

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 9

Nama Mentor: Aditya Bariq Ikhsan

Nama:

- Chairunisa Az Zahra Arifin
- Oktaviani Nurlinda Sari

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



- 1. Latar Belakang**
- 2. Explorasi Data dan Visualisasi**
- 3. Modelling**
- 4. Kesimpulan**

Latar Belakang

Latar Belakang Project

Sumber Data: <https://www.kaggle.com/datasets/barun2104/telecom-churn?datasetId=567482>

Problem: **classification**

Tujuan:

- Memprediksi dan mengelompokkan *target columnn*
- Menganalisis untuk mendapatkan *Interesting Insights* dari data set, meliputi faktor-faktor apa saja yang memengaruhi *Churn*.
- Mendapatkan model Machine Learning yang terbaik
- Memberikan rekomendasi yang dapat dilakukan perusahaan untuk mengurangi *Churn* dari para pelanggan mereka.

Explorasi Data dan Visualisasi

Business Understanding

Churn adalah pemutusan layanan jasa telekomunikasi oleh pelanggan atau perusahaan. Perusahaan lebih memutuskan mempertahankan pelanggan, karena dibutuhkan biaya lebih sedikit daripada mencari pelanggan baru.

Alasan utama mengapa *customer churn rate* penting adalah persentase pelanggan yang hilang tersebut sangat memengaruhi *growth rate* perusahaan.



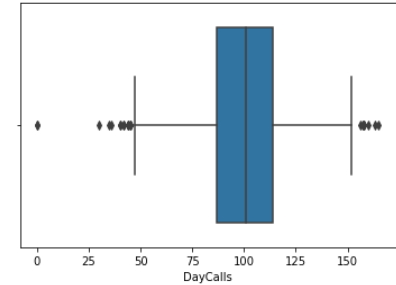
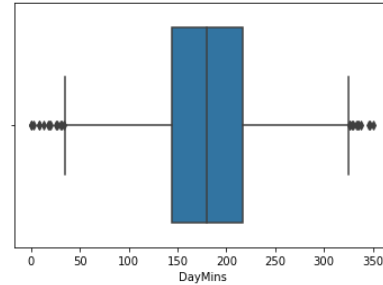
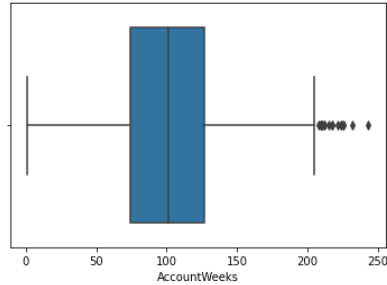
Churn prediction **bertujuan** untuk memprediksi peluang seorang pelanggan untuk churn sebelum pelanggan tersebut benar-benar melakukannya, untuk menentukan strategi *marketing* yang tepat agar dapat menekan persentase *customer churn*.

Data Cleansing

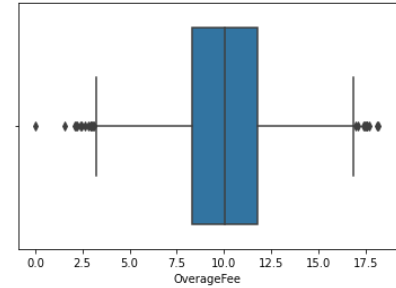
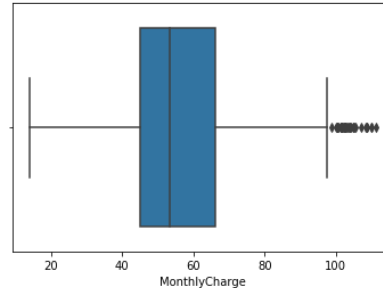
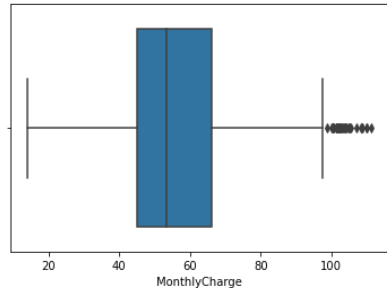
- **Dataframe shape** terdiri atas 1 target column dan 10 feature column, dengan dimensi 3333 x 11.
- Semua feature column **tidak memiliki missing value**
- **Datatype** sudah benar yaitu 'int64' untuk kolom yang berisi bilangan bulat, tanpa desimal. Sedangkan, data type 'float64' adalah kolom yang berisi bilangan desimal.
- Diketahui semua kolom merupakan kolom bertipe **numerik**, sehingga bisa saling dibandingkan.

