

HOME CREDIT SCORECARD MODEL

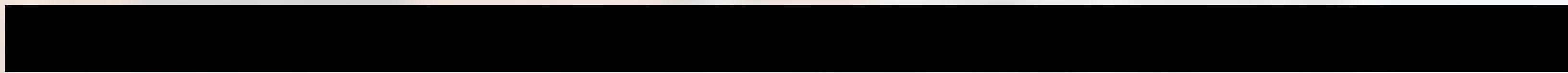
By: Nelly Yulia



HOME CREDIT INDONESIA DATA SCIENTIST PROJECT BASED INTERNSHIP PROGRAM

PROBLEMS

Home Credit Indonesia bertujuan membangun model *machine learning* yang dapat membantu tim dalam memprediksi apakah pengajuan pinjaman dari pelanggan akan menghadapi kesulitan dalam pembayaran kredit atau tidak. Dengan menganalisis data yang tersedia, Home Credit Indonesia berupaya untuk mengidentifikasi karakteristik pelanggan yang cenderung lancar dalam proses pembayaran kredit, sehingga dapat memberikan kontribusi positif terhadap peningkatan pendapatan perusahaan.



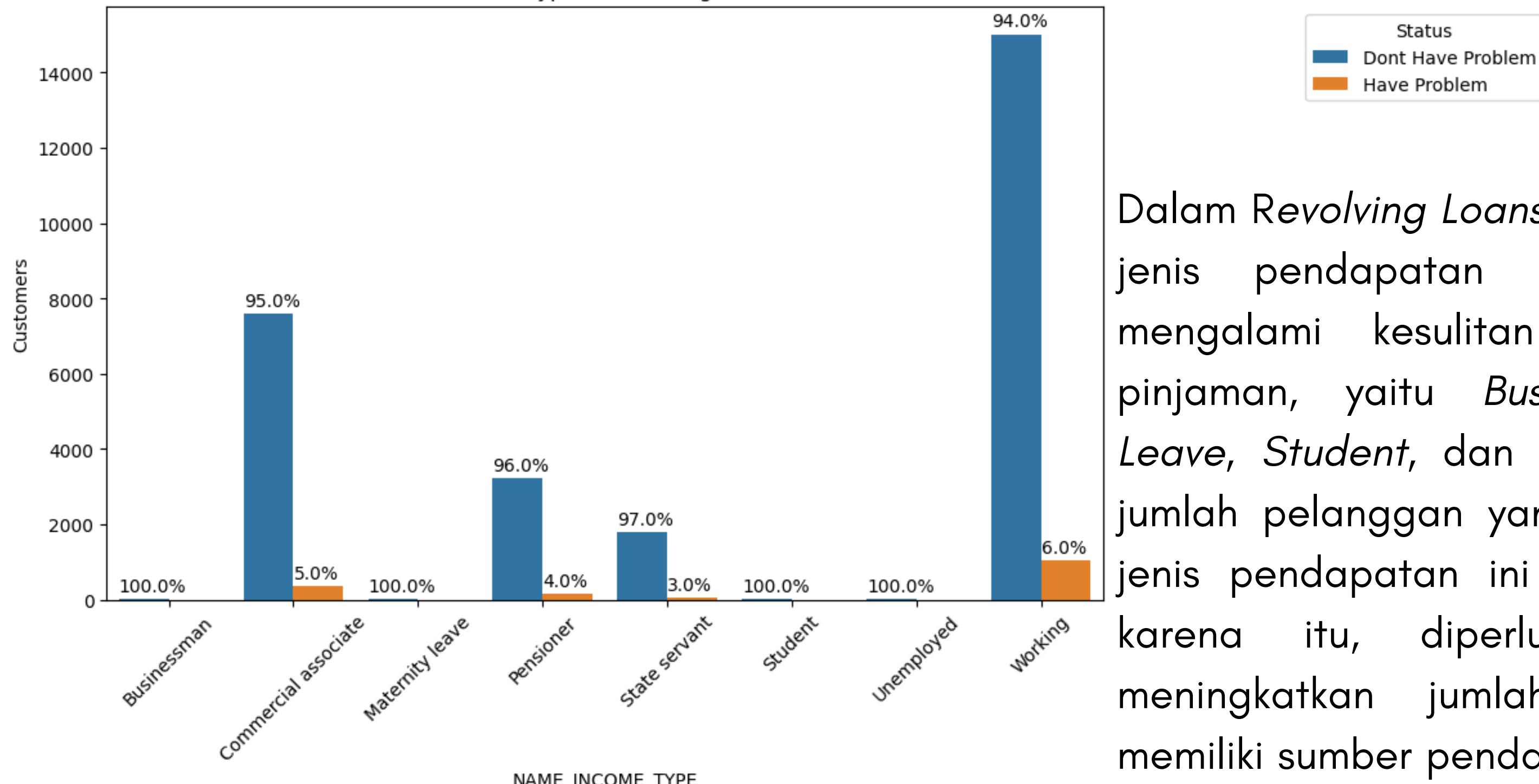
DATASET

Data yang disediakan oleh tim Home Credit Indonesia terdiri dari 122 kolom dan 307 baris yang mencakup informasi mengenai pelanggan yang mengalami masalah atau tidak dapat melunasi pinjaman yang telah diberikan oleh perusahaan. Data ini terbagi menjadi dua bagian, yaitu data pelatihan dan data uji aplikasi. Terdapat sebuah tabel utama yang dipisahkan menjadi 2 file, yaitu untuk pelatihan (*train*) dan untuk pengujian (*test*).



