





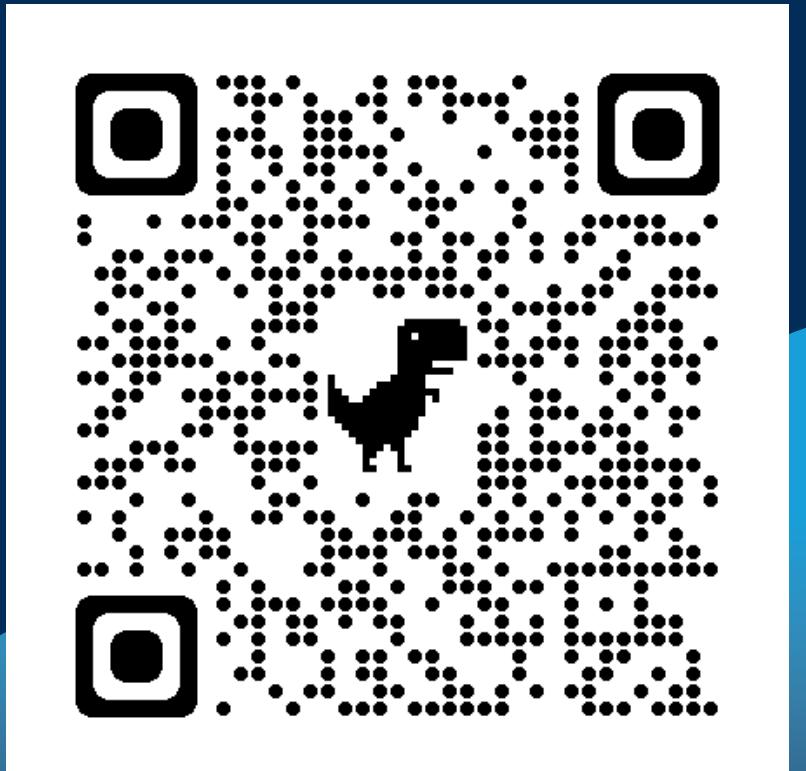
Titanic - ML from Desaster

Panramon Kanhasiri (Jelly)



Overview

- . The goal is to analyze historical data from passengers on the Titanic to predict who survived the disaster and who did not. This is done by using machine learning algorithms, combined with data of passengers such as age, gender, and more to identify patterns that might indicate a passenger's likelihood of survival.





About Titanic

The RMS Titanic, often considered "unsinkable," tragically sank on April 15, 1912, during its maiden voyage after hitting an iceberg. Out of 2,224 passengers and crew, 1,502 lost their lives due to insufficient lifeboats. While survival was partly due to luck, it appears that some groups were more likely to survive than others.

The Data

- PassengerId
- Survived*
- Pclass
- Name
- Sex
- Age
- SibSp
- Parch
- Ticket
- Fare
- Cabin
- Embarked

The Data

train.csv

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599 STON/O2. 3101282	71.2833 7.9250	C85 NaN	C
2	3	1	3 Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	113803	53.1000	C123	S
3	4	1	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
4	5	0
...
886	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

Pre-Process Data

①

Fill missing value

replace missing value in
the data using “mean” or
“median”

②

Encoded categorical data

Change Sex and Embarked
from categorial data to
numerical (Encoded)

③

Select features

Pclass, Sex, Age, SibSp,
Parch, Fare, and Embarked
are selected to used in
training

Pre-Process Data

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	3	1	22.0	1	0	7.2500	2
1	1	0	38.0	1	0	71.2833	0
2	3	0	26.0	0	0	7.9250	2
3	1	0	35.0	1	0	53.1000	2
4	3	1	35.0	0	0	8.0500	2
...
886	2	1	27.0	0	0	13.0000	2
887	1	0	19.0	0	0	30.0000	2
888	3	0	28.0	1	2	23.4500	2
889	1	1	26.0	0	0	30.0000	0
890	3	1	32.0	0	0	7.7500	1

MACHINE LEARNING ALGORITHM

Decision Tree

Random Forest

Support Vector Machine (SVM)