Data Cleansing

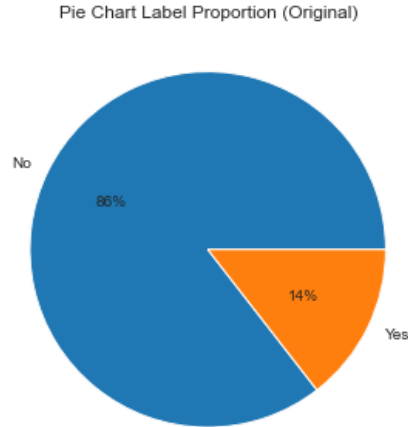
Ditemukan **Outliers**



Maka dilakukan
Handling Outliers
with IQR



Exploratory Data Analysis



Sebaran data secara mayoritas customer tidak melakukan Churn, dengan detail Churn sebanyak 14% dan No Churn sebanyak 86%.

Exploratory Data Analysis

Imbalance menyebabkan output model awal **overfitting**, the model is more biased towards majority class. We apply imbalanced data handling techniques, we use resampling **oversampling**. We see their accuracy and recall results, the recall value of minority class has also improved. This is a good model compared to the previous one. Recall is great.

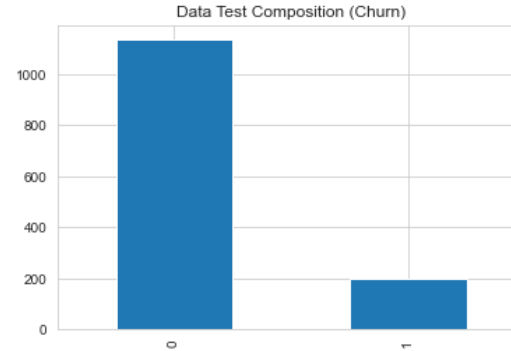
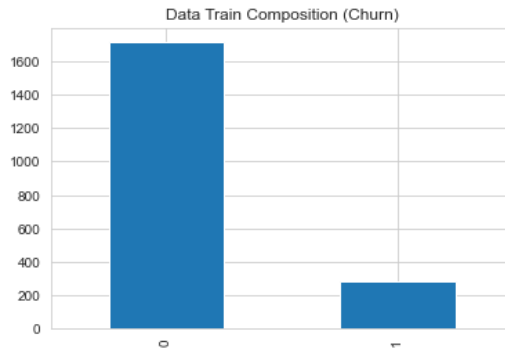
Modelling



Classification Telecom Churn

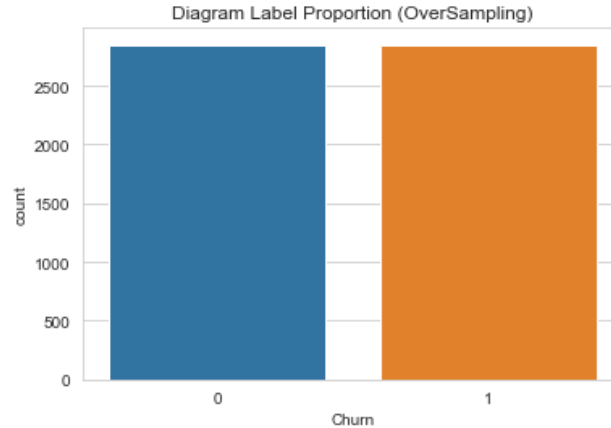
- Metode train test split / cross validation

1. Before resampling



Classification Telecom Churn

- Metode train test split / cross validation
 1. After resampling, teknik oversampling

