FINAL PROJECT

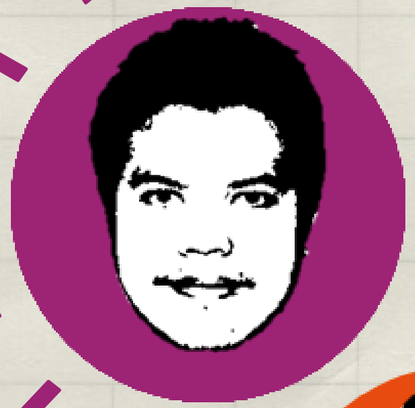
Telco Customer Churn

OMICRON

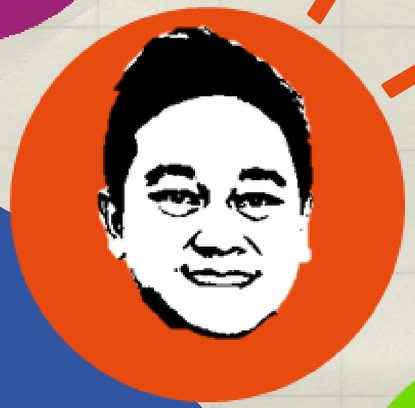
Data Science Bootcamp Batch 11

DigitalSkola

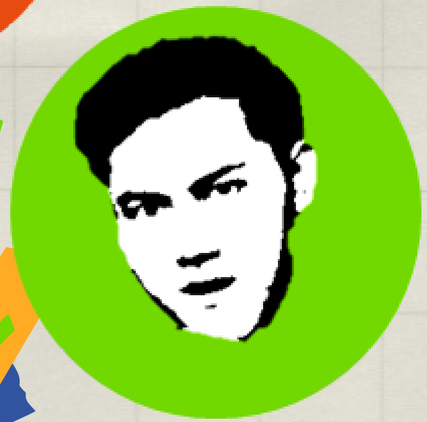
Members



Anugrah Yazid Ghani
https://www.linkedin.com/in/anugrah-yazid-7253bb221/

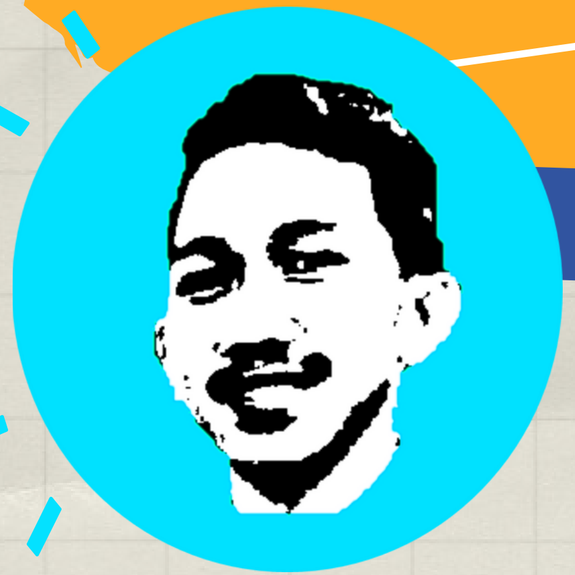


Edo M Hadad Gibran
https://www.linkedin.com/in/edo-gibran-38505a142/



Muhammad Fikri Fadila
https://www.linkedin.com/in/muhammad-fikri-fadila-a551161a6/

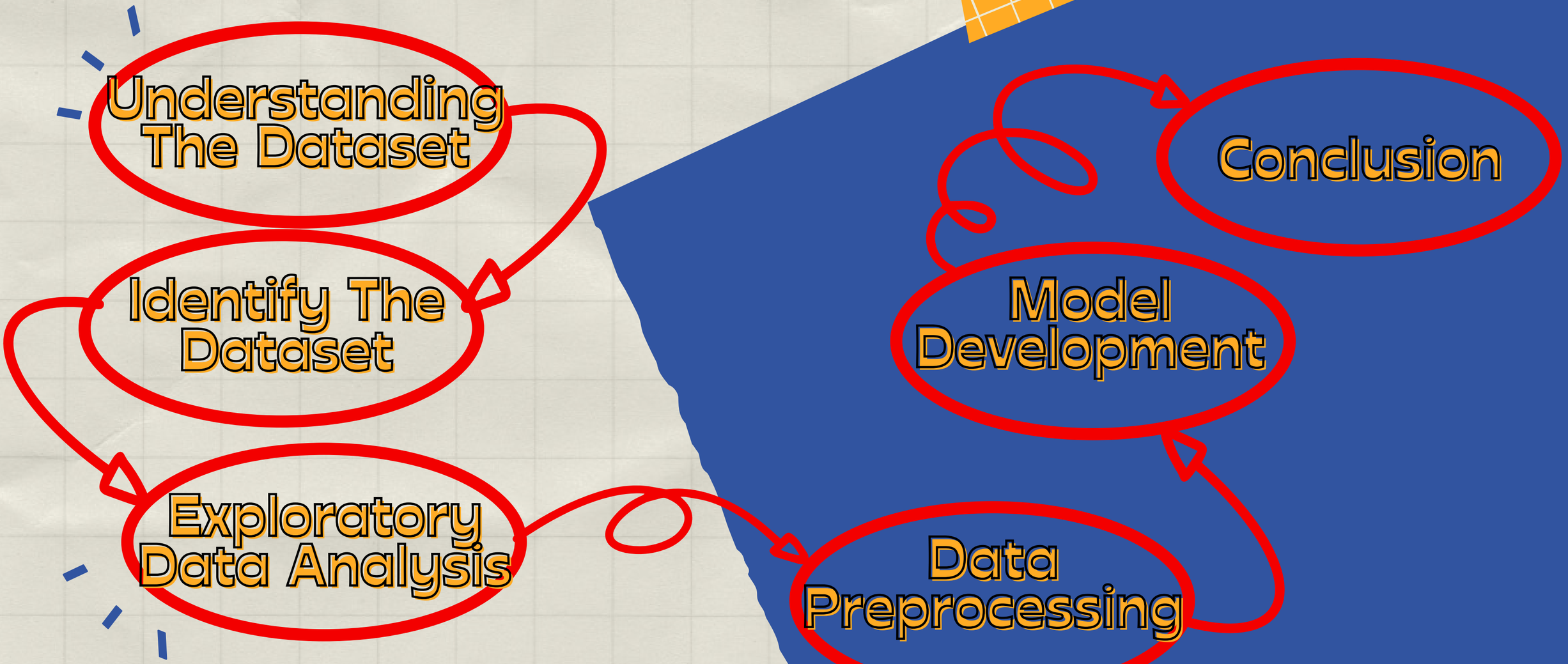
Mentor



Ari Sulistyo Prabowo
https://www.linkedin.com/in/ariprabowo/

DigitalSkola

Modelling Project Stage



DigitalSkola

1

UNDERSTANDING DATASET

DigitalSkola

DigitalSkola

TELCO

- Produsen peralatan telekomunikasi.
- Didirikan pada tahun 1972
- Kantor pusat di Mansfield, Massachusetts.