01

02

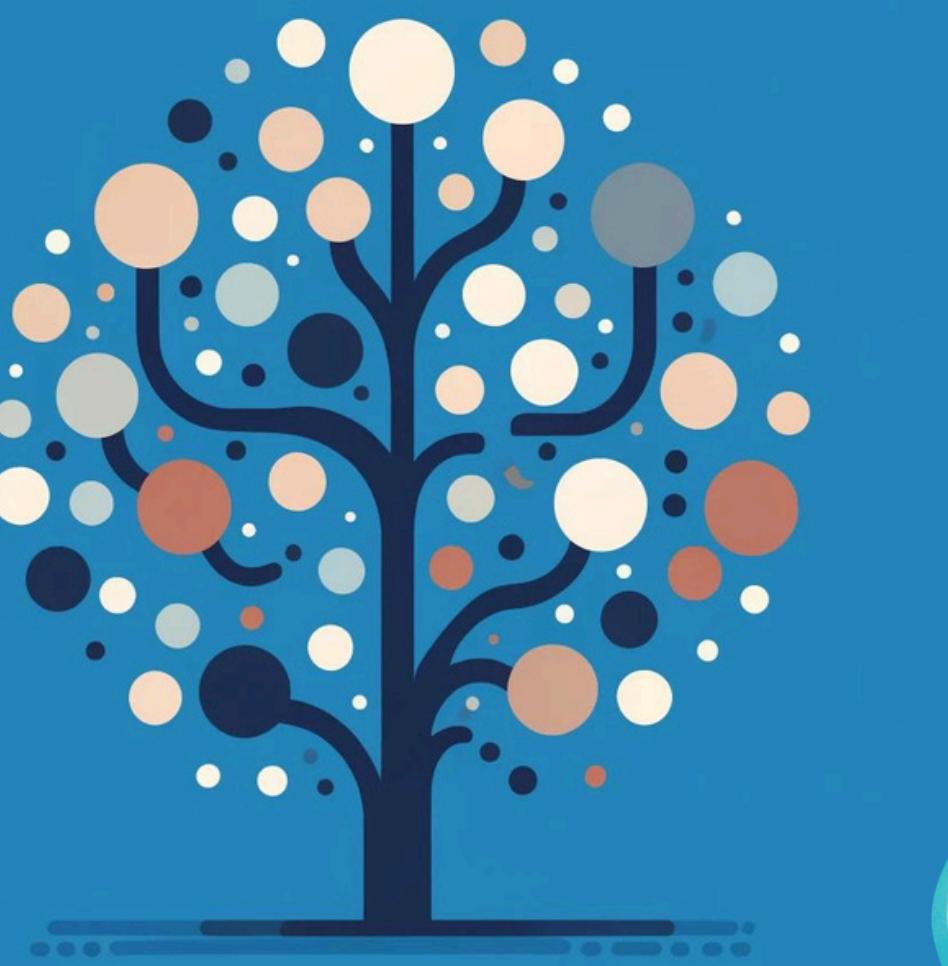
03

04

05

Logistic Regression

K-Nearest Neighbors



Decision Tree

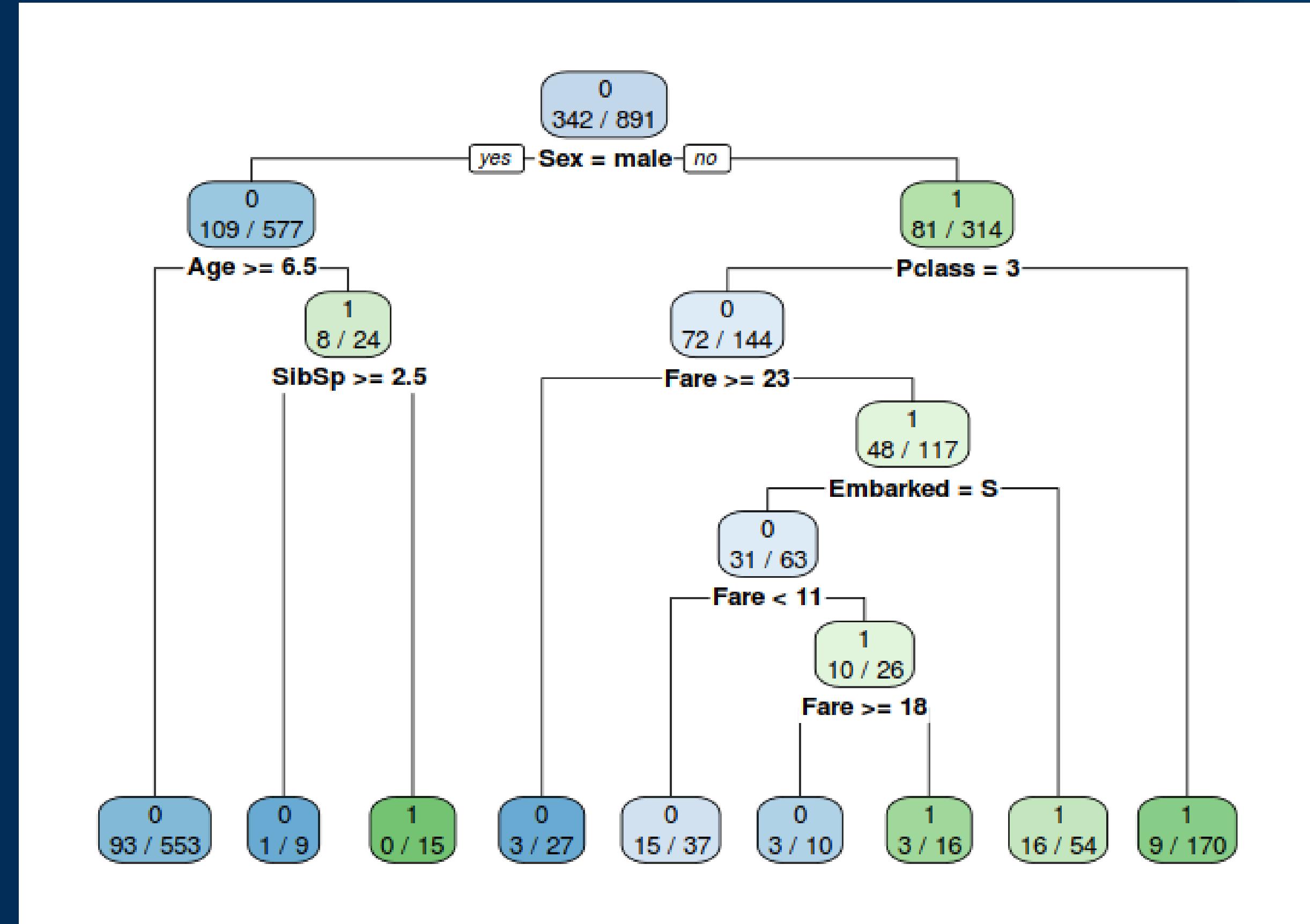
What is it ?

