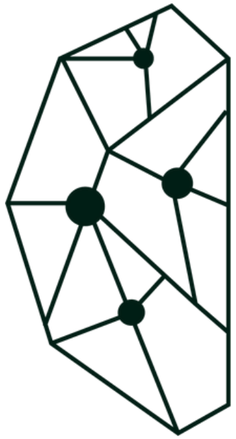


Data Overflow

INSAT ACM Student Chapter

IEEE Computer Society INSAT

SUPERVISED LEARNING

Regression & Classification

LET'S DISCOVER THE SUPERVISED LEARNING WORLD TOGETHER!

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

MEET YOUR WORKSHOPER

Eya Laouini

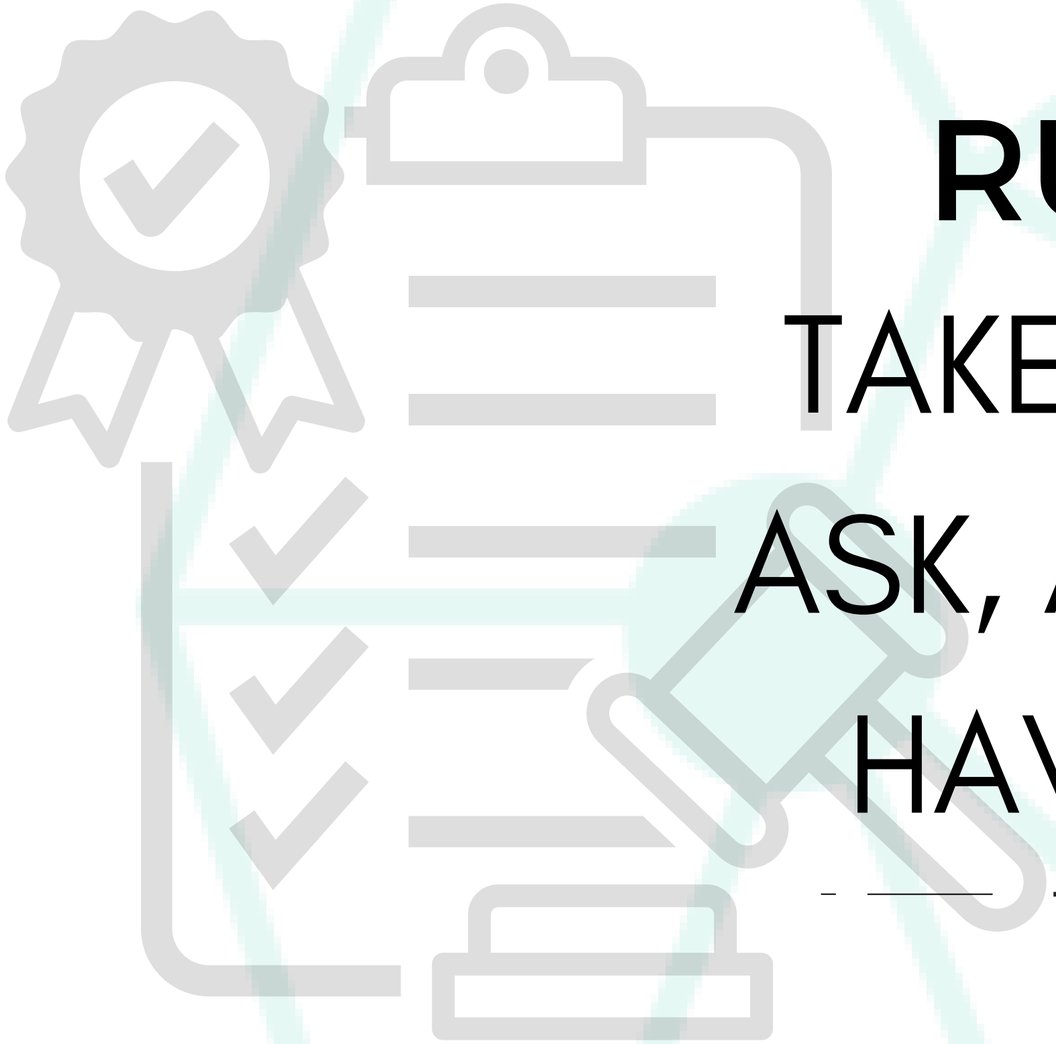
- IT Student @ISET Nabeul
- AI Intern @L-mobile Tunisia
- AI Instructor @GoMyCode

