



CUSTOMER CHURN PREDICTION

DATA MINING AND KNOWLEDGE MANAGEMENT [CS5I3]



TEAM MEMBERS



AISHWARYA
SATHYAKUMAR



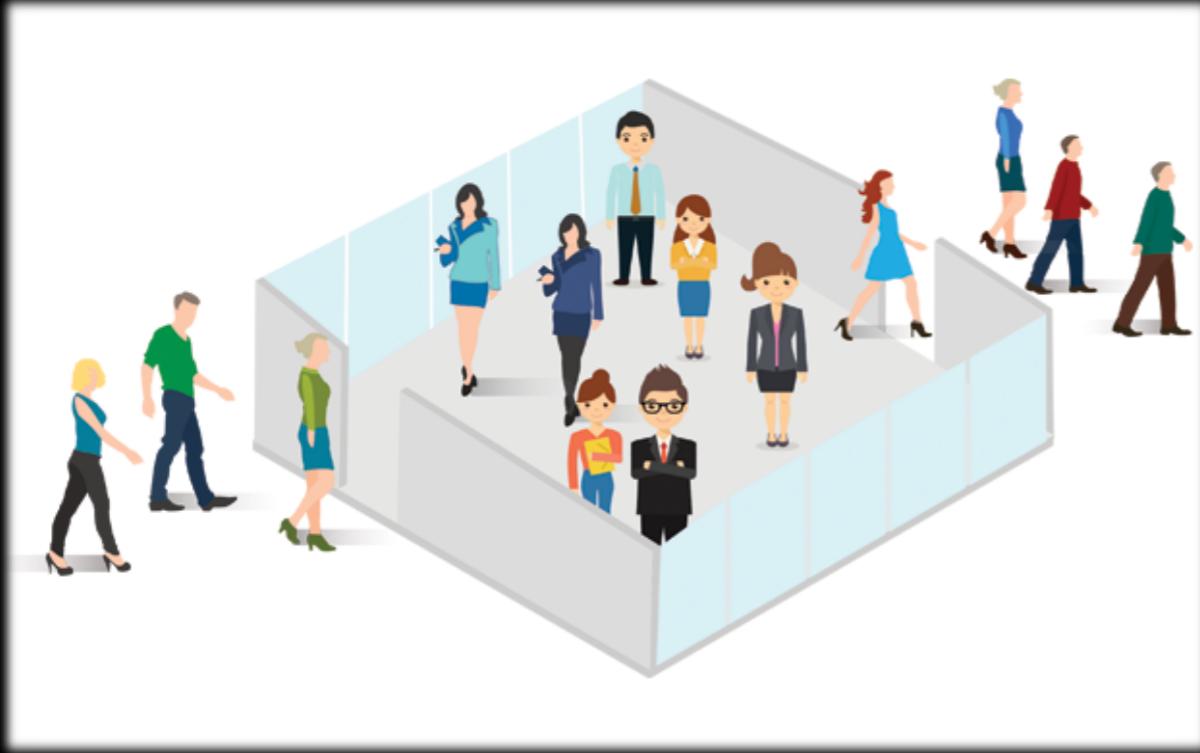
DINESH NADAR



YASH NAVDIWALA



WHAT IS CHURN ?



- A. Existing customer stops using a company's services and/or stops buying their products.
- B. Customer cuts his/her ties with the company.
- C. Health Indicator

