

Rakamin Final Project -- Machine Learning Modeling & Evaluation

Group 1 Members :

Yosafat Respati

Ridho Fajar

Li'izza Diana M

Teguh Tri A

Vito Rihaldijiran

M Supian Noor

Rexy Anggala Putra

1. Split Data Train dan Test

Sebelum kami melakukan pemodelan data, kami membagi data terlebih untuk membedakan antara data training dan data testing. Untuk proporsi data yang kami pakai yaitu 80% dari jumlah data merupakan data training dan 20% untuk data testing. Dari pembagian data tersebut dihasilkan jumlah data training sebanyak 3061, dan data testing sebanyak 766. Karena pada data training jumlah antar label tidak seimbang, yaitu 2506 dan 555 maka dilakukan handle imbalance dengan menggunakan metode SMOTE dengan perbandingan 1:1, sehingga jumlah data antar label menjadi sama.

2. Modeling

Untuk melakukan prediksi churn customer, kami menggunakan beberapa algoritma classification dengan parameter default dengan menghasilkan:

Tabel Metrics Evaluation Algoritma

Model	Accuracy		Recall		Precision		F1-Score	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	0.91	0.86	0.87	0.6	0.95	0.62	0.91	0.64
SGD Classifier	0.871	0.867	0.474	0.532	0.721	0.667	0.572	0.592

Xgboost Classifier	0.944	0.911	0.947	0.735	0.941	0.798	0.944	0.765
--------------------	-------	-------	-------	-------	-------	-------	-------	-------

3. Model Evaluation

Pada kasus ini, metrics evaluation yang kami gunakan yaitu F1-Score, karena semua label dalam data yang digunakan memiliki kepentingan yang sama, sehingga baik *False Positif* maupun *False Negatif* memiliki tingkat bahaya yang sama sehingga perlu diperhatikan kedua-duanya.

4. Model Evaluation: Apakah model sudah best fit?

Dengan melihat tabel metrics evaluation algoritma diatas, dapat kita lihat di metrics F1-Score Logistic Regression dan Xgboost Classifier bersifat overfitting, sedangkan untuk SGD Classifier memiliki performance yang cukup rendah.

5. Hyperparameter Tuning

Pada setiap algoritma yang kami gunakan pada pemodelan, kami lakukan hyperparameter tuning untuk mendapatkan model yang terbaik (tidak overfitting ataupun underfitting).

a. Logistic Regression

Untuk Hyperparameter Tuning yang digunakan dalam pemodelan Logistik Regression ini terdiri dari 3 parameter yaitu:

1. penalty = ['l1', 'l2']
2. C = [0.0001, 0.001, 0.01, 0.02, 0.03]
3. solver = ['newton-cg', 'sag', 'saga']

Dari hasil pengujian model setelah proses randomized Tunning Hyperparamater dihasilkan perubahan metrics f1-score train dari 0.91 turun menjadi 0.86 dan f1-score testing konstan di angka 0.64

b. SGD Classifier

Pada pemodelan SGD Classifier dilakukan hyperparameter tuning

sebagai berikut:

```
'model__loss': ['log', 'hinge', 'modified_huber', 'squared_hinge', 'huber',  
'epsilon_insensitive', 'squared_epsilon_insensitive'],  
'model__penalty': ['elasticnet', 'l2', 'l1'],  
'model__alpha': [10 ** x for x in range(-6, 1)],  
'model__l1_ratio': [0, 0.05, 0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 1]
```

Berdasarkan hasil tuning tersebut, didapatkan metrics f1-score pada data train sebesar 0.591808 dan pada data test sebesar 0.597826.

c. Xgboost Classifier

Pada pemodelan ini dilakukan beberapa percobaan diantaranya:

- Melakukan *randomized* dengan hyperparameter berikut:

```
'max_depth' : [10,20,30,...,110]  
'min_child_weight' : [1,3,5,7,9,11,13,15,17,19]  
'gamma' : [0, 0.1, 0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1],  
'tree_method' : ['auto', 'exact', 'approx', 'hist']  
'colsample_bytree' : [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1],  
'eta' : [0, 0.01, 0.02,...,1],  
'lambda' : [0, 0.1, 0.2, 0.3, 0.4 ,0.5, 0.6, 0.7, 0.8, 0.9, 1],  
'alpha' : [0, 0.1, 0.2, 0.3, 0.4 ,0.5, 0.6, 0.7, 0.8, 0.9, 1]
```

dari percobaan itu dihasilkan metrics f1-score pada data train 1.0 dan f1-score data test 0.888. dengan melihat score tersebut dapat diketahui bahwa model mengalami overfitting, sehingga masih diperlukan tuning hyperparameter ulang.