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

GOAL

UNDERSTAND THE SUPERVISED LEARNING

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT



RULES

TAKE NOTES

ASK, ASK, ASK

HAVE FUN

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

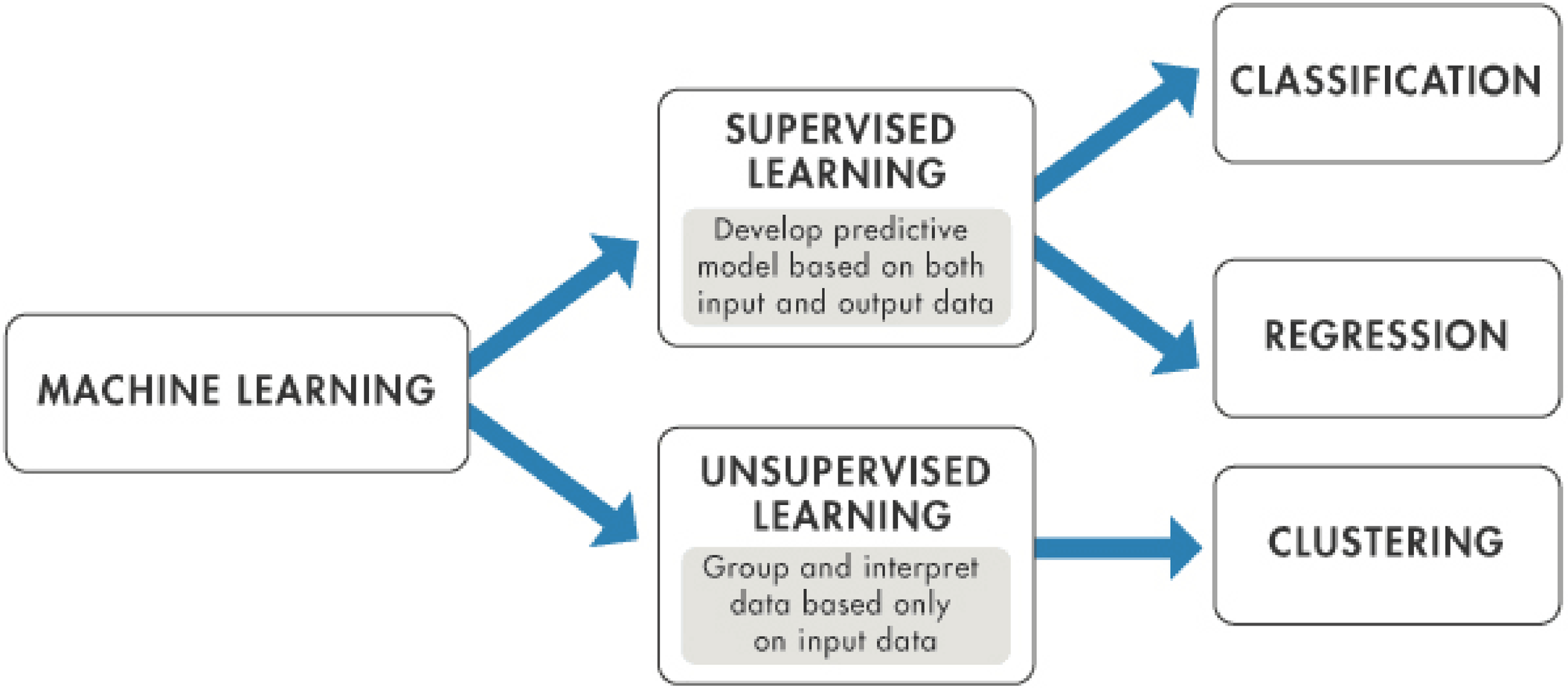
LES'T START OUR JOURNEY!

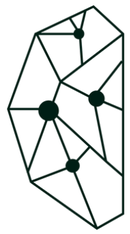
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

WHAT DO YOU KNOW ABOUT SUPERVISED LEARNING?

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

MACHINE LEARNING





Data
Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

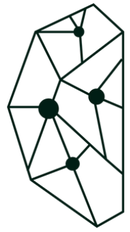
SUPERVISED LEARNING

Supervised learning is a category of machine learning that uses **labeled** datasets to train algorithms to predict outcomes and recognize patterns.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED LEARNING

The data used in supervised learning is **labeled** — meaning that it contains examples of both inputs **(called features)** and correct outputs **(labels)**.



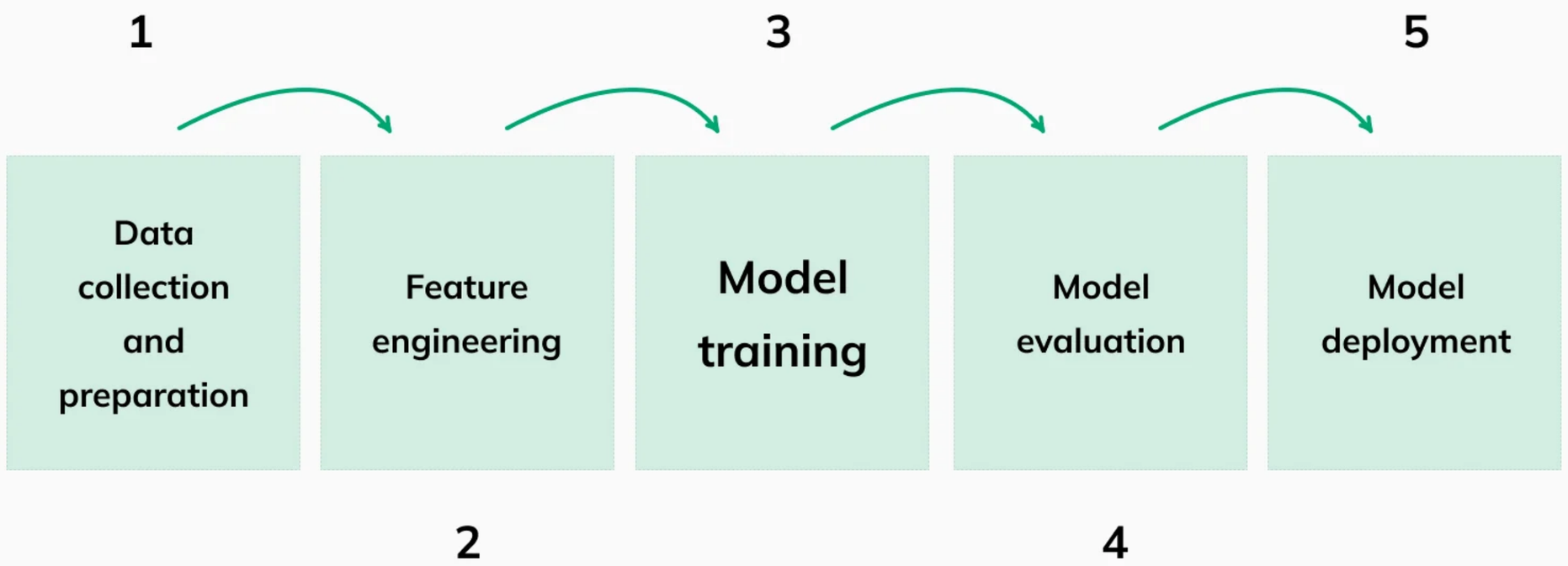
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

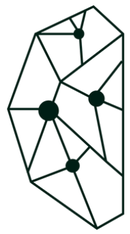
SUPERVISED LEARNING

The **algorithms** analyze a large dataset of these **training pairs** to infer what a desired output value would be when asked to **predict new data**.