PROBLEM STATEMENT



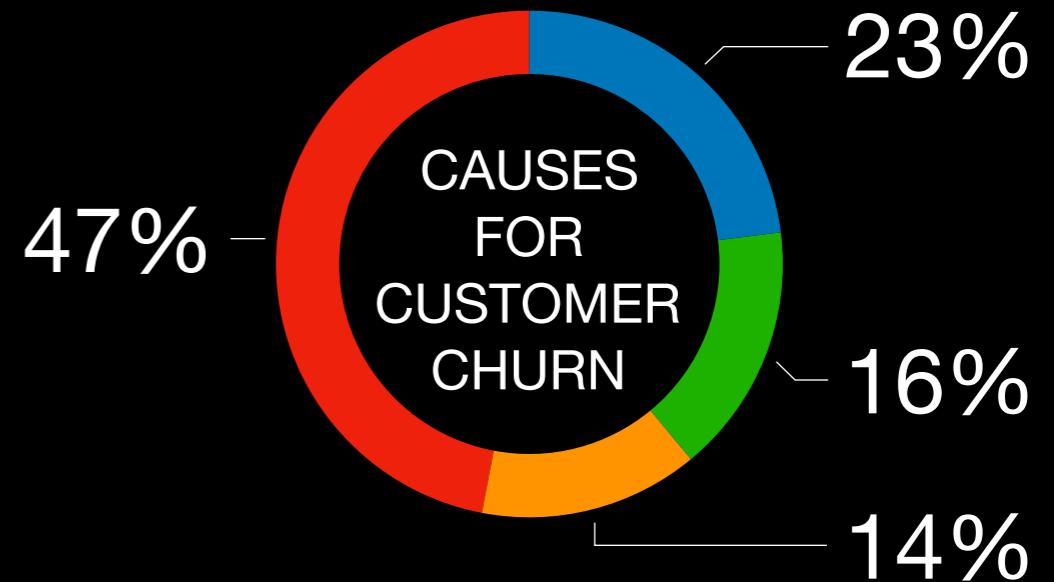
PROBLEM : Churn(attrition) is a huge problem for companies as it contributes to a reduction in the revenue

Additional pressure on teams to make up for the lost revenue

A new relationship with the company's competitors

CAUSES:

- ★ Poor Onboarding
- ★ Weak Relationship Building
- ★ Poor Customer Service



- Poor Onboarding
- Weak Relationship Building
- Poor Customer Service
- Other Causes for customer churn

SOLUTIONS :

Acquiring new customers (Hard and Expensive - Affects the revenue)

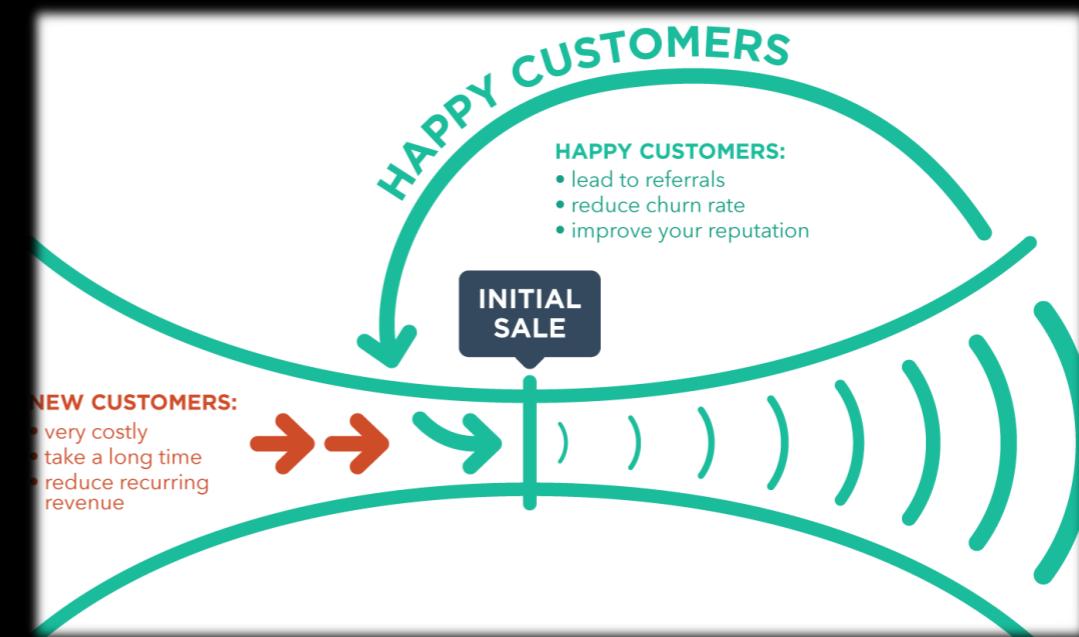
Focus on addressing the specific cause (Time Consuming ?)

Knowledge management and Data Mining ?



