

## Stage 3 - Machine Learning Evaluation & Supervised Learning

### Pembagian Tugas:

Random Forest : Riza, Bilqis

Logistic Regression : Alyani, Windy

Decision Tree : Hanif, Fahru , Rafi

1. **Modeling (70 poin)** Lakukan modeling sesuai task yang ditentukan di awal dari hasil data yang telah dilakukan preprocessing dan cleaning pada tahapan sebelumnya:

#### A. Split Data Train & Test

Dalam setiap algoritma yang digunakan, split data dilakukan dengan persentase 80% data train dan 20% data test.

#### B. Modeling (Algoritma yang diimplementasikan tidak terbatas yang diajarkan di kelas)

Proses Modeling dilakukan dengan menggunakan tiga algoritma Supervised Learning yaitu Logistic Regression, Decision Tree, dan Random Forest pada data yang telah di grouping dan data mentah dari Kaggle yang telah mengalami proses Data Pre-processing.

No.	Algoritma	Data
1	Logistic Regression	Data Group
2		Data Mentah
3	Decision Tree	Data Group
4		Data Mentah
5	Random Forest	Data Group
6		Data Mentah

### C. Model Evaluation: Pemilihan dan perhitungan metrics model

Model Evaluation yang digunakan adalah Confusion Matrix dengan fokus kepada Precision, Recall, dan F1-Score

No	Algoritma	Data	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	Logistic Regression	Data Group	0.52	0.53	0.53
2		Data Mentah	0.88	1	0.93
3	Decision Tree	Data Group	0.65	0.63	0.73
4		Data Mentah	0.51	0.58	0.54
5	Random Forest	Data Group	0.66	0.73	0.69
6		Data Mentah	0.60	0.53	0.57

### D. Model Evaluation: Apakah model sudah best-fit? Hindari Overfit/Underfit. Validasi dengan cross-validation

Logistic Regression :

- Data Group : Nilai cross-validation untuk model ini berkisar antara 0,521 sampai 0,531 yang menunjukkan bahwa model tidak berkinerja baik dengan data dengan akurasi rata-rata sekitar 52,6%. Variabilitas nilai di seluruh folds menunjukkan bahwa model tidak berkinerja dengan baik dan kemungkinan besar underfit. Underfit terjadi karena model terlalu sederhana untuk menangkap pola dari data yang mengakibatkan kinerja yang buruk, maka dari itu diperlukan untuk menggunakan model lain yang
- Data Mentah : Dengan menggunakan cross-validation sebanyak 5 folds dengan kesamaan nilai sekitar 87.7% menunjukkan bahwa model ini stabil, atau bisa dikatakan model ini sudah best-fit.

Decision Tree:

- Data group: Hasil evaluasi model menunjukkan tanda-tanda model yang sudah cukup baik (best-fit) tanpa mengalami overfitting atau underfitting yang signifikan. Dalam hal ini, ROC-AUC adalah metrik yang digunakan untuk mengukur performa model dalam membedakan antara kelas positif dan negatif. Perbedaan yang kecil antara ROC-AUC pada data latih (0.73) dan data uji (0.73) menunjukkan bahwa model tidak

mengalami overfitting, yang dapat mengindikasikan kemampuan model untuk secara konsisten mengidentifikasi kelas positif dan negatif. Selain itu, hasil dari cross-validation pada kedua data latih (0.7337) dan data uji (0.7236) juga menunjukkan tingkat konsistensi yang baik, menandakan bahwa model memiliki umumnya memiliki performa yang stabil di berbagai subset data. Dengan demikian, dapat dianggap bahwa model telah mencapai tingkat best-fit yang memadai.

- Data Mentah: Hasil evaluasi model menunjukkan tanda-tanda model yang sudah cukup baik (best-fit) tanpa mengalami overfitting atau underfitting yang signifikan. Dalam hal ini, metrik ROC-AUC digunakan untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. ROC-AUC pada data latih (0.98) yang mendekati sempurna dan ROC-AUC pada data uji (0.85) yang masih tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan kelas-kelas tersebut. Perbedaan antara kedua nilai ROC-AUC ini tidak terlalu besar, menunjukkan bahwa model tidak mengalami overfitting yang signifikan. Selain itu, hasil cross-validation pada data latih (0.9759) dan data uji (0.8971) menunjukkan tingkat konsistensi yang baik, menandakan bahwa model memiliki performa yang stabil di berbagai subset data. Dengan demikian, dapat dianggap bahwa model ini sudah mencapai tingkat best-fit yang memadai.