- Menggunakan hyperparamter yang sama dengan yg diatas, namun dilakukan pengubahan pada parameter 'max_depth' yaitu menjadi [5,6,7,8,9,...,20] dan dihasilkan f1-score 0.999801 untuk data training dan 0.908425 data testing. Dari hasil ini antara training dan testing mengalami penurunan score, akan tetapi hal itu masih mengandung overfitting.
- Menggunakan hyperparameter 'max_depth' saja dengan range angka 3-20, menghasilkan f1 score untuk training dan testing sebesar 1.0 dan

0.907749. Dari score tersebut juga terlihat bahwa model bersifat overfitting.

- Menggunakan parameter dengan 'max_depth'= 7 dan tree_method='exact, menghasilkan metric f1-score training 0.99 dan f1-score testing 0.89.

berikut adalah tabel metrics setelah dilakukan hyperparameter tuning pada setiap algoritma:

**Tabel Metrics Evaluation Algoritma
dengan Hyperparameter Tuning**

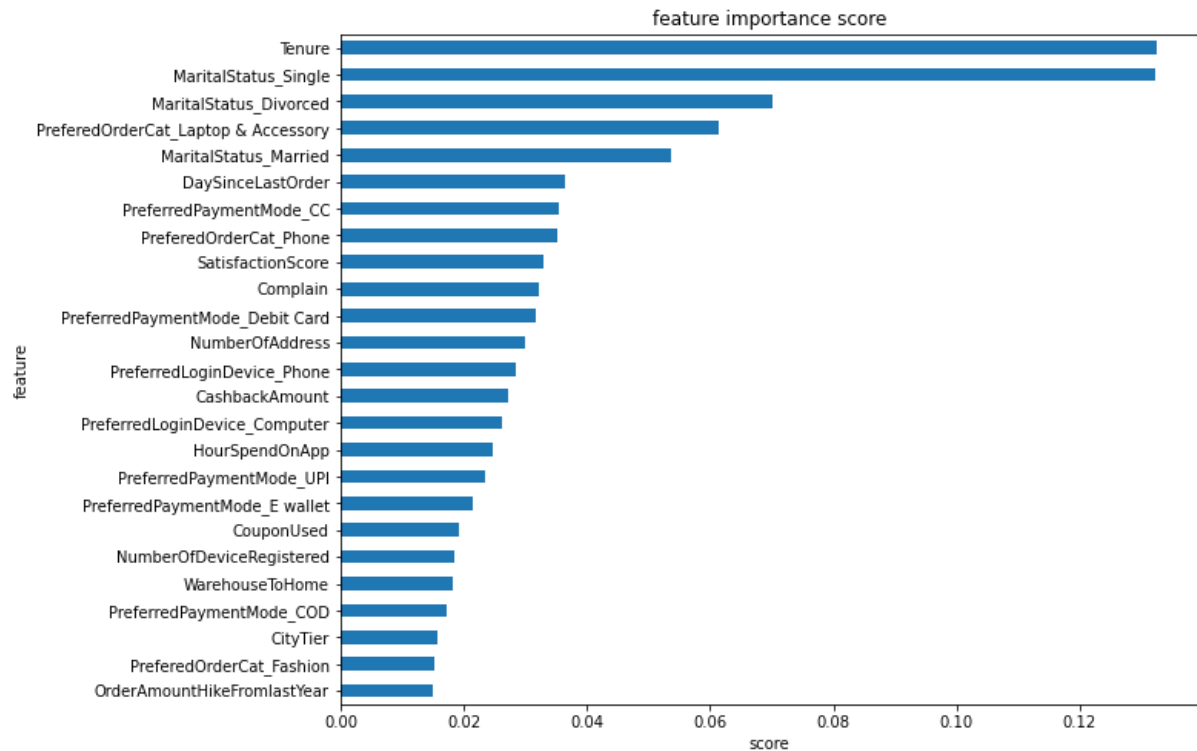
Model	Accuracy		Recall		Precision		F1-Score	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	0.86	0.83	0.85	0.53	0.87	0.83	0.86	0.64
SGD Classifier	0.811	0.807	0.755	0.791	0.487	0.480	0.592	0.598
Xgboost Classifier	0.99	0.96	0.99	0.93	0.99	0.87	0.99	0.899

Jika dilihat metrics F1-Score pada setiap algoritma pada tabel diatas, nilai performance algoritma SGD Classifier cukup rendah dibanding dengan Logistic Regression dan Xgboost Classifier. Sedangkan, kedua algoritma tersebut memiliki selisih score antara training dan testing yang cukup signifikan, sehingga keduanya merupakan data yang overfitting. Namun, untuk pemodelan kasus ini kami memilih algoritma Xgboost sebagai metode machine learning prediction churn karena selisih score antara training dan testing yang lebih rendah yaitu 0.1, jika dibandingkan dengan selisih score training testing pada algoritma Logistic Regression yaitu 0.22.

6. Feature Importance

Setelah menentukan Xgboost Classifier sebagai metode machine learning

prediction churn ini, kami melakukan perhitungan feature importance dengan hasil sebagai berikut:



B. Feature selection

Setelah didapatkan plot feature importance di atas, maka dilakukan feature selection dengan membuang feature dimulai dari feature yang memiliki importance score rendah, yaitu OrderHikeFromlastYear. Setelah dibuat model, didapatkan F1-Score model yang telah dilakukan feature selection, yaitu sebesar 0.998803 pada train set dan sebesar 0.883895 pada test set. Berdasarkan hasil pembuangan satu feature tersebut, terlihat bahwa ada penurunan performa pada test set sebesar 1,5%. Oleh karena itu, pada model ini akan digunakan semua feature yang tersedia karena pembuangan satu feature dapat menurunkan performa model.

3. Git

Berikut link github kelompok 1

https://github.com/to-to-to-to/fp_rakamin_ecommerce/tree/master