

zenius



Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 3

Nama Mentor: Rachmadio Noval L.

Nama:

- Muhammad Ghossani Akhdan
- Asri Rahmi Putri

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

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:

- Perusahaan yang memiliki data behaviour customer yang meninggalkan layanan (churn), setiap baris dan data merepresentasikan behaviour customer
- Memprediksi mengapa customer meninggalkan layanan
- Diperlukan model yang membantu perusahaan agar customer tetap menggunakan layanan

Business Understanding

Apasih “Churn” itu?

→ Tingkat pelanggan yang berpindah services ke layanan lain (unsubscribe)

Churn penting banget ya? Gimana sih pengaruhnya ke bisnis?

→ Survei dari Bain & Company menyebutkan kalau pelanggan yang setia kemungkinan akan menghabiskan 67% lebih banyak uang mereka untuk kembali membeli produk atau layanan. Sementara untuk menggaet pembeli baru lebih banyak membutuhkan biaya operasional.

Penyebabnya churn itu apa?

Churn AccountWeeks ContractRenewal DataPlan DataUsage CustServCalls DayMins DayCalls MonthlyCharge OverageFee RoamMins

Explorasi Data dan Visualisasi

Data Cleansing

- Setelah dilakukan pengecekan, tidak terdapat data yang Null dan semua data bertipe numerik
- Dataset yang terdiri dari 3333 baris dan 11 kolom

	count	mean	std	min	25%	50%	75%	max
Churn	3333.0	0.144914	0.352067	0.0	0.00	0.00	0.00	1.00
AccountWeeks	3333.0	101.064806	39.822106	1.0	74.00	101.00	127.00	243.00
ContractRenewal	3333.0	0.903090	0.295879	0.0	1.00	1.00	1.00	1.00
DataPlan	3333.0	0.276628	0.447398	0.0	0.00	0.00	1.00	1.00
DataUsage	3333.0	0.816475	1.272668	0.0	0.00	0.00	1.78	5.40
CustServCalls	3333.0	1.562856	1.315491	0.0	1.00	1.00	2.00	9.00
DayMins	3333.0	179.775098	54.467389	0.0	143.70	179.40	216.40	350.80
DayCalls	3333.0	100.435644	20.069084	0.0	87.00	101.00	114.00	165.00
MonthlyCharge	3333.0	56.305161	16.426032	14.0	45.00	53.50	66.20	111.30
OverageFee	3333.0	10.051488	2.535712	0.0	8.33	10.07	11.77	18.19
RoamMins	3333.0	10.237294	2.791840	0.0	8.50	10.30	12.10	20.00