BUSINESS INSIGHT

Contract Type = Revolving Loans

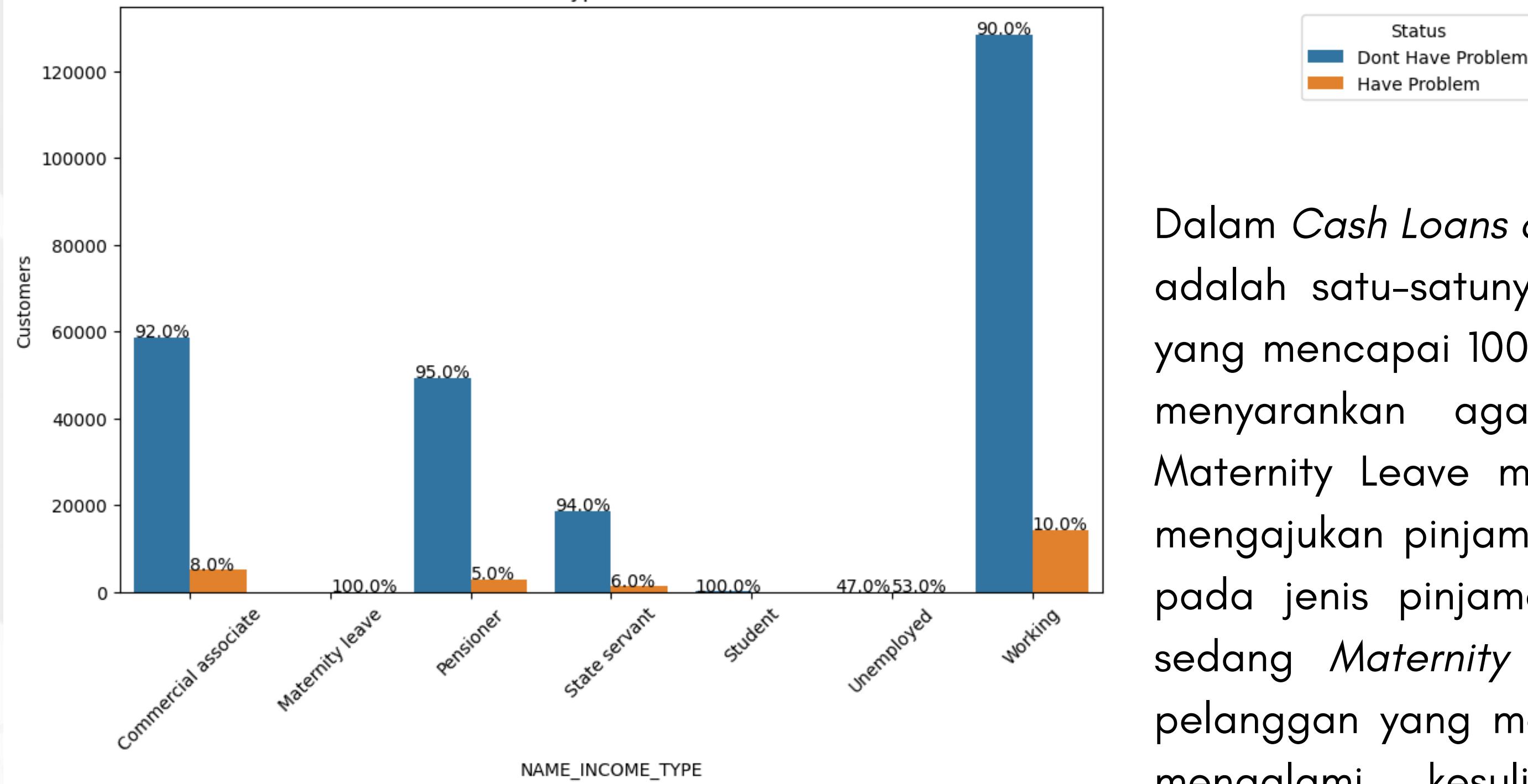


Contract Type (Jenis Kontrak)-(1)

Dalam *Revolving Loans contract*, terdapat 4 jenis pendapatan utama yang tidak mengalami kesulitan dalam membayar pinjaman, yaitu *Businessman*, *Maternity Leave*, *Student*, dan *Unemployed*. Namun, jumlah pelanggan yang berasal dari ke-4 jenis pendapatan ini relatif sedikit. Oleh karena itu, diperlukan upaya untuk meningkatkan jumlah pelanggan yang memiliki sumber pendapatan dari salah satu dari empat jenis pendapatan utama tersebut.

