



We Care About
Your Future

Credit Scoring

Final Project

Oleh : Michael Y Sitanggang
Aji M Zapar
Rizki Setiawan I

Follow our social media on :



@data_bangalore



Data Bangalore



Data Bangalore Id





We Care About
Your Future

Business Understanding

Credit scoring adalah proses penilaian profil peminjam yang dilakukan oleh pihak pemberi pinjaman. Apakah peminjam layak mendapat pendanaan atau tidak.

Pada proses ini, profil calon peminjam akan diulas atau dianalisis, mulai dari profil dasar seperti data diri, usia, status pekerjaan, jabatan, masa kerja, gaji, status pernikahan, beban tanggungan keluarga, pekerjaan pasangan (suami/istri) dan lain sebagainya.



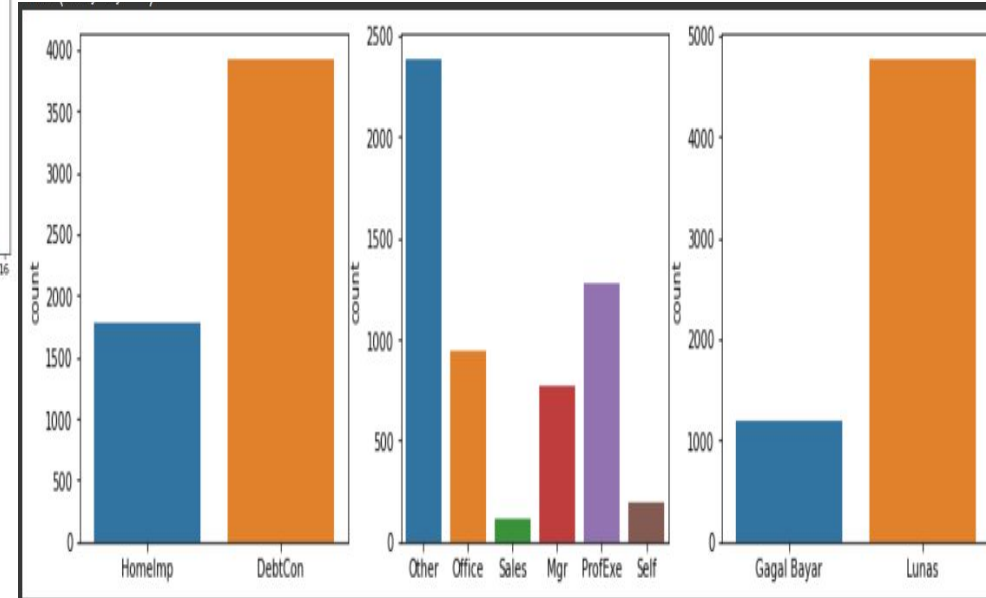
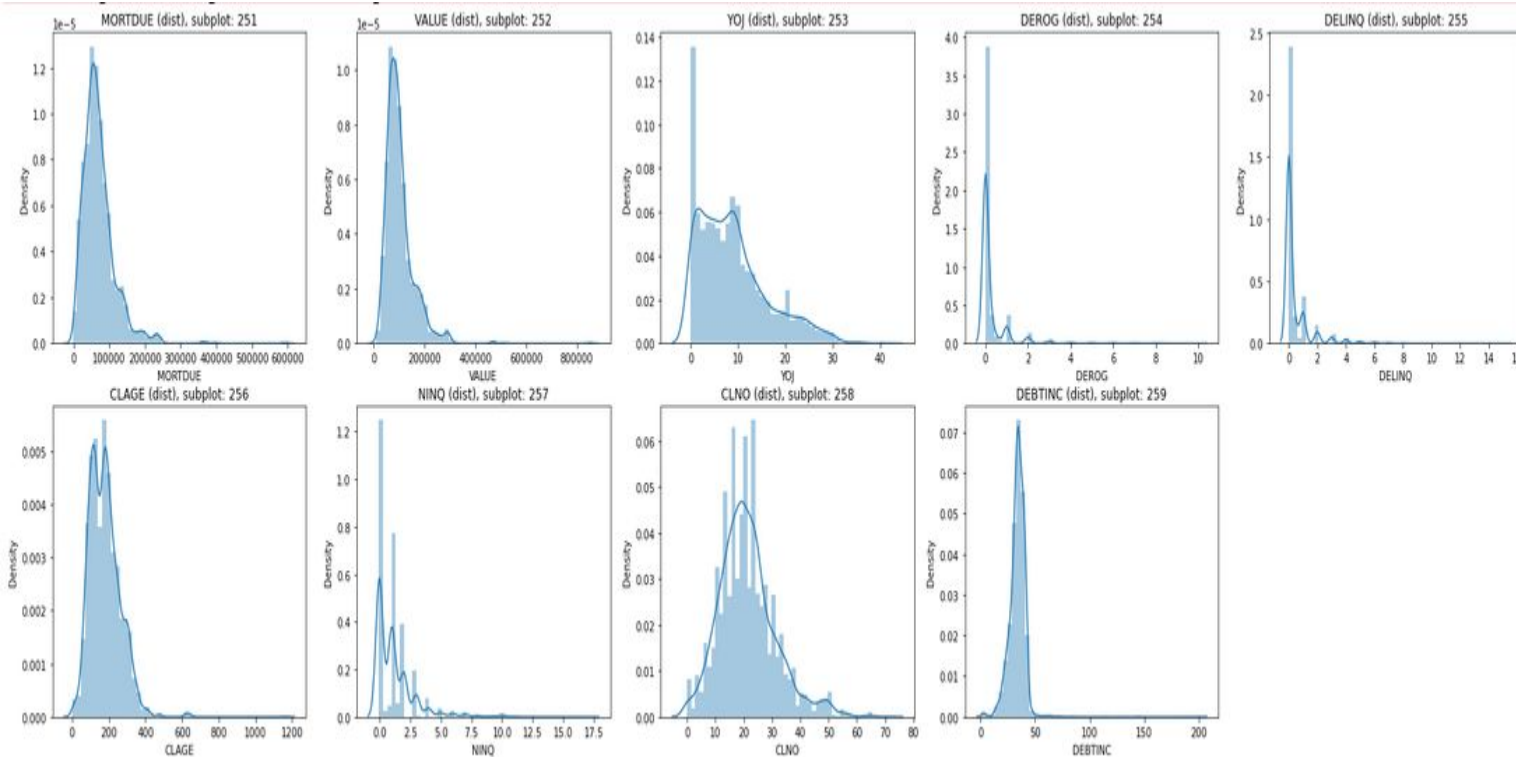