Fokus pada 4 primary vertical market :

Carrier Cloud Networking & Cloud Services

Business Ethernet Services

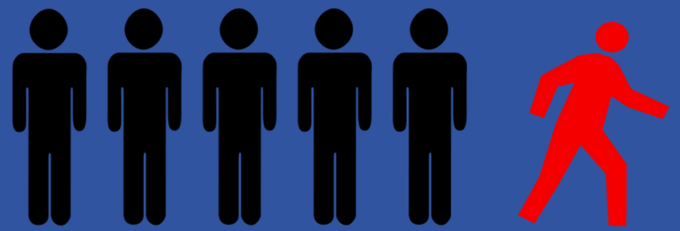
Mobile Backhaul

ATCA Switching Blades

www.telco.com

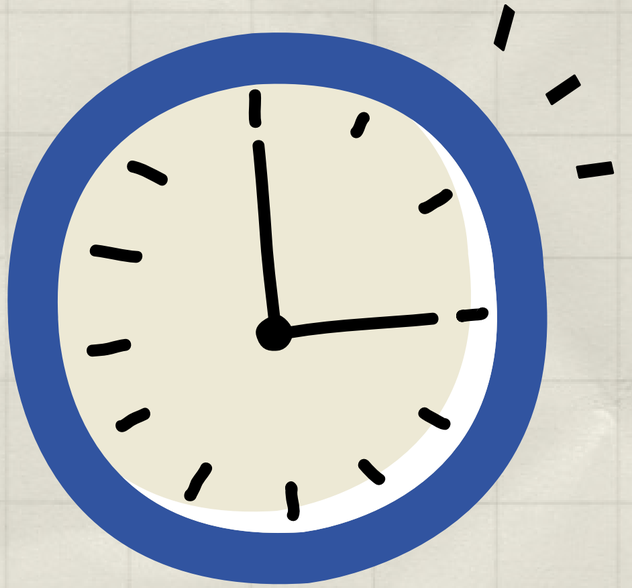
DigitalSkola

CUSTOMER CHURN



Persentase Customer yang berhenti menggunakan layanan atau produk dari perusahaan selama jangka waktu tertentu.

Why is it so important?



CUSTOMER CHURN

FORRESTER®

"cost 5x to attract new customer than keep the existing one"

HARVARD BUSINESS PUBLISHING

"on average, a 5% increase in customer retention rates results in 25% – 95% increase of profits."

KPMG

"customer retention is the main driver of a company's revenue"

DigitalSkola

"Source: https://www.superoffice.com/blog/reduce-customer-churn/

GOALS

DATE

S M T W T F S

* MODEL APA YANG PALING EFEKTIF UNTUK MEMPREDIKSI CUSTOMER CHURN BERDASARKAN BEHAVIOR-NYA?

* VARIABLE APA YANG PALING MEMPENGARUHI CUSTOMER CHURN?