A decision tree splits the dataset into homogeneous sets based on key variables. It makes decisions by dividing data points along feature values, creating clear and interpretable rules.

Titanic ?

A decision tree is built by creating branches from strong predictors of survival, such as sex or passenger class. It continues branching until it effectively represents decision paths that explain survival outcomes.

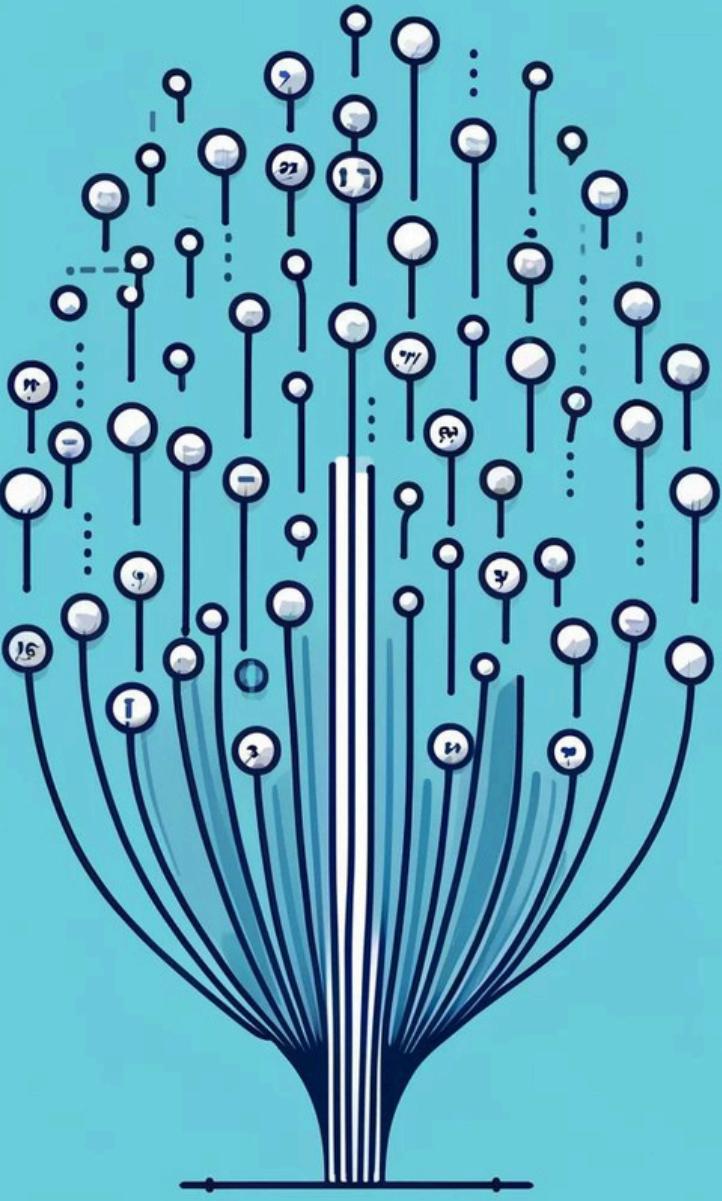
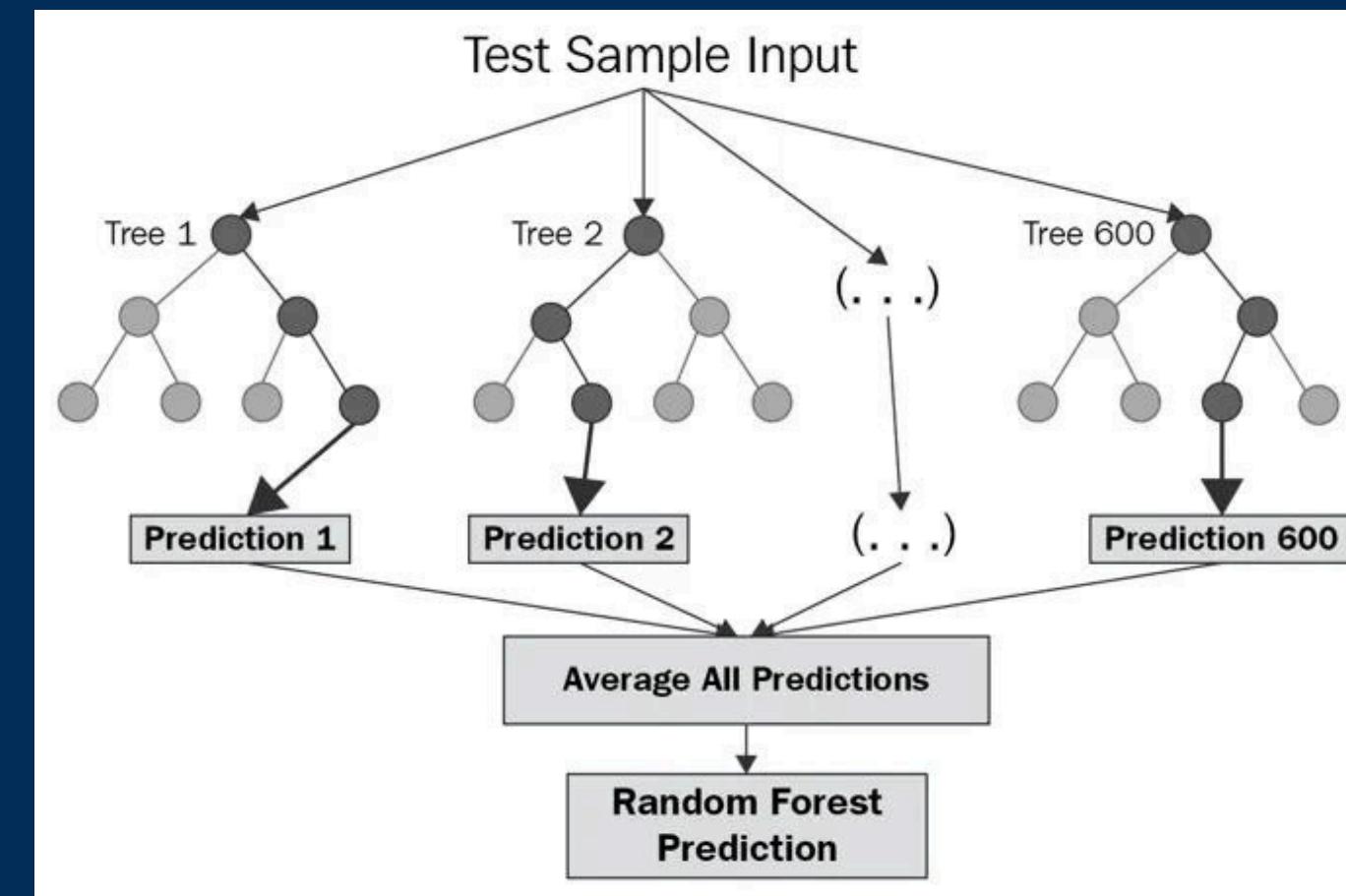




Random Forest

What is it ?