We Care About
Your Future

Data Understanding

BAD : 1 = klien gagal bayar pinjaman; 0 = pinjaman dilunasi
LOAN : Jumlah permintaan pinjaman
MORTDUE : Jumlah yang harus dibayar pada hipotek yang ada
VALUE : Nilai properti saat ini
REASON : DebtCon = konsolidasi hutang; HomImp = perbaikan rumah
JOB : Enam kategori pekerjaan

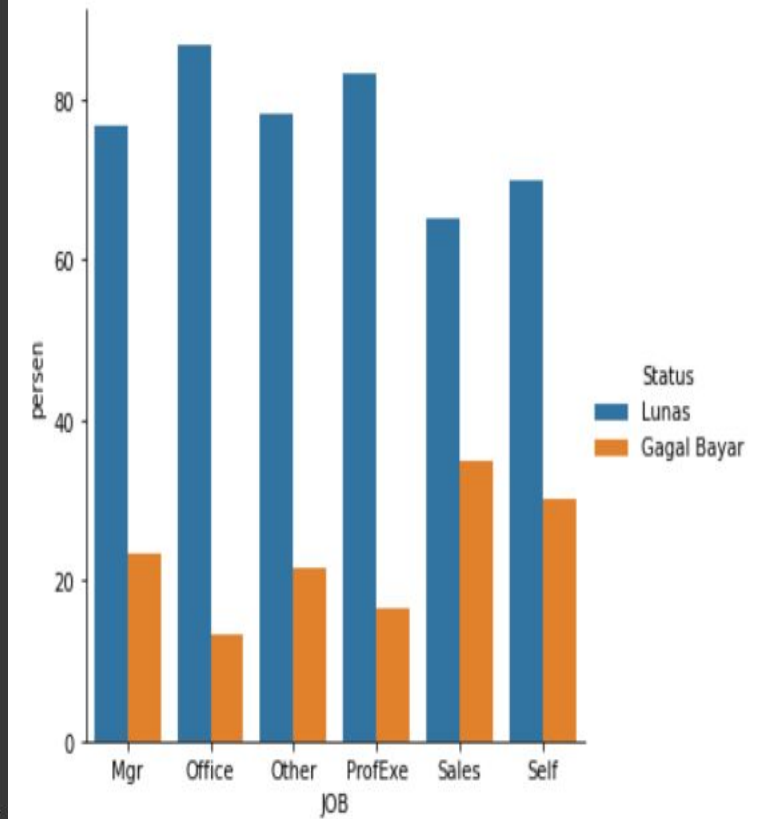
