



Kelompok: Market Insider

Stage: 3

Mentor: Kevin

Pukul/ Tanggal: 20:00/ 10 Januari 2024

Pembagian tugas di stage ini:

- | | |
|--|--|
| 1. Achmad Hilman Shadiqin – Data Analyst | 4. Andreawan Sofian – Data Scientist |
| 2. Riyan Maula – Data Analyst | 5. Figo Akmal Munir – Data Scientist |
| 3. Nabilah Astiarini – Data Analyst | 6. Dzakwan Darussalam – Data Scientist |

Poin Pembahasan:

Sesi mentoring ini membahas mengenai tahap Machine Learning Modeling & Evaluation yang telah didiskusikan oleh tim sebagai berikut:

1. Modelling

Sebelum melakukan pemodelan, telah dilakukan data splitting menjadi data training dan data testing. Selanjutnya dibuat plot Learning Curve, dicari mean dan std serta dilakukan Cross Validation. Pengolahan data dilakukan pada data train dan testing untuk melihat apakah ada overfitting atau underfitting dan dicek selisih antara data train dan data tes.. Pada bagian ini telah dipilih delapan model dengan menghitung Cross Validationnya (belum melakukan Hyperparameter Tuning), yang kemudian akan dievaluasi model mana yang terbaik untuk melakukan prediksi respons pelanggan. Model yang telah dilakukan antara lain Logistic Regression, KNeighbors Classifier, Decision Tree Classifier, SVC, Random Forest Classifier, XGB Classifier, AdaBoost Classifier, dan GaussianNB.

2. Evaluation

Metrics evaluation yang digunakan adalah Precision dan Accuracy. Pemilihan Precision dikarenakan parameter yang akan diminimalisir adalah False Positive. Karena data yang akan dianalisis merupakan beban marketing, kami perlu menggunakan budget seminimum mungkin (cost efficiency). Lalu setelah terpilih dua model yang sama, akan dibandingkan hasil accuracynya.

Model dengan precision terbaik (sebelum dilakukan hyperparameter tuning) adalah Adaboost dan terburuk adalah Random Forest.



Hasil Diskusi:

1. Modelling

- Di tahap awal, cukup memilih lima model, lalu kemudian ambil tiga model terbaik berdasarkan metrics evaluation.

2. Evaluation

- Pemilihan model terbaik berdasarkan nilai Precision sudah tepat.
- Selain nilai Precision, perbedaan precision diff dari data train dan test perlu diperhatikan. Jika nilai diff > 0.1 (10%) bisa dianggap buruk..
- Ambil 3 model dengan Precision tertinggi dengan nilai diff terendah.

3. Hyperparameter Tuning

- Agar tidak mencari parameter satu-satu, bisa menggunakan Grid Search. Solusi untuk mengatasi berat/lamanya proses tuning, bisa dikecilkan parameternya.
- Point dari tahap ini adalah menghasilkan hasil yang lebih baik dibandingkan nilai awal sebelum dilakukan tuning. Umumnya kenaikan nilai ada di kurang lebih 5% setelah ada tahap tuning.

4. Feature Importance, Business Insight & Recommendation

- Dari hasil evaluasi bisa dipilih satu model terbaik dan dilanjutkan dengan Feature Importance (dilakukan sesuai dengan model terpilih).
- Berikan business insight dari fitur tersebut.



Kelompok: Market Insider

Stage: 3

Mentor: Kevin

Pukul/ Tanggal: 20:00/ 10 Januari 2024

Tindak Lanjut:

Setelah mendapatkan feedback dari mentor, berikut adalah perbaikan dan tambahan pada beberapa sub-tahap di tahap Machine Learning Modeling & Evaluation:

1. Modelling

- Dipilih 3 model yang memiliki tingkat kesalahan terkecil dan tingkat presisi yang tinggi berdasarkan hasil pengujian dengan 8 model. Model terpilih adalah Support Vector Machine (SVM), Adaboost, dan XGBoost.
- Hyperparameter Tuning telah dilakukan dengan menggunakan Grid Search

2. Evaluation

- Model terbaik dipilih berdasarkan nilai Precision yang tinggi dan nilai diff dari data train dan test (error) yang terendah ($< 10\%$).

3. Feature Importance, Business Insight & Recommendation

- Dari model terbaik yang dipilih, dilakukan Feature Importance untuk mengetahui fitur yang paling berpengaruh pada model.
- Business insight dan recommendation telah dipaparkan untuk meningkatkan efektifitas campaign.