

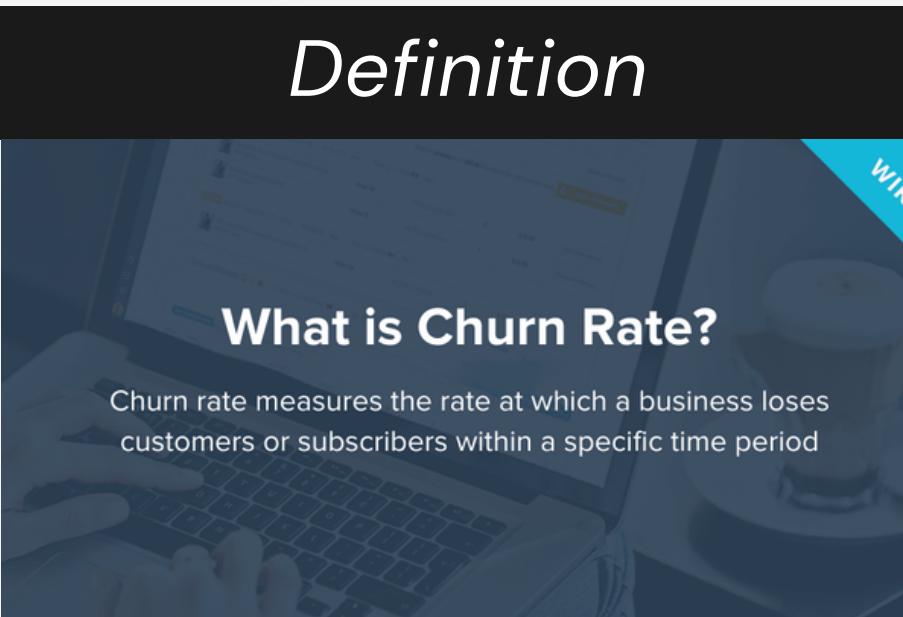


KELOMPOK 3

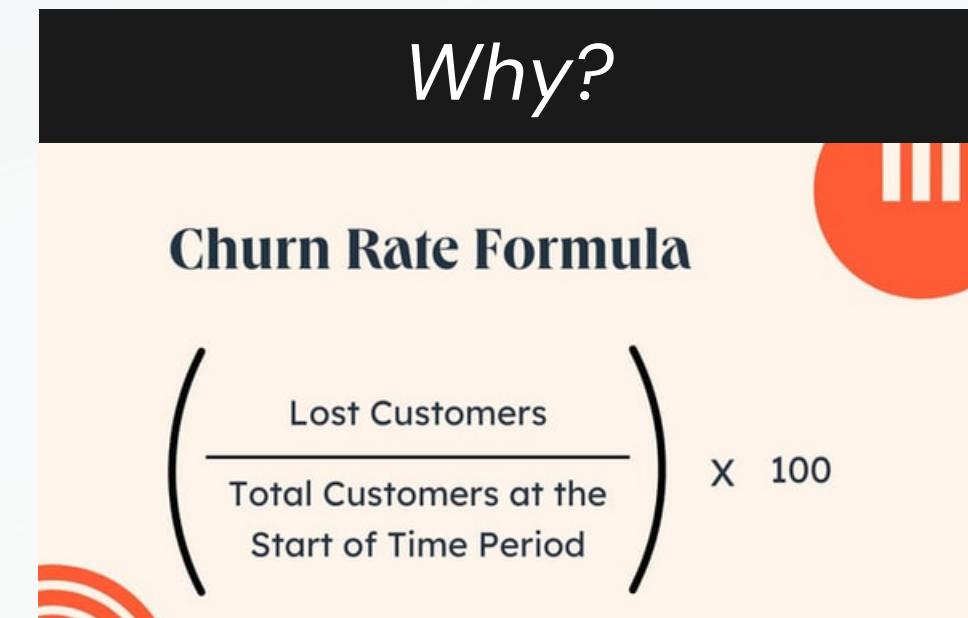
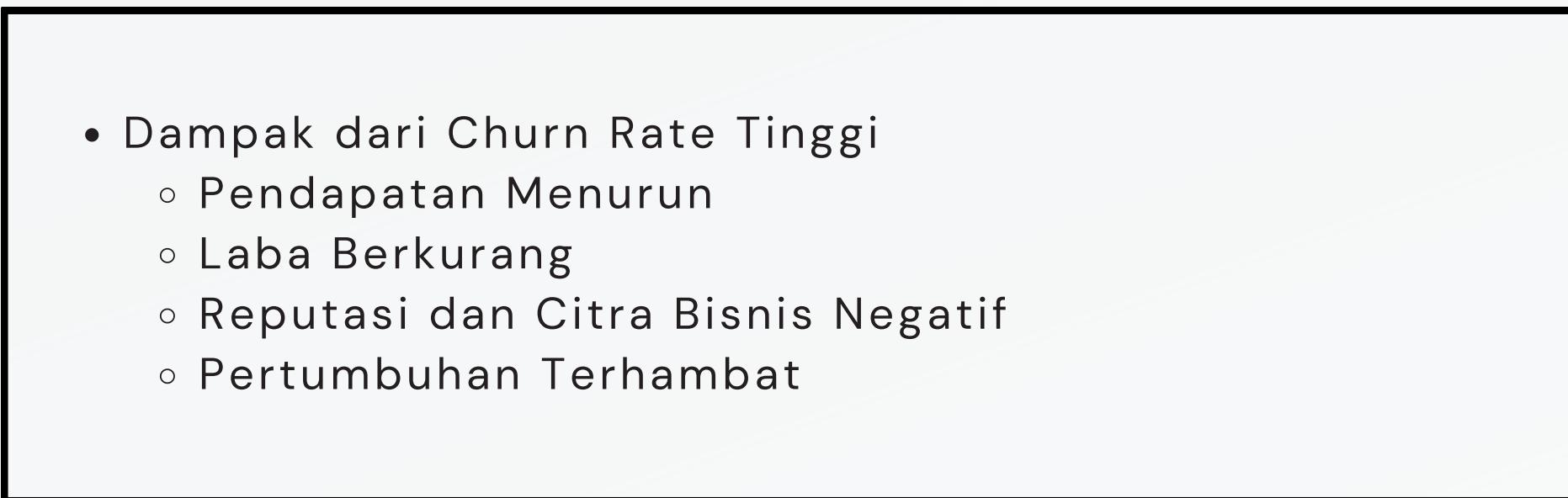
IMPLEMENTASI TEKNIK KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT CHURN PELANGGAN PADA PERUSAHAAN TELEKOMUNIKASI: STUDI KASUS PADA XYZ COMPANY