DigitalSkola

2

IDENTIFY DATASET

DigitalSkola

DigitalSkola

Dataset Info

```
<class 'pandas.core.frame.DataFrame'>
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   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 
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

Customer account Information

- Customer ID
- Tenure
- Contract
- Paperless Billing
- Payment Method
- Monthly Charges
- Total Charges

Demographic Information

- Gender
- Senior Citizen
- Partner
- Dependent

Services

- Phone
- Multiple Lines
- Internet
- Online Security
- Online Backup
- Device Protection
- Tech Support
- Streaming TV
- Streaming Movies

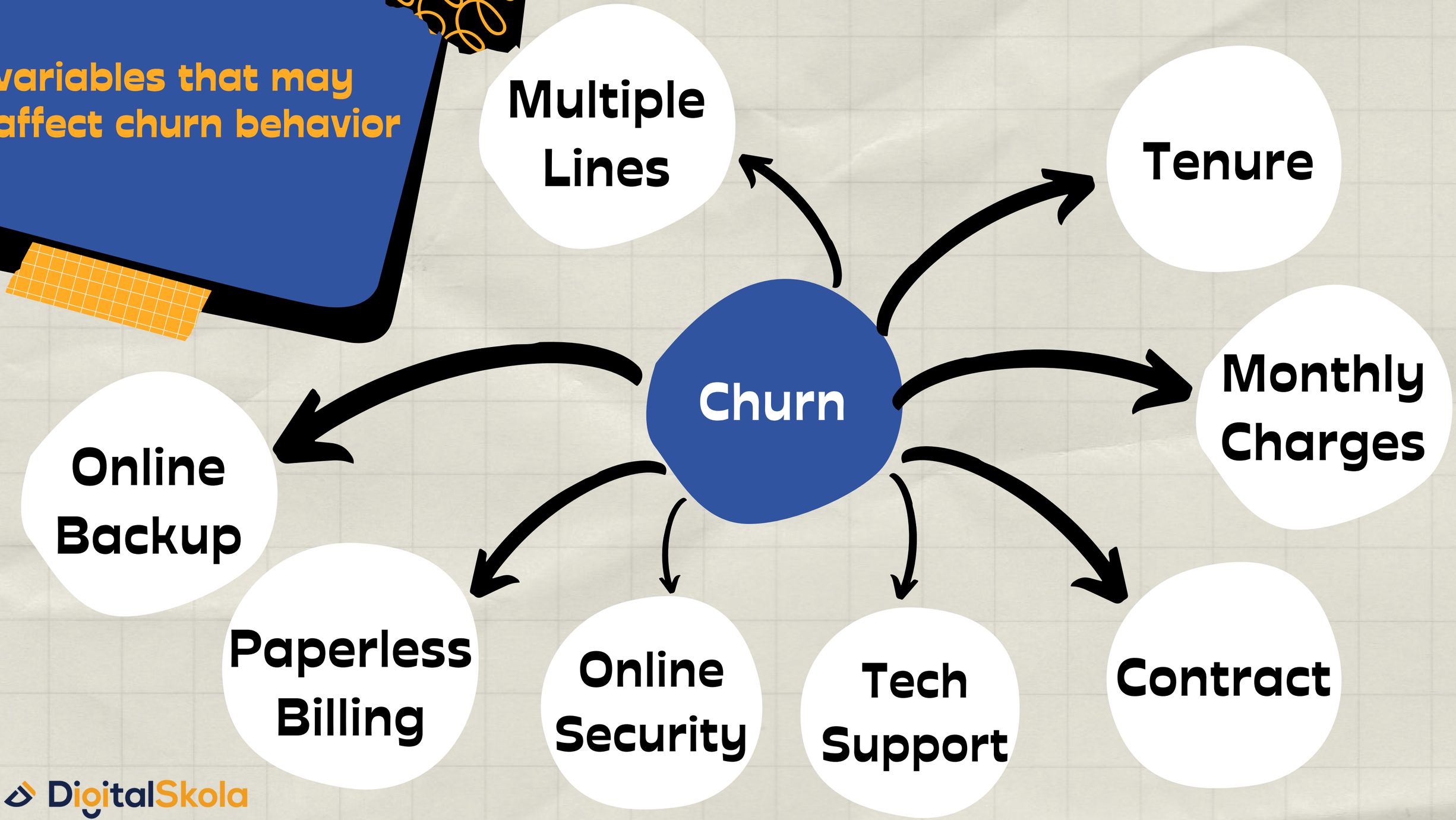
Churn (customer who left within last month)

3

EXPLORATORY DATA ANALYSIS

DigitalSkola

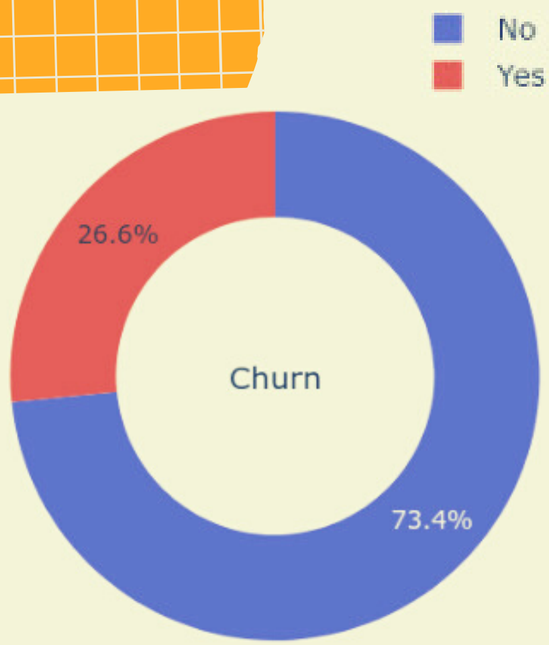
variables that may affect churn behavior



DigitalSkola

DigitalSkola

CHURN CUSTOMER



ALL VARIABLE CORRELATION

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
gender																				
SeniorCitizen	-0.0018																			
Partner	0.0014	-0.017																		
Dependents	0.01	-0.21	-0.45																	
tenure	-0.00026	0.012	-0.1	0.044																
PhoneService	-0.0075	0.0084	-0.018	-0.0011	-0.018															
MultipleLines	-0.01	0.11	-0.12	-0.019	0.065	0.67														