ML WORKFLOW

Machine learning workflow





Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED LEARNING

Features

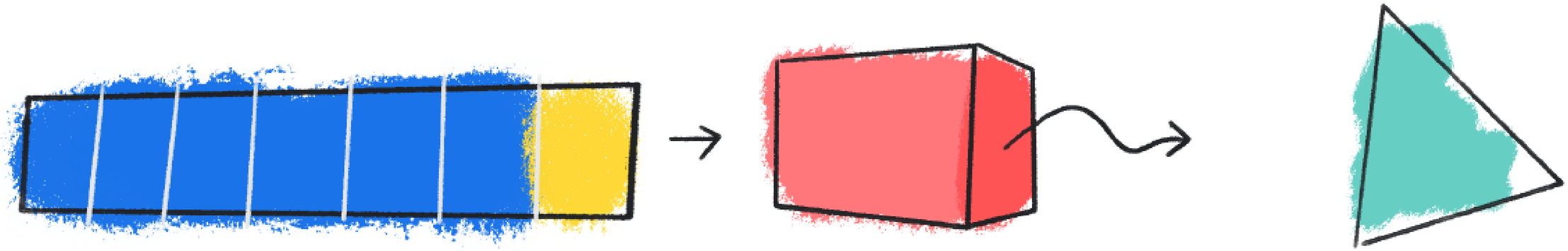
Label

date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure	rainfall
2021-09-09	49.71N	82.16W	74	20	3	N	18.6	.01
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94	.23

Example

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED LEARNING



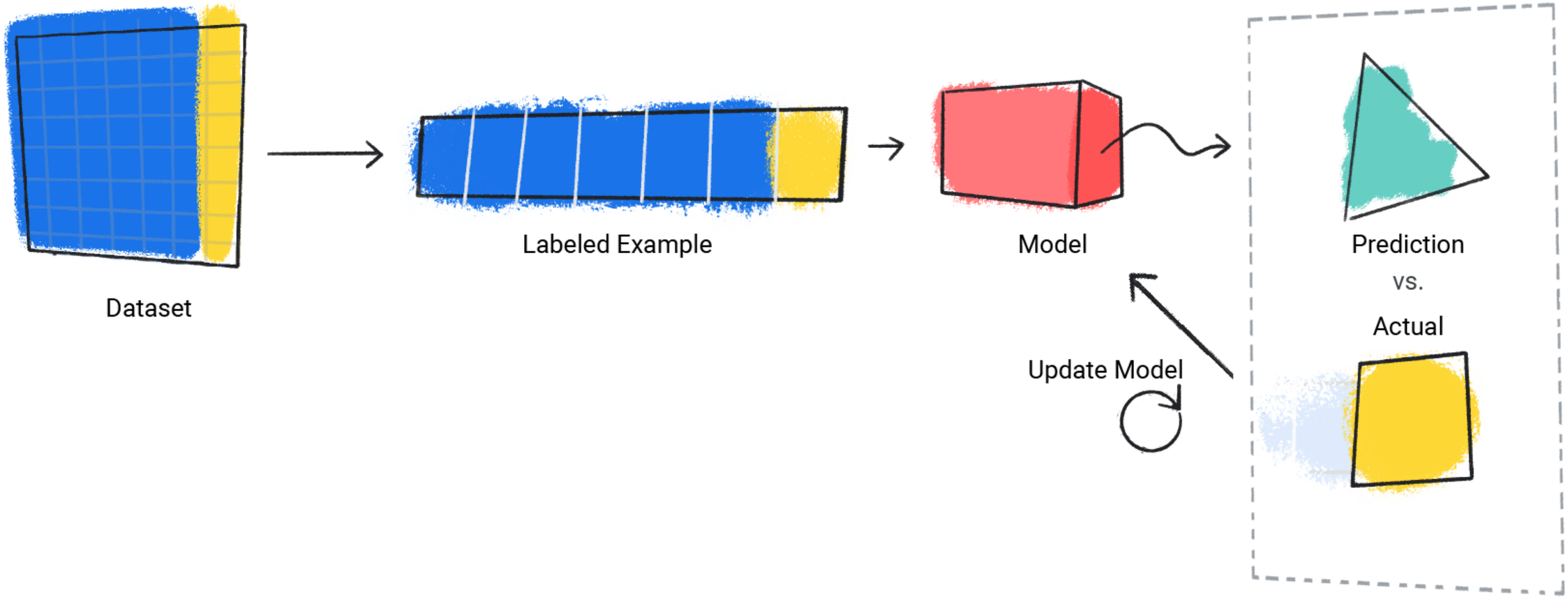
Labeled Example

Model

Prediction

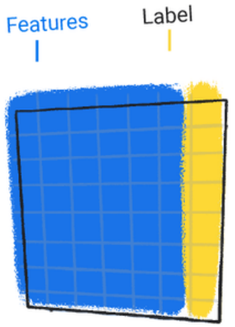
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED LEARNING



