



Kelompok: Fantastic Two

Stage: 0

Mentor: Al Fath Terry

Pukul/ Tanggal: 10.00 wib / 01 April 2023

Pembagian tugas di stage ini:

Ketua Kelompok : Hafidz Alawy **Notulen:** Anshory, F. Artha , Marius Iddo **Moderator :** Annisa Yovinda, Robby Dipomiharjo **Compiler:** Jedi Manullang, Roma Silalahi

Poin pembahasan:

1. Problem dataset
2. Role
3. Goal
4. Objective
5. Busines Metrics

Hasil Diskusi:

Secara keseluruhan untuk hasil analisa Fantastic Two terhadap dataset E-Commerce Shipping Data (baik itu Problem, Role, Goal, Objective dan Business Metrics) sudah baik menurut mentor.

Adapun beberapa saran dari mentor :

Bila terdapat problem lain diluar dari problem yang utama, boleh digunakan dengan catatan masih berhubungan dengan problem yang sudah dijabarkan. Dan dengan adanya problem baru maka akan ada business metrics baru. Untuk secara keseluruhan Bahasa yang digunakan dalam penentuan business metrics tidak harus Bahasa baku, yang terpenting dapat menggambarkan kondisi perusahaan berdasarkan data yang ada.

Tindak Lanjut:

Untuk pengerjaan dan hasil Stage 0 sudah sesuai dengan arahan mentor.



Kelompok: Fantastic Two

Stage: 1

Mentor: Al Fath Terry

Pukul/ Tanggal: 20.00 wib / 05 April 2023

Pembagian tugas di stage ini:

Ketua Kelompok : Hafidz Alawy **Notulen:** F. Artha , Marius Iddo **Moderator :** Annisa Yovinda, Robby Dipomiharjo **Compiler:** Jedi Manullang, Roma Silalahi

Poin pembahasan:

Descriptive Statistic, Univariate Analysis, Multivariate Analysis, Business Insight, Git

Hasil Diskusi:

ID lebih baik di drop saja dan unique value lebih baik ditampilkan menggunakan dot unique. Featured kategorikal dibagi lagi mana yang nominal dan mana yang ordinal, dan reach ontime lebih baik ditampilkan secara bar plot. Karena distribusi data tidak normal jadi disarankan memakai IQR. Tampilan warehouse, mode of shipment, product importance, gender pada bar-plot diperbaiki lagi (vertical saja plot nya). Untuk ordinal memakai ordinal encoding (customer rating) dan numerik dibagi dua lagi mana yang ordinal mana yang numinal. Reach ontime dibuat kategori sendiri. Disarankan untuk memakai anova. Kesalahan pada penulisan purchases priority di perbaiki. Visualisasi dibuat dengan subplot seperti histogram.

Tindak Lanjut:

Untuk pengerjaan dan hasil Stage 1 sudah diperbaiki dan sesuai dengan arahan mentor.



Kelompok: Fantastic Two

Stage: 2

Mentor: Al Fath Terry

Pukul/ Tanggal: 19.00 wib / 14 April 2023

Pembagian tugas di stage ini:

Ketua Kelompok : Hafidz Alawy **Notulen:** Anshory, F. Artha , Marius Iddo **Moderator :** Annisa Yovinda, Robby Dipomiharjo **Compiler:** Jedi Manullang, Roma Silalahi

Poin pembahasan:

Data Pre-Processing

Hasil Diskusi:

Data Cleansing:

1. Handle outlier = nilai threshold dapat disesuaikan menjadi 0,15 dan 0,85 dimana default 0,25 dan 0,75.
2. Feature transform = a. Feature transformation (dengan menggunakan dari beberapa Teknik seperti logarithmic transformation, square-root transformation, dsb. Sampai ditemukan nilai distribusi normal/skewness rendah).
b. Kemudian dilakukan feature scaling (dengan menggunakan beberapa Teknik dan kemudian ditentukan 1 teknik saja yang memiliki hasil skala data yang seimbang/fair).
- 3 . Feature encoding : secara umum Teknik yang digunakan adalah ordinal encoding untuk data categoric dan onehot encoding. Onehot encoder direkomendasikan dikarenakan ada fitur 'drop=first', dimana berguna untuk menghilangkan feature yang sudah di encoding yang pertama sehingga menghindari redundan.
4. Handle class imbalance = data sudah balance.