SOLUTION

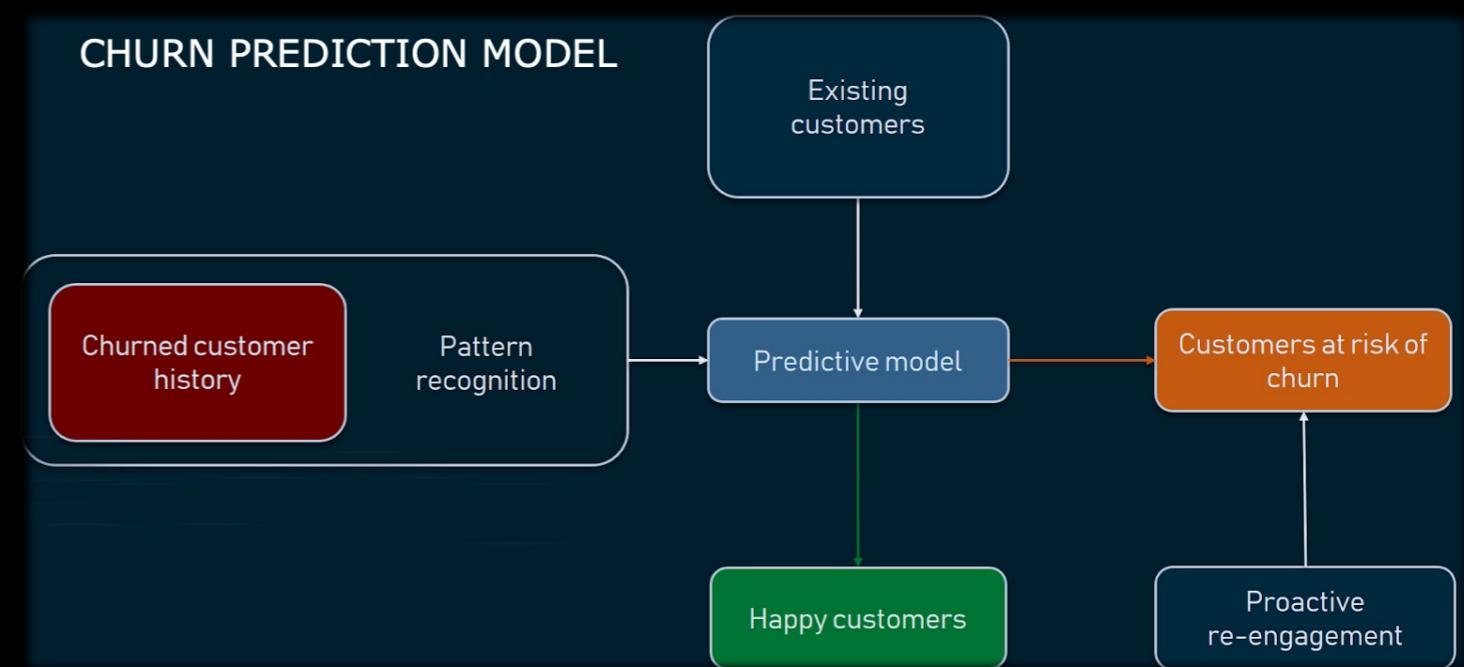
Companies that **constantly monitor** how people engage have greater opportunities to maintain mutually beneficial client relationships

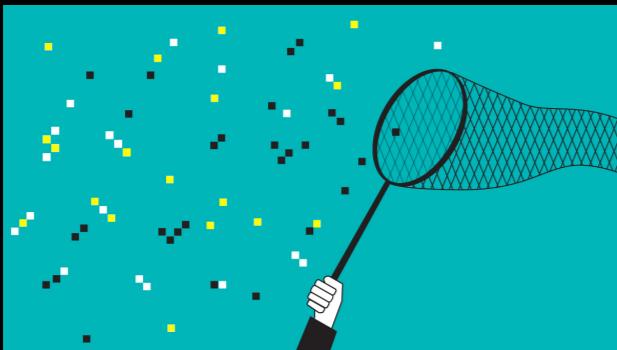


Gather customer data

To identify :

- Behavior patterns of potential churners
- Segment these at-risk customers
- Take appropriate actions





DATA SELECTION

DATA : Telco Customer Churn Dataset
(7043 rows and 21 columns)



kaggle

DATA SOURCE : Kaggle

SAMPLE DATA :