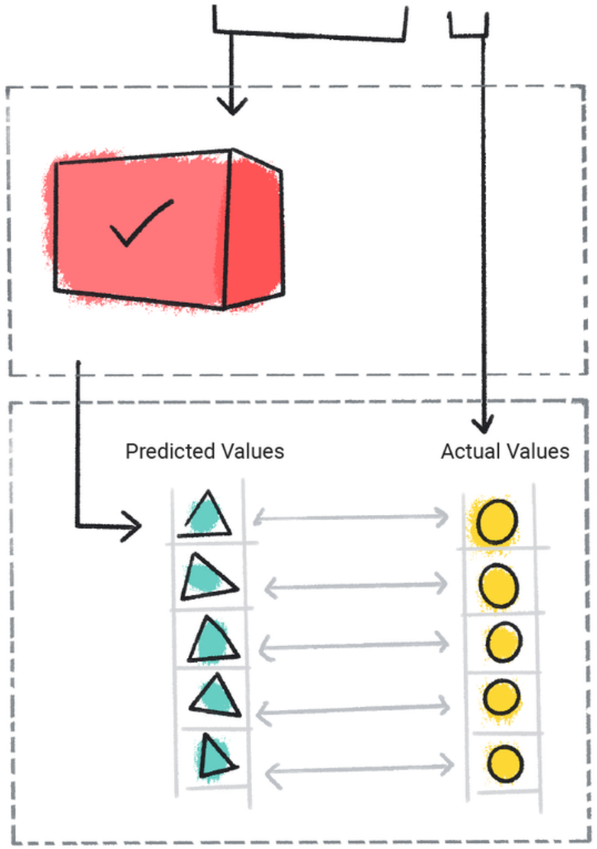
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED LEARNING



1

Use a **trained ML Model** to predict a value from a given dataset with labeled examples.

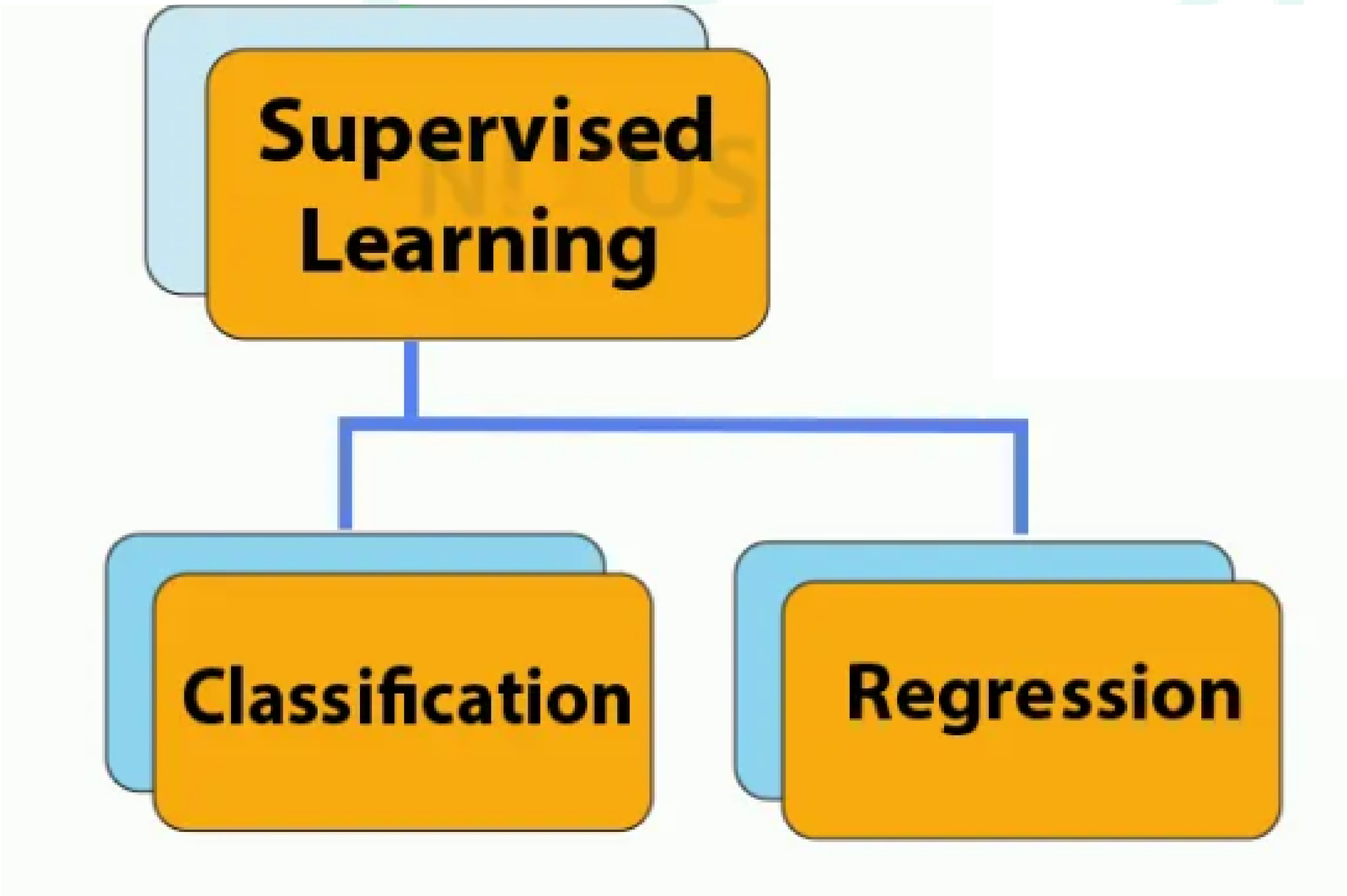


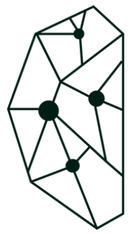
2

Compare **Predicted Values** with **Actual Values**.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED LEARNING



