



**Kelompok: Mandalika**

**Stage: 3**

**Mentor: Hafizh Adi Prasetya**

**Pukul/ Tanggal: 16.00 / 10 April 2022**

**Pembagian tugas di stage ini:**

- |   |  |
|---|--|
| 1. Ardila Safitri (Logistic Regression & Random Forest) | 4. Jonse Kennedy (Decision Tree & XGboost) |
| 2. Arni Cici (Logistic Regression & Random Forest)      | 5. Lise Listiani (KNN & Adaboost)          |
| 3. Fakhry Husnul (KNN & Adaboost)                       | 6. Riska Diah (Decision Tree & XGboost)    |

**Poin pembahasan:**

1. Feedback Stage 3 (Explore Modeling, Tinjau ulang Metrics, Hyperparameter Tuning, Confussion Matrics, Feature engineering)
2. Q&A
3. Target Stage 4

**Hasil Diskusi:**

**1. Feedback Stage 3**

- Keluarkan Confussion Matrics sebagai salah satu gambaran dalam penentuan metrics
- ROC AUC tidak memberikan gambaran akurat ketika datasetnya imbalance dan kurang memperhatikan errornya. Dalam kasus ini lebih baik pakai recall . Dalam error treatment ada istilah yaitu precission recall curve yang merupakan curve rate off antara precission dengan recall ketika recall nya sekian maka sebaiknya score precission nya sekian hal ini bisa menjadi pertimbangan untuk penentuan best performance model.
- Cara untuk meningkatkan score dapat dilakukan dengan data dan model. Data bisa bermain di feature set melalui feature engineering. Sedangkan model bisa mulai menggunakan model yang kompleks. Solusi lainnya yaitu melalui proporsi data smotanya bisa diubah contoh menjadi 1:1

**2. Q & A**

- Jika setelah hyperparameter tuning score nya tidak berubah atau turun ada beberapa kemungkinan yaitu:
  - a. Default sebelumnya sudah memberikan parameter terbaik
  - b. Penggunaan randomized yang perlu diatur sample defaultnya untuk menghitung beberapa kemungkinan
  - c. Catatan bahwa hyperparameter tuning tidak mungkin meningkatkan score diatas 5%
- Akan lebih baik jika score data train dan data testnya untuk setiap metrics ditampilkan
- Penentuan linear dan non linear salah satunya melalui heatmap jika correlationnya mayoritas rendah maka hasil di logistic regressionnya juga jelek karena kemungkinan data non linear.



**Kelompok: Mandalika**

**Stage: 3**

**Mentor: Hafizh Adi Prasetya**

**Pukul/ Tanggal: 16.00 / 10 April 2022**

**Hasil Diskusi:**

**3. Target Stage 4**

- Gladi resik
- Gabungkan Power Point mulai dari stage 1-3
- Jelaskan bagaimana prosedur menangani churn seperti apa sebelum adanya model dan setelah adanya model
- Jelaskan improvement ke business matrices sebelum dan setelah model
- Strategi bisnis apa yang direkomendasikan berdasarkan business insight dan model

**Tindak Lanjut:**

**1. Feedback Stage 3**

- Confussion Matrics dimunculkan dalam proses modeling
- Metrics yang dipilih recall
- Pada awalnya dilakukan feature selection berdasarkan feature importance kemudian merubah proporsi data smote menjadi 1:1 dan menambah model lain yaitu Random Forest, Adaboost dan Xgboost

**2. Q & A**

- Mengecek Kembali parameter yang digunakan dalam hyperparameter tuning
- Menampilkan score data train dan data testnya untuk setiap metrics

**3. Target Stage 4**

- Mempersiapkan poin-poin yang telah diberikan untuk stage 4