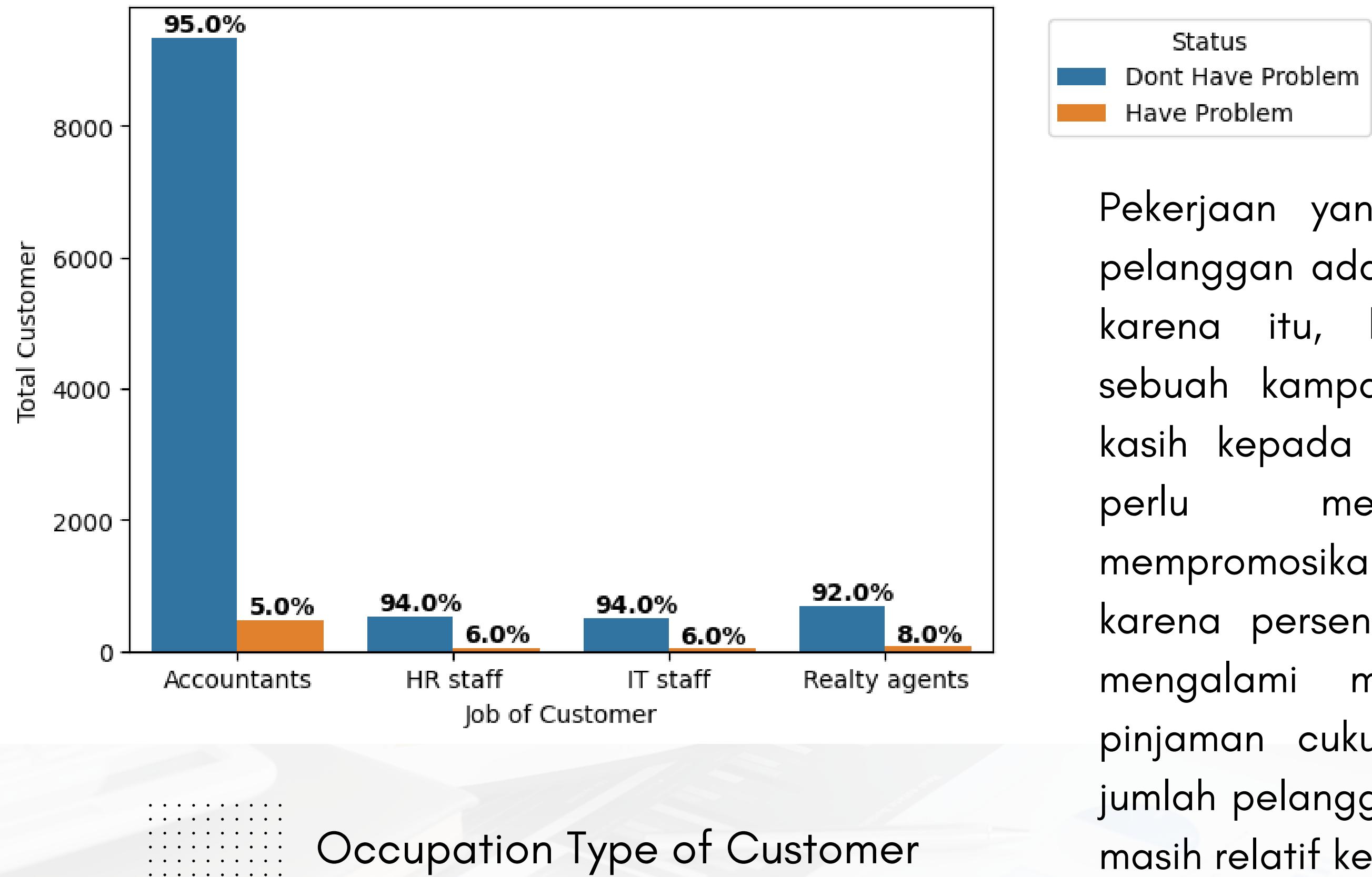
BUSINESS INSIGHT

Contract Type = Cash Loans



Dalam *Cash Loans contract*, *Maternity Leave* adalah satu-satunya yang memiliki tingkat yang mencapai 100%. Oleh karena itu, kami menyarankan agar klien yang sedang *Maternity Leave* mempertimbangkan untuk mengajukan pinjaman jenis bergulir, karena pada jenis pinjaman tersebut, klien yang sedang *Maternity Leave* memiliki tingkat pelanggan yang mencapai 100% dan tidak mengalami kesulitan dalam melunasi pinjaman.

.....
Contract Type (Jenis Kontrak)-(2)



Pekerjaan yang paling umum di antara pelanggan adalah profesi Accountant, oleh karena itu, kami harus merencanakan sebuah kampanye sebagai tanda terima kasih kepada mereka. Namun, kami juga perlu mempertimbangkan untuk mempromosikan pekerjaan lain dalam grafik, karena persentase pelanggan yang tidak mengalami masalah dalam membayar pinjaman cukup tinggi. Meskipun begitu, jumlah pelanggan dari 3 pekerjaan tersebut masih relatif kecil.

MACHINE LEARNING MODEL

DATA PREPARATION

1. Make a new column (age)
2. Feature selection
3. Remove Outlier
4. Feature Encoding (One hot encoding and label encoding)
5. Feature Transformation

HANDLING IMBALANCE DATA

1. Oversampling
2. Undersampling

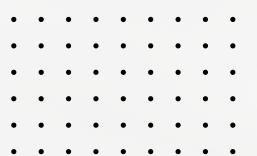
TRAIN MACHINE LEARNING MODEL

1. Logistic Regression
2. XGBoost
3. Random Forest
4. Desicion Tree

RESULT

No.	Model	Accuracy Tran Data	Accuracy Test Data	Precision	Recall	F1-Score
1.	Logistic Regression with Oversampling	62%	62%	62%	62%	62%
2.	Logistic Regression with <u>Undersampling</u>	62%	58%	62%	62%	62%
3.	XGBoost with Oversampling	96%	95%	96%	96%	96%
4.	Random Forest	100%	96%	96%	96%	96%
5.	Decision Tree	100%	91%	91%	91%	91%