**SANDY REYNALDO YANUARDHA HAULUSSY
DAFFA RIFQI KANZ
MUHAMMAD RAFIF BAIHAQI**

LATAR BELAKANG



- Churn rate mengacu pada tingkat kehilangan pelanggan atau langganan dari suatu bisnis dalam jangka waktu tertentu.
- Churn rate dihitung dengan membagi jumlah pelanggan yang berhenti menggunakan produk atau layanan dengan total jumlah pelanggan pada awal periode tersebut.



RUANG LINGKUP

Penelitian ini akan difokuskan pada implementasi data mining untuk memprediksi tingkat churn pelanggan dengan menggunakan tiga metode klasifikasi utama, yaitu decision tree, random forest, dan logistic regression

TUJUAN

1

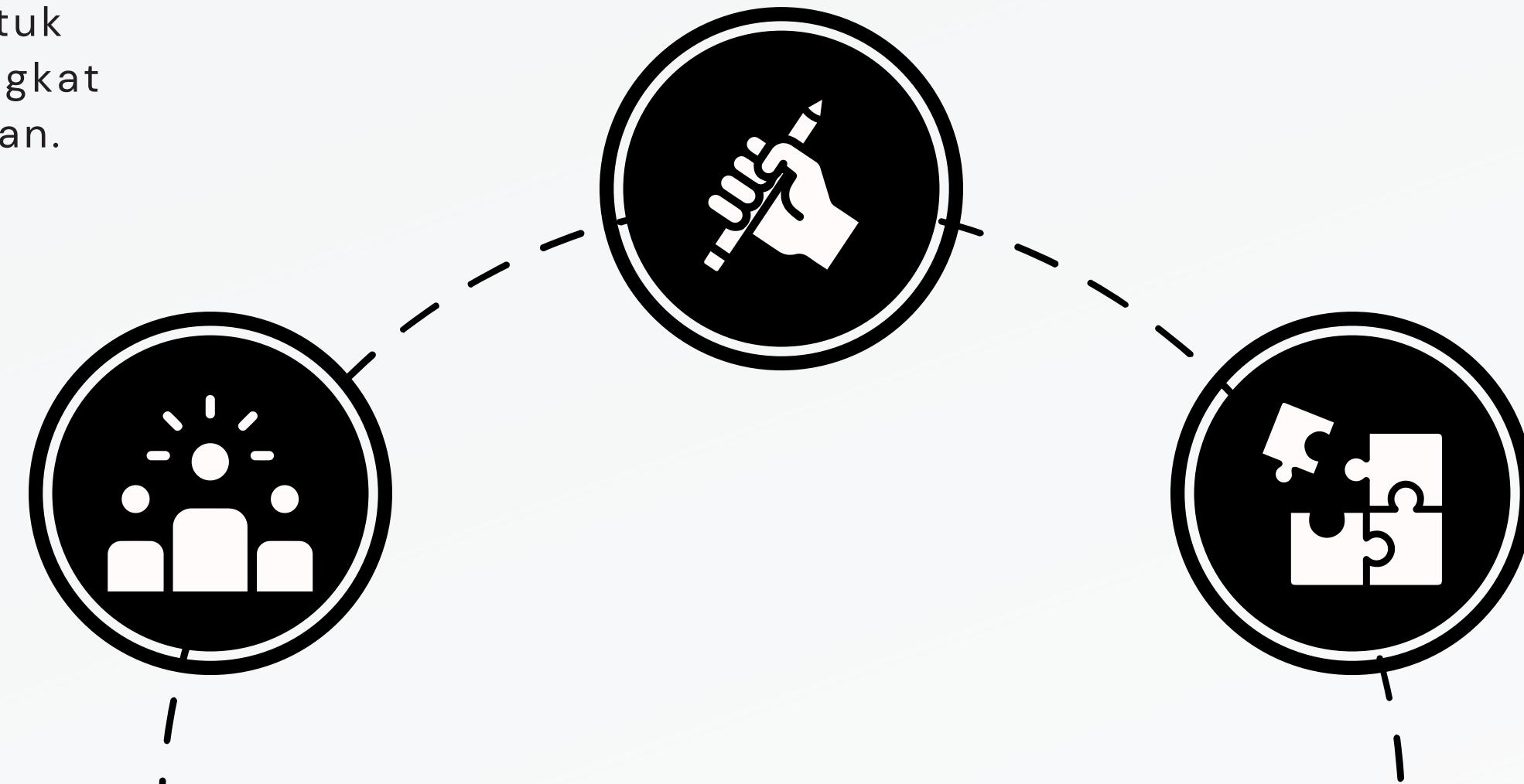
Menerapkan metode data mining, yaitu decision tree, random forest, dan logistic regression, untuk memprediksi tingkat churn pelanggan.

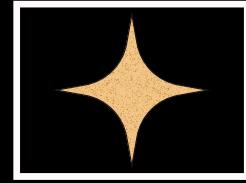
2

Mengevaluasi dan membandingkan kinerja ketiga metode klasifikasi dalam memprediksi tingkat churn pelanggan.