InternetService	-0.0022	-0.032	-0.00051	0.044	-0.013	0.39	0.19													
OnlineSecurity	-0.0044	-0.21	-0.081	0.19	0.014	0.13	-0.067	0.61												
OnlineBackup	0.011	-0.14	0.092	0.062	-0.066	0.13	-0.13	0.65	0.62											
DeviceProtection	0.0045	-0.16	-0.093	0.15	0.035	0.14	-0.013	0.66	0.75	0.6										
TechSupport	5.7e-05	-0.22	-0.068	0.18	0.03	0.12	-0.067	0.61	0.79	0.62	0.77									
StreamingTV	0.00058	-0.13	-0.079	0.14	0.025	0.17	0.031	0.71	0.7	0.6	0.76	0.74								
StreamingMovies	-0.0013	-0.12	-0.075	0.13	0.03	0.16	0.028	0.71	0.7	0.61	0.77	0.74	0.81							
Contract	9.5e-05	-0.14	-0.29	0.24	0.12	0.003	0.084	0.1	0.39	0.035	0.39	0.42	0.33	0.33						
PaperlessBilling	0.012	-0.16	-0.014	0.11	-0.013	-0.017	-0.13	0.14	0.33	0.26	0.28	0.31	0.2	0.21	0.18					
PaymentMethod	-0.0049	-0.094	-0.13	0.12	0.076	-0.0031	0.026	0.0089	0.21	0.0038	0.19	0.22	0.12	0.12	0.36	0.1				
MonthlyCharges	-0.008	0.049	-0.037	-0.029	0.043	-0.14	0.024	-0.29	-0.22	-0.28	-0.22	-0.21	-0.23	-0.24	-0.0068	-0.087	-0.0099			
TotalCharges	-0.012	0.023	-0.044	0.0097	0.11	-0.031	0.015	-0.038	-0.025	-0.054	-0.023	-0.02	-0.017	-0.025	0.055	-0.0098	0.0086	0.27		
Churn	-0.0085	0.15	0.15	-0.16	-0.14	0.012	0.036	-0.047	-0.33	-0.074	-0.28	-0.33	-0.21	-0.21	-0.4	-0.19	-0.26	0.02	-0.029	