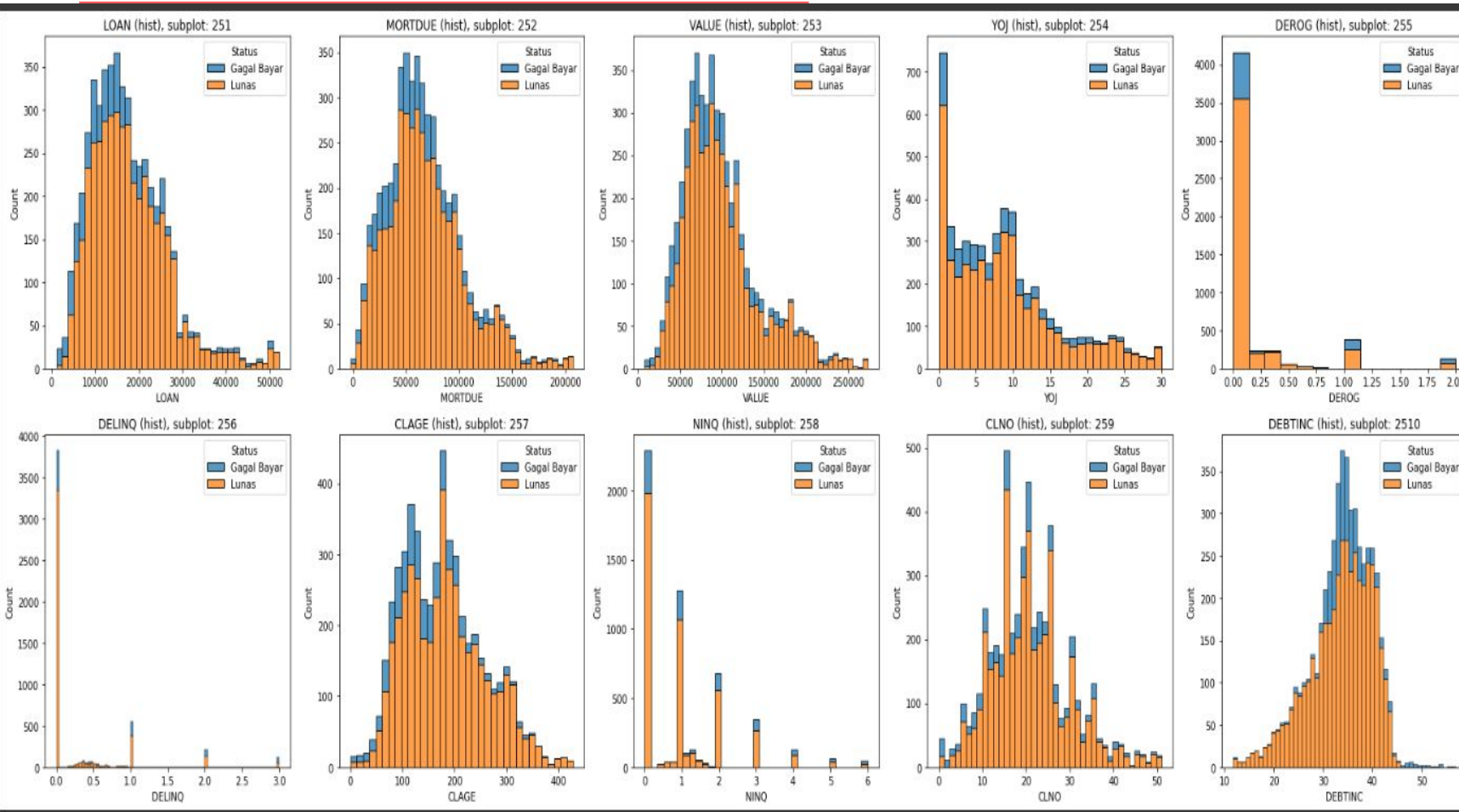
YOJ : Tahun bekerja saat ini
DEROG : Jumlah laporan penghinaan/ketidaksukaan utama
DELINQ : Jumlah kredit macet
CLAGE : Usia jalur perdagangan tertua dalam beberapa bulan
NINQ : Jumlah kredit terbaru yang ditanyakan
CLNO : Jumlah perdagangan
DEBTINC : Rasio utang terhadap pendapatan