3

Mengidentifikasi faktor-faktor yang paling mempengaruhi churn pelanggan pada industri telekomunikasi.





MANFAAT

- 1 Membantu perusahaan telekomunikasi, seperti XYZ Company, mengidentifikasi pelanggan yang berpotensi churn sehingga dapat mengambil aksi pencegahan yang tepat.
- 2 Mengoptimalkan strategi retensi pelanggan dengan memanfaatkan hasil prediksi churn yang akurat.
- 3 Memberikan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi churn pelanggan pada industri telekomunikasi.
- 4 Memberikan kontribusi terhadap pengembangan pengetahuan dan aplikasi data mining dalam domain prediksi churn pelanggan pada perusahaan telekomunikasi.

TINJAUAN PUSTAKA



Tingkat churn mengacu pada persentase pelanggan yang berhenti menggunakan layanan atau berpindah ke penyedia layanan lain dalam periode waktu tertentu dan mempengaruhi keuntungan sebuah perusahaan (Suryana 2017).

CHURN RATE



Menurut Mardi (2017). Data mining adalah suatu proses untuk menemukan pola atau informasi menarik dalam kumpulan data yang telah dipilih dengan menggunakan berbagai teknik atau metode

DATA MINING



Klasifikasi adalah teknik data mining yang digunakan untuk mengelompokkan data ke dalam kelas yang sudah ditentukan berdasarkan fitur-fitur tertentu (Mardi 2017).

KLASIFIKASI

TINJAUAN PUSTAKA



Decision tree dapat digunakan dalam memprediksi sebuah klasifikasi pelanggan berdasarkan tingkat perbandingan confidence antara prediksi loyal dan churn (Nurzahputra et al. 2017)

DECISION TREE



Menurut Husein et al. (2021), Random forest memiliki akurasi yang paling baik dibandingkan dengan KNN classifier, Logistic Regression, dan Linear SVC dalam menentukan prediksi churn

RANDOM FOREST



Menurut Mujaddid et al. (2017), Logistic regression dapat digunakan dalam memprediksi tingkat churn dengan bentuk biner dan memiliki akurasi sangat baik, dengan SMOTE memiliki akurasi sebesar 90%

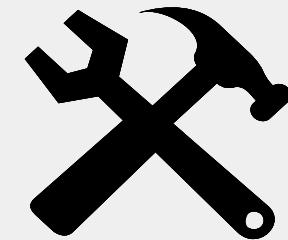
LOGISTIC REGRESSION

METODE



Data yang digunakan:
Telco Customer Churn ([Kaggle](#))

Teknik Data Mining:
Klasifikasi (Logistic Regression, Decision Tree, Random Forest)



Lingkungan Pengembangan:

- Google Collaboratory, Weka
- Python
- sklearn, pandas.

WA_Fn-UseC_Telco-Customer-Churn.csv (977.5 kB)						
Detail	Compact	Column				
About this file						
Telcom Customer Churn						
Each row represents a customer, each column contains customer's attributes described on the right.						
The raw data contains 7043 rows (customers) and 21 columns (features).						
The "Churn" column is our target.						
▲ customerID	≡	▲ gender	≡	# SeniorCitizen		
Customer ID		Whether the customer is a male or a female		Whether the customer is a senior citizen or not (1, 0)		
7043 unique values		Male 50%	50%	✓ Partner		
		Female		Whether the customer has a partner (Yes/No)		
7590-VHVEG		Female	0	Yes		
5575-GNVDE		Male	0	No		
3668-QPYBK		Male	0	No		
7795-CFOCW		Male	0	No		
9237-HQITU		Female	0	No		
9305-CDSKC		Female	0	No		
1452-KIOVK		Male	0	No		

TAHAPAN PENELITIAN



PENGUMPULAN
& EKSPLORASI
DATA

PRA PROSES
DATA

IMPLEMENTASI
ALGORITME
KLASIFIKASI

EVALUASI &
PENARIKAN
KESIMPULAN

HASIL EKSPLORASI DATA

Karakteristik Data

```
RangeIndex: 7043 entries, 0 to 7042  
Data columns (total 21 columns):  
 #   Column           Non-Null Count Dtype  
 ---  --  
 0   customerID      7043 non-null  object  
 1   gender          7043 non-null  object  
 2   SeniorCitizen    7043 non-null  int64  
 3   Partner          7043 non-null  object  
 4   Dependents       7043 non-null  object  
 5   tenure           7043 non-null  int64  
 6   PhoneService     7043 non-null  object  
 7   MultipleLines    7043 non-null  object  
 8   TotalCharges     7043 non-null  float64  
 9   MonthlyCharges  7043 non-null  float64  
 10  Total账单          7043 non-null  object  
 11  PaymentMethod    7043 non-null  object  
 12  PaperlessBilling 7043 non-null  object  
 13  MonthlyCharges  7043 non-null  float64  
 14  TotalCharges     7043 non-null  float64  
 15  Contract         7043 non-null  object  
 16  InternetService  7043 non-null  object  
 17  PaymentMethod    7043 non-null  object  
 18  MonthlyCharges  7043 non-null  float64  
 19  TotalCharges     7043 non-null  float64  
 20  Churn            7043 non-null  object  
 dtypes: float64(1), int64(2), object(18)
```