1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00

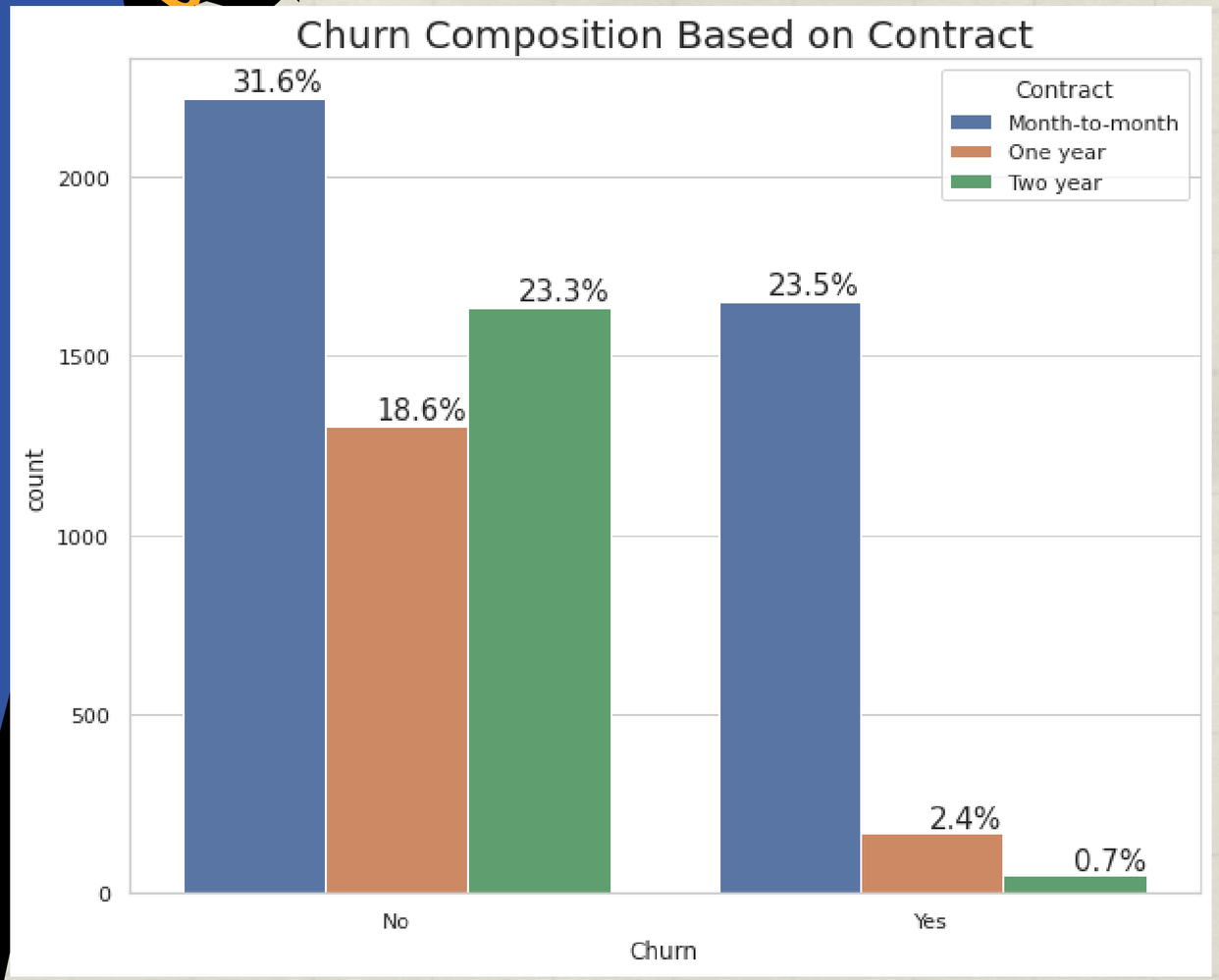
Tech Support (-0.33)

Online Security (-0.33)

Contract (-0.4)

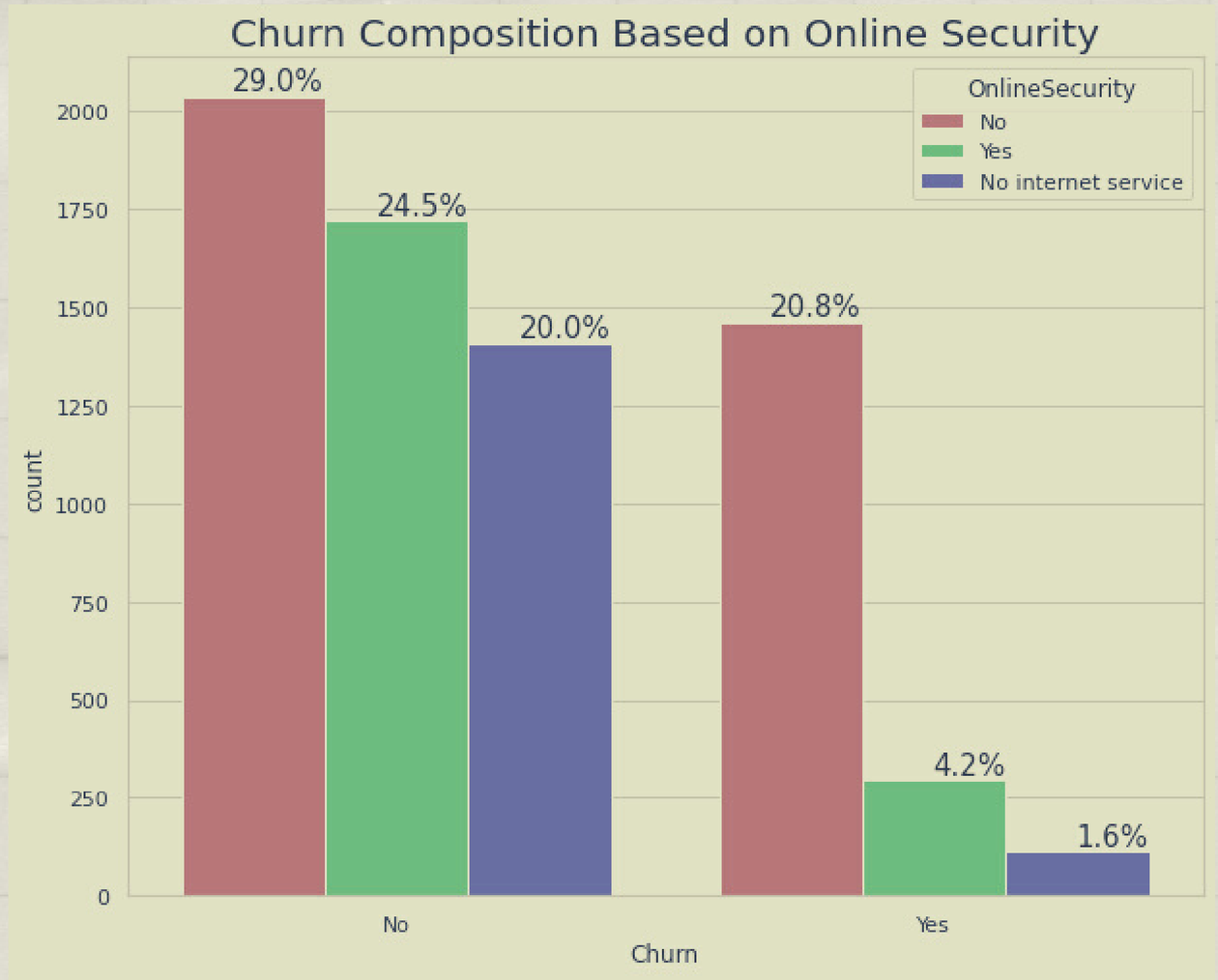
CHURN CUSTOMER BY CONTRACT

Sekitar 23,5% pelanggan dengan Kontrak Bulanan memilih untuk churn dibandingkan dengan 2,4% pelanggan dengan Kontrak Satu Tahun dan 0,7% dengan Kontrak Dua Tahun



DigitalSkola

Churn by Online Security Service

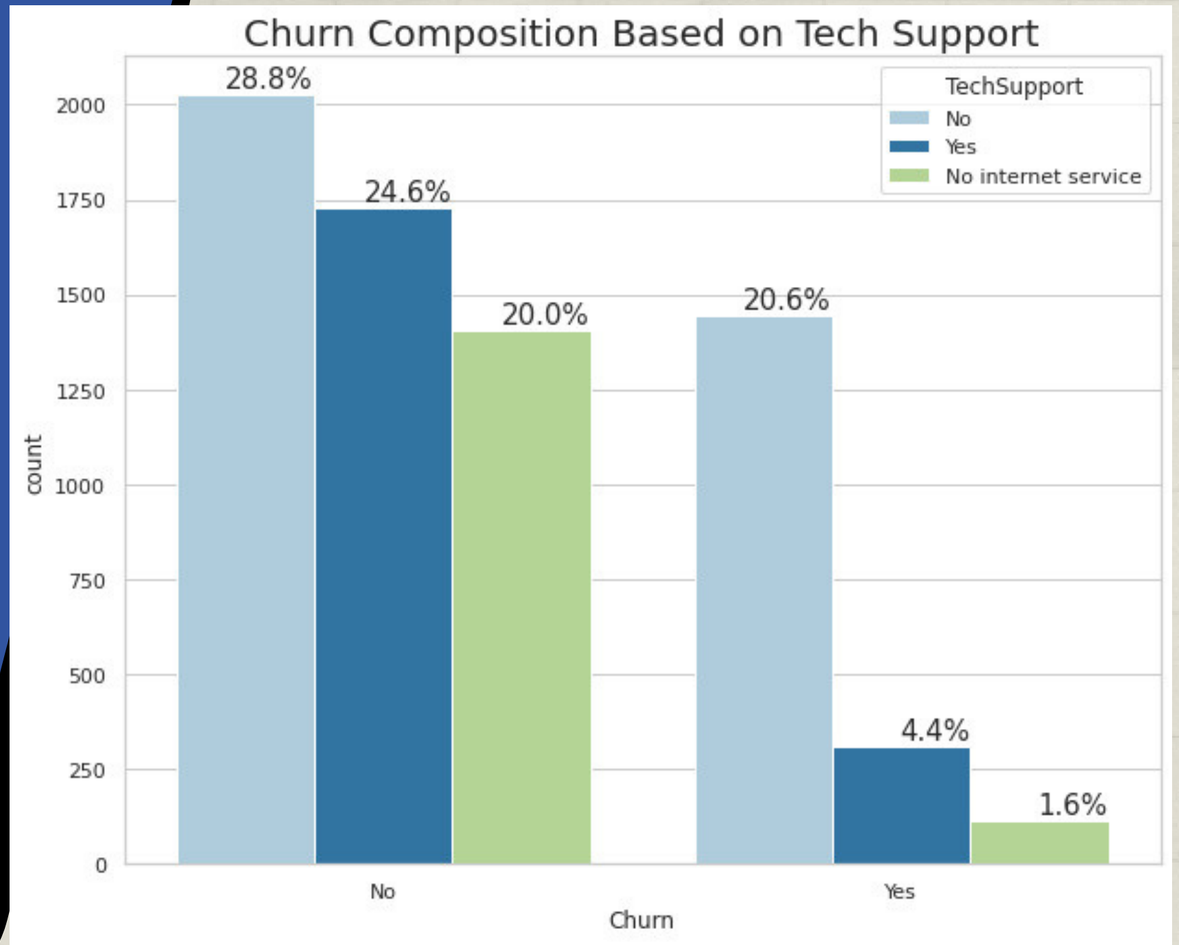


Terdapat sekitar 1400 pelanggan yang tidak menggunakan layanan *online security* telah berhenti berlangganan.

DigitalSkola

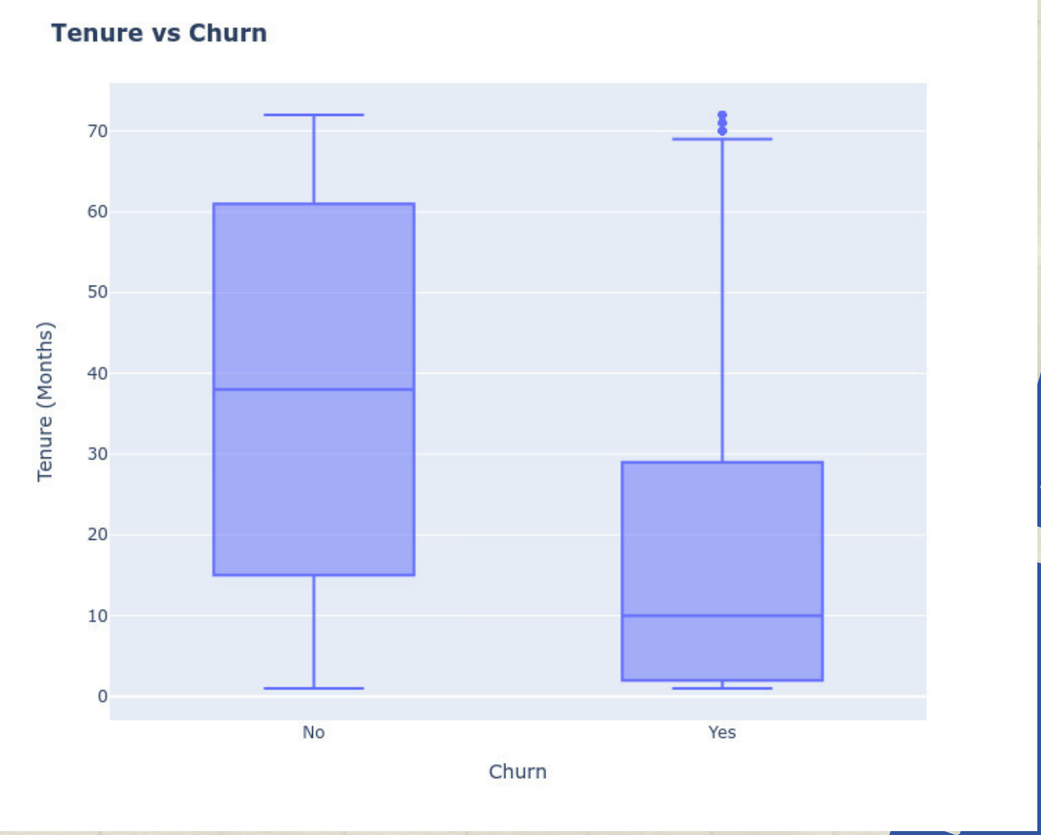
CHURN CUSTOMER BY TECH SUPPORT

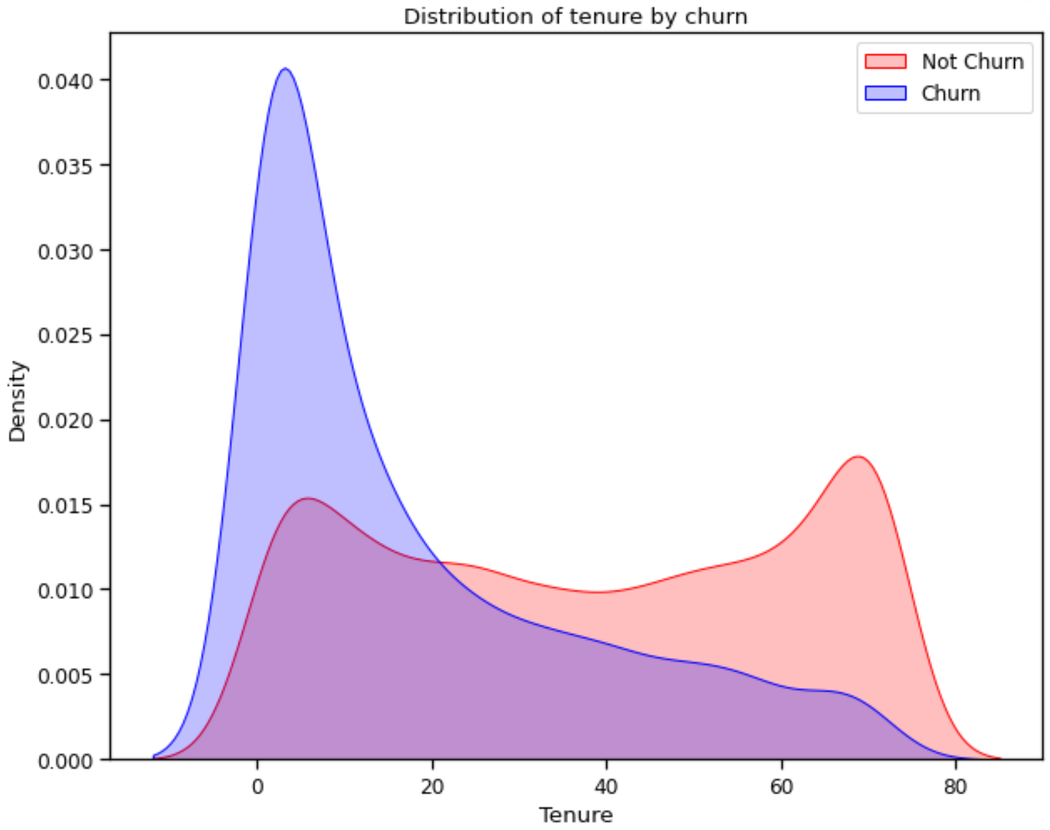
Sekitar 1400 customer yang tidak menggunakan layanan *technical support* telah berhenti berlangganan.



CHURN CUSTOMER BY TENURE

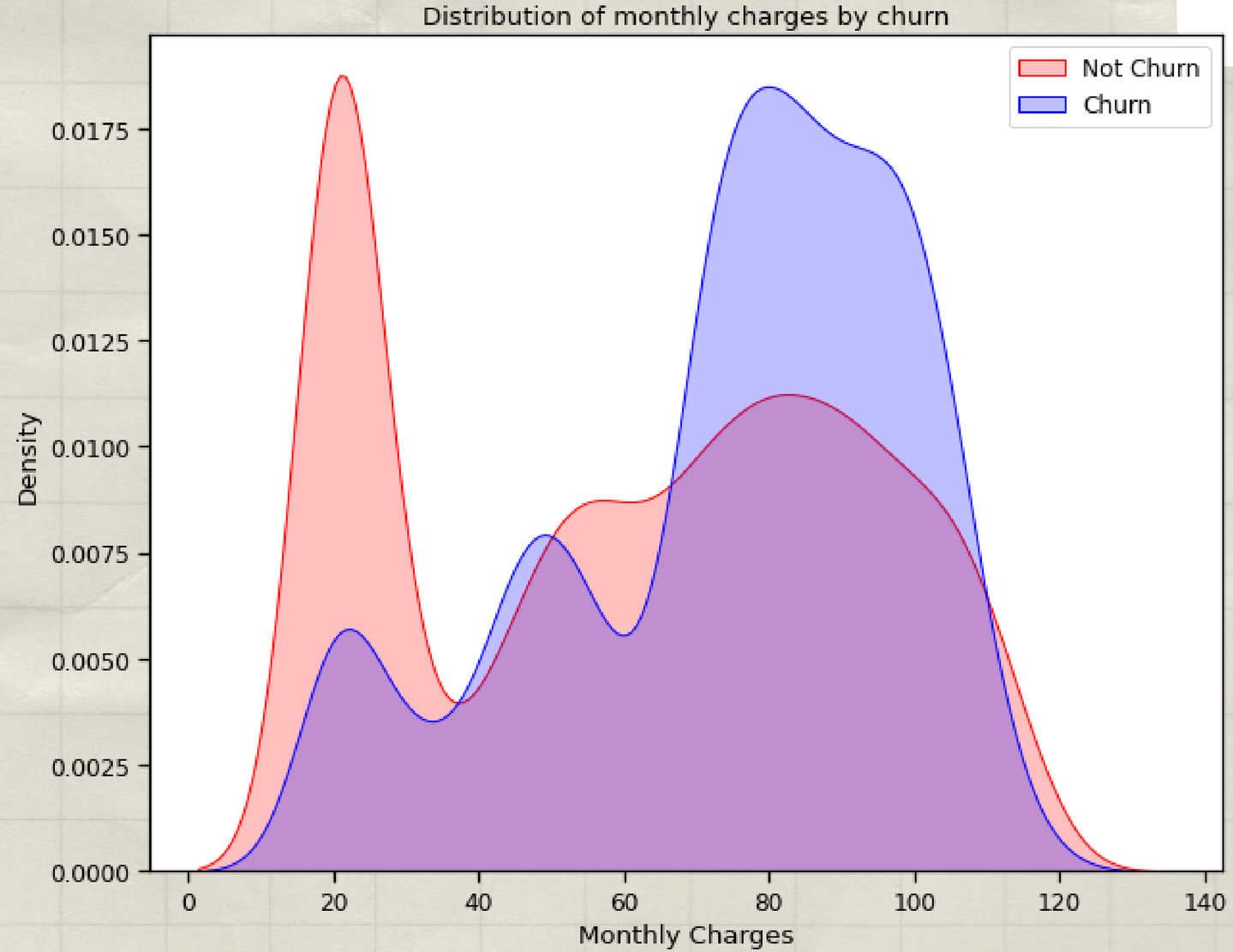
Pelanggan baru lebih cenderung churn.
karena masih dalam masa percobaan apakah akan tetap menggunakan layanan tersebut atau mencari layanan lain yang lebih ekonomis dengan penawaran yang menarik





DigitalSkola

Churn by Monthly Charges



Pelanggan dengan Biaya Bulanan yang lebih tinggi juga lebih cenderung untuk berhenti berlangganan

DigitalSkola

4

DATA PRE-PROCESSING