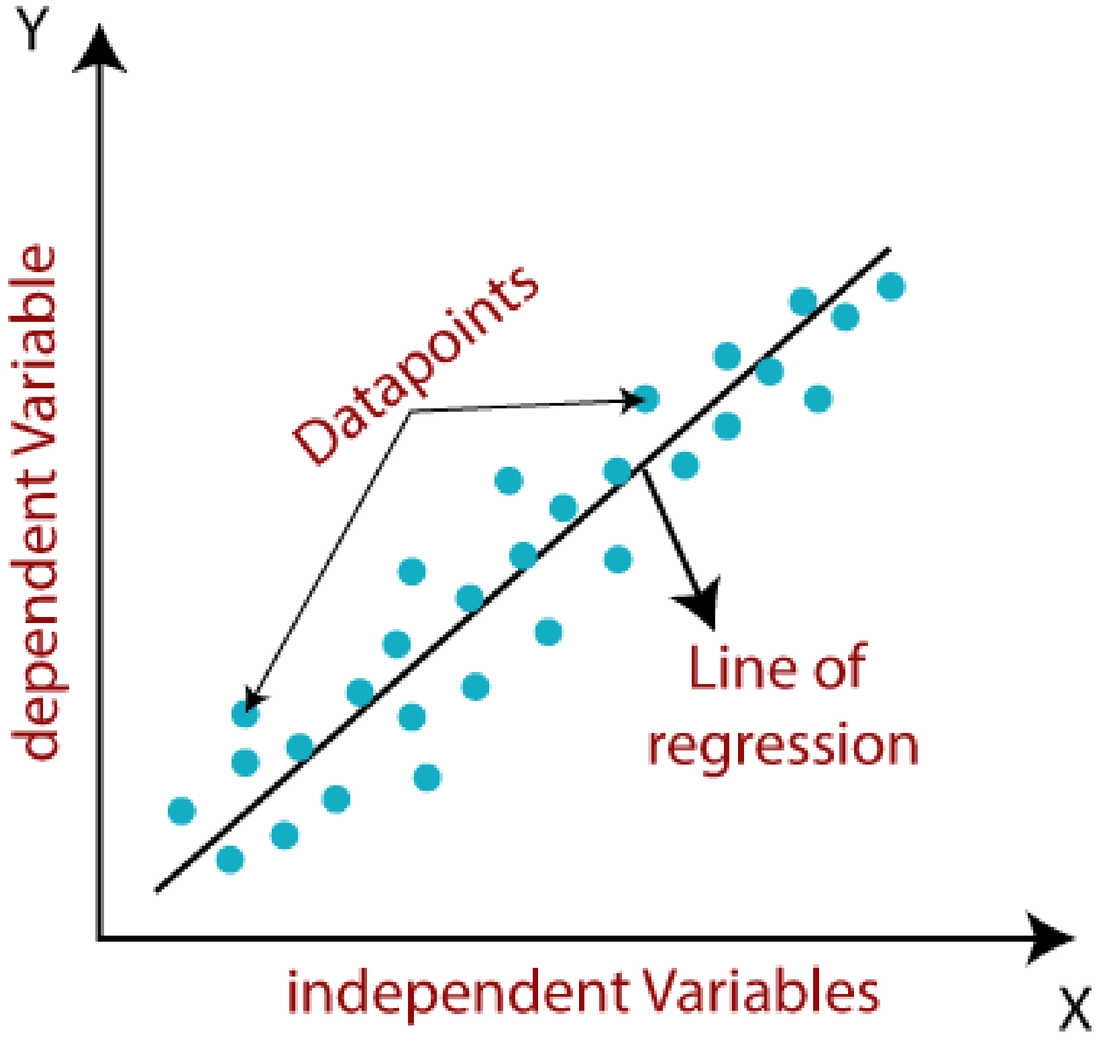


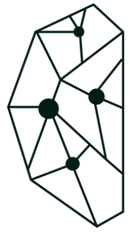
Data
Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

REGRESSION

Regression algorithms are used to predict a real or **continuous value**, where the algorithm detects a relationship between two or more variables.

