

# Virtual Internship Experience

Home Credit Score Card Model

By: Trianto Haryo Nugroho

**HOME  
CREDIT**  
*Kamu Bisa!*



**HOME  
CREDIT**  
*Kamu Bisa!*

# Outline

Problem Research

Data Pre-Processing

Data Visualization & Business Insight

Machine Learning Implementation

Machine Learning Evaluation

Business Recommendation

# Problem Research



- Home Credit saat ini sedang menggunakan berbagai macam metode statistik dan Machine Learning untuk membuat prediksi skor kredit. Berdasarkan data aplikasi pinjaman kredit yang diajukan dibuatlah 10 model prediksi dengan algoritma klasifikasi kemudian membandingkannya untuk memperoleh model prediksi terbaik dalam menentukan apakah pengajuan pinjaman diterima atau tidak.
- Dataset yang digunakan adalah `application_train.csv`
- 10 Model Machine Learning yang digunakan yaitu Random Forest Classifier, Support Vector Classifier (SVC), Logistic Regression, Decision Tree Classifier, KNeighbors Classifier, Gradient Boosting Classifier, AdaBoost Classifier, XGBoost Classifier, CatBoost Classifier, dan ExtraTrees Classifier.

# Data Pre-Processing



**HOME  
CREDIT**  
*Kamu Bisa!*

- Mengecek missing values dan terdapat banyak missing values
- Menghapus kolom yang mengandung missing values
- Menghapus kolom yang semua isinya bernilai nol
- Membuat kolom Age dengan rumus:  $(\text{DAYS\_REGISTRATION} - \text{DAYS\_BIRTH}) / 365$
- Mengecek data duplikat, tidak terdapat data duplikat
- Menghapus kolom yang tidak diperlukan yaitu: SK\_ID\_CURR, DAYS\_BIRTH, DAYS\_EMPLOYED, DAYS\_REGISTRATION, DAYS\_ID\_PUBLISH, WEEKDAY\_APPR\_PROCESS\_START, dan HOUR\_APPR\_PROCESS\_START.
- Mengubah data kategorik menjadi biner



# Date Pre-Processing



**HOME  
CREDIT**  
*Kamu Bisa!*

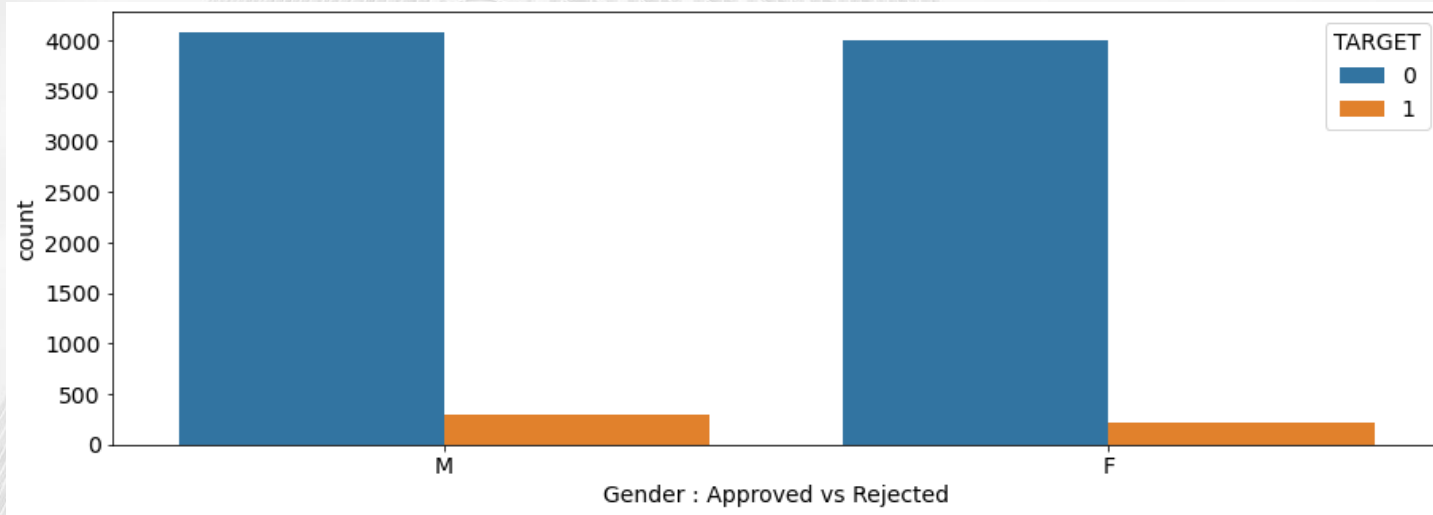
- Mengatasi kolom yang terdapat outliers yaitu: CNT\_CHILDREN, AMT\_INCOME\_TOTAL, AMT\_CREDIT, AMT\_ANNUITY, AMT\_GOODS\_PRICE, REGION\_POPULATION\_RELATIVE, OWN\_CAR\_AGE, CNT\_FAM\_MEMBERS, dan EXT\_SOURCE\_2
- Membagi dataset menjadi 80% data training dan 20% data testing
- Menormalisasi data numerik dengan MinMaxScaler()
- Melakukan oversampling (SMOTE) untuk balancing label
- Mengurangi overfitting dengan menggunakan PCA