WA_Fn-UseC_-Telco-Customer-Churn.csv (954.59 KB)											20 of 21 columns	Views	CSV	JSON	HTML	PDF	SQL	Excel	Word
	A custom...	A gender	# Senior...	✓ Partner	✓ Depen...	# tenure	✓ PhoneS...	A Multipl...	A Interne...	A Online...									
1	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No									
2	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes									
3	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes									

A Device...	A TechSu...	A Stream...	A Stream...	A Contract	✓ Paperle...	A Payme...	# Monthl...	# TotalC...
No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85
Yes	No	No	No	One year	No	Mailed check	56.95	1889.5
No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15

CASE STUDY - Telecom Churn

- Customers choose from **multiple service providers** and **actively switch** from one operator to another
- An average of **15-25% annual churn rate**
- Costs **5-10 times** more to **acquire a new customer** than to retain an existing one
- **Customer retention** has now become even more important than customer acquisition
- **Retaining high profitable customers** is the number one business goal
- To reduce customer churn, telecom companies need to **predict customers** are at **high risk of churn**

DATA PREPROCESSING

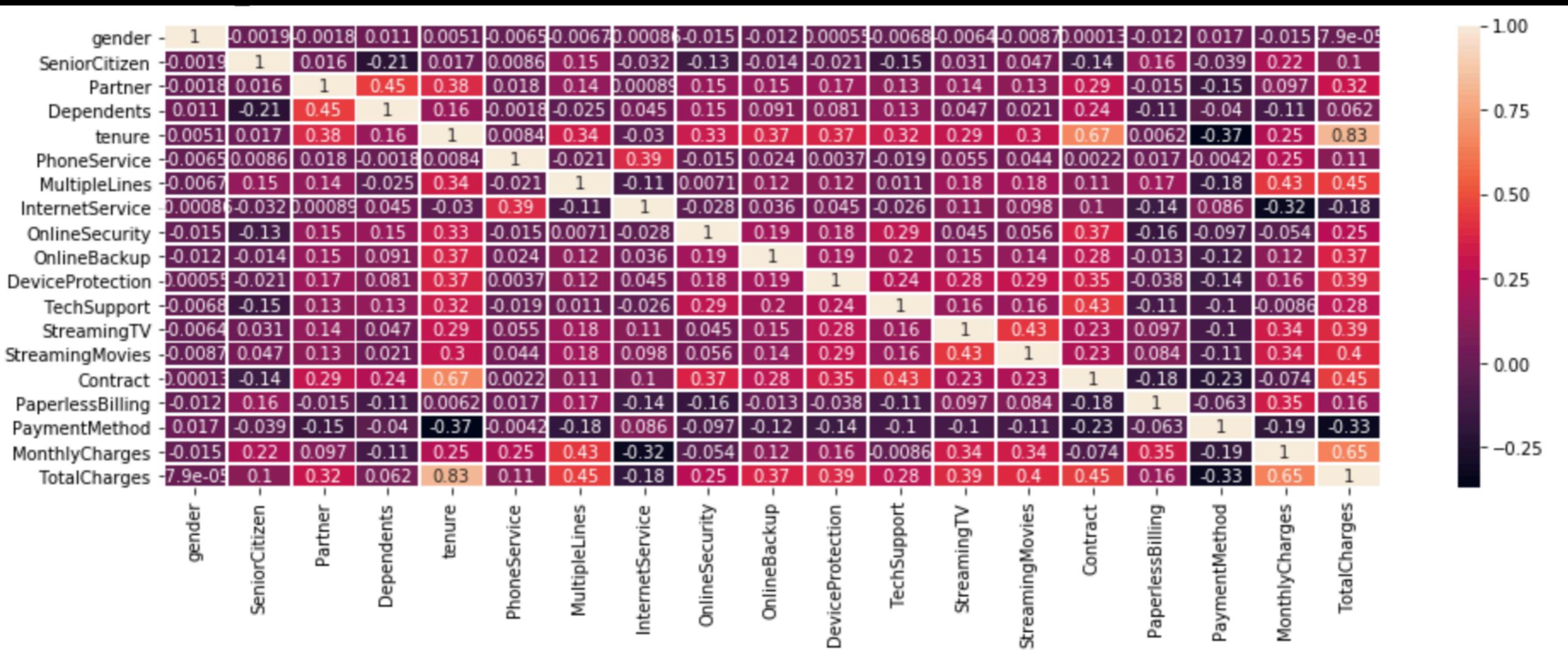
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
customerID      7043 non-null object
gender          7043 non-null object
SeniorCitizen   7043 non-null int64
Partner          7043 non-null object
Dependents       7043 non-null object
tenure           7043 non-null int64
PhoneService     7043 non-null object
MultipleLines    7043 non-null object
InternetService 7043 non-null object
OnlineSecurity   7043 non-null object
OnlineBackup      7043 non-null object
DeviceProtection 7043 non-null object
TechSupport       7043 non-null object
StreamingTV      7043 non-null object
StreamingMovies   7043 non-null object
Contract          7043 non-null object
PaperlessBilling  7043 non-null object
PaymentMethod     7043 non-null object
MonthlyCharges    7043 non-null float64
TotalCharges      7043 non-null object
Churn             7043 non-null object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