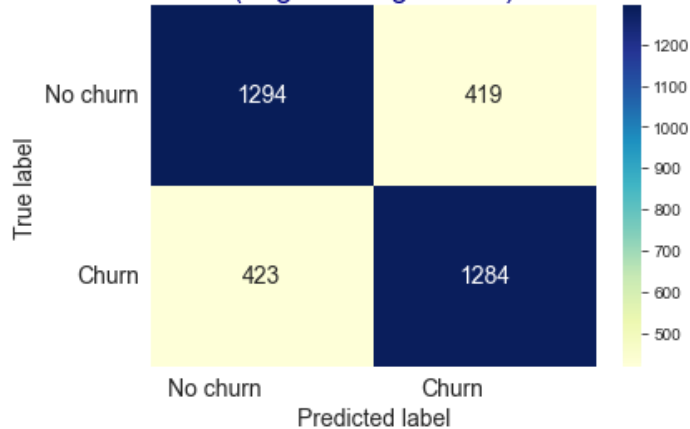


Classification Telecom Churn

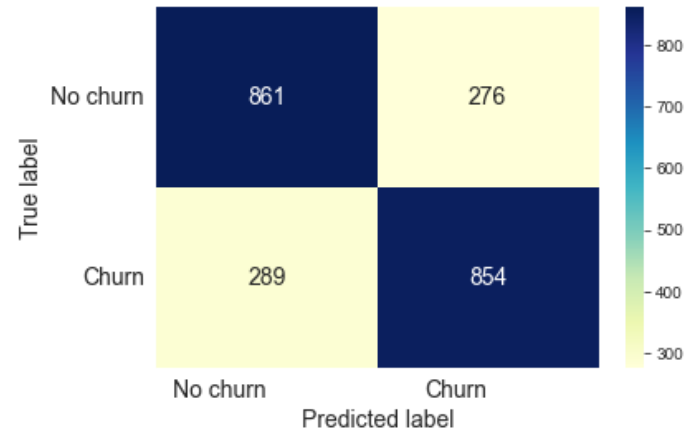
- Jenis model yang dicoba
 1. Logistic Regression:
akurasi data train sebesar 75%, akurasi data set sebesar 75%
 2. Decision Tree:
akurasi data train sebesar 96%, akurasi data set sebesar 96%
 3. Random Forest:
akurasi data train sebesar 100%, akurasi data set sebesar 97%
 4. Xgboost:
akurasi data train sebesar 91%, akurasi data set sebesar 89%

1. Logistic Regression

Confusion Matrix untuk Training Model
(Logistic Regression)



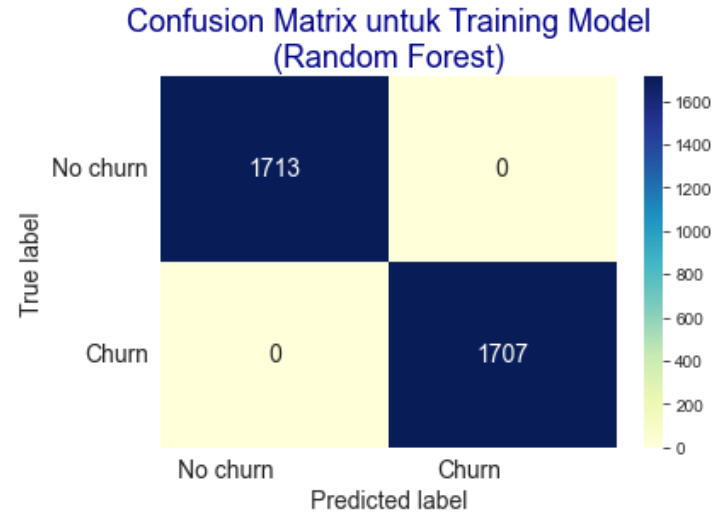
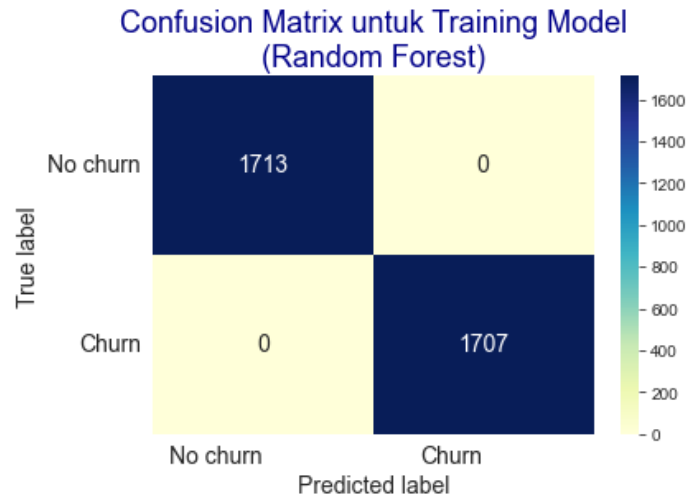
Confusion Matrix untuk Testing Model
(Logistic Regression)



