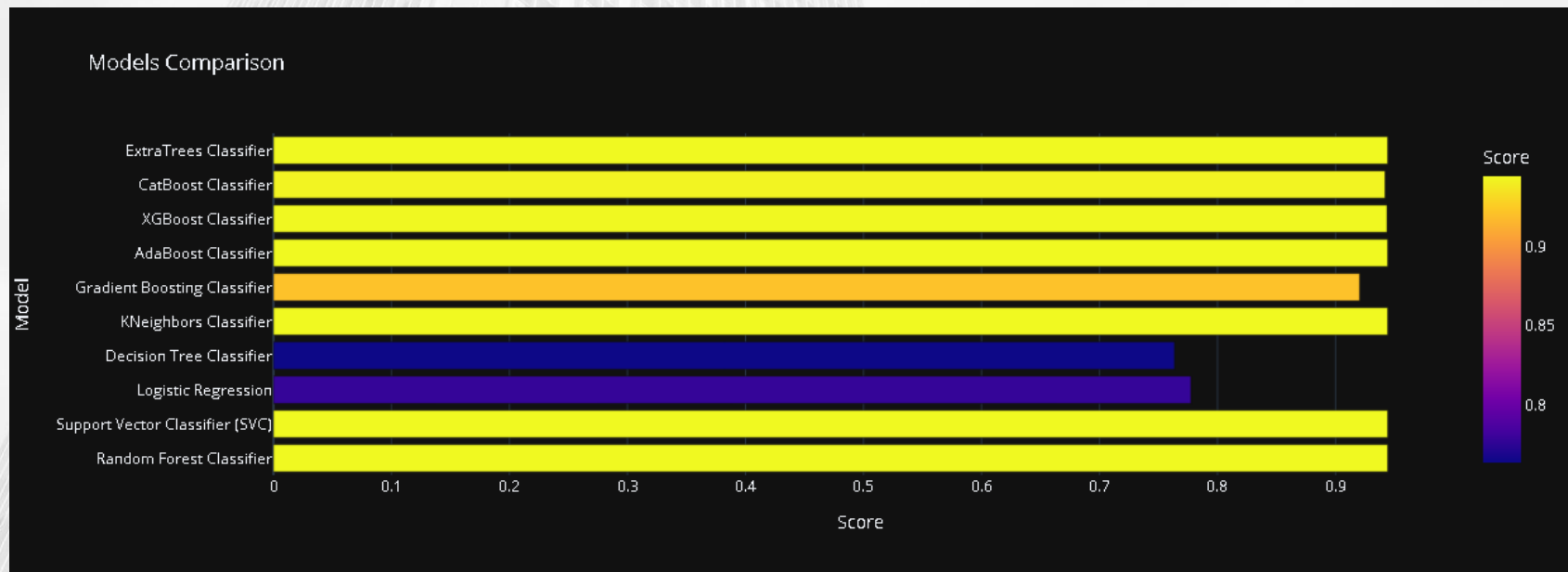
Business Recommendation

- The type of loan contract that is more widely accepted is Cash Loans, so it is necessary to increase the reserve of loan funds large enough to be able to accommodate the approved Cash Loans.

- Fewer types of Revolving Loans are accepted, so it is necessary to improve marketing strategies related to Revolving Loans products.

- The type of loan contract that is rejected the most is Cash Loans, therefore it is necessary to review the investigation process of prospective customers, for this reason, they can use the Random Forest machine learning model with an accuracy of 94.42%.

Thank You!

LinkedIn	https://www.linkedin.com/in/trianto-haryo-nugroho/
GitHub	https://github.com/triantonugroho/Home-Credit-Score-Card-Model

HOME
CREDIT
Kamu Bisa!