



Kelompok

:1

Stage

:3

Mentor

: Fadilah Nur Imani

Pukul/ Tanggal

: 20.00 WIB / January 26th 2023

Pembagian tugas di stage ini:

- | | |
|--|------------------------------------|
| - Devriansyah Sya'ban : management tim & problem solving | - Kun Anggiar : feature importance |
| - Vira Diana : feature importance | - Laurenzius Julio : modelling |
| - Farih Afdhalul : modelling | - Adinda Dita : modelling |
| - Dignu Akbar : modelling | - Ramadhani A : feature importance |

Poin pembahasan:

1. Machine learning model
2. Pemilihan model evaluation
3. Homework unsupervised learning

Hasil Diskusi:

Machine learning model

1. Pada saat menjalankan algoritma machine learning harus ditambahkan random state di setiap algoritma yang dicoba nya agar hasil model yang ditampilkan selalu stabil
Contohnya LogisticRegression(random_state=42) atau RandomForestClassifier(random_state=42) dll.
2. Pada saat melakukan transformasi data misalkan normalisasi atau standardisasi, data harus di split terlebih dahulu antara data train dan data test.
Hal ini untuk menghindari data leak pada saat melakukan modelling.
3. Coba untuk lakukan standardisasi, tapi perlu diingat jika data sensitive terhadap perubahan mean dan standard deviasi nya maka standardisasi harus dilakukan setelah split data test dan data train.
Standardisasi juga dilakukan pada data test saja.
4. Sedangkan untuk data encoding seperti label encoding atau one hot encoding bisa dilakukan sebelum split data train dan data test. Namun jika encoding sudah menggunakan teknik yang advance, maka split data train dan test perlu dilakukan terlebih dahulu.



Kelompok

:1

Stage

:3

Mentor

: Fadilah Nur Imani

Pukul/ Tanggal

: 20.00 WIB / January 26th 2023

Pemilihan model evaluation

1. Pemilihan model evaluation harus sesuai dengan goal yang mau dilihat saat melakukan pemodelan
2. Evaluation metrics accuracy kurang pas jika dipakai untuk melakukan evaluasi model pada dataset ini, karena accuracy dipakai jika data balance (jumlah label seimbang).
3. Jika tujuannya untuk mengurangi false positive (hasil prediksi menunjukkan bahwa customer A potential, sedangkan data actual menunjukkan customer A tidak potential) yang menyebabkan adanya kemungkinan untuk mengeluarkan cost marketing, maka evaluation metrics yang bisa dipakai adalah precision.
4. Jika tujuannya untuk mengurangi false negative (hasil prediksi menunjukkan bahwa customer A tidak potential, sedangkan data actual menunjukkan customer A potential) yang menyebabkan adanya kemungkinan penurunan revenue, maka evaluation metrics yang bisa dipakai adalah recall.
5. Untuk dataset ini, lebih baik menggunakan metrics F1 atau ROC-AUC.
6. Hasil evaluasi model setidaknya harus berada di 60% – 70%. Namun perlu dicurigai jika performa model mencapai 90%. Bisa jadi ada data processing yang kurang tepat atau terdapat data leaked.
7. Untuk melakukan evaluasi model, lakukan modelling step by step dari perubahan yang dilakukan pada data agar bisa dianalisa impact dari perubahan data terhadap performa model yang dihasilkan.
8. Jika berdasarkan performa model lebih bagus memakai seluruh feature dibandingkan menggunakan feature selection, maka lebih baik memakai model dari hasil feature selection. Hal ini karena pemilihan feature nya lebih relevan, selain itu juga bisa dijelaskan business insightnya.

Homework unsupervised learning

Pada dataset di HW week 4 terdapat feature FFP_Tier yaitu data class penerbangan setiap customer. Unique value nya adalah 4,5, dan 6.

Jika demikian, data tersebut adalah data numerik, karena data tersebut menunjukkan level class penerbangan setiap customer. Terdapat meaning yang berarti untuk setiap unique valuenya.



Kelompok

: 1

Stage

: 3

Mentor

: Fadilah Nur Imani

Pukul/ Tanggal

: 20.00 WIB / January 26th 2023

Tindak Lanjut

1. Akan dicoba kembali melakukan machine learning model dengan menambahkan random state di setiap algoritma yang dipakai
2. Akan dicoba kembali melakukan machine learning model dengan melakukan transformasi data pada data train
3. Akan dicoba kembali melakukan machine learning model dengan mengevaluasi metrics F1 atau ROC AUC nya
4. Perlu dilakukan asses overfitting dan underfitting
5. Perlu dikeluarkan feature importance nya dari hasil model untuk ditarik insight
6. Perlu ditarik rekomendasi bisnis dari insight yang jelas penerapan pada bisnisnya