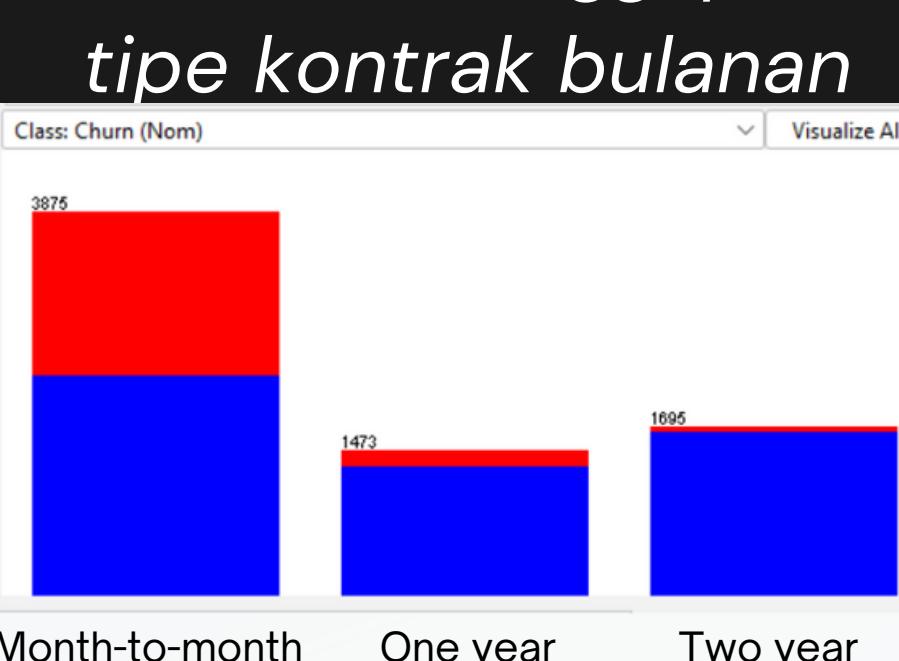
Terdiri atas:

- 21 Kolom: 20 Atribut predictor, 1 Target ("Churn")
- 7043 Baris: 7043 record pelanggan perusahaan'

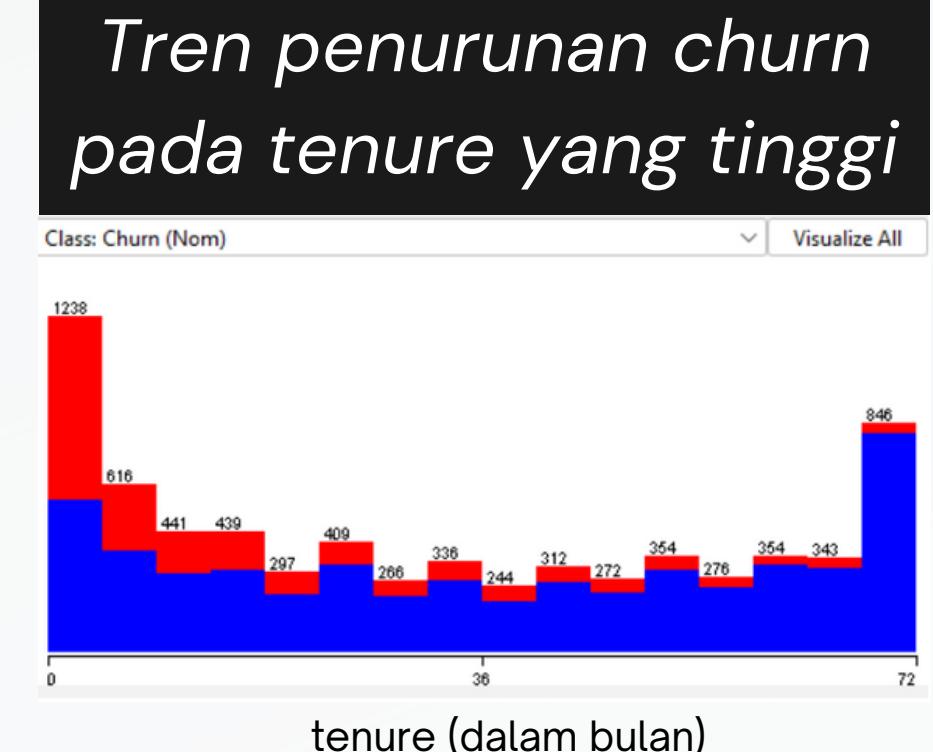
Tipe Data:

- object --> Kategorik
- int64 & float64 --> Numerik

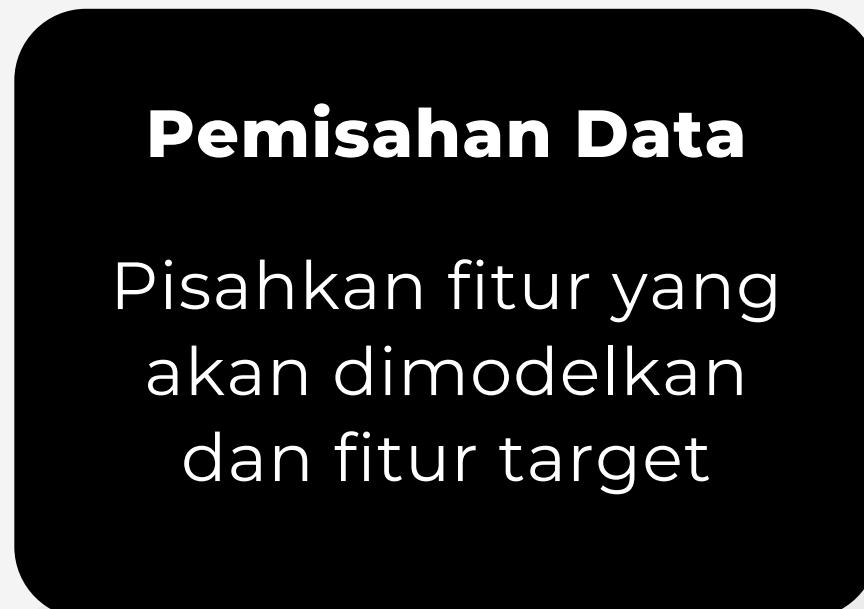
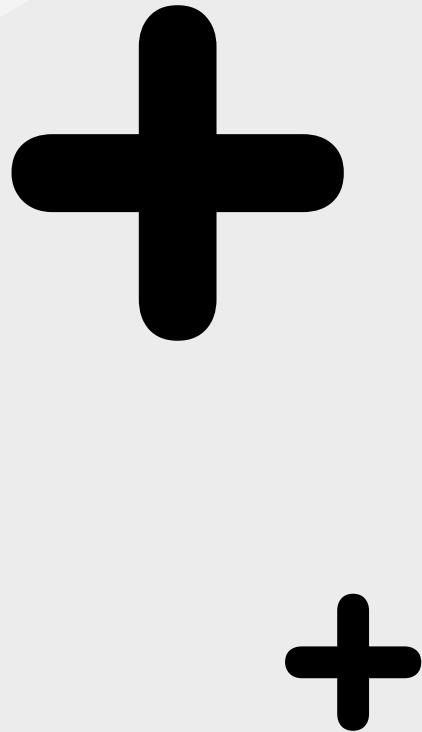
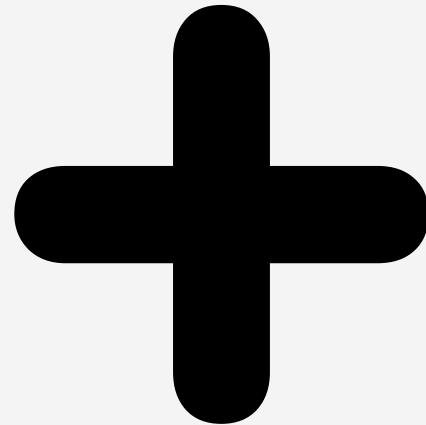
Tren churn tinggi pada tipe kontrak bulanan



Tren penurunan churn pada tenure yang tinggi



PRA-PROSES

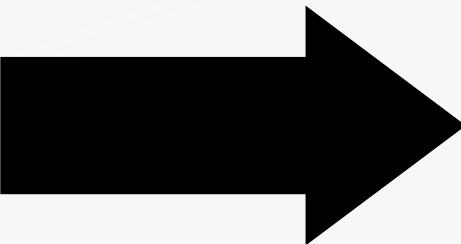


PEMISAHAN DATA

Dataset Awal

Data columns (total 21 columns):			
#	Column	Non-Null Count	Dtype
0	customerID	7043 non-null	object
1	gender	7043 non-null	object
2	SeniorCitizen	7043 non-null	int64
3	Partner	7043 non-null	object
4	Dependents	7043 non-null	object
5	tenure	7043 non-null	int64
6	PhoneService	7043 non-null	object
7	MultipleLines	7043 non-null	object
8	InternetService	7043 non-null	object
9	OnlineSecurity	7043 non-null	object
10	OnlineBackup	7043 non-null	object
11	DeviceProtection	7043 non-null	object
12	TechSupport	7043 non-null	object
13	StreamingTV	7043 non-null	object
14	StreamingMovies	7043 non-null	object
15	Contract	7043 non-null	object
16	PaperlessBilling	7043 non-null	object
17	PaymentMethod	7043 non-null	object
18	MonthlyCharges	7043 non-null	float64
19	TotalCharges	7043 non-null	object
20	Churn	7043 non-null	object

Buang
customerID
(tidak
digunakan)



Pisahkan
variabel
target
(Churn)

Setelah Pemisahan Data

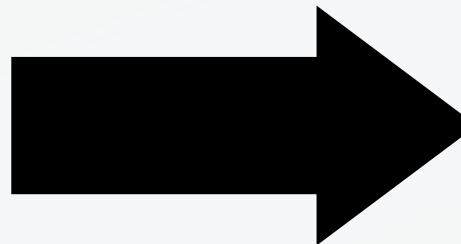
Data columns (total 19 columns):			
#	Column	Non-Null Count	Dtype
0	gender	7043 non-null	object
1	SeniorCitizen	7043 non-null	int64
2	Partner	7043 non-null	object
3	Dependents	7043 non-null	object
4	tenure	7043 non-null	int64
5	PhoneService	7043 non-null	object
6	MultipleLines	7043 non-null	object
7	InternetService	7043 non-null	object
8	OnlineSecurity	7043 non-null	object
9	OnlineBackup	7043 non-null	object
10	DeviceProtection	7043 non-null	object
11	TechSupport	7043 non-null	object
12	StreamingTV	7043 non-null	object
13	StreamingMovies	7043 non-null	object
14	Contract	7043 non-null	object
15	PaperlessBilling	7043 non-null	object
16	PaymentMethod	7043 non-null	object
17	MonthlyCharges	7043 non-null	float64
18	TotalCharges	7043 non-null	object

KONVERSI FITUR

Sebelum Konversi

Data columns (total 19 columns):			
#	Column	Non-Null Count	Dtype
0	gender	7043 non-null	object
1	SeniorCitizen	7043 non-null	int64
2	Partner	7043 non-null	object
3	Dependents	7043 non-null	object
4	tenure	7043 non-null	int64
5	PhoneService	7043 non-null	object
6	MultipleLines	7043 non-null	object
7	InternetService	7043 non-null	object
8	OnlineSecurity	7043 non-null	object
9	OnlineBackup	7043 non-null	object
10	DeviceProtection	7043 non-null	object
11	TechSupport	7043 non-null	object
12	StreamingTV	7043 non-null	object
13	StreamingMovies	7043 non-null	object
14	Contract	7043 non-null	object
15	PaperlessBilling	7043 non-null	object
16	PaymentMethod	7043 non-null	object
17	MonthlyCharges	7043 non-null	float64
18	TotalCharges	7043 non-null	object

