

zenius



Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 7

Nama Mentor: Erwin Fernanda

Nama:

- Mohammad Rizky Putra
- Evrida Indra Pramesti

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: Customer Churn

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

Problem: **classification**

Tujuan:

- Mengetahui faktor yang menyebabkan pelanggan Churn.
- Memprediksi model machine learning mengenai pelanggan yang melakukan Churn dengan aspek-aspek yang terkait

# Explorasi Data dan Visualisasi



# Business Understanding

Dengan pesatnya perkembangan industri telekomunikasi, penyedia layanan lebih condong ke arah perluasan basis pelanggan. Untuk memenuhi kebutuhan bertahan dalam lingkungan yang kompetitif, retensi pelanggan yang ada telah menjadi tantangan besar. Disebutkan bahwa biaya untuk mendapatkan pelanggan baru jauh lebih mahal daripada mempertahankan pelanggan yang sudah ada.

Customer churn atau perputaran pelanggan adalah metrik bisnis yang mengukur jumlah pelanggan yang telah berhenti menggunakan produk atau layanan Anda. Kadang-kadang disebut sebagai pergantian pelanggan atau pengurangan pelanggan, pergantian pelanggan bisnis adalah alat analitik yang kuat dan diukur sebagai persentase.

Oleh karena itu, sangat penting bagi industri telekomunikasi untuk menggunakan analisis canggih untuk memahami perilaku konsumen dan pada gilirannya memprediksi asosiasi pelanggan apakah mereka akan meninggalkan perusahaan atau tidak.

# Data Cleansing

Data Customer Churn memiliki 3333 baris dan 11 kolom dengan penjelasan sebagai berikut :

1. Churn = Pelanggan yang berhenti berlangganan, dengan 1 jika berhenti dan 0 jika tidak
2. AccountWeeks = lama berlangganan (dalam satuan minggu)
3. ContractRenewal = Pembaruan kontrak (1 jika memperbarui, 0 jika tidak)
4. DataPlan = paket data (1 jika punya 0 jika tidak)
5. DataUsage = penggunaan data (dalam gigabytes)



# Data Cleansing

- 6. CustServCalls = jumlah pelanggan menghubungi customer service
- 7. DayMins = rata-rata menit penggunaan per bulan
- 8. DayCall = rata-rata penggunaan telepon
- 9. MonthlyCharge = rata-rata tagihan perbulan
- 10. OverageFee = biaya kelebihan dalam 12 bulan terakhir
- 11. RoamMins = rata-rata telepon roaming

Berdasarkan analisis kelompok kami, data tersebut sudah bersih tidak ada data kosong serta data sudah sesuai tipe data dan saat dicek tidak ada outliers pada data tersebut.

# Exploratory Data Analysis

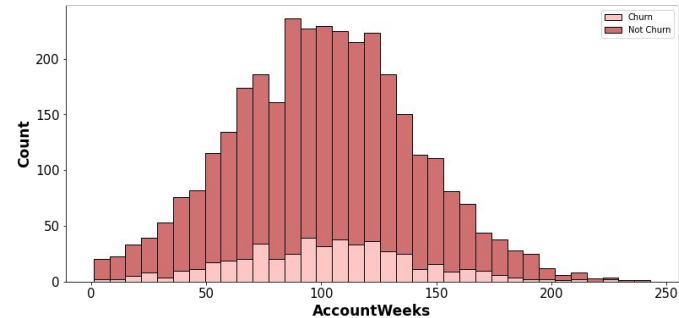
Berdasarkan visualisasi dan EDA yang telah dilakukan dapat diketahui beberapa hal berikut :

- Terdapat sekitar 14% pelanggan yang Churn atau sekitar 483 pelanggan.
- Durasi berlangganan tidak menjamin pelanggan Churn atau tidak.

**Perhitungan Churn**



**Durasi terhadap Churn**

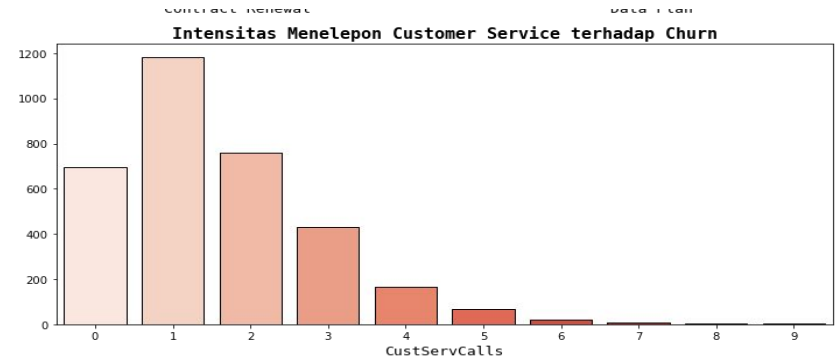
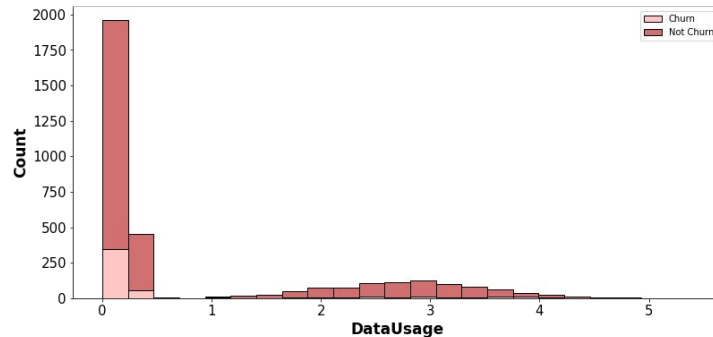


# Exploratory Data Analysis

Berdasarkan visualisasi dan EDA yang telah dilakukan dapat diketahui beberapa hal berikut :

- Pelanggan yang Churn rata-rata menggunakan data (Data Usage) kurang 1 GB.
- Rata-rata pelanggan melakukan panggilan kepada Cust. Service sebanyak 1-3 kali.