REGRESSION


Y
dependent Variable
Datapoints
Line of
regression
independent Variables
X



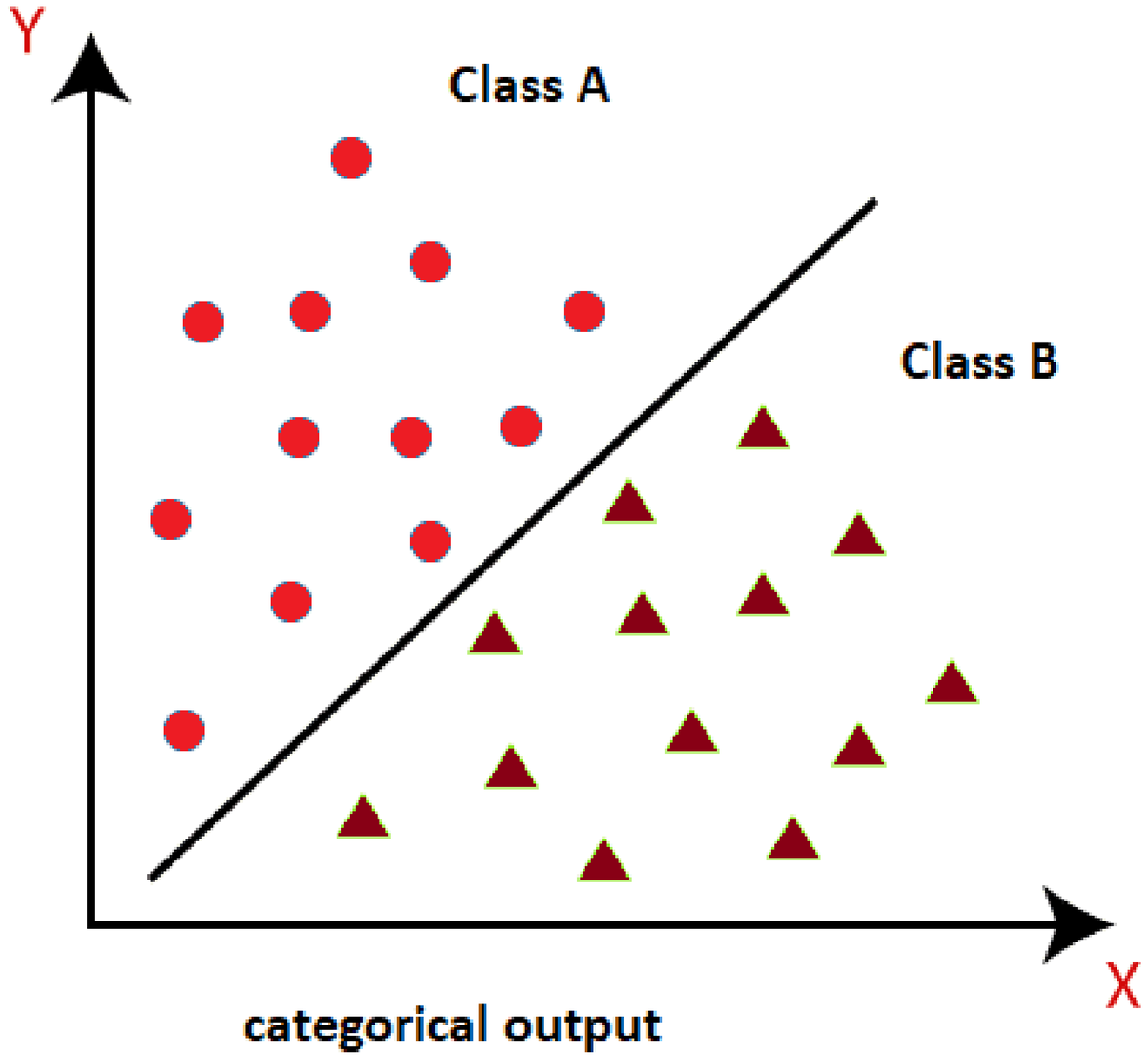
Data
Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

CLASSIFICATION

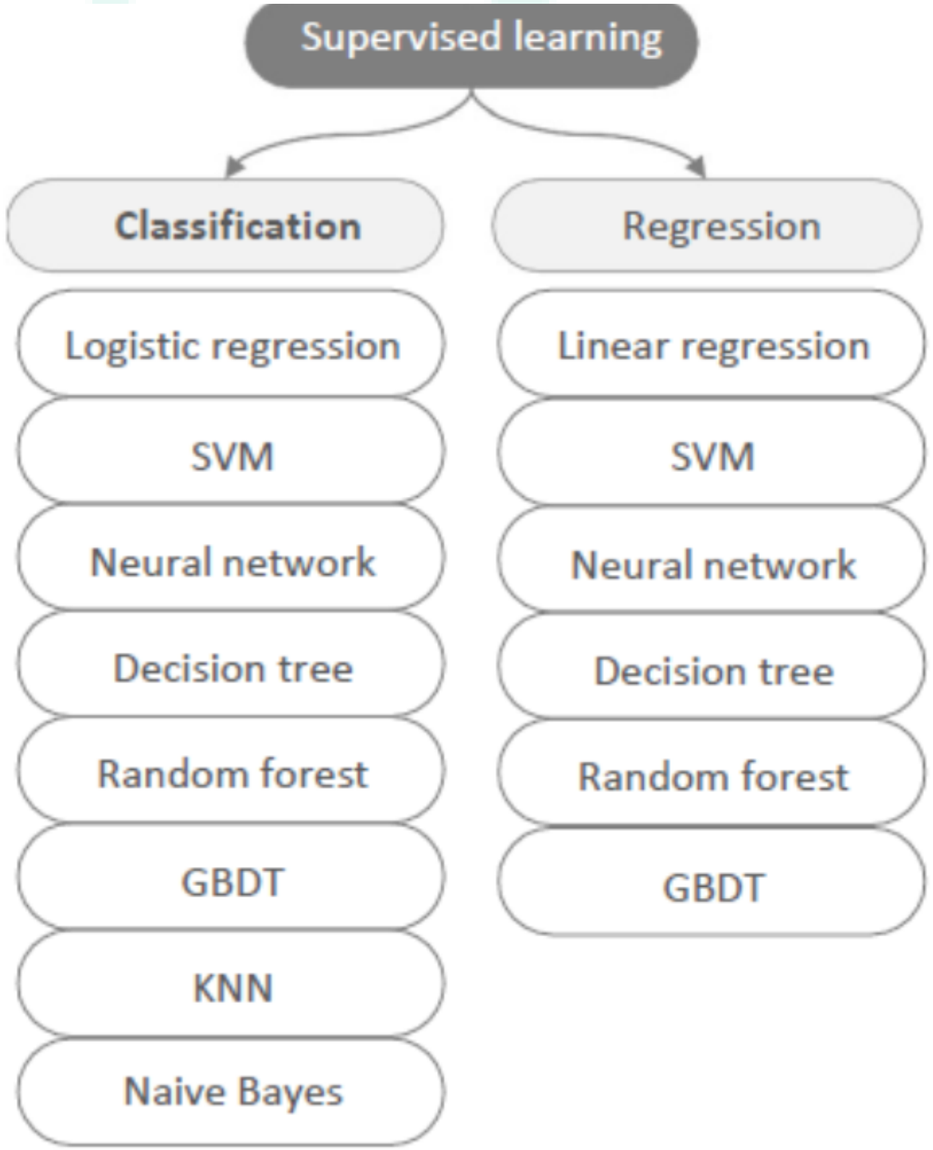
Classification algorithms are used to **group data** by predicting a **categorical label** or output variable based on the input data. Classification is used when output variables are categorical, meaning there are two or more **classes**.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