**Konversi Fitur
Kategorik
menjadi
Numerik**



Setelah Konversi

Data columns (total 19 columns):			
#	Column	Non-Null Count	Dtype
0	gender	7043 non-null	int64
1	SeniorCitizen	7043 non-null	int64
2	Partner	7043 non-null	int64
3	Dependents	7043 non-null	int64
4	tenure	7043 non-null	int64
5	PhoneService	7043 non-null	int64
6	MultipleLines	7043 non-null	int64
7	InternetService	7043 non-null	int64
8	OnlineSecurity	7043 non-null	int64
9	OnlineBackup	7043 non-null	int64
10	DeviceProtection	7043 non-null	int64
11	TechSupport	7043 non-null	int64
12	StreamingTV	7043 non-null	int64
13	StreamingMovies	7043 non-null	int64
14	Contract	7043 non-null	int64
15	PaperlessBilling	7043 non-null	int64
16	PaymentMethod	7043 non-null	int64
17	MonthlyCharges	7043 non-null	float64
18	TotalCharges	7043 non-null	object

NORMALISASI

Sebelum Normalisasi Data

TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges
0	0	0	0	1	2	29.85	29.85
0	0	0	1	0	3	56.95	1889.50
0	0	0	0	1	3	53.85	108.15
2	0	0	1	0	0	42.30	1840.75
0	0	0	0	1	2	70.70	151.65

Tiap fitur memiliki rentang & skala yang berbeda-beda

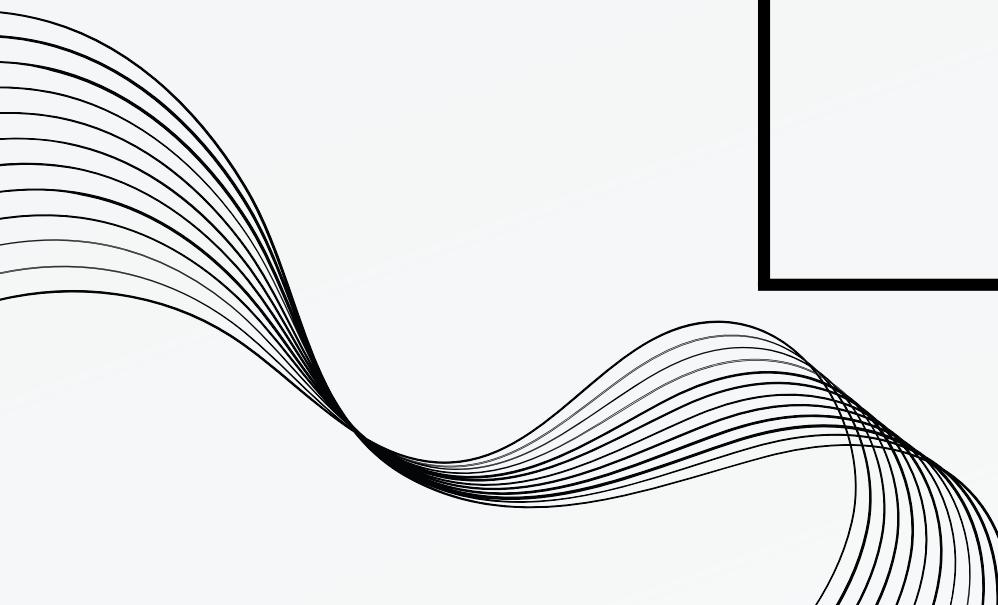
Semua fitur memiliki skala yang sama di rentang 0-1

Setelah Normalisasi Data

TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges
0.0	0.0	0.0	0.0	0.0	1.0	0.666667	0.115423
0.0	0.0	0.0	0.5	0.0	1.000000	0.385075	0.215867
0.0	0.0	0.0	0.0	1.0	1.000000	0.354229	0.010310
1.0	0.0	0.0	0.5	0.0	0.000000	0.239303	0.210241
0.0	0.0	0.0	0.0	1.0	0.666667	0.521891	0.015330

HASIL FEATURE SELECTION

LassoCV	SelectKBest
<p>Fitur terpilih:</p> <ul style="list-style-type: none">1. MonthlyCharges2. tenure3. TotalCharges	<p>Fitur terpilih:</p> <ul style="list-style-type: none">1. Contract2. OnlineSecurity3. TechSupport4. tenure5. SeniorCitizen6. Dependents7. OnlineBackup8. PaperlessBilling



HASIL & PEMBAHASAN

Matriks Evaluasi	Logistic Regression			Decision Tree	Random Forest
	Tanpa Feature Selection	Seleksi dengan Lasso	Seleksi dengan KBest		
Validation Accuracy	80.82%	79.12%	78.55%	77.41%	78.27%
Testing Accuracy	82.41%	80.57%	80.14%	78.87%	79.72%

CATATAN

Sebagai benchmark, kriteria akurasi yang dapat diterima (acceptable) pada penelitian-penelitian sebelumnya adalah 80%

HASIL & PEMBAHASAN

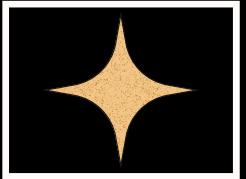
**Mengapa seleksi
fitur tidak
meningkatkan
akurasi?**

Hipotesis Awal:

- Akurasi mengalami penurunan karena model Lasso
- KBest kurang berhasil dalam memilih semua fitur yang penting bagi model, sehingga akurasi menurun.

Tetapi:

- Nilai akurasi tersebut tetap lebih tinggi dibandingkan klasifikasi dengan metode Decision Tree & Random Forest.
- Berhasil mengurangi kompleksitas model dari 19 fitur menjadi hanya 3 fitur.



ANALISIS LANJUTAN

Semua fitur pada dataset memiliki kontribusi positif terhadap akurasi prediksi churn.

Score importance positif pada SelectKBest mengimplikasikan bahwa fitur tersebut memiliki relevansi tinggi terhadap target prediksi.

