



Kelompok

: 1

Stage

: 4

Mentor

: Fadilah Nur Imani

Pukul/ Tanggal

: 21.00 WIB / February, 9th 2023

Pembagian tugas di stage ini:

- | | |
|--|------------------------------|
| - Devriansyah Sya'ban : management tim | - Kun Anggiar : design |
| - Vira Diana : visualisasi | - Laurenzius Julio : insight |
| - Farih Afdhalul : interpretasi | - Adinda Dita : design |
| - Dignu Akbar : interpretasi | - Ramadhani A : insight |

Poin pembahasan:

1. Hasil pengerjaan dan masukan untuk stage 3
2. Perbandingan decision tree dan random forest
3. Slide presentasi

Hasil Diskusi:

Hasil pengerjaan dan masukan untuk stage 3

1. Pada score crossval, nilai yang ditampilkan masih merupakan nilai mean dari keseluruhan part. Jika score cross validation (cv) displit menjadi 5 part perlu dilihat hasil score validation setiap part nya.
2. Agar hasil train test split nya seimbang untuk label 0 dan 1, di bagian train test split bisa ditambahkan stratify = y
3. Selain menampilkan feature important, lebih baik dimunculkan juga shap value nya untuk melihat pengaruh dari feature – feature yang ada. Hal ini untuk memastikan bahwa model yang dipakai cukup konsisten terhadap performa model apa tidak.



Kelompok

: 1

Stage

: 4

Mentor

: Fadilah Nur Imani

Pukul/ Tanggal

: 21.00 WIB / February, 9th 2023

Perbandingan decision tree dan random forest

Keuntungan menggunakan decision tree dibandingkan dengan random forest :

- a. Pada decision tree lebih mudah diinterpretasikan dibandingkan dengan pada random forest. Hal ini karena pada decision tree hanya menggunakan 1 tree, sedangkan pada random forest terdiri dari banyak tree di mana hasil model yang dipakai merupakan hasil voting dari seluruh tree,
- b. Pada kasus data travel insurance yang dipakai, performa model menggunakan decision tree dan random forest tidak jauh berbeda. Artinya pemilihan algoritma decision tree bisa dibilang tepat karena algoritma random forest terlalu kompleks.

Slide presentasi

1. Pada slide presentasi, better ditambahkan flowchart pengerjaan. Misalkan step 1 data cleansing, step 2 data encoding, dll. Hal ini untuk mengurangi penggunaan text dan memudahkan audience untuk memahami isi dari presentasi..
2. Pada slide presentasi, harus dipastikan semua diagram atau grafik yang ditampilkan jelas, penjelasan sumbu x dan sumbu y nya jelas, serta pemakaian warna dan segala macamnya harus tepat guna.



Kelompok

: 1

Stage

: 4

Mentor

: Fadilah Nur Imani

Pukul/ Tanggal

: 21.00 WIB / February, 9th 2023

Tindak Lanjut

1. Akan dimunculkan hasil score cross validation untuk setiap part untuk melihat konsistensi model yang dipakai.
2. Akan dimunculkan class report untuk melihat probabilitas label 1 dan 0 agar dapat dipastikan probabilitas nya di atas 50%.
3. Memunculkan leave and gain analysis
4. Pada slide akan dimunculkan simulasi bisnis agar lebih mudah dipahami interpretasi model yang sudah dilakukan