- ONE HOT ENCODING CATEGORICAL DATA
- NORMALIZATION

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	I
0	7590-VHVEG	0	0	1	0	0.013889	0	1	
1	5575-GNVDE	1	0	0	0	0.472222	1	0	
2	3668-QPYBK	1	0	0	0	0.027778	1	0	
3	7795-CFOCW	1	0	0	0	0.625000	0	1	
4	9237-HQITU	0	0	0	0	0.027778	1	0	
...
7038	6840-RESVB	1	0	1	1	0.333333	1	2	
7039	2234-XADUH	0	0	1	1	1.000000	1	2	

	treamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges
0	0	0	0	1	2	0.11	0.003339
0	0	1	0	0	3	0.38	0.217526
0	0	0	1	1	3	0.35	0.012437
0	0	1	0	0	0	0.24	0.211884
0	0	0	1	1	2	0.52	0.017388
...
2	2	1	1	1	3	0.66	0.229157

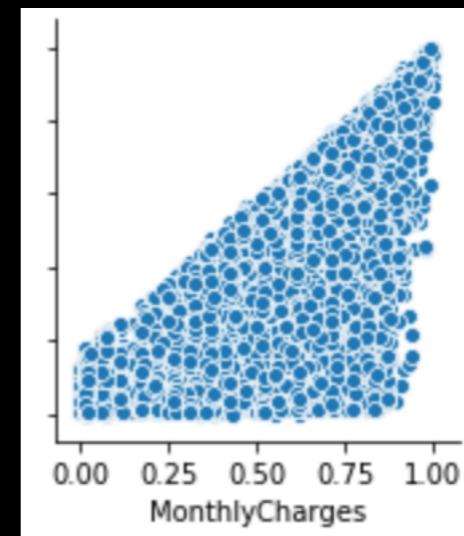
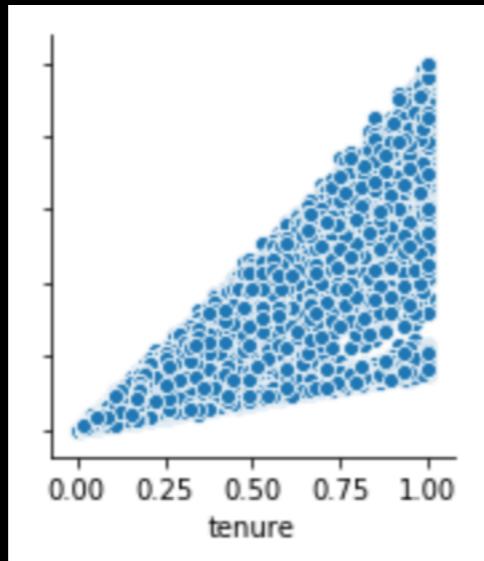
HEAT MAP



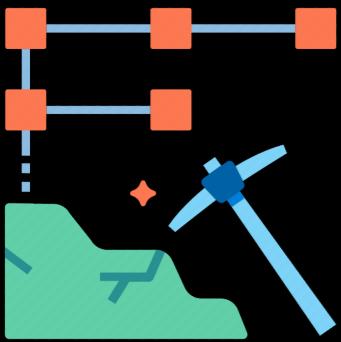
OBSERVATIONS

POSITIVE CORRELATION

TOTAL CHARGES AND
TENURE



TOTAL CHARGES AND
MONTHLY CHARGES



DATA MINING - ALGORITHMS

- CLASSIFICATION
 - ★ K-NEAREST NEIGHBOUR
 - ★ DECISION TREE
 - ★ NEURAL NETWORKS(ANN)
- CLASSIFIER ENSEMBLE
 - ★ RANDOM FOREST