We Care About
Your Future

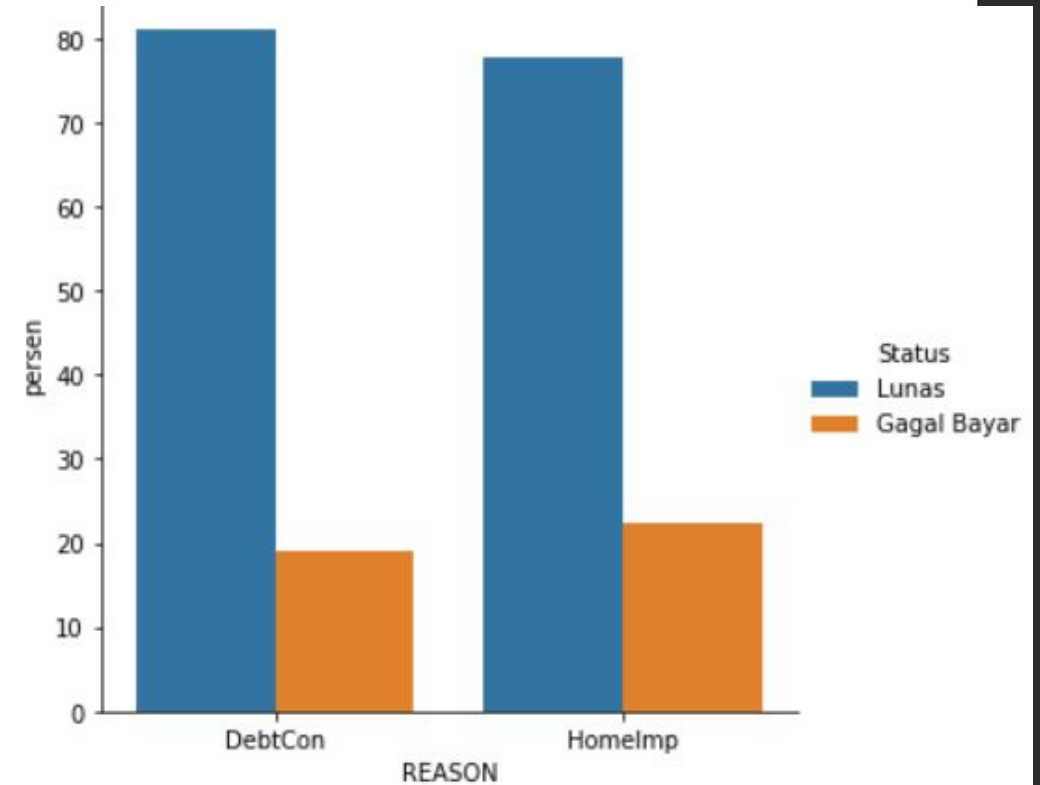
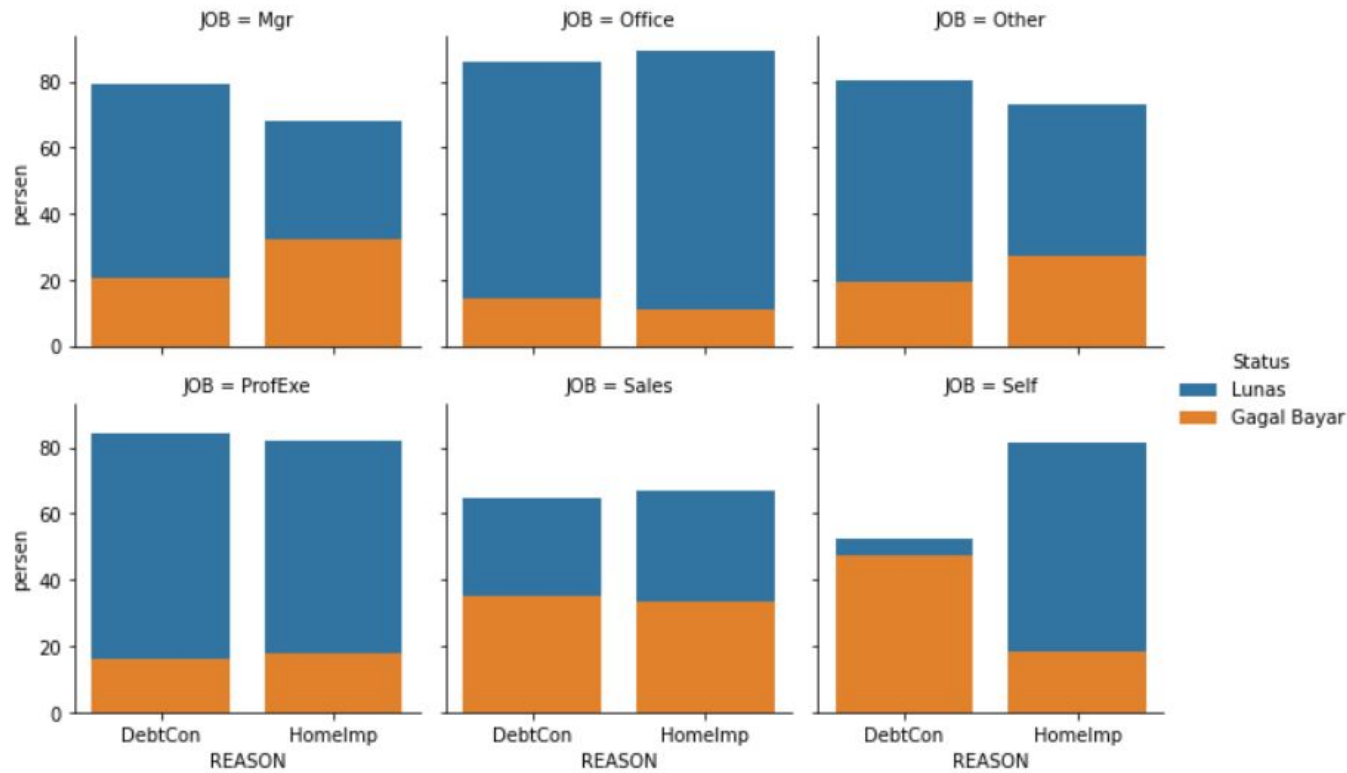
Data Understanding





