**Laporan Final Project Stage 3**

**Kelompok Kitasabi.com**

Nama:

1. Mochamad Choiril Iman
2. Yehezkiel Novianto
3. Ahmad Reza
4. Celestial Randy
5. Risca Naquitasia
6. Sonia Epifany Sandah
7. Oky Hariawan

**1. Modeling**

Dalam modeling ini kami melakukan beberapa eksperiment modeling dengan menggunakan:

1. Logistic Regression
2. KNN
3. XG Boost
4. Adabosst
5. LightGBM

**A. LightGBM**

1. **Split Feature & Label**



Modeling menggunakan default parameter didapat score sebagai berikut:

* Accuracy (Test Set): 0.8641
* Precision (Test Set): 0.8494
* Recall (Test Set): 0.7716
* F1-Score (Test Set): 0.8086
* AUC: 0.9417
* Training set score: 0.8709
* Test set score: 0.8641

Dapat dilihat kita mendapatkan score presisi sebesar 84% yang menjadi acuan metrics. Terlihat pula jika skor training model LightGBM sebesar 87% sangat dekat dengan skor testing 86,4%, yang berarti bahwa model tidak overfit atau underfit.

**B. Logistic Regression**

Modeling menggunakan default parameter didapat score sebagai berikut:

* Accuracy (Test Set): 0.80
* Precision (Test Set): 0.81
* Recall (Test Set): 0.62
* F1-Score (Test Set): 0.70
* AUC: 0.88
* Train score: 0.8062223285868134
* Test score:0.8031605986151441

Dapat dilihat kita mendapatkan score presisi sebesar 81% yang menjadi acuan metrics. Terlihat pula jika skor training model logistic regression sebesar 80.6% sangat dekat dengan skor testing 80.3%, yang berarti bahwa model tidak overfit atau underfit.

**2. Model Evaluation**



Sebagaimana dinyatakan dalam pernyataan masalah kami, kami mencari model yang memberikan nilai presisi tertinggi sebagai acuan metrics. Model terbaik dari hasil evaluasi yang dipilih adalah LightGBM yang memberikan performa terbaik.

**3. Hyperparameter Tuning LightGBM** **Text

Description automatically generated**

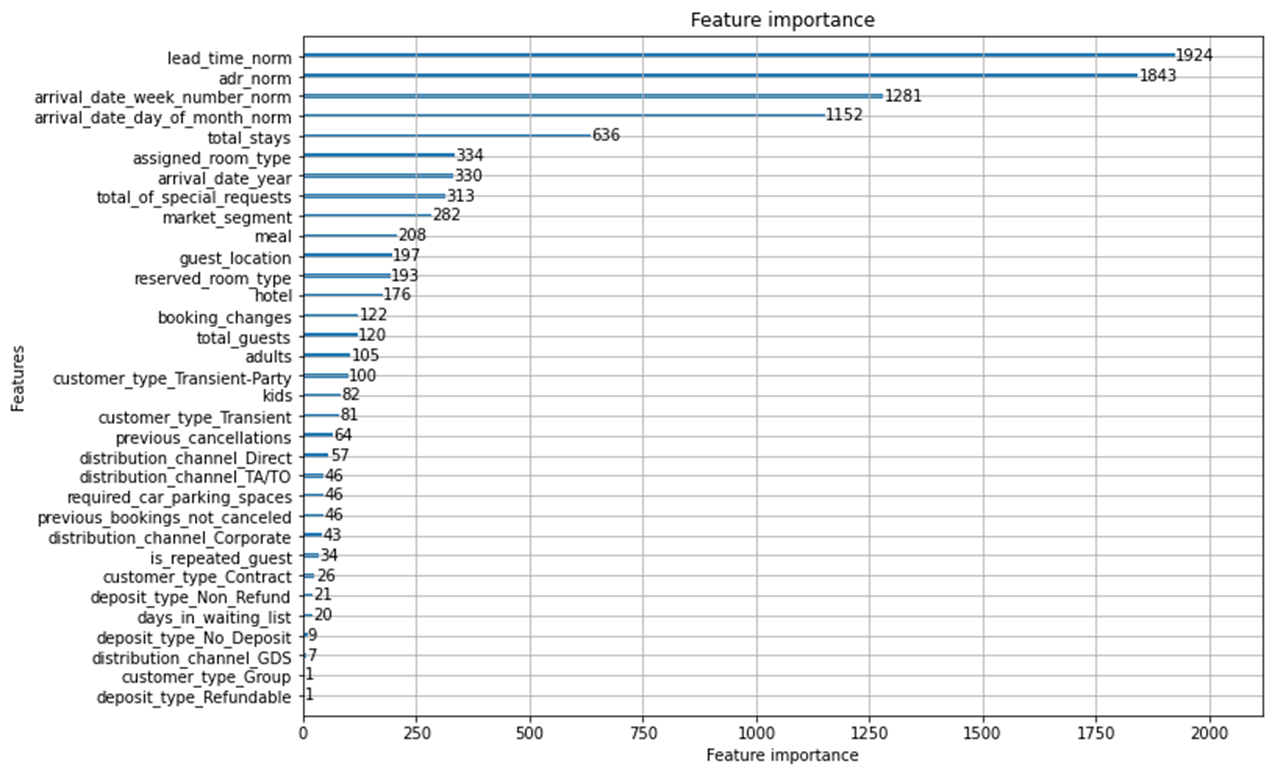
Kami mencoba hyperparameter tuning dengan parameter *objective, metric, num\_boost\_round, dan learning rate*. Nantinya optuna akan mencoba coba kombinasi parameter yang tepat dan melakukan trial sejumlah yang kita masukkan (disini n\_trial yang digunakan 100).

Dengan menggunakan hyperparameter tuning didapat score sebagai berikut:

* Accuracy (Test Set): 0.8812
* Precision (Test Set): 0.8607
* Recall (Test Set): 0.8123
* F1-Score (Test Set): 0.8358
* Training set score: 0.9335
* Test set score: 0.8812

Dapat dilihat setelah melakukan hypertuning score presisi meningkat menjadi 86% dari nilai sebelumnya 84%. Terlihat pula jika skor training model LightGBM setelah hyperparameter tuning sebesar 93.3% dekat dengan skor testing 88,1%, yang berarti bahwa model tidak overfit atau underfit.

**4. Feature Importance**



Top Feature:

1. Lead Time
2. Adr
3. Arrival date week number
4. Arrival date day of month
5. Total stays

Business Insight & Recommendation

Dari hasil feature importance dapat terlihat bahwa Lead time mempunyai pengaruh terbesar dalam pembatalan pesanan hotel. Maka dari itu pihak hotel dapat menerapkan batasan maksimal waktu pemesanan kamar dan menerapkan doposit/uang muka untuk reservasi dengan jangka waktu lama guna menurunkan tingkat pembatalan pesanan.

**4. Feature selection**

Dengan Melakukan iterasi modeling dengan 10 top feature yang dipilih didapat score sebagai berikut:

* Precision (Test Set): 0.8312
* Recall (Test Set): 0.7628
* F1-Score (Test Set): 0.7955
* Training set score: 0.8852
* Test set score: 0.8311

Dapat dilihat setelah melakukan feature selection score presisi sebesar 83%. Terlihat pula jika skor training model LightGBM dengan feature selection sebesar 88.5% dekat dengan skor testing 83,1%, yang berarti bahwa model tidak overfit atau underfit.