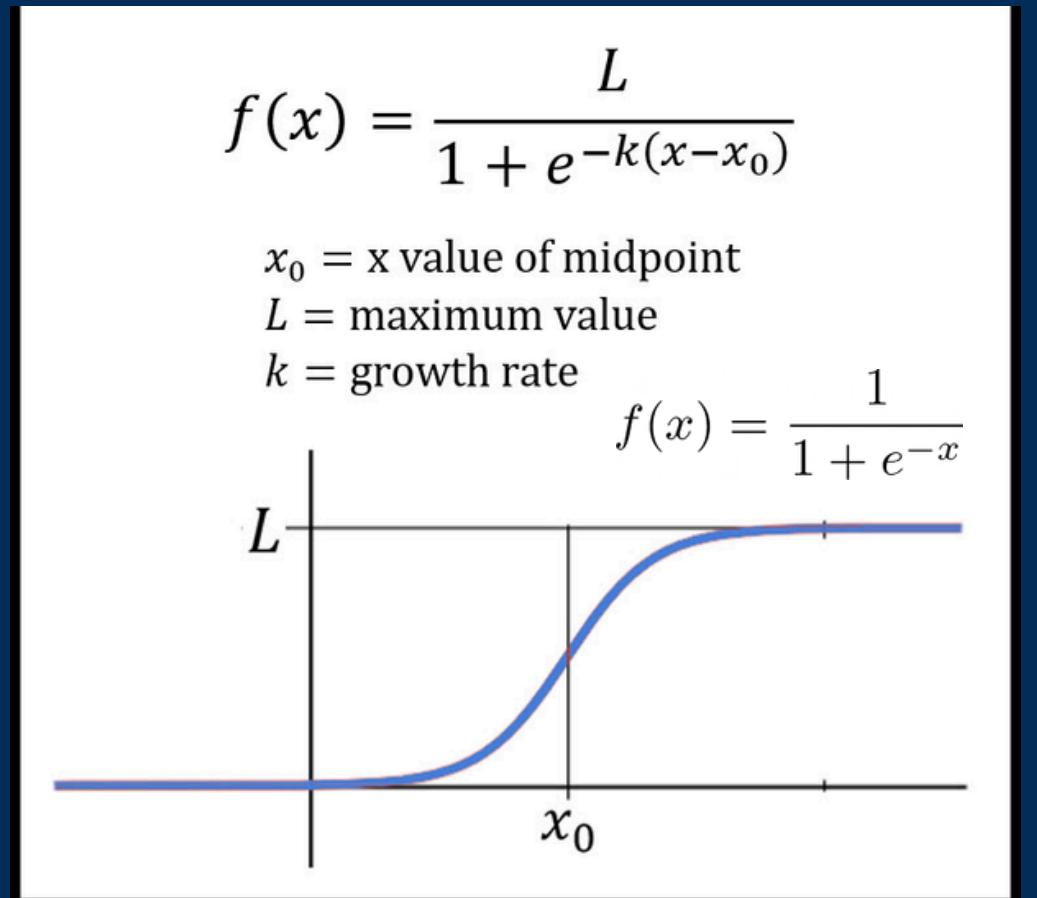
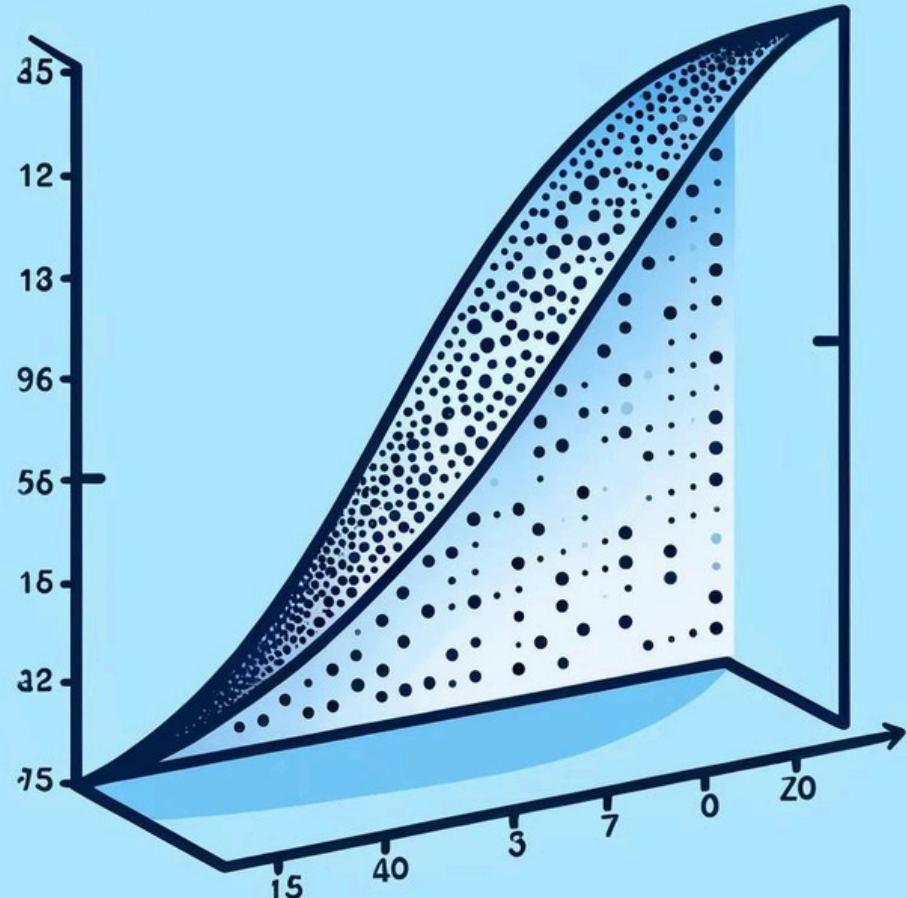
Random Forest constructs several decision trees and combines them for a more accurate and stable prediction. It randomly selects subsets of the training data, builds a tree for each subset, and averages the results to determine the final classification.



Logistic Regression

What is it ?