We Care About
Your Future

Data Understanding



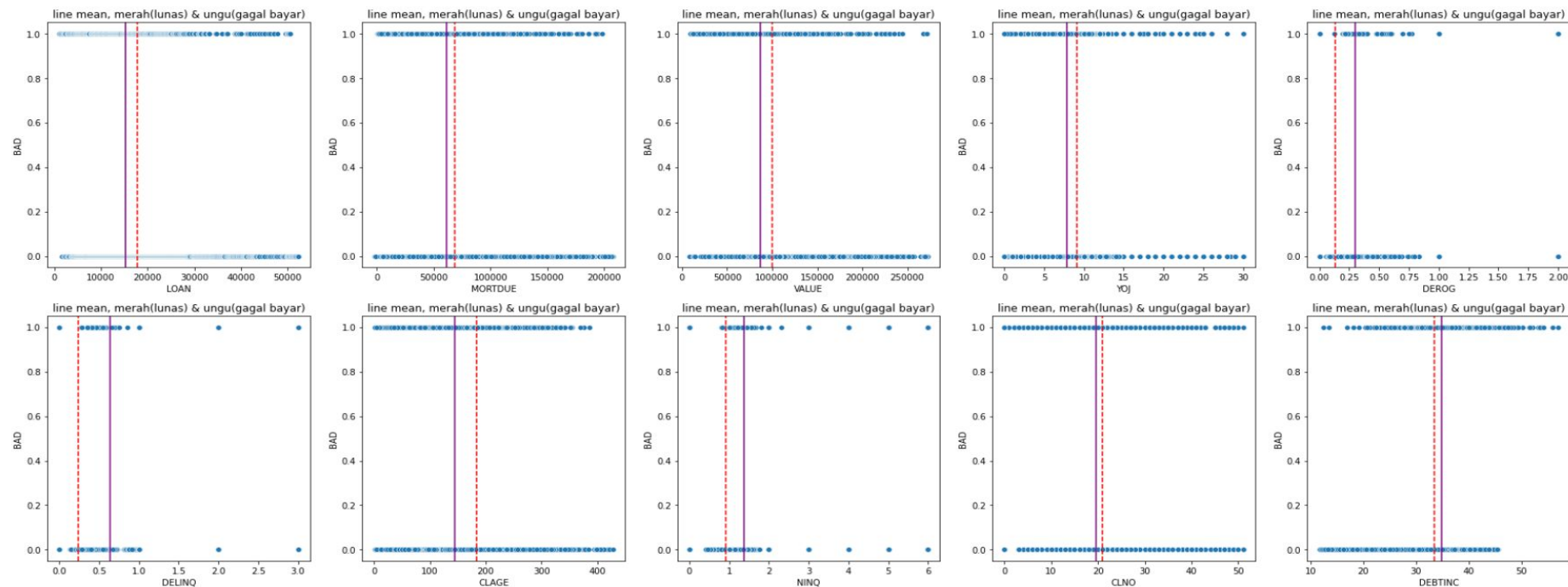


We Care About
Your Future

Data Understanding

- * LOAN : rata-rata jumlah permintaan pinjaman lunas > gagal bayar
- * MORTDUE : tidak ada perbedaan rata-rata jumlah hipotek yang harus dibayar pada pinjaman lunas dan gagal bayar
- * VALUE : tidak ada perbedaan rata-rata nilai properti pinjaman lunas dan gagal bayar
- * YOJ : rata-rata tahun bekerja pinjman lunas > gagal bayar

- * DEROG : rata-rata jumlah laporan penghinaan utama pinjaman gagal bayar > lunas
- * DELINQ : rata-rata jumlah kredit macet pinjaman gagal bayar > lunas
- * CLAGE : Usia jalur perdagangan tertua dalam beberapa bulan
- * NINQ : Jumlah kredit terbaru yang ditanyakan
- * CLNO : Jumlah perdagangan
- * DEBTINC : Rasio utang terhadap pendapatan





We Care About
Your Future

Data Preparation

Total missing values

Field	Nilai
DEBTINC	21.258389 %
DEROG	11.879195 %
DELINQ	9.731544 %
MORTDUE	8.691275 %
YOJ	8.640940 %
NINQ	8.557047 %
CLAGE	5.167785 %
JOB	4.681208 %
REASON	4.228188 %
CLNO	3.724832 %
VALUE	1.879195 %
BAD	0.000000 %
LOAN	0.000000 %