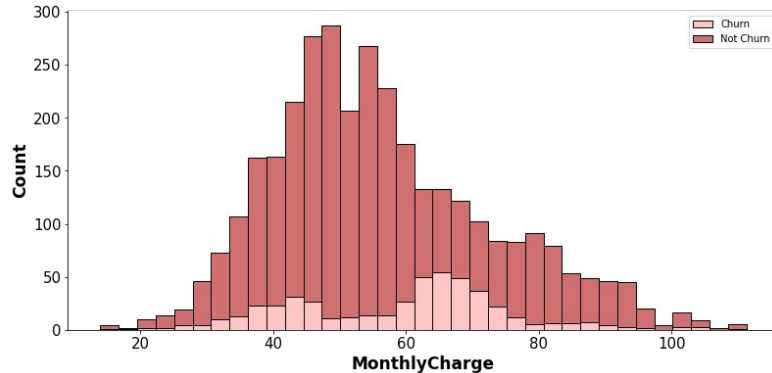
Penggunaan data terhadap Churn



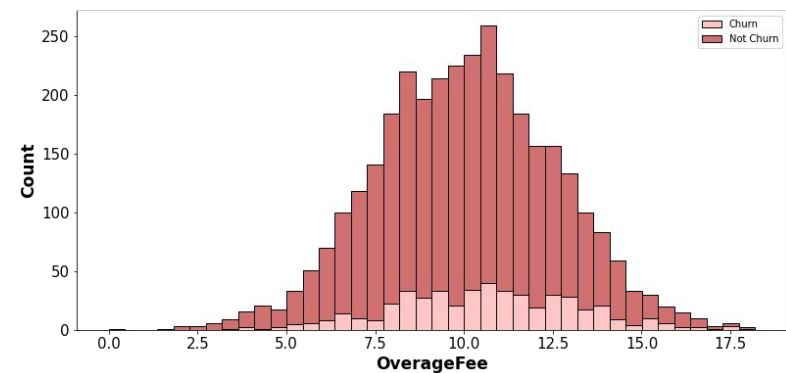
# Exploratory Data Analysis

Tagihan dan overpayment pada pelanggan yang Churn lebih besar daripada yang tidak Churn. Hal ini dapat menjadi alasan utama pelanggan tersebut Churn.

MonthlyCharge by churn



OverageFee by churn



# Modelling



# Metrik

Metrik yang digunakan adalah metrik confusion untuk mengevaluasi akurasi klasifikasi dengan menghitung matriks confusion dengan setiap baris yang sesuai dengan kelas yang sebenarnya. Terdiri dari :

- (TN) True Negative
- (FP) False Positive
- (FN) False Negative
- (TP) True Positive

# Train Test Split

Train/test split adalah salah satu metode yang dapat digunakan untuk mengevaluasi model dengan membagi dataset menjadi dua bagian yakni bagian yang digunakan untuk training data dan untuk testing data dengan proporsi tertentu (train size = 60%, test size = 40%)

# Jenis Model Awal Yang Dicoba (Logistic Regression)

- Evaluasi dengan Classification Report

Dari classification report disebutkan bahwa model memiliki rata-rata sekitar mulai dari precision, recall, f1-score, dan support. Accuracy juga menunjukkan akurasi 84%

- Evaluasi dengan AUC

Untuk score AUC dapat disebutkan bahwa nilainya sekitar 55,4%



# Logistic Regression (Hyperparameter Turning)

- Evaluasi dengan Classification Report

Setelah dilakukan Hyperparameter tuning nilai akurasi meningkat menjadi 86% dari sebelumnya 84%

- Evaluasi dengan AUC

Untuk score AUC disebutkan bahwa nilainya **tetap** sekitar 55,4%

# Decision Tree

- Evaluasi dengan Classification Report

Dari classification report disebutkan bahwa model memiliki rata-rata sekitar mulai dari precision, recall, f1-score, dan support. Accuracy juga menunjukkan akurasi 92%

- Evaluasi dengan AUC

Untuk score AUC dapat disebutkan bahwa nilainya sekitar 75%

# Random Forest

- Evaluasi dengan Classification Report

Dari classification report disebutkan bahwa model memiliki rata-rata sekitar mulai dari precision, recall, f1-score, dan support. Accuracy juga menunjukkan akurasi 93%

- Evaluasi dengan AUC

Untuk score AUC dapat disebutkan bahwa nilainya sekitar 80%

# Conclusion

## Kesimpulan

Berdasarkan evaluasi yang telah dilakukan pada ketiga model, model terbaik adalah Random forest karena memiliki akurasi dan AUC tertinggi diantara ketiga model

Dalam pemodelan ini, kelompok kami menggunakan semua variabel yang ada pada data karena dirasa variabel dalam data mempengaruhi pelanggan yang Churn.

<pada bagian ini, silakan paparkan kesimpulan Anda, apa saja insights/trend yang menarik, dan sertakan saran Anda kepada stakeholder>

<misal, proyek Anda adalah tentang properti, maka Anda bisa memberikan saran kepada calon pembeli properti, kira-kira properti yang seperti apa yang paling worth it>

<misal, proyek Anda adalah tentang churn, maka Anda bisa memberikan saran kepada perusahaan bagaimana untuk menurunkan churn, faktor-faktor apa yang harus diperhatikan, dst>

# Terima kasih!

Ada pertanyaan?

zenius



Kampus  
Merdeka  
INDONESIA JAYA