Oleh karena itu, jika fitur tersebut diseleksi, dapat terjadi penurunan tingkat akurasi.

	Feature	Score
14	Contract	557.890083
8	OnlineSecurity	275.805765
11	TechSupport	261.651933
4	tenure	226.096162
1	SeniorCitizen	134.351545
3	Dependents	133.036443
9	OnlineBackup	115.043260
15	PaperlessBilling	105.680863
10	DeviceProtection	95.651570
2	Partner	82.412083
18	TotalCharges	72.992350
17	MonthlyCharges	50.995371
16	PaymentMethod	19.497417
7	InternetService	4.910514
6	MultipleLines	4.873460
13	StreamingMovies	4.117700
12	StreamingTV	3.745102
0	gender	0.258699
5	PhoneService	0.097261

Score Importance dari SelectKBest

FITUR TERPENTING UNTUK PREDIKSI CHURN

Berdasarkan koefisien fungsi regresi (pendekatan Logistic Regression):

- tenure (-3.08330268246702)
- MonthlyCharges (2.3117954997474692)
- Contract (-1.4521506422703168)
- TotalCharges (1.2612188720884403)
- PhoneService (-0.9536136782369321)

Semakin koefisien fitur dalam fungsi regresi menjauhi 0, semakin besar peran fitur tersebut dalam prediksi.

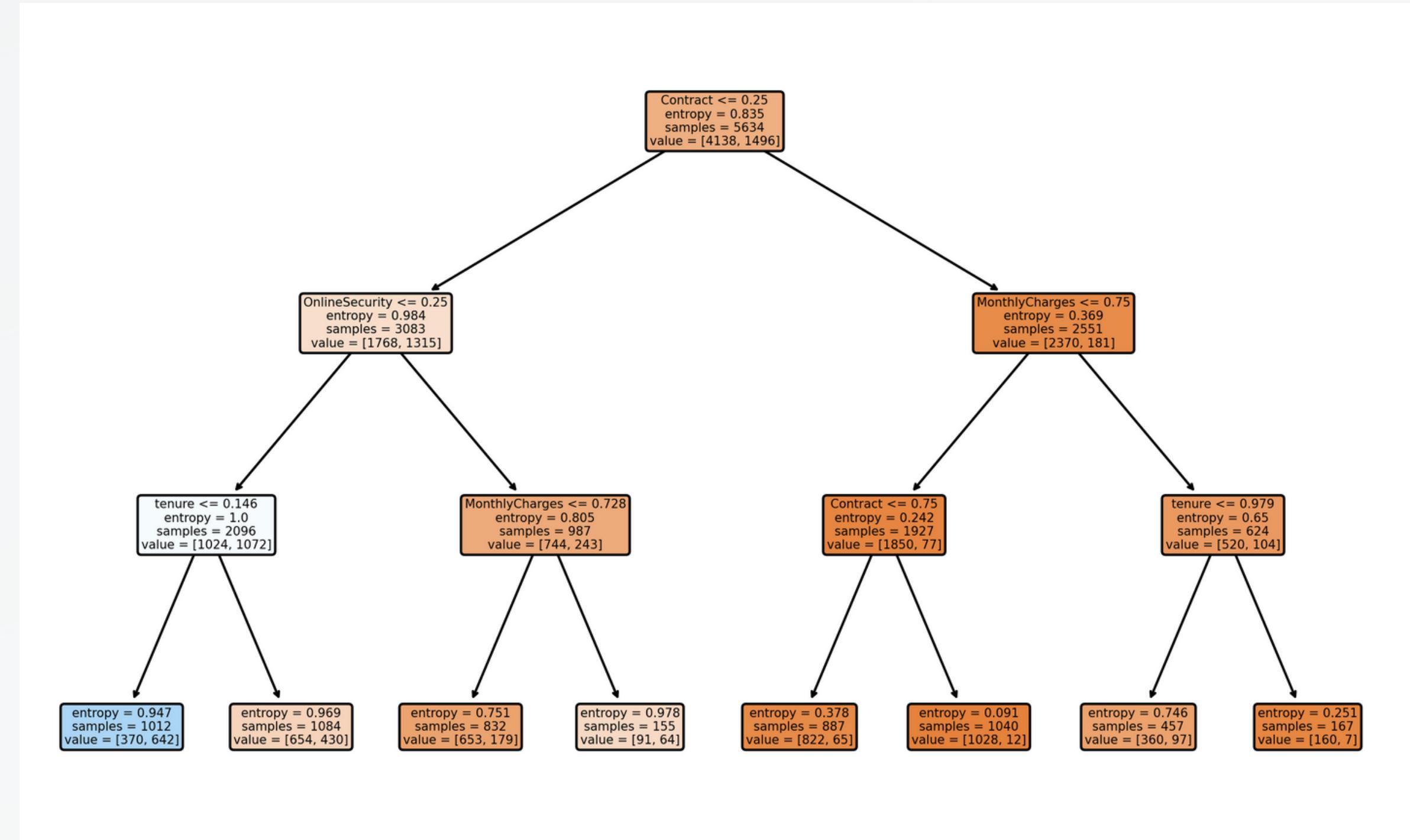
```
Regression function:  
Intercept: [-0.25082335]  
Coefficient for gender: -0.05353614322050116  
Coefficient for SeniorCitizen: 0.1848241387056489  
Coefficient for Partner: 0.05950153052809574  
Coefficient for Dependents: -0.1880955620394571  
Coefficient for tenure: -3.08330268246702  
Coefficient for PhoneService: -0.9536136782369321  
Coefficient for MultipleLines: 0.15458323693266884  
Coefficient for InternetService: 0.45025687093741773  
Coefficient for OnlineSecurity: -0.5841800227508095  
Coefficient for OnlineBackup: -0.3023510049020858  
Coefficient for DeviceProtection: -0.15615847331372082  
Coefficient for TechSupport: -0.5076453578443438  
Coefficient for StreamingTV: -0.010392861112327127  
Coefficient for StreamingMovies: 0.08843524249838351  
Coefficient for Contract: -1.4521506422703168  
Coefficient for PaperlessBilling: 0.36041528243499554  
Coefficient for PaymentMethod: 0.18988408385829372  
Coefficient for MonthlyCharges: 2.3117954997474692  
Coefficient for TotalCharges: 1.2612188720884403
```