#### Random Forest

- Data Mentah: Dalam data hasil train menunjukkan rerata nilai akurasi yang cukup baik dengan Cross Validation Score juga baik sehingga model dapat dengan baik mengidentifikasi kelas Risk\_Flag dengan cukup baik.

Cross-Validation Scores: [0.89910714 0.90300099 0.90019841 0.89789187 0.89952877]

Rata-rata Akurasi: 0.8999

Dari data yang disajikan, terlihat bahwa model telah memberikan akurasi yang baik pada data latih. Akurasi rata-rata adalah sekitar 89.99%, sementara Cross-Validation Scores menunjukkan hasil yang konsisten dan baik pula, berada pada kisaran sekitar 89.7% hingga 90.3%.

Hasil ini menandakan bahwa model cenderung baik dalam mengidentifikasi kelas Risk\_Flag dari data yang diberikan. Namun, untuk memahami secara menyeluruh performa model, sangat penting untuk mempertimbangkan metrik lain, terutama jika distribusi kelas tidak seimbang (imbalanced classes) atau jika karakteristik khusus lainnya perlu dieksplorasi lebih lanjut.

- Data Group:

Rata-rata akurasi sekitar 0.6753 menunjukkan tingkat akurasi yang konsisten di seluruh lipatan dan model memiliki tingkat akurasi yang stabil dalam prediksi. Tidak terlihat adanya perbedaan yang signifikan antara skor akurasi pada setiap lipatan. Hal ini menunjukkan bahwa model tidak cenderung overfitting atau underfitting. Meskipun demikian, akurasi sekitar 0.6753 mungkin masih belum cukup tinggi sehingga diperlukan adanya peningkatan performa model.

## E. Hyperparameter Tuning

Di laporan homework, tuliskan eksperimen apa saja yang telah dilakukan dan metode yang digunakan, dan mengapa memilih dan mengimplementasikan algoritma tersebut. Jelaskan masing-masing hasil dari eksperimen model yang telah dilakukan, alasan menggunakan metrics pada model tersebut, serta hyperparameter yang digunakan dan pengaruhnya terhadap model.

1. Logistic Regression

- a. Data Mentah

Hasil penelitian menunjukkan bahwa parameter optimal yang ditemukan adalah 'C' dengan nilai 0.001. Nilai ini berkontribusi pada performa terbaik model dalam konteks algoritma yang digunakan. Kinerja model saat pelatihan mencapai akurasi sekitar 87.7%, menandakan kemampuan model dalam memprediksi dengan benar sebagian besar data train yang digunakan. Sementara itu, akurasi pada data test mencapai sekitar 87.6%, mengindikasikan kemampuan model dalam melakukan prediksi yang baik bahkan pada data yang belum pernah dilihat sebelumnya.

- b. Data Group

Ditemukan parameter terbaik, yaitu 'C' dengan nilai 0.001, yang kemudian digunakan dalam pembentukan model. Akurasi saat melatih model menunjukkan angka sekitar 52.7%, mengindikasikan bahwa model berhasil memprediksi sebagian

besar data train. Meski demikian, angka akurasi ini tidak terbilang tinggi. Begitu juga dengan akurasi pada data test, yang sekitar 52.8%. Akurasi pada kedua set data yang bernilai rendah dapat menandakan bahwa model ini mungkin mengalami kendala dalam melakukan prediksi dengan baik..

## 2. Decision Tree

### a. Data Mentah

Setelah melakukan hyperparameter tuning dengan metode Randomized Search pada model Decision Tree, terjadi peningkatan performa. Sebelum tuning, model memiliki akurasi sebesar 0.88, yang meningkat menjadi 0.89 setelah tuning. Namun, yang paling signifikan adalah peningkatan dalam Recall (recall rate) dari 0.58 menjadi 0.42, yang berarti model lebih baik dalam mengidentifikasi positif sejati. F1-Score juga mengalami peningkatan dari 0.54 menjadi 0.48, menunjukkan keseimbangan yang lebih baik antara Presisi dan Recall. Meskipun demikian, ROC-AUC tetap relatif stabil pada sekitar 0.87, menunjukkan bahwa kemampuan model dalam memisahkan kelas positif dan negatif tidak banyak berubah. Hasil Cross-Validation juga menunjukkan konsistensi model dalam performa. Dalam keseluruhan, hyperparameter tuning memberikan peningkatan performa pada model, terutama dalam hal kemampuan mengenali positif sejati dan keseimbangan antara Presisi dan Recall. ROC-AUC dan konsistensi model dalam penilaian tetap relatif stabil.

