



**Kelompok**  
**Stage**  
**Mentor**  
**Pukul/ Tanggal**

**: peerData**  
**: 3**  
**: Ridho Aryo Pratama**  
**: 20.00-21.00/Jum'at, 8 September 2023**

**Pembagian tugas di stage ini:**

1. Muhammad Iqbal : Modeling, Hyperparameter Tuning, Feature Importance
2. A Nahda La Roiba : Modeling, Laporan Project
3. Ilham Maulana : Laporan Project
4. Firstandy Edgar Dhafa : Laporan Project
5. Clara Natalie S : Modeling, Hyperparameter Tuning, Feature Importance
6. R. Rani Indah Salamah : Laporan Project
7. Eka Apriyani : Laporan Project
8. Sekar Ayu Larasati : Notulensi, Laporan Project

**Poin Pembahasan:**

- Menampilkan progress, yaitu model yang telah diuji cobakan
- Score cross validation dan precision
- ROC-AUC
- Precision Recall Curve
- Hyperparameter Tuning
- Evaluasi dan perbaikan

### Hasil Diskusi :

- Menampilkan berbagai macam model yang telah dicoba (6 model), dan meng-evaluasi score dari masing-masingnya. Berdasarkan hal tersebut didapatkan hasil bahwa score dari cross validation tinggi, namun score precision rendah.
- Pemilihan model yang terlalu banyak sehingga poin pembahasannya menjadi kompleks dan bisa menimbulkan kebingungan.
- Hyperparameter Tuning, setelah melakukan hyperparameter tuning, sebaiknya dilakukan check cross validation lagi, dibandingkan diantara beberapa jenis model yang dipertimbangkan untuk dipilih.
- Sebaiknya setting threshold setelah didapatkan score  $X_{test}$  dan  $Y_{test}$

### Tindak Lanjut:

- Setelah dievaluasi terdapat kesalahan yaitu pada saat cross validation kami menggunakan keseluruhan data, yang mana seharusnya hanya menggunakan data  $X_{train}$  dan  $Y_{train}$ . Sehingga nanti score yang dibandingkan adalah score precision dan score training.
- Score cross validation yang tinggi dikarenakan adanya kemungkinan dari beberapa nilai K-fold yang kombinasinya bagus dan tinggi, sehingga rata-ratanya mempengaruhi model, atau fluktuasi nilai, yaitu nilai yang sangat tinggi atau sangat rendah.
- Disarankan memakai Precision Recall Curve, supaya lebih mudah dalam meng-interpretasikan score, dan berfokus pada performa model dalam mengidentifikasi kelas positif.
- Apabila telah dilakukan, hal diatas namun hasilnya masih tetap sama, maka kami harus meng-evaluasi tahap pre-processing, kami harus memahami kembali data dan membersihkannya kembali.
- Disarankan menggunakan 3 jenis model saja, yaitu Logistic Regression, Random Forest, dan XGBoost.
- Disarankan dibuat data frame, dari model yang dipilih, dan score dari precision dan crossvalidation (data test dan data train), supaya mudah untuk dilihat dan dievaluasi.