EXECUTION AND ANALYSIS

ANN

CONFUSION MATRIX:

```
[[188 180]
 [118 923]]
```

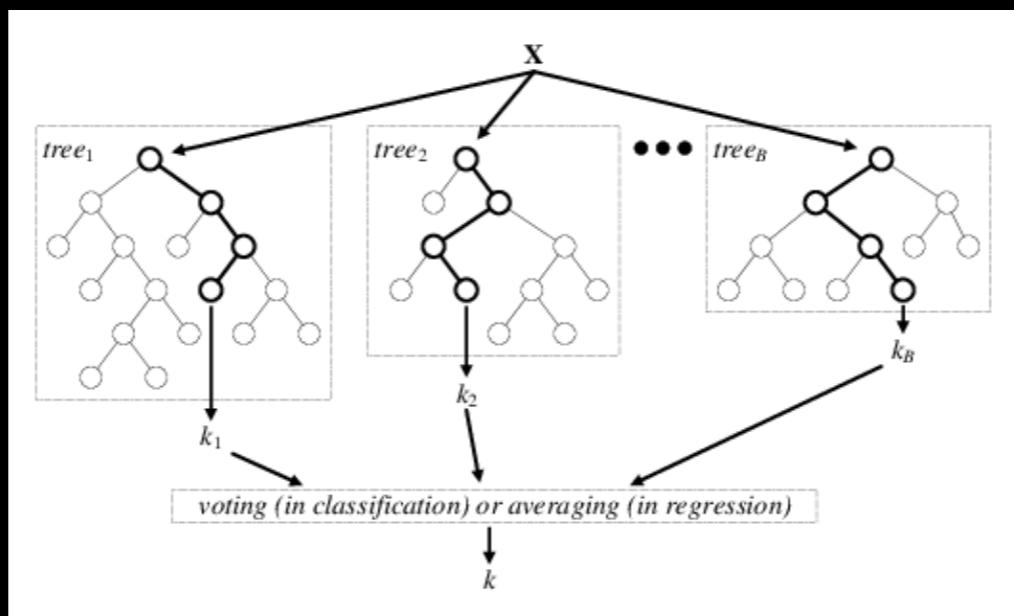
CLASSIFICATION REPORT:

	precision	recall	f1-score	support
Customers did not churn	0.61	0.51	0.56	368
Customers churned	0.84	0.89	0.86	1041
accuracy			0.79	1409
macro avg	0.73	0.70	0.71	1409
weighted avg	0.78	0.79	0.78	1409