DigitalSkola

DigitalSkola

Tahapan Data Pre-processing

Data Cleansing

- Cek *Duplicated Data*
- Hapus tenure == 0
- Ganti *Null value* dengan *Mean*

Feature Encoding

Menggunakan *Label Encoder*

Train Test Split Data

70% *Train Data*
30% *Test Data*

Standardizing Numeric Attributes

Menggunakan *Standardization* pada kolom:

- *tenure*
- *MonthlyCharges*
- *TotalCharges*

5

MODEL DEVELOPMENT

DigitalSkola

Model Development

Machine Learning Model Evaluations and Predictions

Handling Imbalanced Data using **ADASYN**

Feature Selection

- Filter Method
- Wrapped Method
- Embedded Method

Hyperparameter Tuning using **Grid Search CV**

DigitalSkola

BENCHMARK ML MODEL SELECTION

Confusion Matrix

		Predicted	
		No Churn (0) Negative	Churn (1) Positive
Actual	No Churn (0) Negative	TN	FP
	Churn (1) Positive	FN	TP

Where:

- FN: The actual is churn, BUT predicted not churn
- FP: The actual is not churn, BUT predicted churn

$$Recall = \frac{(TruePositive)}{(TP+FN)}$$

→ Minimalize it!

&

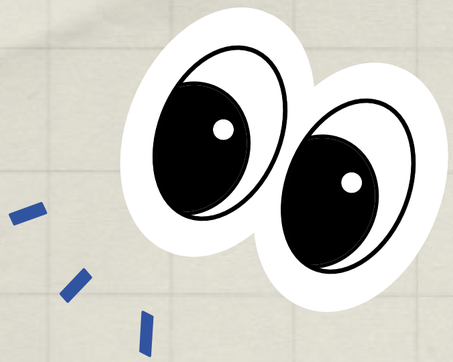
Execution Time

DigitalSkola

Machine Learning Model Evaluations and Predictions

	Accuracy	Precision	Recall	F1-Score	ROC
Logistic Regression	80.8%	75.6%	73.6%	74.5%	85.8%
XG Boost	81.0%	76.1%	72.7%	74.0%	85.9%
AdaBoost	80.6%	75.5%	72.2%	73.5%	85.9%