CLASSIFICATION



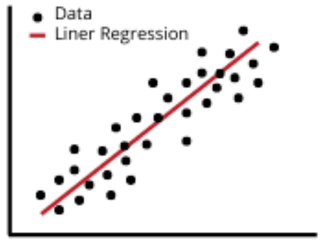
SUPERVISED ALGOS



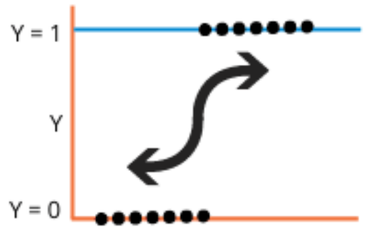
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPERVISED ALGOS

Linear Regression

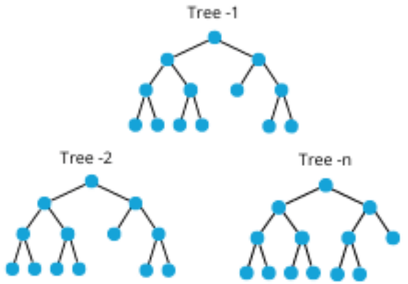


Logistic Regression

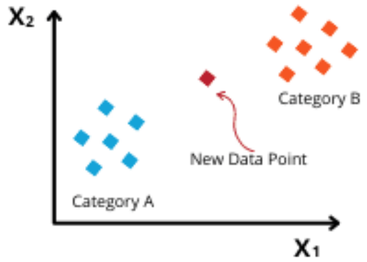


Decision Trees

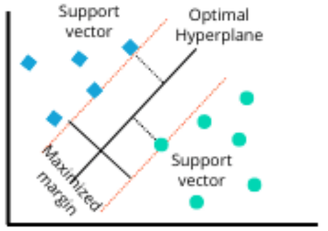
Random Forest

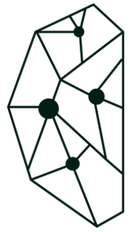


K-Nearest Neighbor



Support Vector Machine





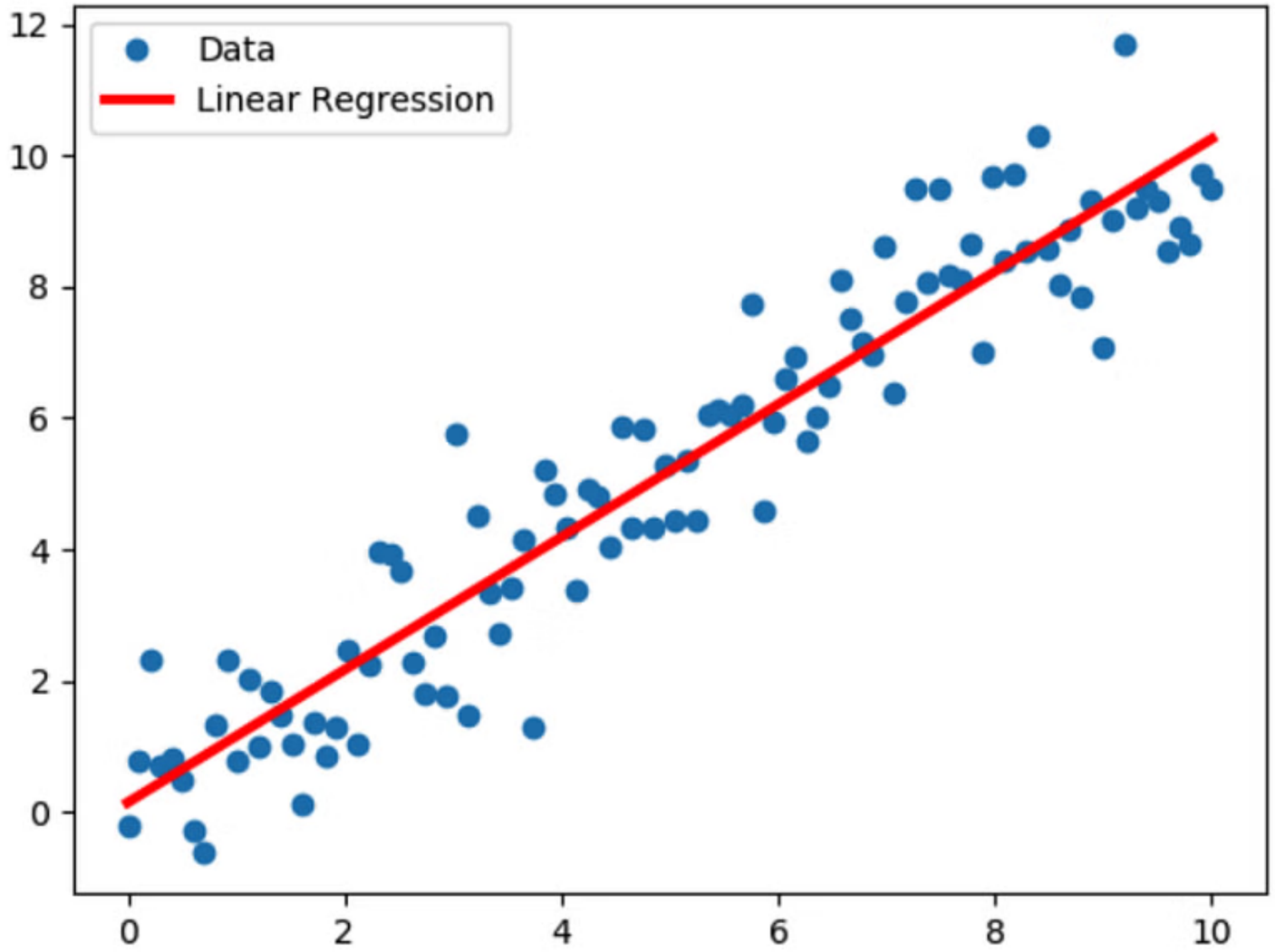
Data
Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