EXECUTION AND ANALYSIS

RANDOM FOREST

Accuracy : 0.72 %



EXECUTION AND ANALYSIS

DECISION TREE

CONFUSION MATRIX:

```
[[869 169]
 [215 156]]
```

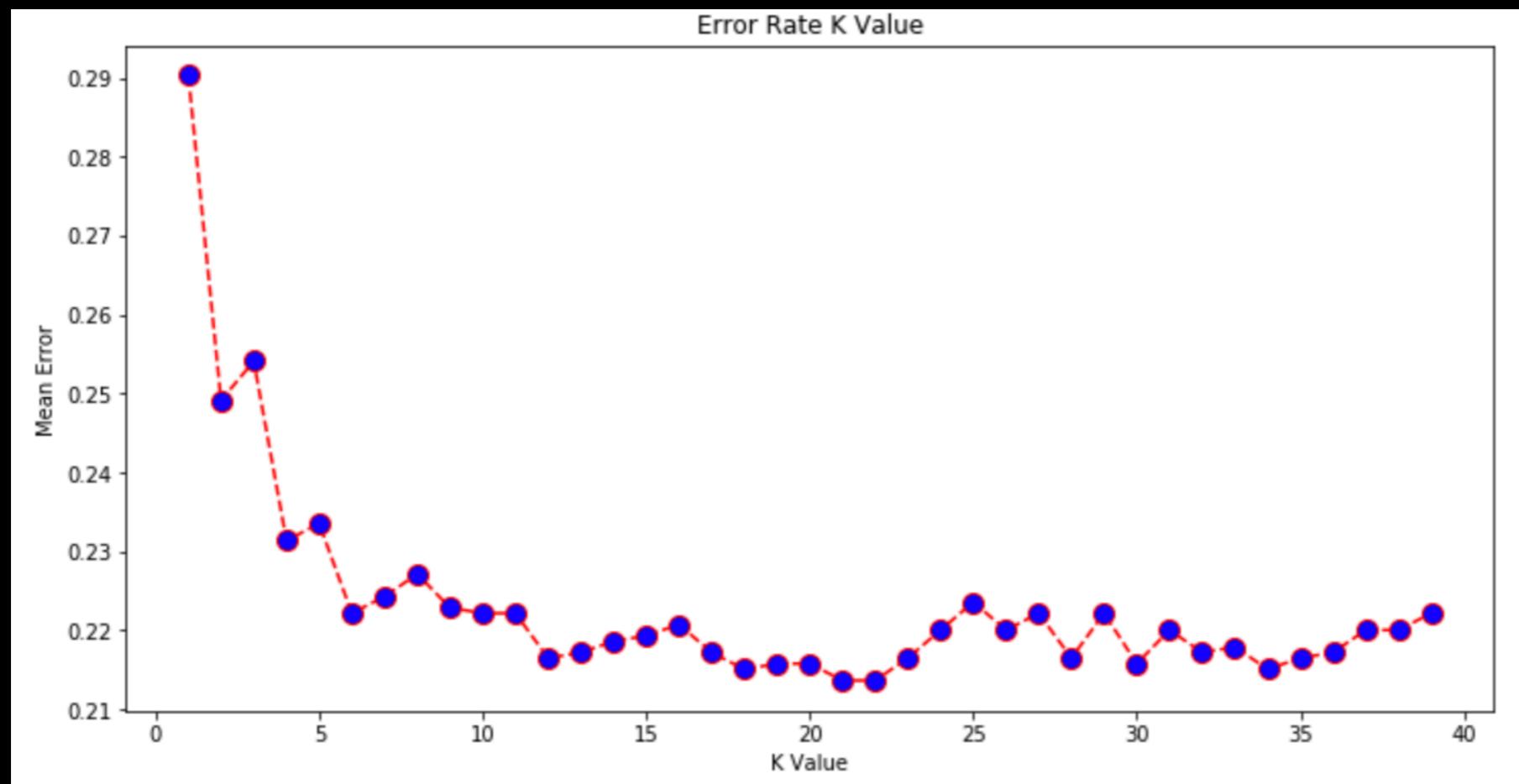
CLASSIFICATION REPORT:

	precision	recall	f1-score	support
Customers did not churn	0.80	0.84	0.82	1038
Customers churned	0.48	0.42	0.45	371
accuracy			0.73	1409
macro avg	0.64	0.63	0.63	1409
weighted avg	0.72	0.73	0.72	1409

