|  |  |  |
| --- | --- | --- |
| Nama | : | Anton Wijaya |
| Universitas | : | Universitas Tidar |

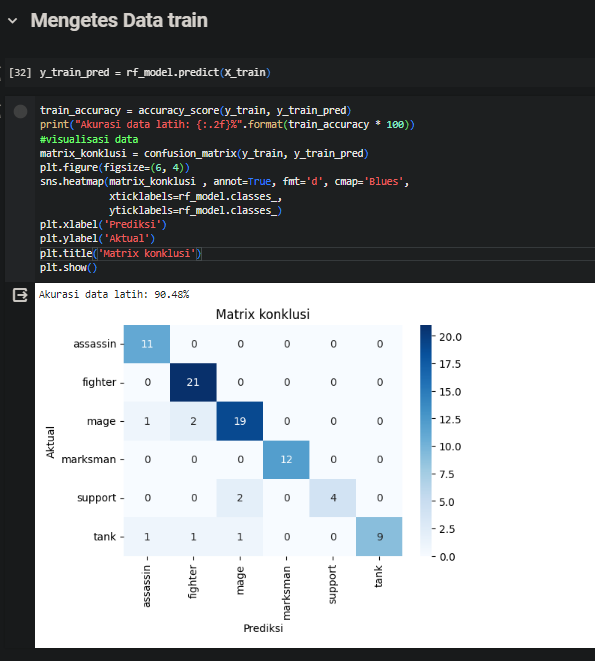
1. Deskripsi Dataset

* Sumber: <https://github.com/antonwijayacakra/MACHINE-Learning-Steqhoc>
* Dataset *“MLBB\_kompetitif\_Using”* merupakan dataset yang ada pada game MLBB yang telah dimodifikasi pada tugas sebelumnya*.* Kolom data yang ada pada dataset saya yaitu nama hero, role hero, pick rate, win rate, dan ban rate. Data- data tersebut dipilih karena data tersebutlah yang digunakan untuk kompetisi MLBB. Dataset ini memiliki jumlah kolom yang lebih sedikit yaitu 5 kolom dibandingkan dataset lama yaitu 19 kolom.



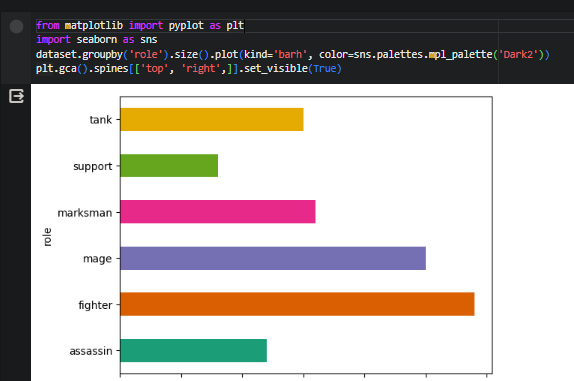
1. Deskripsi model ML yang dibangun

Melalui dataset ini, saya membangun sebuah model AI yang dapat mengklasifikasikan Role Hero menggunakan Random Forest. Hasil yang didapatkan adalah saat melakukan pengujian data train mendapatkan akurasi 90.48%, sedangkan data test memiliki akurasi klasifikasi 14,26%. Maka dari itu model pertama yang dibuat terjadi overfitting, sehingga perlu dilakukan peningkatan model

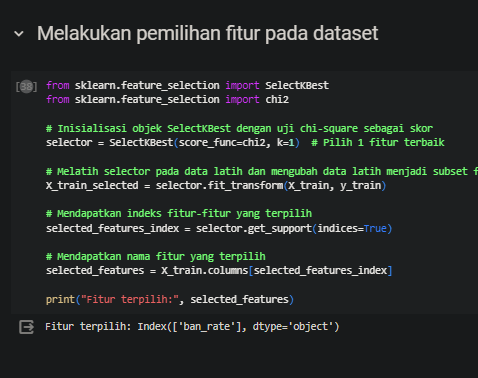
 

1. Memilih fitur yang paling berpengaruh dan melakukan undersampling

* Diharapkan dengan memilih fitur yang paling relevan dapat meningkatkan akurasi model dan melakukan undersampling agar menyamakan jumlah role data yang terlalu mendominasi ke fighter.
* Data imbalnce:

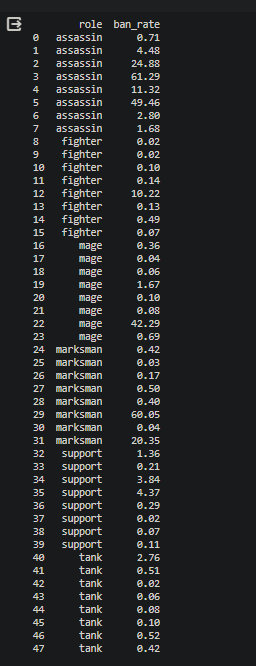


* Pemilihan fitur terbaik adalah ban\_rate, maka akan dilakukan penghapusan kolom selain role hero dan ban rate.





* Berikut ini merupakan dataset yang telah diunder sampling dan dihapus kolom- kolom tertentu, hingga hanya menyisakan kolom role dan ban\_rate



1. Pengujian model kembali dengan menggunakan random forest.

Setelah menggunakan dataset yang telah diatur ulang fitur nya, hasil yang didapatkan bahwa data train mengalami peningkatan akurasi menjadi 28,04 persen. Selain melakukan pengolahan data kolom, nilai random state diubah menjadi 100 dan jumlah pohon random forest adalah 100. Nilai akurasi tersebut masih rendah, tetapi telah meningkatkan akurasi pengklasifikasian data test sehingga model berhasil ditingkatkan akurasinya.