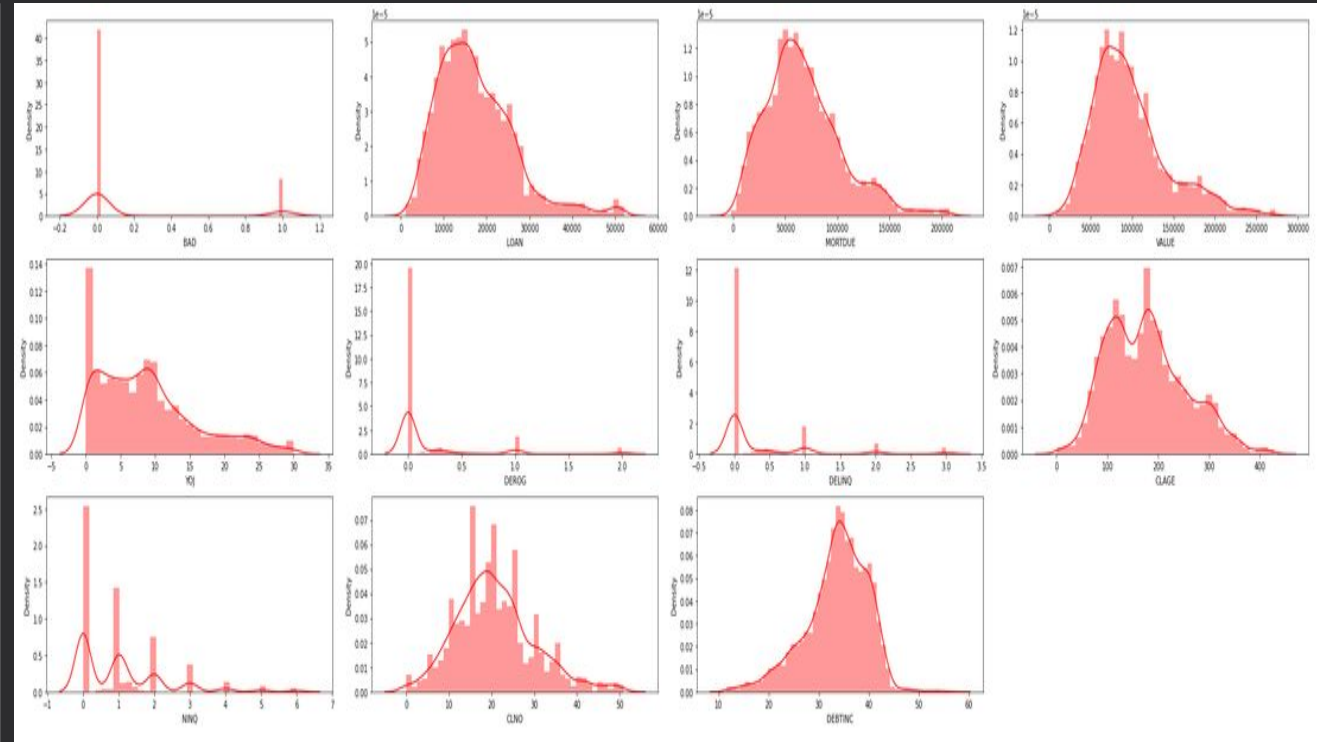




We Care About
Your Future

Data Preparation

- ❑ Mengisi missing values menggunakan MICE imputation
- ❑ Terlihat banyak data mendekati dist normal tetapi memiliki banyak outlier
- ❑ Menghapus outlier dengan zscore
- ❑ Mengisi missing value var kategori menggunakan modus
- ❑ Dataset dengan select feature
- ❑ Dataset tanpa select feature
- ❑ Check embalance var target (perlu dilakukan embalance data untuk data target)
- ❑ Memisahkan/partisi data latih dan test untuk model
- ❑ Melakukan teknik oversampling dalam menyeimbangkan variabel target
- ❑ Menggunakan Standart Scaller untuk yang distribusinya skew





Modelling

- > Melakukan pengukuran model performance dengan selected feature anova, chi square, dan non-selected feature
- > Non select feature memiliki akurasi lebih tinggi dibanding selected feature
- > Melakukan pengukuran ensemble modelling performance dari non-selected feature
- > KNN memiliki akurasi tertinggi namun XTREE memiliki sensitivitas dan presisi tertinggi, Bagging Random forest berada di urutan ke-3
- > Melakukan tuning parameter dari 3 model terbaik
- > Menggunakan XTREE sebagai parameter terbaik
- > Check overfitting dengan hasil training set accuracy : 1.0 dan test set accuracy 0.9258555133079848

	Model	Akurasi	Presisi	Sensitivitas
1	k-Nearest Neighbors	0.955239	0.948761	0.962517
7	extra tree	0.955086	0.989234	0.989234
9	bagging randomforest	0.950397	0.975517	0.975517
6	bagging	0.921794	0.956466	0.956466
2	Decision Tree	0.901754	0.919203	0.877849
8	gradient boosting	0.880755	0.918800	0.918800
11	xgboosting	0.879302	0.914301	0.914301
5	ada boosting	0.857320	0.878508	0.878508
10	stochastic gradient	0.818699	0.886333	0.886333
0	Logistic Regression	0.815628	0.886613	0.723698
4	Support Vector Machine	0.802223	0.946403	0.650668
3	Naive Bayes	0.757446	0.765285	0.737901

Tuning Parameter

Extra Tree Classifier - ROC-AUC Score: 0.99519

with best parameters: {'criterion': 'gini', 'max_depth': None, 'max_features': 'auto', 'min_samples_leaf': 1}.

Extra Tree Classifier - Accuracy Score: 0.95622

with best parameters: {'criterion': 'entropy', 'max_depth': None, 'max_features': 'log2', 'min_samples_leaf': 1}.

k-Nearest Neighbors - ROC-AUC Score: 0.99819

with best parameters: {'metric': 'manhattan', 'n_neighbors': 7, 'weights': 'distance'}.

k-Nearest Neighbors - Accuracy Score: 0.98788

with best parameters: {'metric': 'manhattan', 'n_neighbors': 1, 'weights': 'uniform'}.