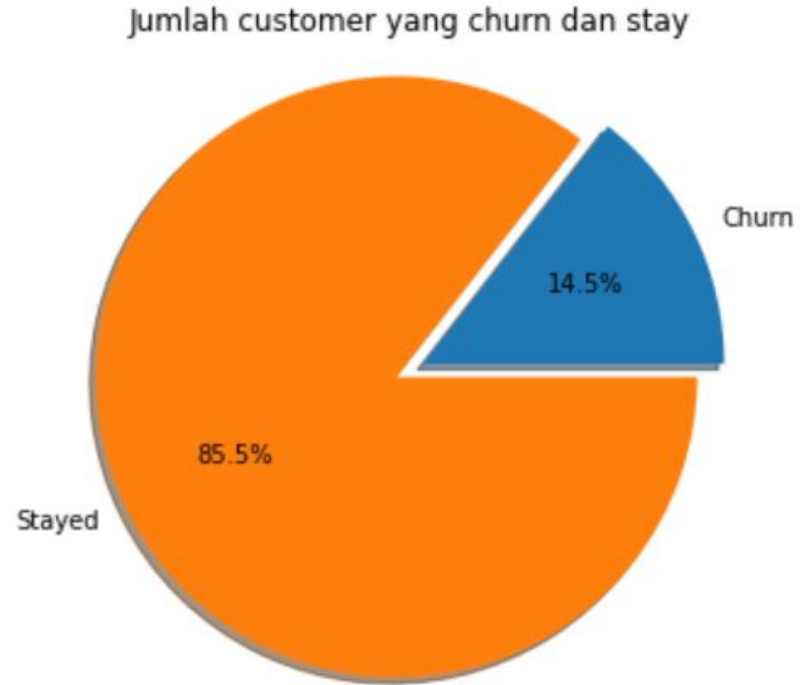
Exploratory Data Analysis

- Pada Heatmap disamping, terdapat beberapa korelasi diantaranya: Data Plan - Data Usage, Data plan - Monthly Charge, DayMins - Monthly Charge, Monthly Charge - Overage Fee.

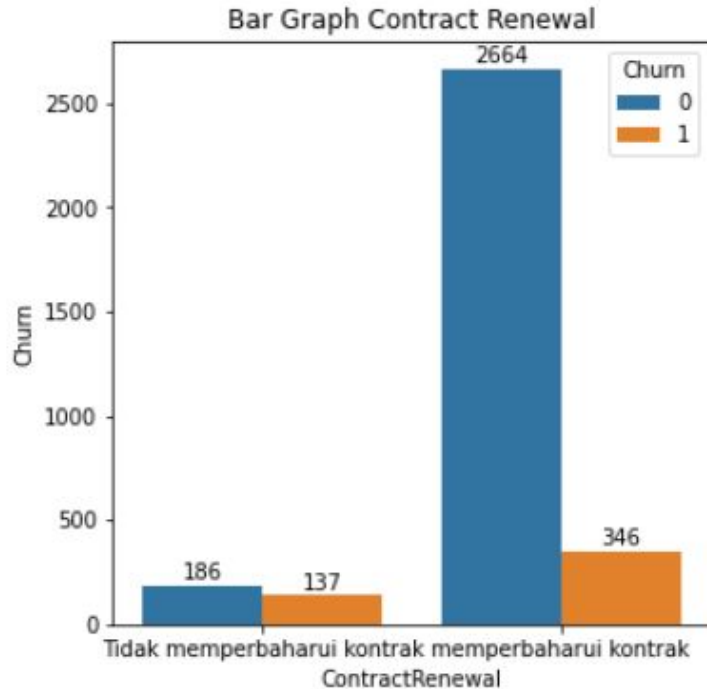


Exploratory Data Analysis

- Dapat dilihat pada visualisasi disamping, dataset yang sangat tidak seimbang akan menghasilkan model yang bias terhadap 'Churn'.
- 85% customer stayed atau masih tetap menggunakan jasa dan 14,5% lainnya "churn" atau sebanyak 483 orang