# Data Visualization & Business Insight



**HOME  
CREDIT**  
*Kamu Bisa!*

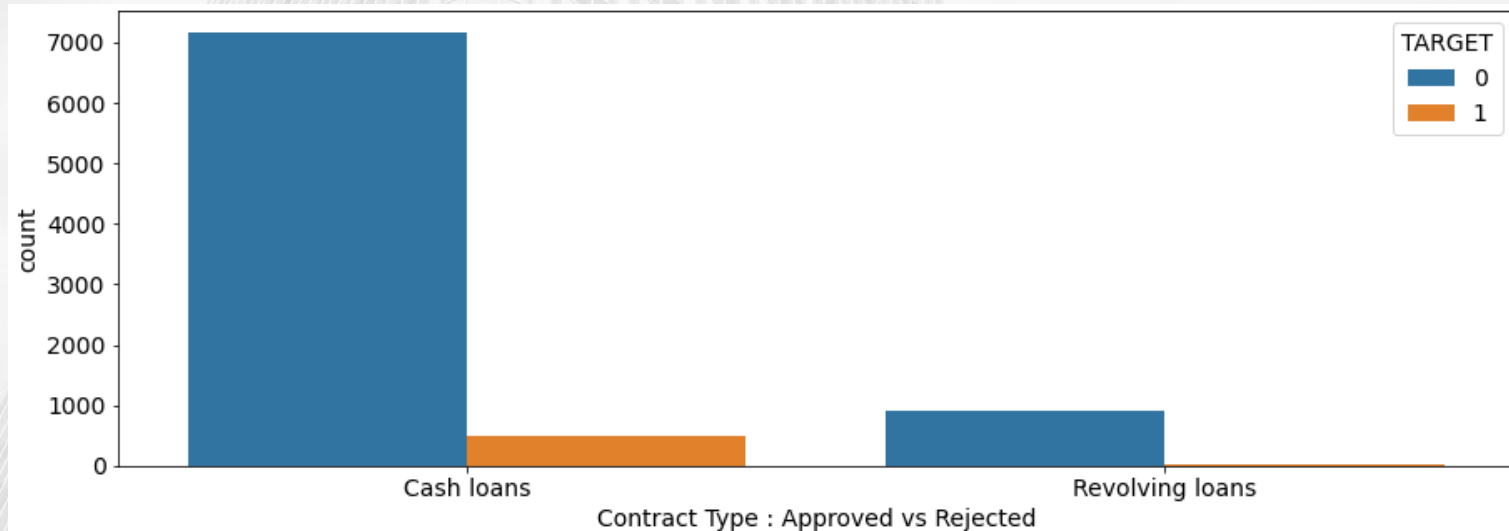
Nasabah yang lebih banyak diterima dan ditolak pinjamannya berjenis kelamin laki-laki.





# Data Visualization & Business Insight

Tipe kontrak pinjaman yang lebih banyak diterima dan ditolak adalah Cash Loans.