### b. Data Group

Eksperimen ini menunjukkan bahwa setelah melakukan hyperparameter tuning menggunakan Randomized Search, tidak ada peningkatan signifikan dalam performa model. Hasil sebelum dan sesudah tuning tetap serupa. Dalam kasus ini, metrik evaluasi seperti Accuracy, Precision, Recall, dan F1-Score tetap pada tingkat yang sama. Hal ini menunjukkan bahwa setelah eksplorasi kombinasi hyperparameter, tidak ada perubahan signifikan dalam kinerja model. Namun, ROC-AUC memiliki sedikit perbedaan antara sebelum dan sesudah hyperparameter tuning, meskipun perbedaannya sangat kecil. Dari hasil ini, dapat disimpulkan bahwa hyperparameter tuning mungkin tidak signifikan untuk model Decision Tree pada kasus ini, atau mungkin kombinasi hyperparameter yang kami jelajahi tidak memberikan perbaikan yang substansial. Itu sebabnya hasil sebelum dan sesudah tuning relatif sama. Dalam

eksperimen ini, kami lebih fokus pada Presisi, tetapi Presisi tetap tidak berubah dalam kasus ini. Hal ini menunjukkan bahwa hyperparameter yang kami jelajahi dalam Randomized Search belum mempengaruhi presisi dengan signifikan.

### 3. Random Forest

#### a. Data Mentah

Data hyperparameter tuning menunjukkan parameter terbaik yang dihasilkan setelah melakukan pencarian yang sistematis pada kumpulan hyperparameter dalam algoritma pembelajaran mesin (machine learning). Dalam kasus ini, parameter terbaik untuk model yang digunakan adalah:

max\_depth: Tidak terbatas (None)  
min\_samples\_leaf: 1  
min\_samples\_split: 5  
n\_estimators: 100

F1 score terbaik yang dicapai oleh model setelah menggunakan parameter-parameter ini adalah sebesar 0.5685. F1 score mengukur presisi dan recall model secara bersamaan dan merupakan metrik umum yang digunakan untuk evaluasi kinerja model klasifikasi. Semakin tinggi nilai F1 score, semakin baik performa modelnya.

#### b. Data Group

Data hyperparameter tuning menunjukkan parameter terbaik yang dihasilkan setelah melakukan pencarian yang sistematis pada kumpulan hyperparameter dalam algoritma pembelajaran mesin (machine learning). Dalam kasus ini, parameter terbaik untuk model yang digunakan adalah:

max\_depth: Tidak terbatas (None)  
min\_samples\_leaf: 2  
min\_samples\_split: 2  
n\_estimators: 100

F1 score terbaik yang dicapai oleh model setelah menggunakan parameter-parameter ini adalah sebesar 0.6947. F1 score mengukur presisi dan recall model secara bersamaan dan merupakan metrik umum yang digunakan untuk evaluasi

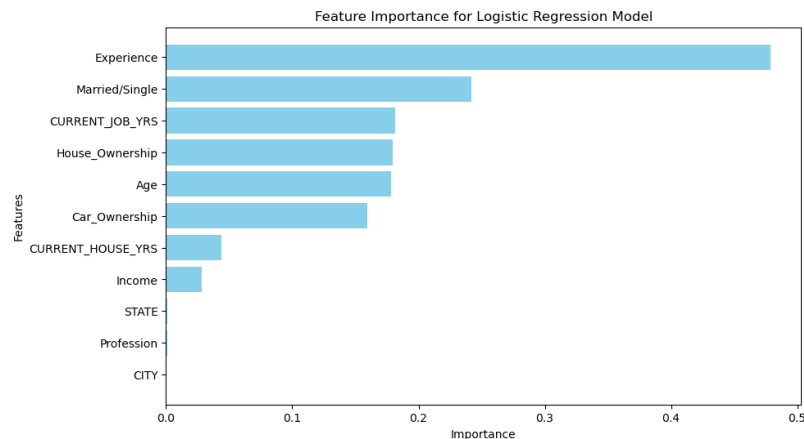
kinerja model klasifikasi. Semakin tinggi nilai F1 score, semakin baik performa modelnya.