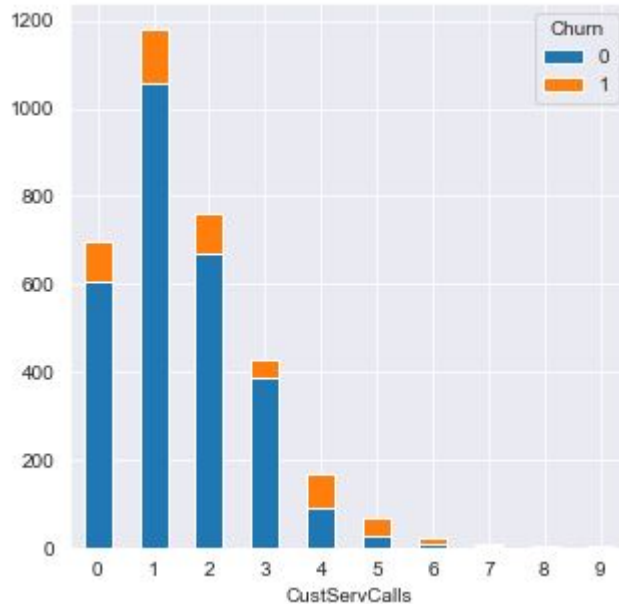


Exploratory Data Analysis



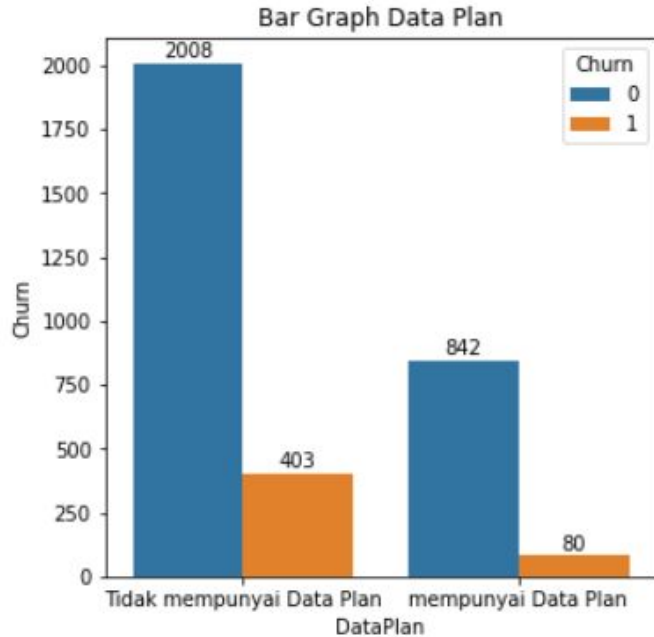
- 90% dari customer memperbaharui kontrak
- Jumlah customer churn yang memperbaharui kontrak lebih tinggi dibandingkan churn yang tidak memperbaharui kontrak (0).

Exploratory Data Analysis



- Semakin banyak customer melakukan panggilan ke CS, maka akan semakin rentan untuk churn.

Exploratory Data Analysis



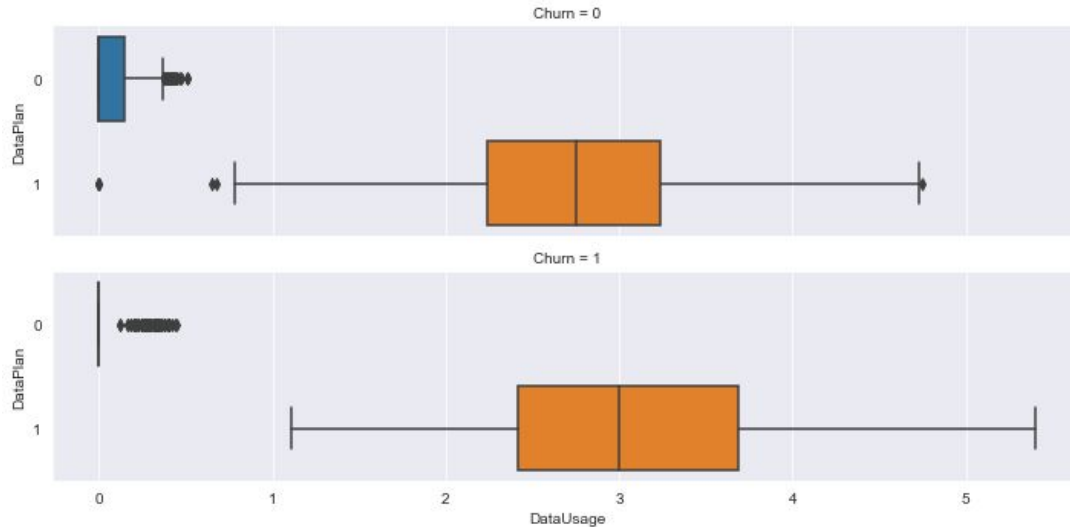
1. Dari keseluruhan dataset, 60% dari customer tidak mempunyai data plan
2. pada grup yang tidak mempunyai data plan, churnnya (1) lebih tinggi dibanding yang mempunyai data plan
3. 0 di DataUsage kemungkinan banyak karena faktor customer yang tidak mempunyai data plan

Exploratory Data Analysis



Dapat kita lihat pada scatterplot di samping, customer yang memiliki tagihan bulanan antara 60 - 80\$ dan dengan dayMins di atas 200 lebih cenderung untuk churn.