2. Decision Tree

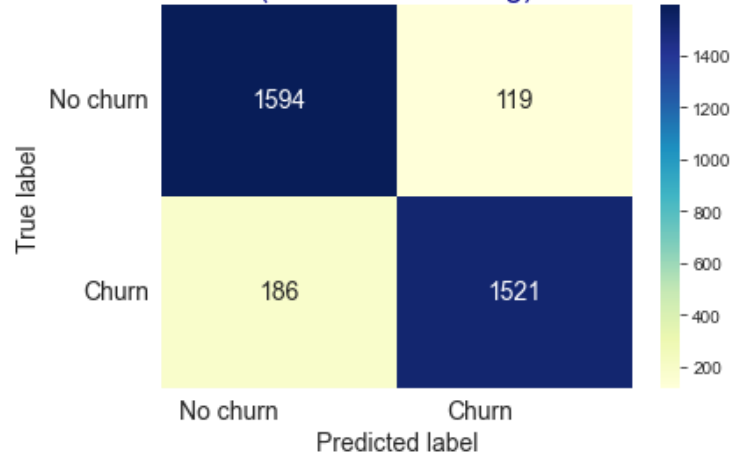
```
[[1058  79]  
 [  16 1127]]
```

3. Random Forest

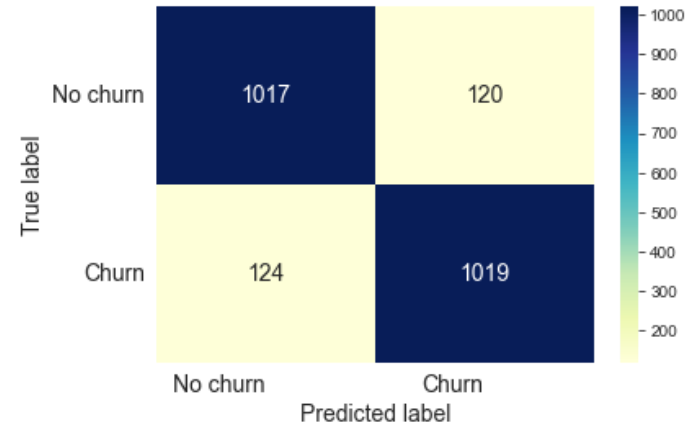


4. XGBoost

Confusion Matrix untuk Training Model
(Gradient Boosting)



Confusion Matrix untuk Testing Model
(Gradient Boosting)



Classification Telecom Churn

- Tindakan-tindakan untuk mencoba menambah akurasi model
 1. **Random forest hyperparameter tuning**
menaikkan 1% akurasi data test dari 97% menjadi 98%
- Model final
- Model yang baik adalah model yang mampu memberikan performa bagus di fase training dan testing model. Sehingga dapat disimpulkan model yang terbaik dari keempat model di atas adalah model dengan metode **Random Forest**.
- Semua predictor kolom digunakan untuk pemodelan target kolom

Conclusion

Kesimpulan

Hasil analisis tersebut menghasilkan strategi pengurangan pelanggan *Churn* berupa paket penawaran yang tepat sasaran dan sesuai kebutuhan pelanggan.

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