## 2. Feature Importance (30 poin)

Setelah mendapatkan model yang paling baik, lakukan interpretasi pada model dengan melihat feature importance-nya.

### 1. Logistic Regression

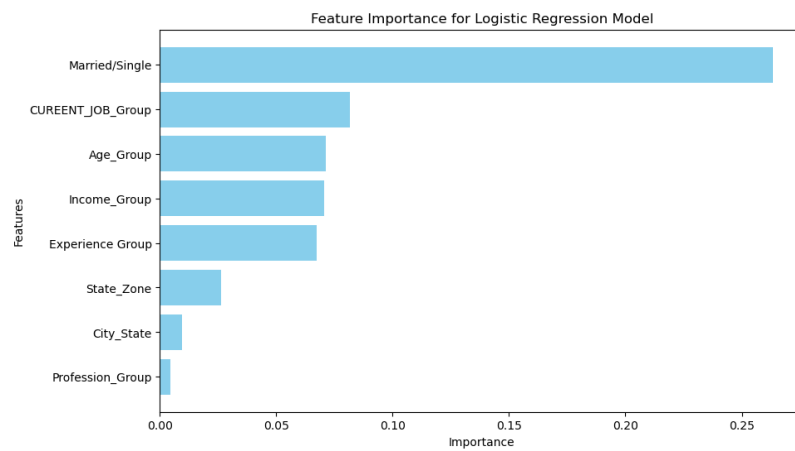
#### a. Data Mentah



Interpretasi:

- Feature 'Experience' merupakan feature dengan nilai importance tertinggi, semakin tinggi jabatan seseorang bekerja semakin besar approval loannya karena dianggap memiliki pendapatan yang stabil.
- Feature 'Married/Single' merupakan feature dengan nilai importance tertinggi kedua, nasabah dengan status married akan lebih dianggap bertanggungjawab dengan pinjamannya.
- Feature 'CURRENT\_JOB\_YRS', 'House\_Ownership', dan 'Age' merupakan feature dengan nilai importance tertinggi ketiga dengan nilai yang sama. Lama bekerja, kepemilikan rumah, dan umur juga menjadi pertimbangan dalam loan approval.
- Feature 'CURRENT\_HOUSE\_YRS' menjadi feature dengan nilai importance tertinggi keempat karena mempertimbangkan apakah orang tersebut berpindah-pindah tempat tinggal atau tidak.
- Feature 'Income' menjadi feature dengan importance tertinggi kelima karena menunjukkan kemampuan seseorang dalam membayar pinjaman sesuai pendapatannya. Tetapi nilai yang didapatkan dalam menggunakan model ini sangat rendah sehingga kurang bisa dikatakan berpengaruh.