EXECUTION AND ANALYSIS

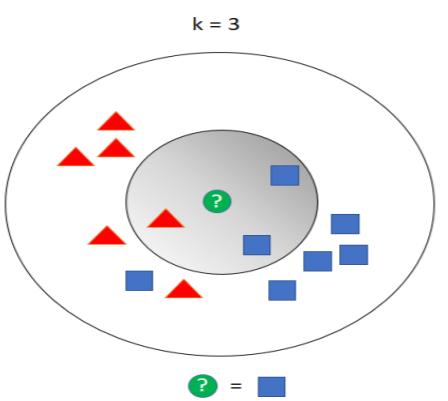
K-NEAREST NEIGHBOUR

Accuracy : 0.77%



COMPARISON OF TECHNIQUES

ALGORITHM

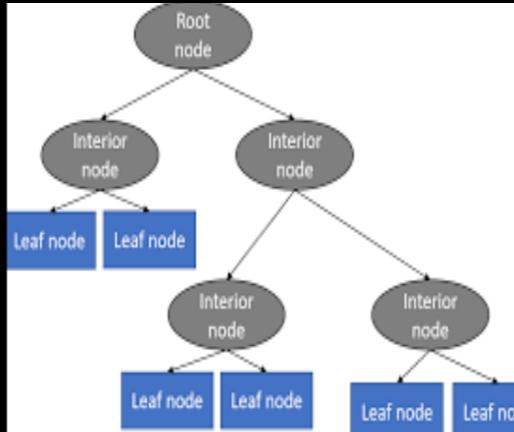


ACCURACY

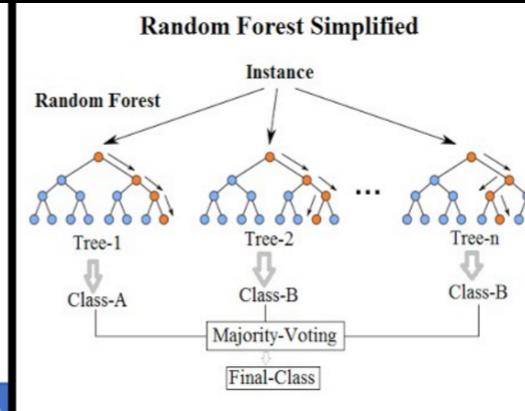
0.77%

METHOD

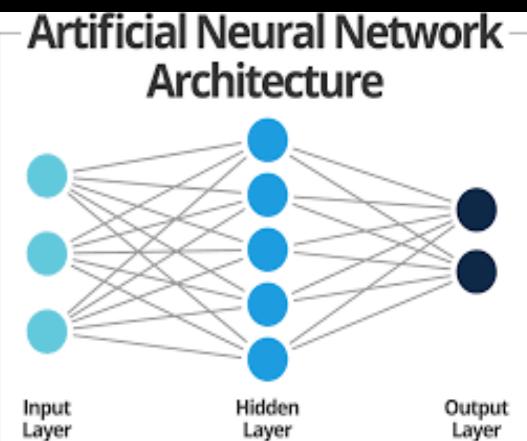
K-NEAREST



DECISION
TREE



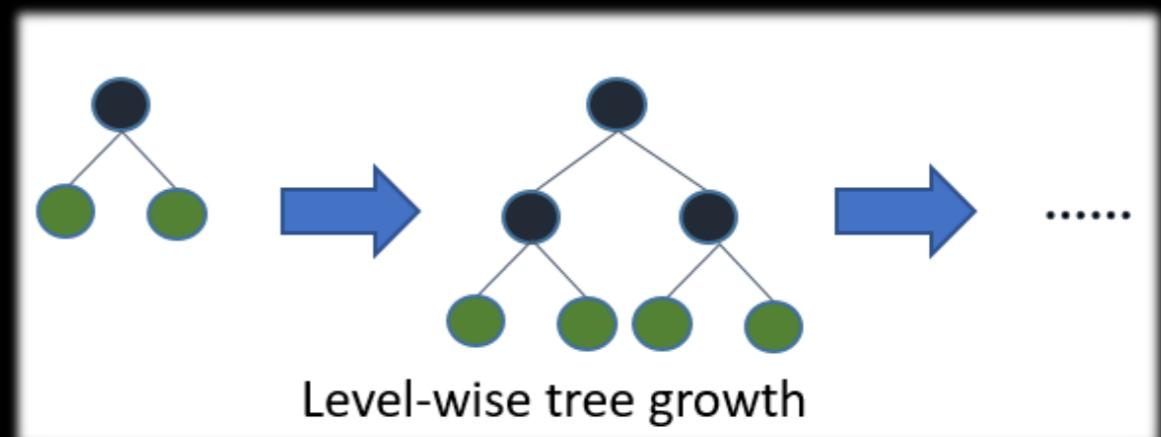
RANDOM
FOREST



ARTIFICIAL
NEURAL
NETWORKS

ADDITIONAL TECHNIQUES

XGBoost - implementation of the gradient boosted tree algorithms used for classification and regression problems consisting of a group of weaker models (trees), which sums up their estimates to predict a target variable with more accuracy.



CONCLUSION & QUESTIONS

Ability to classify customers who aren't happy with provided solutions allows businesses to learn about

- Product or pricing plan weak points operation issues
- Customer preferences
- Expectations to proactively reduce reasons for churn

Of all the models implemented ANN model delivers highest accuracy

