

### Kelompok 3B: Insight Miner

1. Purwantiyo
2. Regina Aprilia Sembiring
3. Anisa Agustiana
4. Yolana Faiz Jamahsyari
5. Hendra Hermawan
6. Gideon Kurniawan Sugiarto
7. Faradila Suwandi
8. Widya Angela

## STAGE 3

---

### Laporan Homework Stage 3: Machine Learning Evaluation dan Supervised Learning

#### 1. Modeling

##### A. Split Data Train & Test

Pada bagian ini, dilakukan splitting dataset menjadi data train dan test pada data yang sudah dipastikan bersih/sudah melalui preprocessing. Pertama-tama dilakukan pemisahan antara feature dengan target (feature =  $x$ , target =  $y$ ). Lalu dataset dipisah menjadi antara data train dan data test, dengan pembagian  $\frac{1}{3}$  dari dataset merupakan data test dan sisanya merupakan data train. Hasilnya adalah terdapat dua set data terpisah: satu untuk melatih model ( $X_{train}$  dan  $y_{train}$ ) dan satu untuk menguji model ( $X_{test}$  dan  $y_{test}$ ). Data test dapat digunakan untuk menguji kinerja model machine learning yang Anda latih dengan data pelatihan.

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

Pada bagian ini, dilakukan beberapa modeling, seperti Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, AdaBoost, dan k-Nearest Neighbor. Beberapa modeling tersebut ditujukan untuk mendapatkan score yang paling baik dalam prediksi churn. Dari beberapa model, dapat dilihat bahwa score paling tinggi dimiliki oleh modeling random forest dan decision tree yang setelah diatur kedalaman maksimumnya masing-masing memiliki score 0,95 dan 0,93.

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

Model-model tersebut kemudian dievaluasi dengan beberapa metrics. Setelah Overfitting data di random forest dan decision tree dapat dihindari dengan mengatur kedalaman maksimumnya, diketahui bahwa random forest dan decision tree memiliki akurasi yang paling baik (di atas 95%). Diukur pula nilai accuracy, precision, dan recall pada data test dari semua model tersebut dan didapatkan bahwa random forest dan decision tree merupakan model yang paling baik. Pada pengujian accuracy dan precision data test, model random forest menunjukkan hasil yang lebih baik, namun pada pengujian recall, model decision tree menunjukkan hasil yang lebih baik.

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

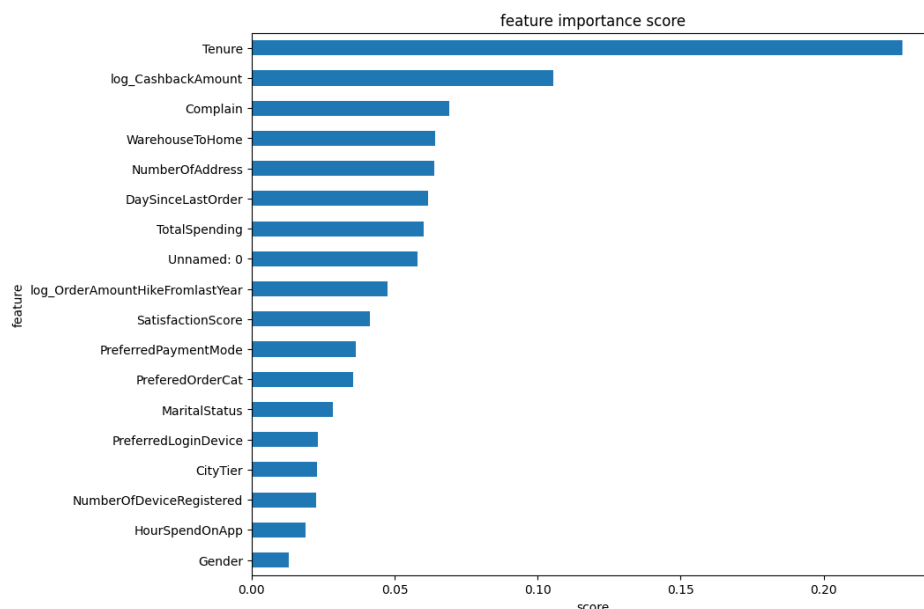
Setelah dilakukan modeling diatas didapatkan dua modeling yang memiliki performance terbaik yaitu Random Forest dan Decision Tree. Selanjutnya dua modeling ini akan diuji cross-validation untuk mengetahui cross valitaion score nya. Nantinya, modeling yang memiliki score tertinggi yang akan digunakan dalam tahapan selanjutnya. Setelah dilakukan cross validation didapatkan rata-rata score dari masing-masing modeling adalah untuk Random Forest sebesar 0.9392 dan Decision Tree sebesar 0.9050 sehingga modeling yang diambil adalah modeling dengan score tertinggi yaitu Random Forest.

#### E. Hyperparameter Tuning

Dari Hasil Cross-Validation ternyata Random Forest mempunyai nilai rata-rata score tertinggi. Oleh karena itu model tersebut dilakukan tuning untuk mencari kombinasi Hyperparameter terbaik yang dapat meningkatkan performa model.

## 2. Feature Importance

A. Setelah mendapatkan model yang paling baik, dapat dilihat bahwa feature importance sebagai berikut.



B. Berdasarkan plot di atas dapat ditarik business insight berupa:

- Churn pelanggan sangat dipengaruhi oleh Usia Tenure User
- Dari hasil analisis juga diperoleh bahwa Gender merupakan salah satu feature yang kurang mempengaruhi angka Churn

C. Rekomendasi

- Usia Tenure atau lamanya pelanggan memakai aplikasi E-Commerce ini memiliki dampak yang signifikan terhadap tingkat churn. Oleh karena itu, sangat direkomendasikan untuk memahami dengan baik pada usia tenure berapa yang potensial menyebabkan pelanggan untuk churn? Dengan pemahaman ini, kita dapat merancang strategi pencegahan yang efektif untuk mengurangi tingkat Churn pelanggan pada usia tenure tersebut.