## b. Data Group

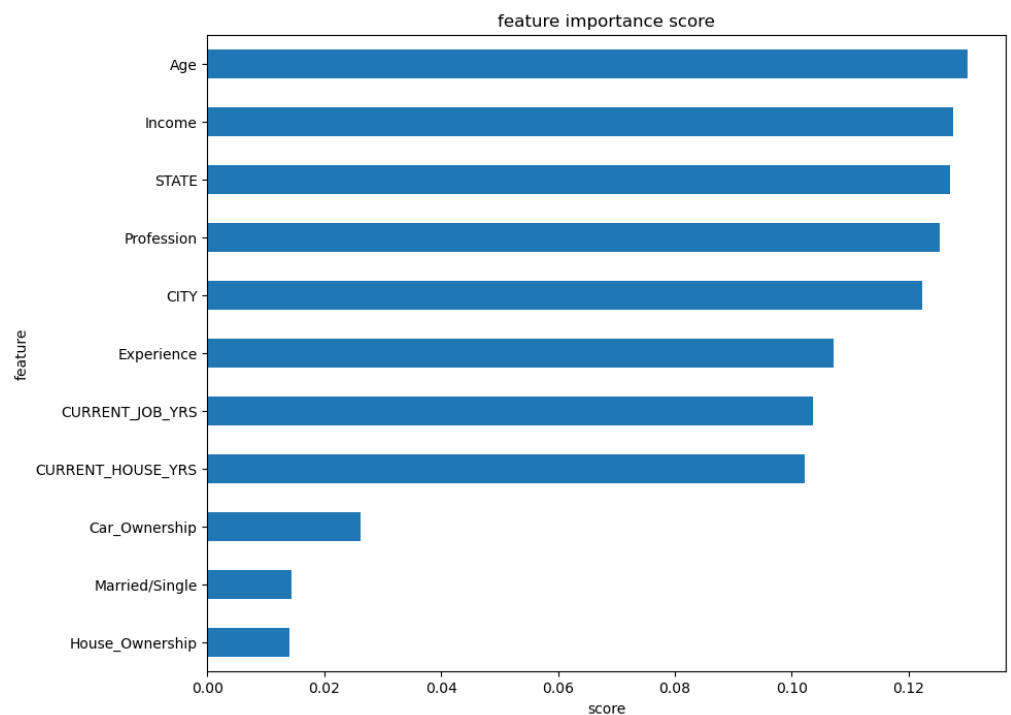


### Interpretasi:

- Feature ‘Married/Single’ memiliki pengaruh tinggi dalam approval loan karena nasabah dengan status menikah akan lebih bertanggungjawab atas pinjamannya.
- Feature lain tidak begitu berpengaruh karena perbedaannya terlalu signifikan terhadap marriage status.

## 2. Decision Tree

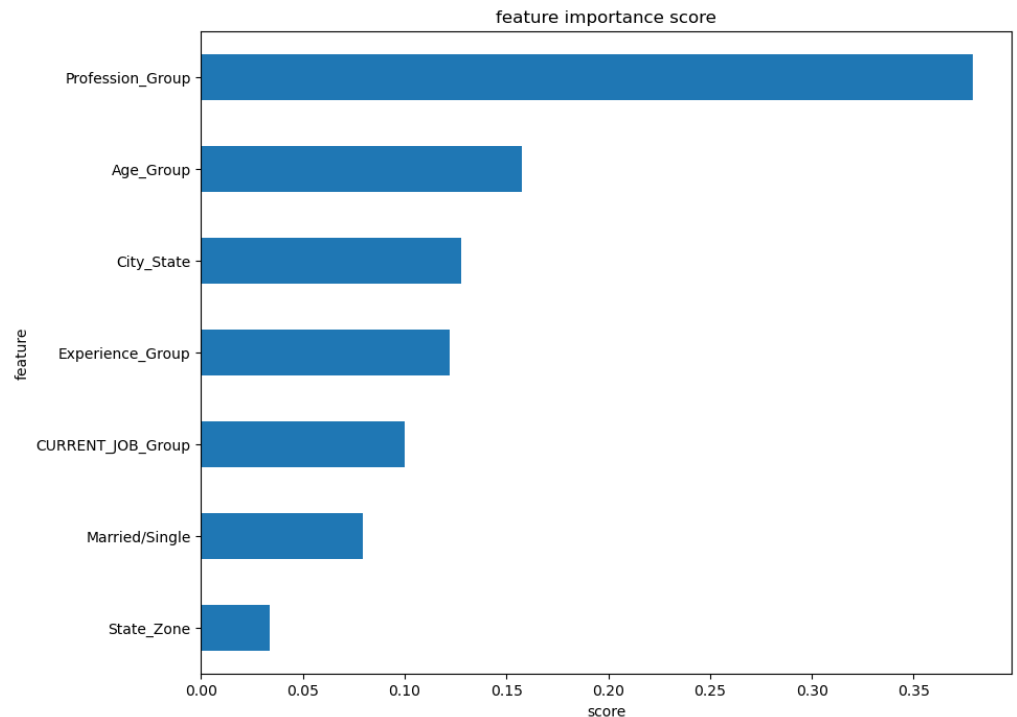
### a. Data Mentah





Interpretasi:

- a. Feature “Age” memiliki pengaruh yang tinggi dalam approval loan karena semakin dewasa seseorang maka sifat tanggung jawabnya juga tinggi.
  - b. Feature “Income” juga menjadi salah satu feature yang memiliki pengaruh dalam approval loan karena nasabah dengan income yang tinggi maka resiko untuk gagal bayarnya rendah.
  - c. Feature “State” memiliki pengaruh yang tinggi juga mungkin karena suatu state tertentu memiliki rata-rata gaji pekerja yang lebih tinggi
  - d. Feature “Profession” memiliki pengaruh yang tinggi juga karena status pekerjaan seseorang akan memiliki pengaruh dalam hal tanggung jawab dalam membayar pinjaman.
- b. Data Group



Interpretasi:

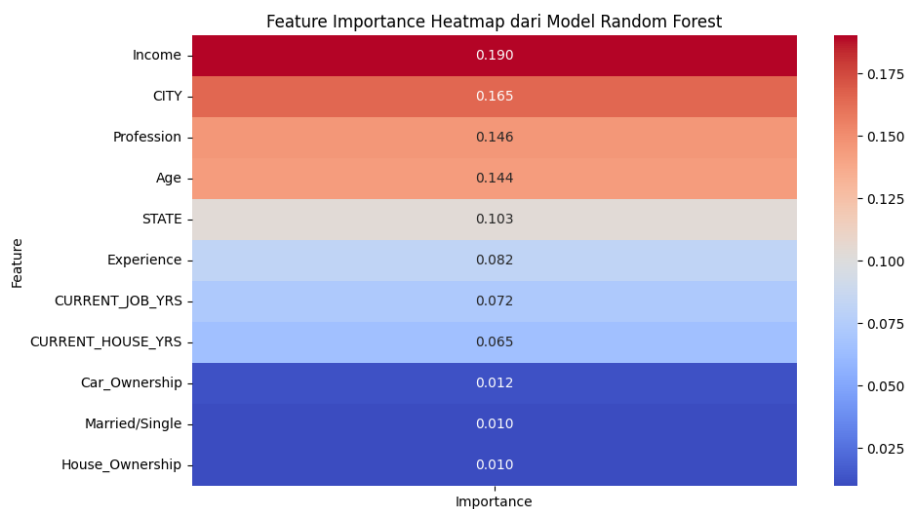
- a. Profession\_Group: Fitur ini memiliki skor Feature Importance tertinggi, yang menunjukkan bahwa kelompok profesi pelanggan memiliki dampak signifikan pada hasil prediksi model. Misalnya, pelanggan dengan profesi tertentu mungkin memiliki pola tertentu yang mempengaruhi hasil prediksi.
- b. Age\_Group: Fitur ini memiliki skor Feature Importance kedua tertinggi, yang berarti kelompok umur pelanggan

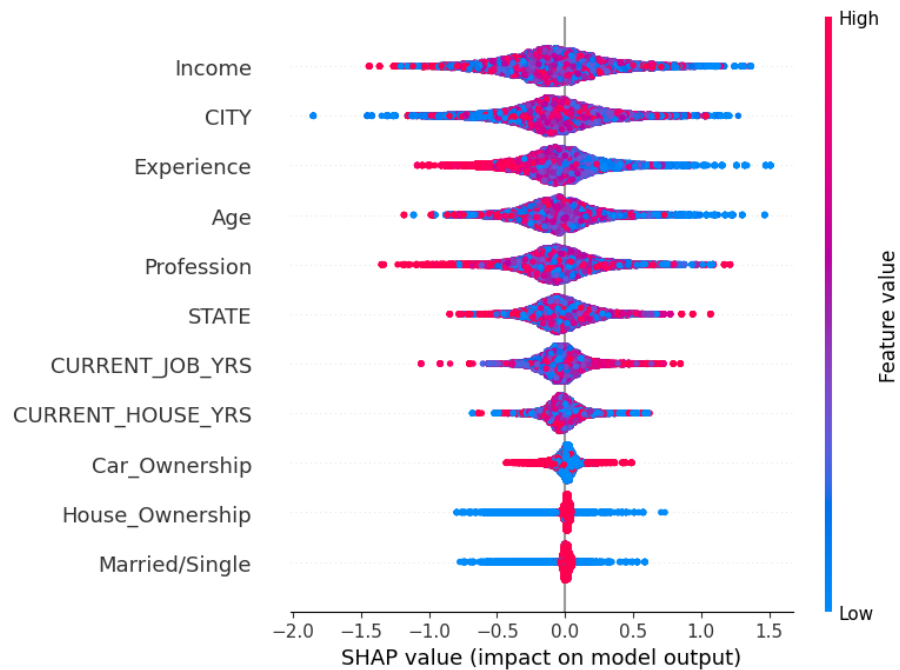
berpengaruh terhadap hasil prediksi model. Misalnya, pelanggan dalam kelompok umur tertentu mungkin menunjukkan perilaku yang berbeda yang mempengaruhi hasil prediksi.