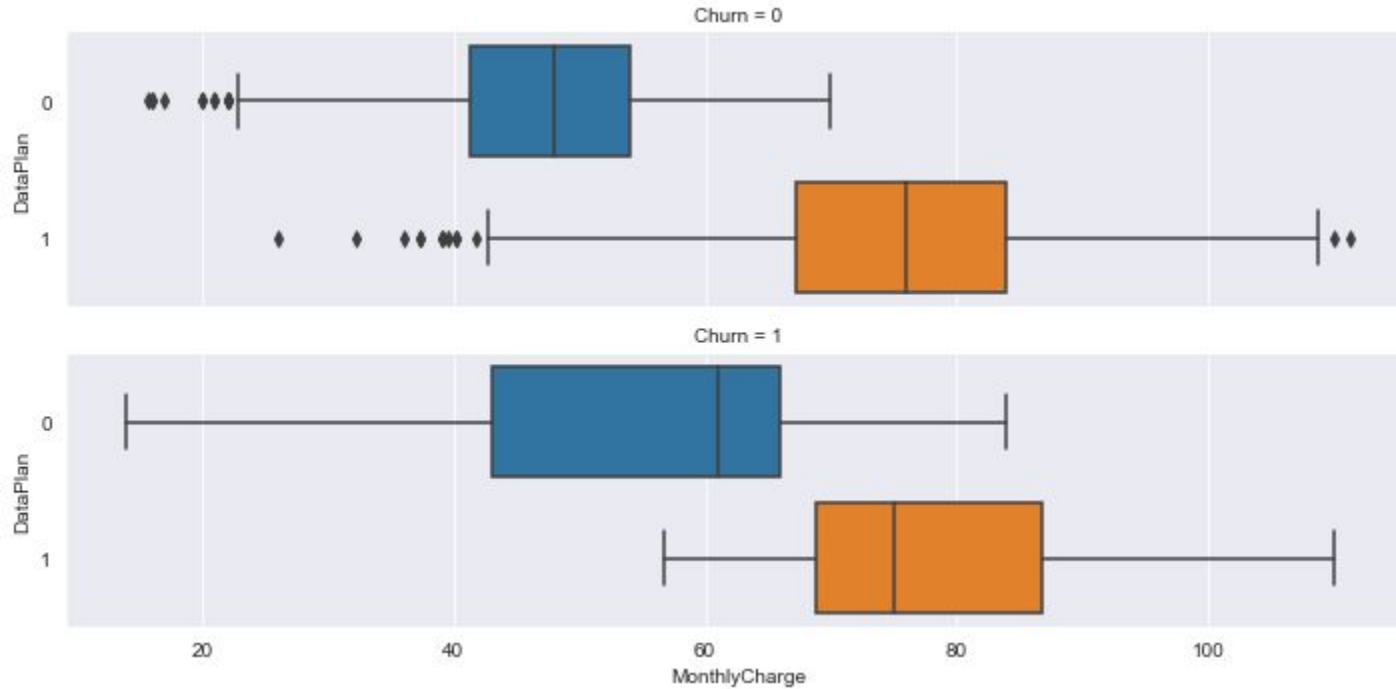
Exploratory Data Analysis



Customer yang mempunyai data plan mempunyai tingkat penggunaan data rata-rata yang lebih tinggi daripada yang tidak memiliki dataplan.

Dapat dilihat, customer yang churn menggunakan rata-rata data lebih banyak daripada yang tidak churn.

Exploratory Data Analysis



Biaya perbulan yang dibebankan untuk yang menggunakan data plan cenderung tinggi.