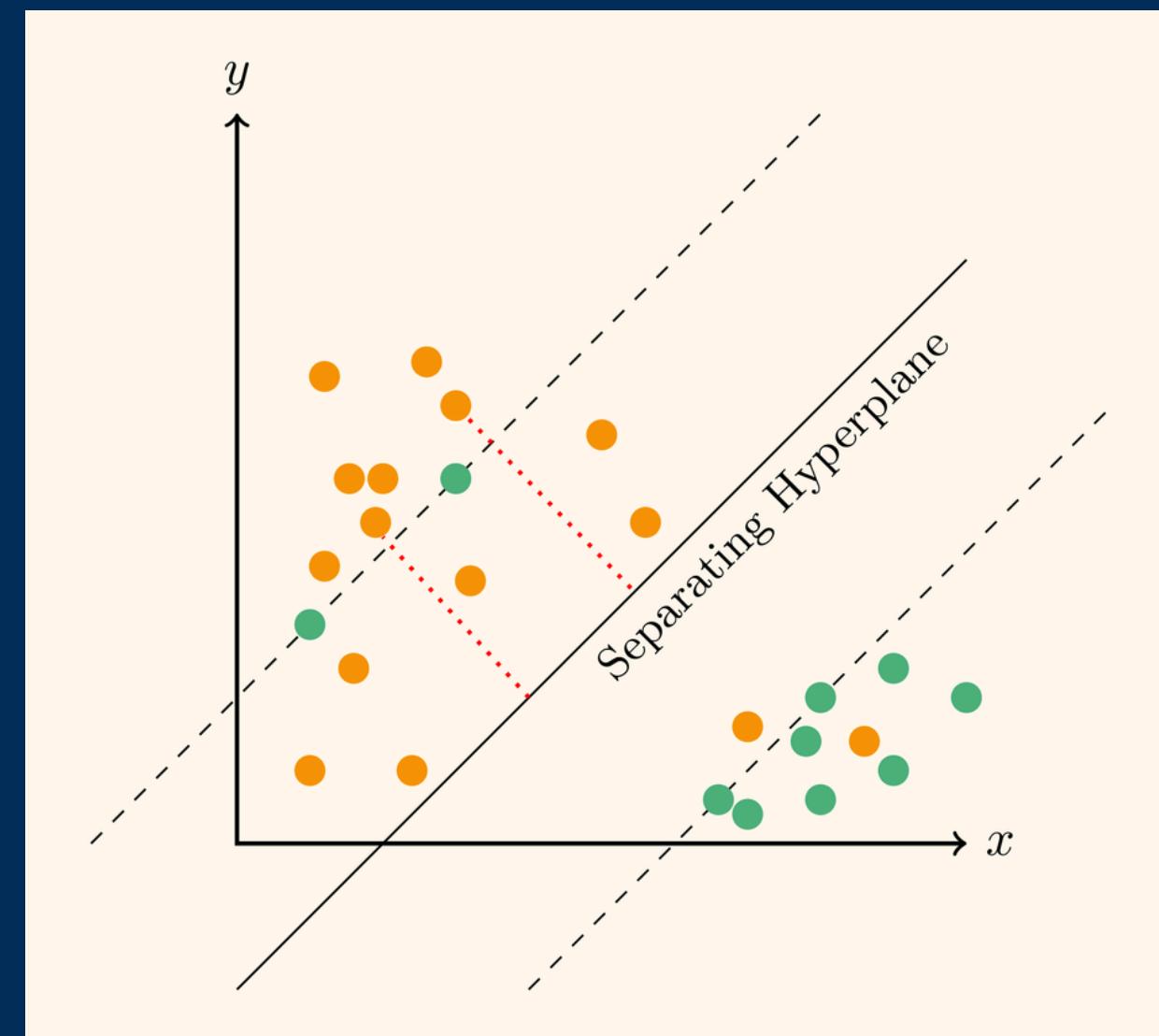
Estimates probabilities using a logistic function, which is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits. It's used primarily for binary classification.



Support Vector Machine (SVM)

What is it ?