- c. City\_State: Fitur ini memiliki skor Feature Importance yang lebih rendah dibandingkan dengan Profession\_Group dan Age\_Group, tetapi masih cukup tinggi. Ini berarti lokasi geografis pelanggan juga berkontribusi dalam melakukan prediksi, meskipun dampaknya mungkin tidak sebesar dua fitur sebelumnya.
- d. Experience\_Group, CURRENT\_JOB\_Group, Married/Single, dan State\_Zone: Fitur-fitur ini memiliki skor Feature Importance yang lebih rendah dibandingkan dengan fitur-fitur sebelumnya. Ini berarti fitur-fitur ini memiliki kontribusi yang lebih kecil dalam melakukan prediksi.

### 3. Random Forest

#### a. Data Mentah



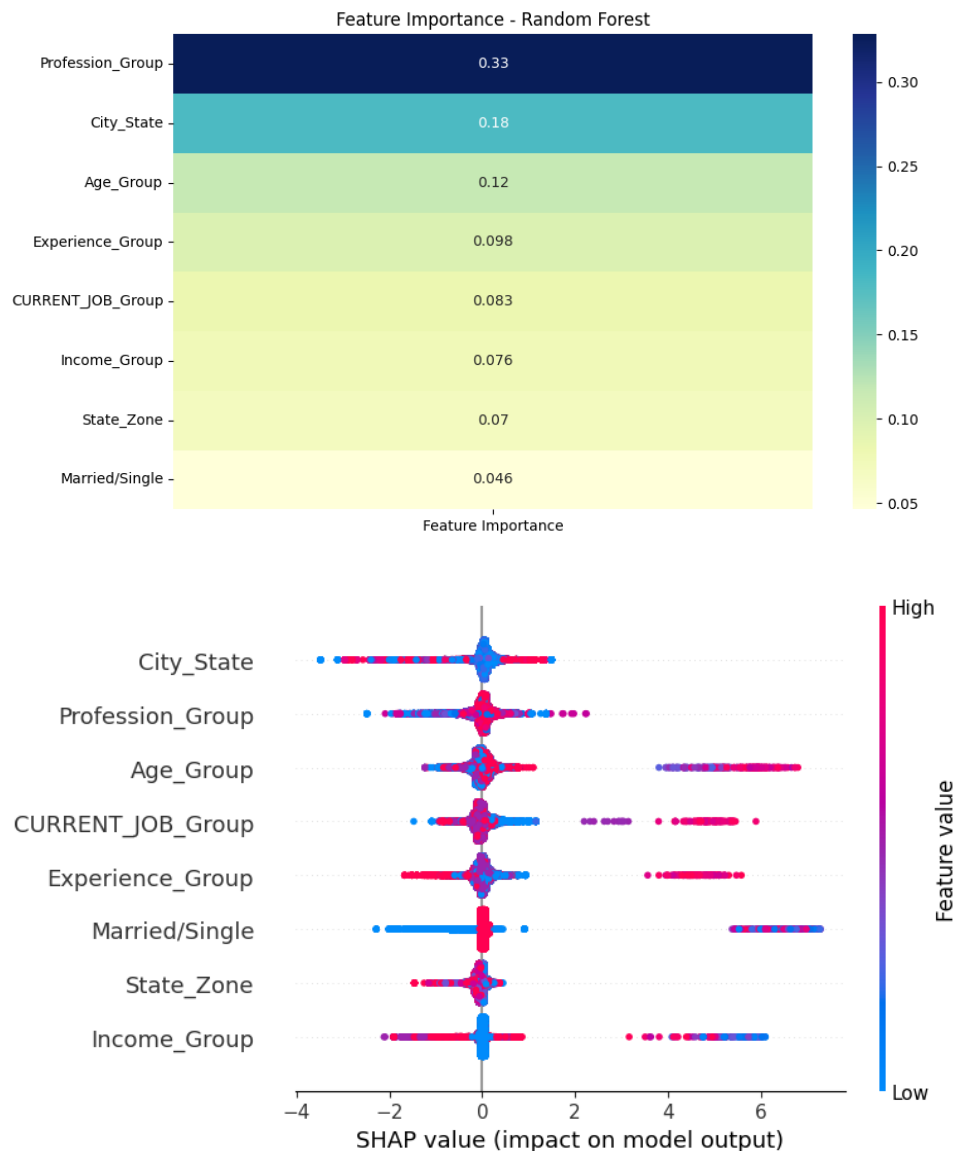


Feature importance Heatmap dan Shap value memiliki interpretasi sebagai berikut,

- Income : fitur dengan nilai tertinggi yang menunjukkan bahwa pendapatan pelanggan memiliki dampak signifikan pada keputusan loan approval. Pelanggan dengan pendapatan lebih tinggi cenderung akan lebih stabil dan lebih mudah dalam persetujuan
- City: fitur tertinggi kedua yang berarti menunjukkan bahwa lokasi geografis pelanggan berpengaruh terhadap keputusan persetujuan pinjaman. Misalnya kawasan kota tertentu memiliki risiko kredit lebih tinggi daripada yang lain
- Profession: fitur tertinggi ketiga pada heatmap dan kelima pada shap value menunjukkan bahwa jenis pekerjaan berpengaruh terhadap keputusan churn. Profesi tertentu dianggap lebih stabil atau lebih beresiko dari sudut pandang keuangan.
- Age : fitur tertinggi keempat pada heatmap dan shap value menunjukkan persetujuan pinjaman akan lebih mudah pada usia matang daripada pelanggan usia yang muda
- Experience : fitur tertinggi ketiga pada shap value yang menunjukkan bahwa pengalaman kerja lebih lama dianggap lebih stabil dan memiliki kemampuan membayar pinjaman
- State, Current job years dan current house years memiliki pengaruh menengah
- Car ownership, Married/Single dan House Ownership : fitur yang kurang berpengaruh