FITUR TERPENTING UNTUK PREDIKSI CHURN

Berdasarkan posisi cabang pada Decision Tree:

- Contract (root)
- OnlineSecurity (tingkat 1)
- MonthlyCharges (tingkat 1)
- tenure (tingkat 2)

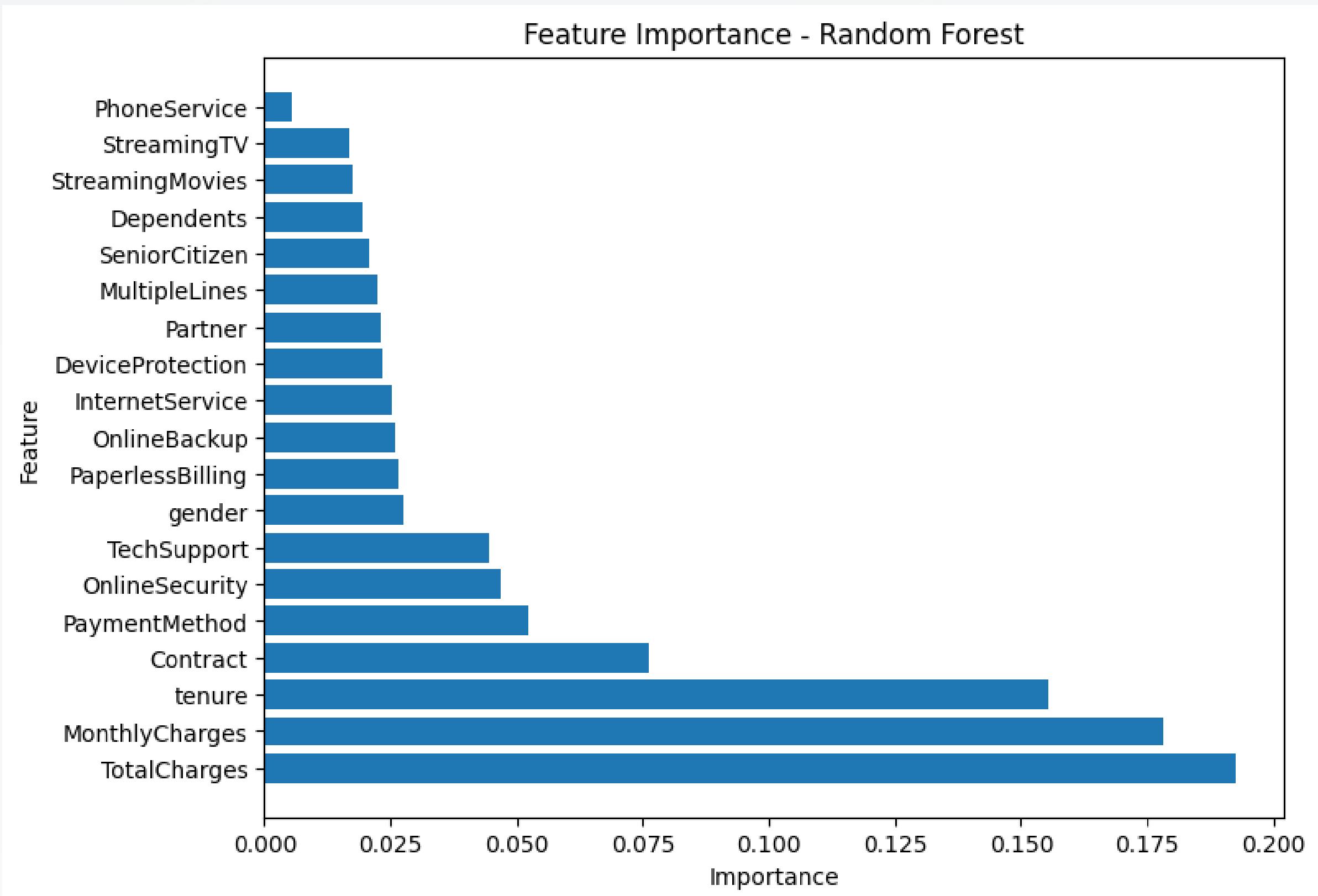
Semakin tinggi posisi suatu fitur pada decision tree, semakin signifikan pengaruhnya dalam prediksi.



FITUR TERPENTING UNTUK PREDIKSI CHURN

Berdasarkan plot skor importance Random Forest:

- TotalCharges
- MonthlyCharges
- tenure
- Contract
- PaymentMethod



KESIMPULAN

Algoritme Logistic Regression tanpa seleksi fitur menghasilkan model yang paling akurat untuk memprediksi churn pelanggan dengan nilai akurasi testing 82.41% dibandingkan algoritme Decision Tree (78.87%) dan Random Forest (79.72%).

Penggunaan feature selection baik dengan Lasso maupun KBest menyebabkan penurunan nilai akurasi menjadi 80.57% (Lasso) dan 80.14% (KBest). Namun, dengan memilih fitur-fitur yang penting saja, performa model saat training maupun saat inferensi dapat ditingkatkan.

Beberapa fitur yang memiliki peran penting dalam memprediksi keputusan churn pelanggan diantaranya tenure, MonthlyCharges, Contract, dan TotalCharges

TERIMA KASIH



KELOMPOK 3