Gradient Boosting	80.3%	75.0%	72.1%	73.3%	85.7%
Random Forest	81.4%	77.4%	71.6%	73.6%	86.0%
SVC	80.8%	76.6%	70.4%	72.4%	81.4%
KNN	77.7%	71.3%	67.2%	68.5%	80.8%
Decision Tree	73.0%	65.4%	65.4%	65.4%	65.4%

HANDLING IMBALANCED DATA USING ADASYN



	Logistic Regression		XG Boost		AdaBoost	
	Before	After	Before	After	Before	After
Accuracy	80.8%	74.8%	81.0%	76.9%	80.6%	75.3%
Precision	75.6%	71.2%	76.1%	72.0%	75.5%	70.9%
Recall	73.6%	76.3%	72.7%	75.8%	72.2%	75.3%
F1-Score	74.5%	71.8%	74.0%	73.1%	73.5%	71.8%
ROC	85.8%	84.0%	85.9%	84.0%	85.9%	83.6%

Recall value improves about 3% each!

DigitalSkola

DigitalSkola

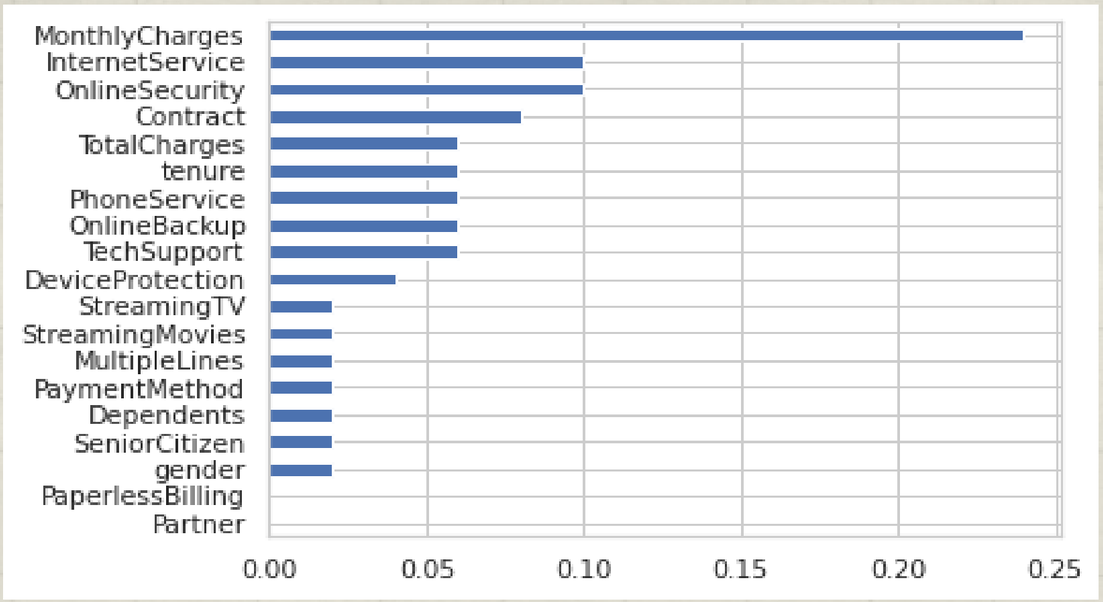
Feature Selection

FILTER METHOD ADASYN			
	Logistic Regression	XG Boost	AdaBoost
Accuracy	70.7%	72.6%	69.9%
Precision	69.3%	68.1%	68.7%
Recall	74.7%	72.0%	73.9%
F1-Score	68.5%	68.8%	67.7%
ROC	82.0%	81.4%	81.5%

WRAPPER METHOD ADASYN			
	Logistic Regression	XG Boost	AdaBoost
Accuracy	70.9%	76.8%	74.1%
Precision	68.2%	71.3%	69.9%
Recall	72.9%	74.2%	74.3%
F1-Score	68.1%	72.3%	70.6%
ROC	80.5%	82.2%	83.1%

EMBEDDED METHOD ADASYN			
	Logistic Regression	XG Boost	AdaBoost
Accuracy	71.3%	76.0%	74.7%
Precision	69.1%	70.6%	71.4%
Recall	74.3%	73.8%	76.7%
F1-Score	68.8%	71.6%	71.9%
ROC	81.0%	82.2%	84.9%