Jadi, faktor faktor seperti income, profesi, usia, pengalaman, dan lokasi geografis berpengaruh terhadap penilaian risiko kredit.

## b. Data Group



Pada fitur “Profession group” dan “city state” memiliki tingkatan paling tinggi pada heatmap dan shap value sehingga menunjukkan bahwa jenis profesi dan lokasi geografis pelanggan memiliki pengaruh yang signifikan pada persetujuan pinjaman. Fitur lain seperti “Age group”, “experience group”, “current job group”, “income group”, “state zone”, dan “married/single” memiliki pengaruh yang rendah. Jadi, persetujuan pinjaman akan cenderung diberikan kepada kelompok profesi yang dianggap memiliki risiko kredit rendah dan lokasi geografis yang dianggap lebih stabil.

### **Apa yang menyebabkan hasil prediksi model demikian.**

Hasil prediksi demikian dihasilkan karena data yang diperoleh atau data yang digunakan memiliki fitur data dengan persebaran nilai nya lebih tinggi pada kondisi Risk Flag tertentu. Dan pada data Profession Group dan Income menjadi data yang paling penting dan selaras persebaran nilainya dengan hasil prediksi Risk\_Flag.

### **Feature importance Evaluasi feature yang paling penting, Tarik business insight-nya, Berikan action items berupa rekomendasi terhadap insight tersebut**

Dari proses feature importance yang telah dilakukan, didapatkan insight pada setiap data yang digunakan, dari data yang telah di-grouping, **Profession group** menjadi salah satu fitur yang memiliki kepentingan tinggi dalam pengambilan keputusan pemberian kredit. Lalu pada data mentah, **Income** menjadi fitur yang cukup penting dalam pengambilan keputusan pemberian kredit selanjutnya (Pada seluruh algoritma yang digunakan).

Sehingga, calon pendaftar kredit harus dinilai berdasarkan **Profession dan Income** nya terlebih dahulu untuk diberikan kredit atau tidak.

### **3. Git**

Link Repository:

<https://github.com/AlyaniNS/Loan-Prediction-Based-on-Customer-Behavior>