Random Forest Classifier - ROC-AUC Score: 0.99532

with best parameters: {'bootstrap': False, 'criterion': 'gini', 'max_depth': 20, 'max_features': 'log2', 'n_estimators': 100}.

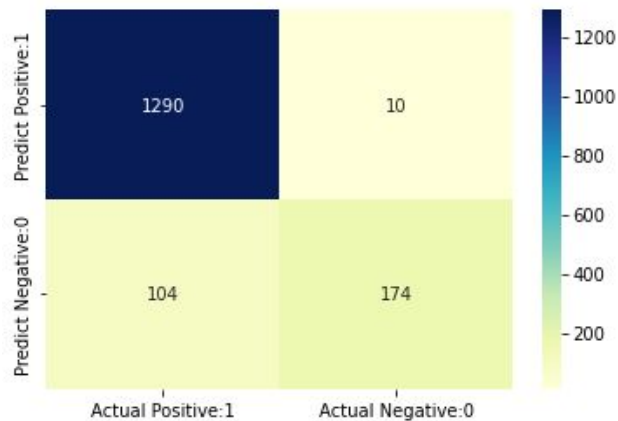
Random Forest Classifier - Accuracy Score: 0.95444

with best parameters: {'bootstrap': False, 'criterion': 'gini', 'max_depth': 100, 'max_features': 'auto', 'n_estimators': 100}.

- > Terlihat KNN lebih baik dari Xtree dan RF namun karena semua model memiliki tingkat FN lebih tinggi setelah dilakukan uji coba, maka Xtree dan Random Forest lebih baik untuk dilakukan pengoptimalan nilai FP dan FN.
- > Xtree lebih memiliki tingkat sensitivitas lebih baik dari Random Forest, maka tim memilih Xtree untuk dieksekusi

Evaluation

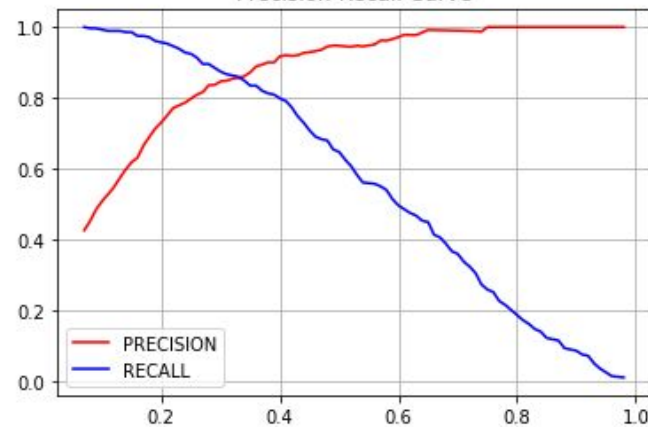
Confusion matrix



Hasil confusion matriks dari xtree menghasilkan:

True Positives(TP) = 1290
True Negatives(TN) = 174
False Positives(FP) = 10
False Negatives(FN) = 104

Precision-Recall Curve



Karena memilih FP lebih baik dari FN maka model akan dioptimalkan :

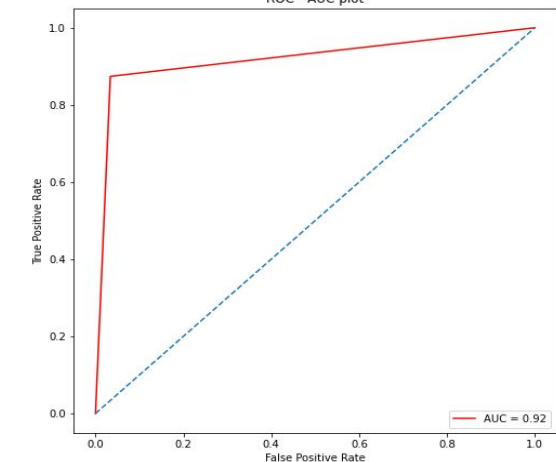
melakukan threshold dan didapatkan nilai 0.3

Confusion matrix

```
[[1256  44]  
 [ 35 243]]
```

True Positives(TP) = 1256
True Negatives(TN) = 243
False Positives(FP) = 44
False Negatives(FN) = 35

ROC - AUC plot



ROC-AUC = 0,92



We Care About
Your Future

Thanks For Your Attention.

Follow our social media on :



@data_bangalore



Data Bangalore



Data Bangalore Id