Exploratory Data Analysis

Modelling



Modelling

Pada tahap pemodelan, dataset dibagi menjadi dua bagian:

1. Data train 80% sebesar dari total data
2. Data test 20% sebesar dari total data

Tahap 1: Melakukan pemodelan

- ❖ Logistic Regression
- ❖ Decision Tree
- ❖ Random Forest
- ❖ KNN
- ❖ Gradient Boosting

Dan parameter terbaik didapatkan oleh Gradient Boosting dengan akurasi 0.9355 atau setara dengan 93.55% dan disusul dengan Random Forest senilai 93.40%

Modelling

Tahap 2: Melakukan Peningkatan akurasi

- ❖ Peningkatan akurasi dilakukan dengan melakukan resampling pada data serta dengan melakukan hyperparameter tuning pada model.
- ❖ Hyperparameter dilakukan dengan modul SearchGridCV

```
✓ [108] display(cv)
```

```
Best parameters are: {'learning_rate': 0.01, 'max_depth': 1, 'n_estimators': 50}
```

```
0.728 + or -0.001 for the {'learning_rate': 0.01, 'max_depth': 1, 'n_estimators': 5}
```

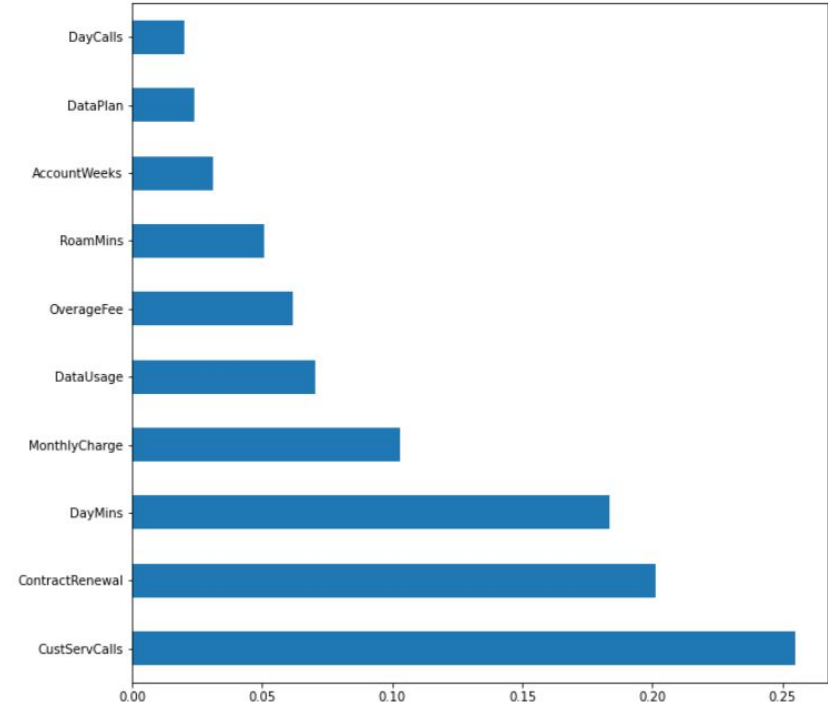
```
0.999 + or -0.001 for the {'learning_rate': 0.01, 'max_depth': 1, 'n_estimators': 50}
```

Hasil dari proses *GridSearchCV* dilihat pada gambar diatas, kita dapat memilih parameter yang terbaik dari hasil yang tertera. Pada hasil diatas parameter terbaiknya adalah `{'learning_rate': 0.01, 'max_depth': 1, 'n_estimators': 50}` yaitu dengan skor 0.999 untuk dilewatkan ke dalam *pipeline* yang telah kita buat.

Conclusion

Feature Selection

1. DayCalls adalah fitur yang paling sedikit untuk churn
2. CustServCalls adalah fitur yang paling banyak mempengaruhi pelanggan untuk churn, lalu disusul oleh Contract Renewal



Kesimpulan

Dari data hasil modeling didapatkan beberapa faktor yang harus diperhatikan agar dapat menurunkan tingkat churn:

- Mengoptimalkan biaya panggilan dan biaya data.
- Membuat paket data telepon bagi yang sering melakukan panggilan.
- Menawarkan dataplan bagi yang menggunakan data namun belum memiliki data plan.
- Memberikan dan menawarkan paket data yang lebih menarik.

Terima kasih!

Ada pertanyaan?

zenius



Kampus
Merdeka
INDONESIA JAYA