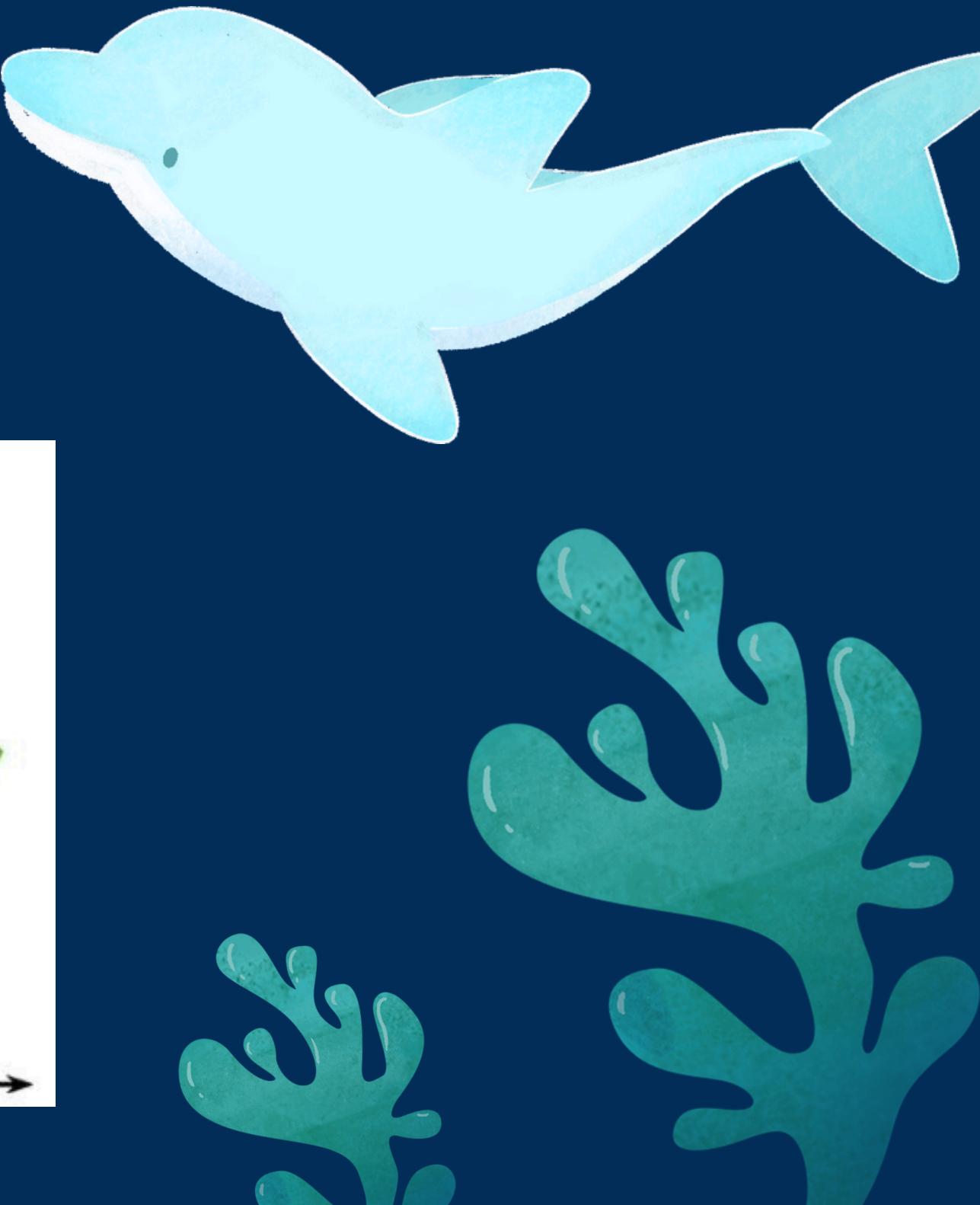
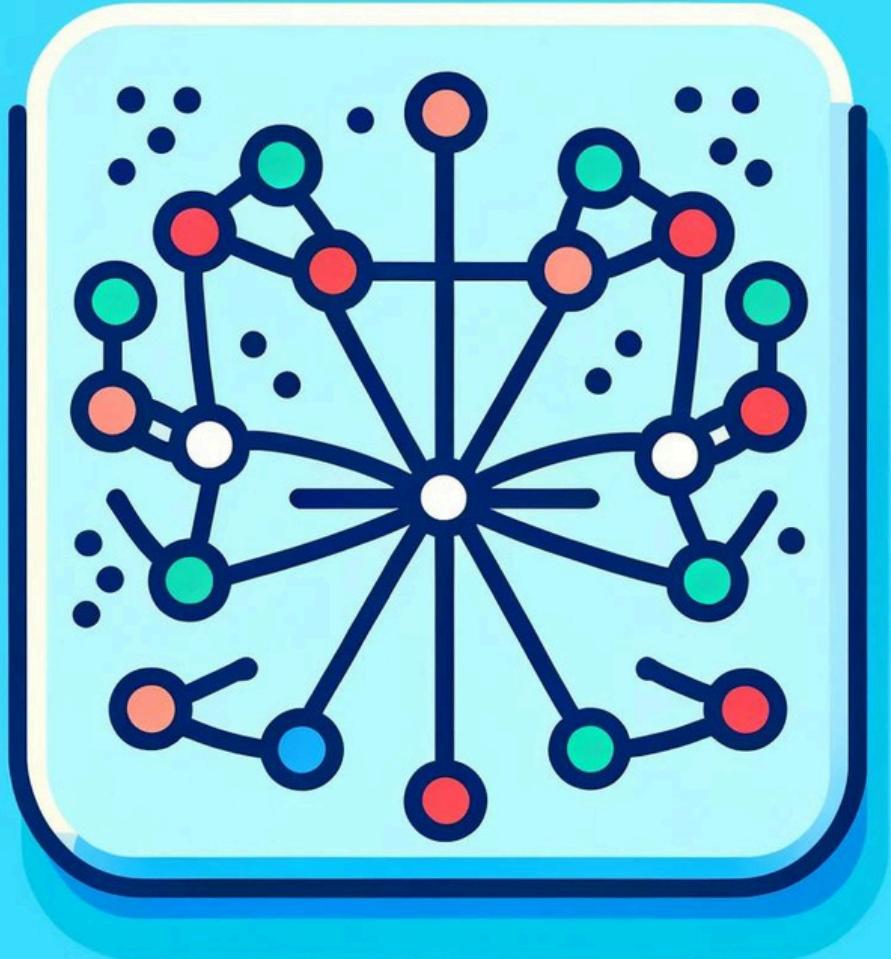
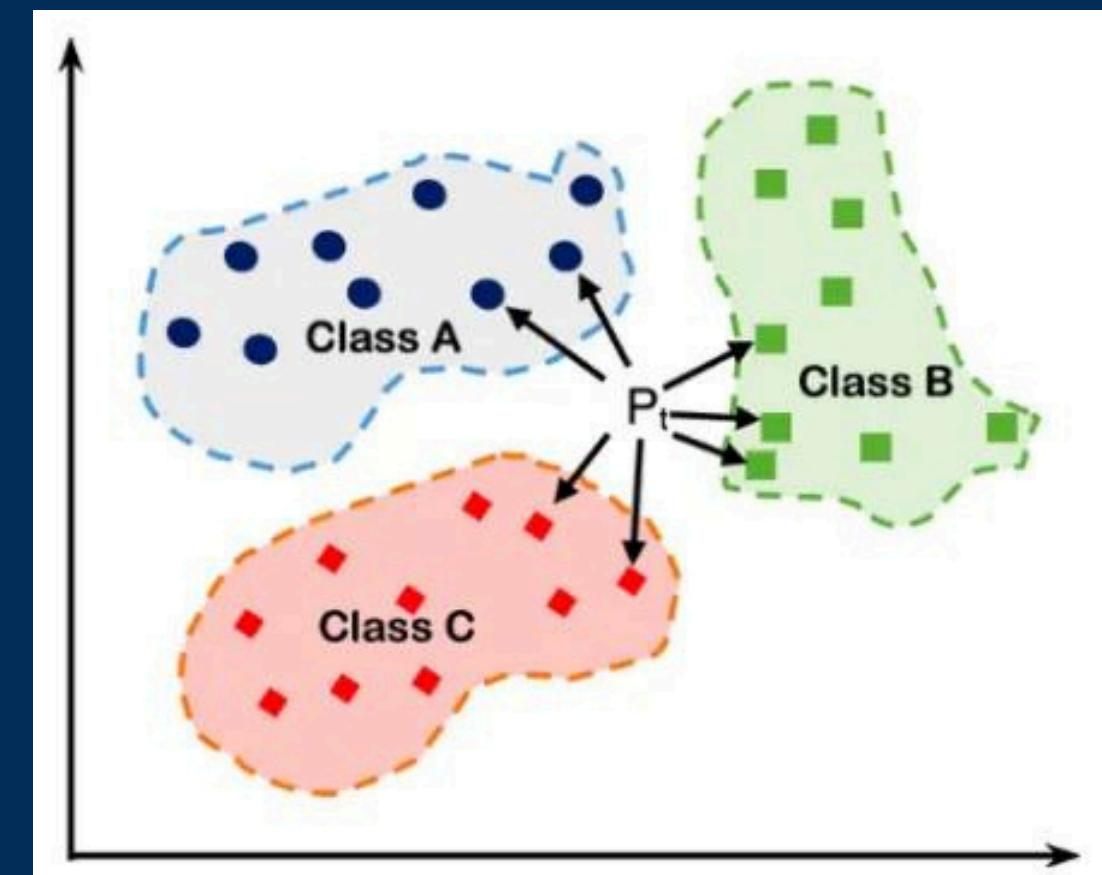
Aims to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To enhance the performance, it can use the kernel trick to handle nonlinear input spaces.



K-Nearest Neighbors

What is it ?

Predicts the classification of a point by a majority vote of its nearest neighbors, with the point being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small).



EVALUATION MODEL

01

02

04

03

F1 Score

$$\frac{TP + TN}{TP + FP + TN + FN}$$

Accuracy Score

Recall Score

$$\frac{TP}{TP + FP}$$

Precision Score

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

The Code



A dark blue background featuring stylized white outlines of pink fish swimming at the top and bottom. The bottom half is filled with various types of blue and teal sea plants, including large leafy ones and branching corals.

THANK YOU!

Any Questions?