Variabel penting berdasarkan Embedded Method



Di pilih Embedded Method karena memiliki nilai Recall terbesar = 76,7%

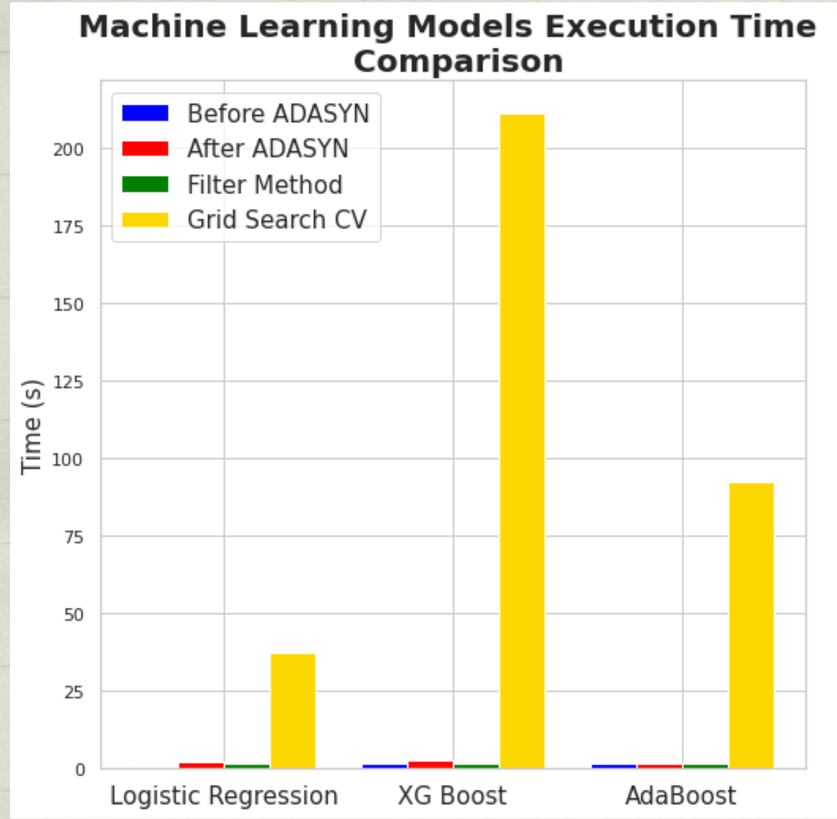
DigitalSkola

Hyperparameter Tuning using Grid Search CV

GRID SEARCH CV WITH EMBEDDED METHOD ADASYN			
	Logistic Regression	XG Boost	AdaBoost
Accuracy	73.7%	75.6%	74.7%
Precision	71.1%	69.9%	71.4%
Recall	76.6%	72.4%	76.7%
F1-Score	71.1%	70.7%	71.9%
ROC	84.8%	81.1%	84.9%

BEFORE ADASYN, EMBEDDED, & GRID SEARCH CV

	Logistic Regression	XG Boost	AdaBoost
Accuracy	80.8%	81.0%	80.6%
Precision	75.6%	76.1%	75.5%
Recall	73.6%	72.7%	72.2%
F1-Score	74.5%	74.0%	73.5%
ROC	85.8%	85.9%	85.9%



Logistic Regression Time 2X FASTER than AdaBoost Time

6

CONCLUSION

DigitalSkola

CONCLUSION

Model yang dipakai adalah Logistic Regression yang memiliki nilai:

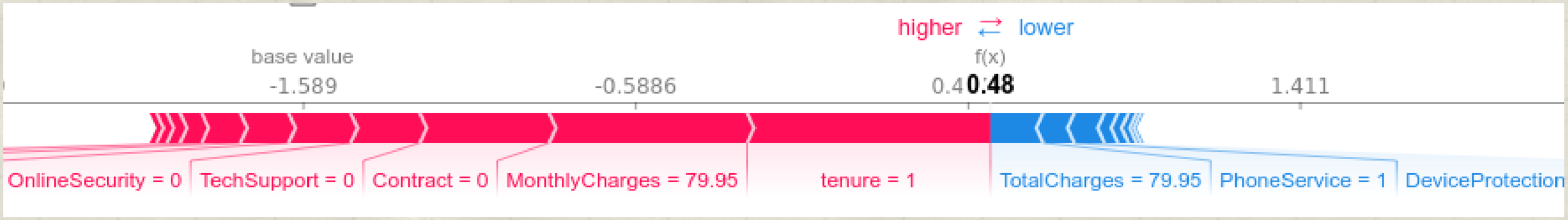
- Accuracy = 73,7%
- Precision = 71,1%
- Recall = 76,6%
- F1-Score = 71,1%
- ROC Score = 84,8%
- Execution Time ± 37.6 s

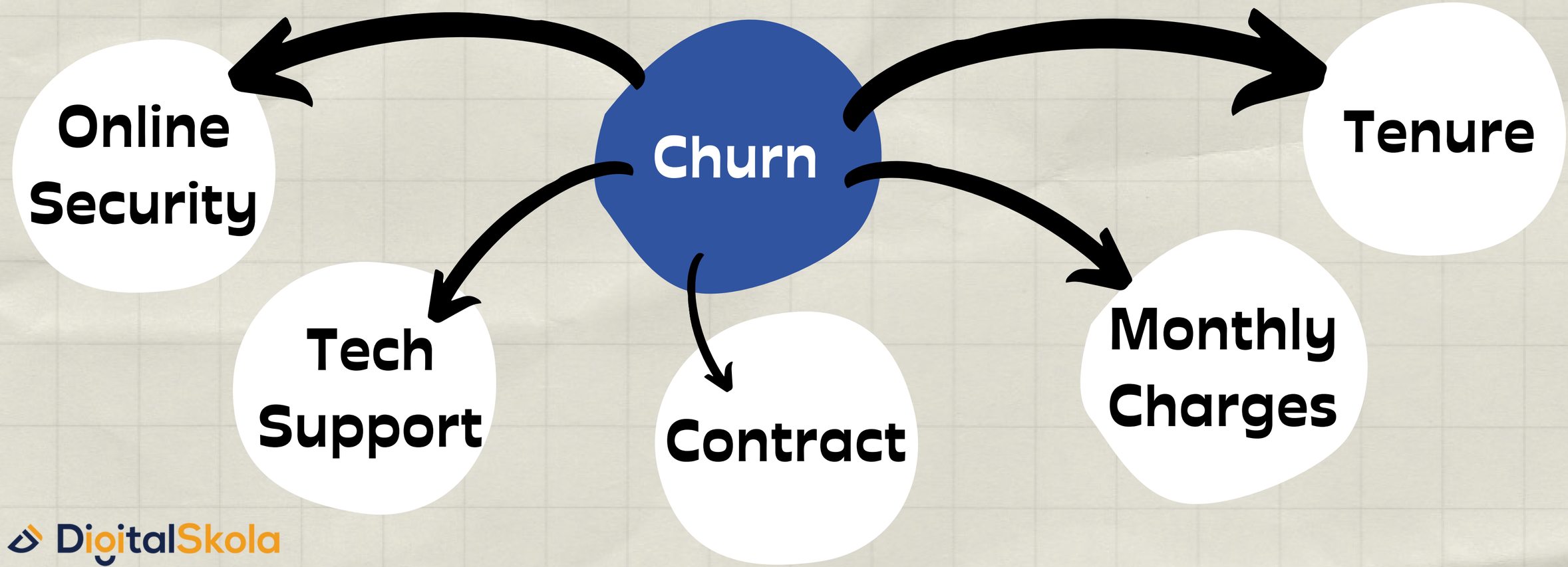
Kita memilih Logistic Regression karena:

- Memiliki nilai Recall yang terbesar setelah AdaBoost (hanya berbeda 0,1%)
- Execution Time 2 kali lebih cepat dibandingkan model AdaBoost (93.9 s)

DigitalSkola

VARIABEL YANG MEMPENGARUHI CHURN





DigitalSkola

THANK YOU!

DigitalSkola