# Machine Learning Implementation

Model Machine Learning menggunakan hyperparameter tuning ada 12, yaitu:

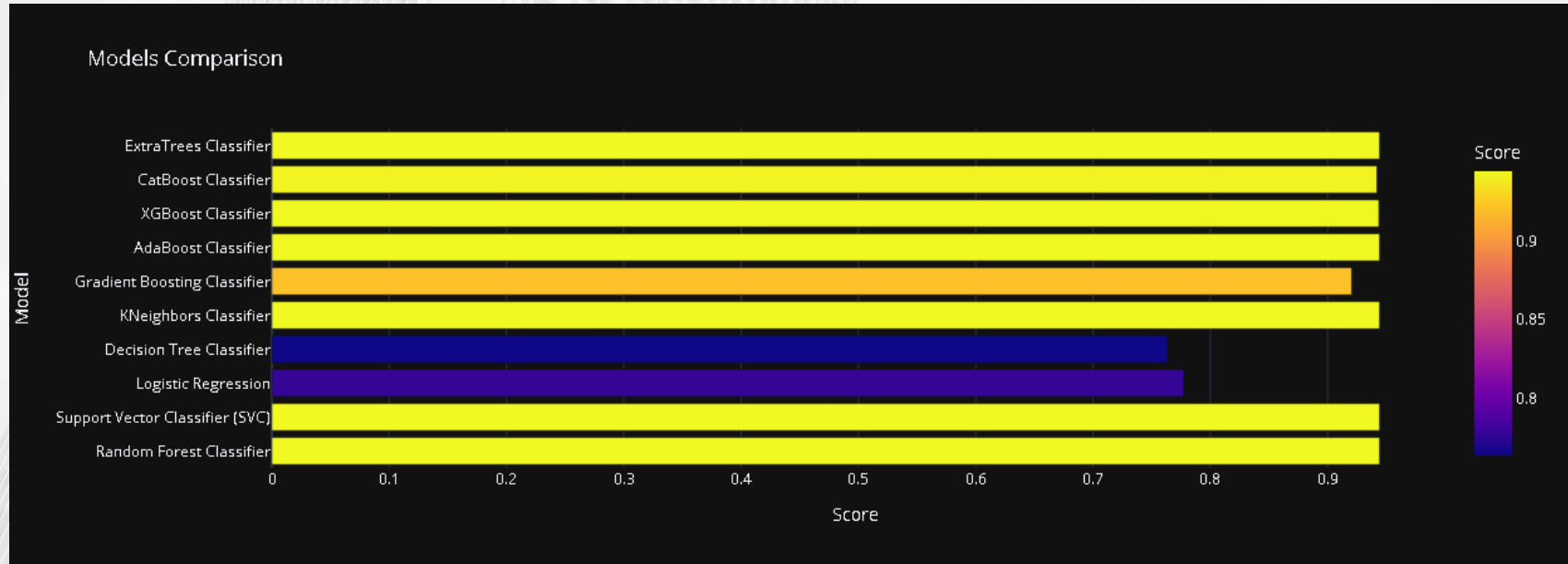
1. Random Forest Classifier
2. Support Vector Classifier (SVC)
3. Logistic Regression
4. Decision Tree Classifier
5. KNeighbors Classifier
6. Gradient Boosting Classifier
7. AdaBoost Classifier
8. XGBoost Classifier
9. CatBoost Classifier
10. ExtraTrees Classifier

	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

Model dengan akurasi terbaik adalah Random Forest Classifier dengan akurasi 94,42%.



# Business Recommendation

- Tipe kontrak pinjaman yang lebih banyak diterima adalah Cash Loans sehingga perlu ditingkatkan lagi cadangan dana pinjaman cukup besar untuk dapat mengakomodasi Cash Loans yang disetujui.
- Tipe kontrak pinjaman Revolving Loans lebih sedikit diterima sehingga perlu peningkatan strategi pemasaran terkait produk Revolving Loans.
- Tipe kontrak pinjaman yang lebih banyak ditolak adalah Cash Loans untuk itu perlu ditinjau kembali proses investigasi calon nasabah, untuk itu dapat menggunakan model machine learning Random Forest dengan akurasi sebesar 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!*