Feature Engineering:

- a. Feature selection = setelah mendapatkan rekomendasi fitur mana saja, kemudian dilakukan train_test_split. *disarankan untuk menggunakan data frame copy untuk setiap test (contoh df_anova dan df_chi. Bila memakai data frame asli maka dapat terjadi data leakage/ kebocoran data).
- b. Feature Extraction = contoh kolom date DD/MM/YY → dipecah menjadi kolom baru menjadi dd, mm, yy. Dapat juga pada kolom yang terdapat korelasi dari ada kolom tsb di drop. Dapat dilakukan PCA(principal component analysis) yaitu menggabungkan / disatukan kolom yang korelasi.
- c. Berhubungan dengan feature extraction, membuat feature lain dengan catatan selama masih relevan sudah oke.

Tindak Lanjut:

Untuk pengerjaan dan hasil Stage 2 sudah diperbaiki dan sesuai dengan arahan mentor.



Kelompok: Fantastic Two

Stage: 3

Mentor: Al Fath Terry

Pukul/ Tanggal: 20.00 wib / 05 May 2023

Pembagian tugas di stage ini:

Ketua Kelompok : Hafidz Alawy **Notulen:** Anshory, F. Artha , Marius Iddo **Moderator :** Annisa Yovinda, Robby Dipomiharjo **Compiler:** Jedi Manullang, Roma Silalahi

Poin pembahasan:

Machine Learning Evaluation & Supervised Learning

Hasil Diskusi:

1. tahapan dalam melakukan modeling :
 - a. crossvalidasi = digunakan untuk memilih model terbaik.
 - b. hyperparameter tuning = melakukan tuning untuk model supaya lebih baik.
 - c. prediksi data test.
2. data test baru akan dipakai ketika sudah dapat model yang sudah melewati hyperparameter tuning.
3. jangan pernah otak atik data test di tahap crossvalidasi dan hyperparameter tuning.(jadi memanfaatkan data train).
4. kalau misalnya membuat beberapa model, nanti akan dipilih mana yang akan layak masuk ke tahap hyperparameter tuning.
5. cara memilih nya adalah dengan cara memilih score.
6. model mana yang menghasilkan score cv akurasi mean terbaik bukan train nya.
7. kemudian bandingkan score cv akurasi mean dengan score training akurasi, untuk melihat model kita mengalami under fit atau overfit.
8. underfit kondisi dimana score data train dan score cross validasi sama-sama rendah.
9. score ideal yang diinginkan rentang 0,92- 0,99
10. overfit kondisi dimana score data train dan score cv itu jauh.



Kelompok: Fantastic Two

Stage: 3

Mentor: Al Fath Terry

Pukul/ Tanggal: 20.00 wib / 05 May 2023

11. score cv standard deviasi menggambarkan kestabilan model.
12. semakin kecil standar deviasi model akan semakin stabil.
13. kalau standard deviasi rendah , model sudah layak masuk ke tahap hyperparameter tuning.
14. ketika split data classification harap menambahkan parameter stratify=y. untuk menjaga proporsional data.
15. crossvalidation hanya menggunakan data train.
16. semua feature dilakukan standarisasi. semuanya di transform dan di scaling.
17. mau mendekati normal juga di transform dan scaling biar semakin bagus.
18. untuk scaling, semua feature numeric wajib di scaling.
19. recall digunakan jika kita ingin ngejar false negatif serendah rendahnya.

Tindak Lanjut:

Untuk pengerjaan dan hasil Stage 3 sudah diperbaiki dan sesuai dengan arahan mentor.



Kelompok: Fantastic Two

Stage: 4

Mentor: Al Fath Terry

Pukul/ Tanggal: 20.30 wib / 12 May 2023

Pembagian tugas di stage ini:

Ketua Kelompok : Hafidz Alawy **Notulen:** F. Artha , Marius Iddo **Moderator :** Annisa Yovinda, Robby Dipomiharjo **Compiler:** Jedi Manullang, Roma Silalahi

Poin pembahasan:

Machine Learning - Deep Learning

Hasil Diskusi:

- ppt jangan terlalu banyak hal hal teknis, lebih membahas segi bisnis.
- menguasai busines metrics.
- menguasai mengapa memakai roc_auc.
- memahami lebih jelas lagi pengertian akurasi itu apa.
- pelajari lebih dalam rekomendasi bisnis dan simulasi nya.
- model seharusnya bisa lebih baik lagi, tidak offerfitt, namun setelah di hyperparameter sudah lebih baik.
- ideal angka yang diinginkan diantara 0.90-0.95
- standar scaling dengan normalisasi, masukkan ke sesi hyperparameter tuning.
- split data kurang, ditambahkan parameter, stratify=y . agar data proporsional.

Tindak Lanjut:

Pengerjaan dan hasil pada stage 4 sudah dilakukan sesuai dengan arahan mentor.