Dari beberapa model tersebut, **Random Forest** menonjol sebagai yang **terbaik** dengan tingkat akurasi, presisi, dan recall yang lebih tinggi dibandingkan dengan model lainnya. Random Forest menjadi pilihan terbaik dari segi akurasi dan prediksi. Meskipun XGBoost memiliki akurasi yang tinggi (95%), Random Forest memiliki akurasi yang sama (96%) namun memiliki prediksi yang lebih baik (100%), menunjukkan kemampuan superior dalam mengklasifikasikan data.



SUMMARY

Lorem ipsum dolor sit amet,
consectetur adipiscing elit.

Suspendisse quis enim pretium,
bibendum ante ullamcorper,
tincidunt augue. Nunc sed lorem
aliquam, malesuada lectus eu,
placerat lorem.



+123-456-7890



hello@reallygreatsite.com



www.reallygreatsite.com



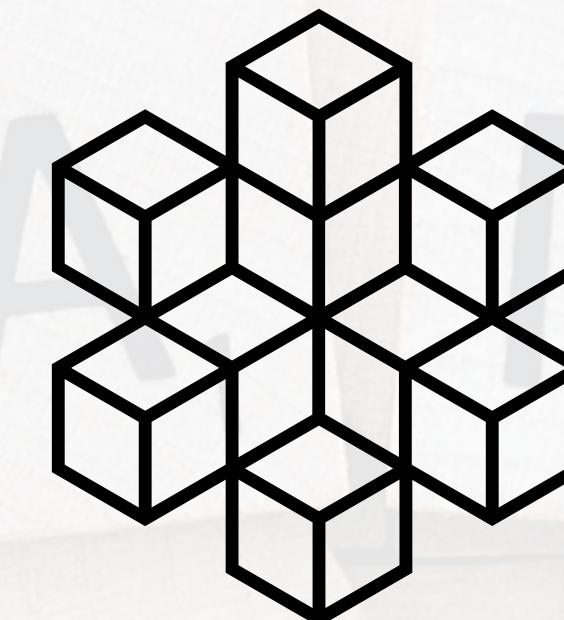
123 Anywhere St., Any City

SUMMARY

1. Dalam Revolving Loans Contract, perlu ditemukan pelanggan dengan berbagai jenis pendapatan seperti Businessmen, Maternity Leave, Students dan Unemployed.
2. Disarankan untuk menginisiasi kampanye khusus bagi pelanggan yang bekerja sebagai Akuntan karena profesi tersebut memiliki jumlah pelanggan terbanyak dan tingkat keberhasilan pembayaran yang tinggi.
3. Pengembangan iklan atau promosi yang ditargetkan kepada staf HR, staf IT, dan agen properti untuk meningkatkan aplikasi kredit.
4. Model Machine Learning yang dipilih adalah Random Forest. Meskipun XGBoost memiliki akurasi yang tinggi (95%), Random Forest memiliki akurasi yang sama (96%) tetapi memiliki prediksi yang lebih baik (100%), menunjukkan kemampuan superior dalam mengklasifikasikan data. Decision Tree memiliki akurasi 91% dan prediksi sempurna (100%), tetapi lebih rendah daripada XGBoost dan Random Forest yang memiliki akurasi lebih tinggi. Meskipun prediksi sempurna, akurasi yang rendah menunjukkan bahwa model tersebut mungkin tidak dapat memahami pola yang kompleks dengan baik seperti XGBoost dan Random Forest. Dengan akurasi yang lebih rendah, Decision Tree tidak seunggul XGBoost dan Random Forest.

THANK YOU

FOR YOUR ATTENTION



LINK GITHUB: [HTTPS://GITHUB.COM/NELYYULIAA/MODEL-HCI](https://github.com/nellyyuliaa/model-hci)