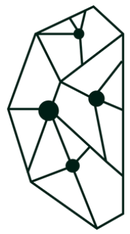
LINEAR REGRESSION

A simple algorithm that models a linear relationship between inputs and a continuous numerical output variable.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

LINEAR REGRESSION





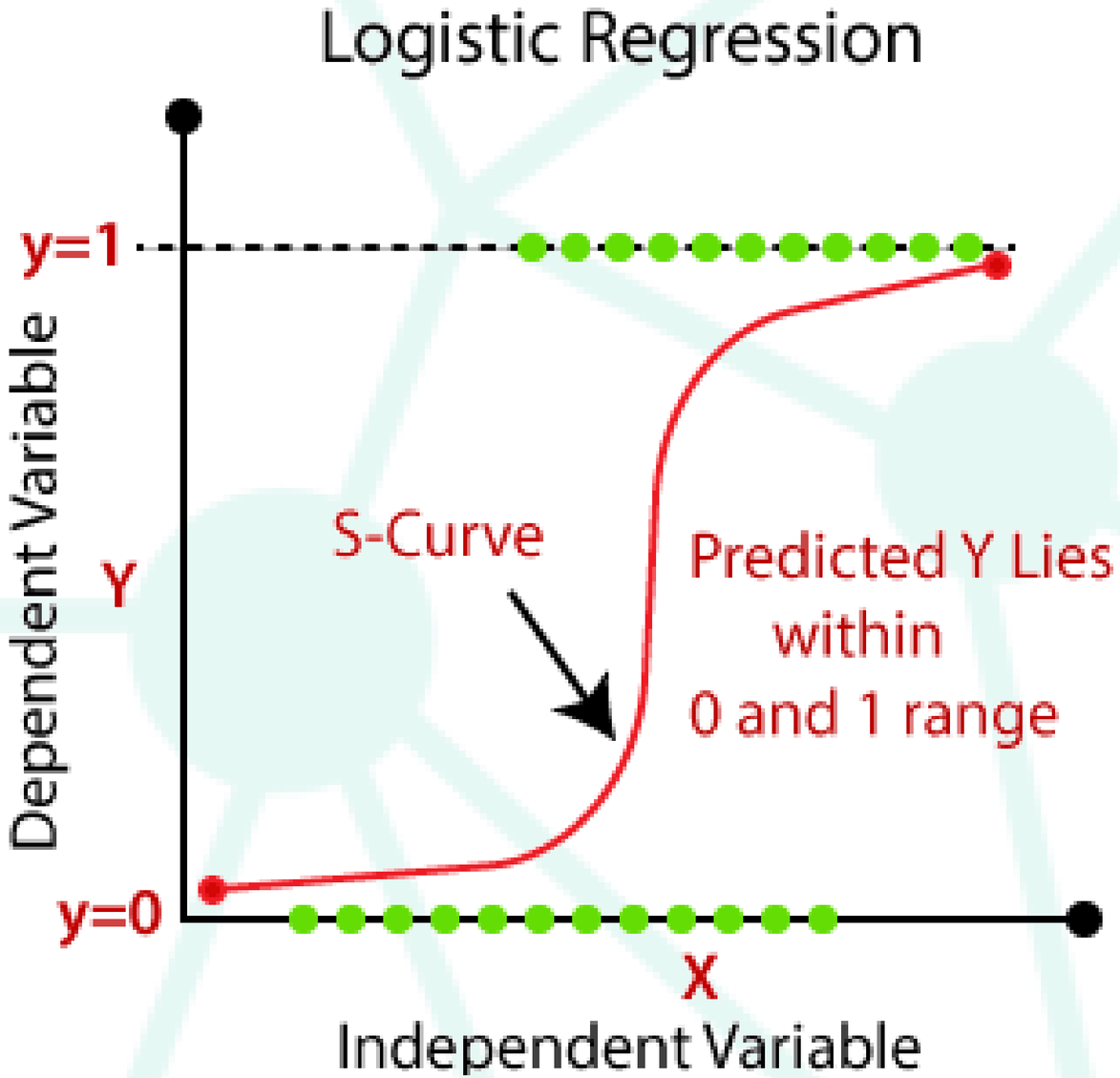
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

LOGISTIC REGRESSION

Logistic Regression is a special case of Linear Regression where the target variable (y) is discrete/categorical such as 1 or 0, True or False, Yes or No, Default or No Default.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

LOGISTIC REGRESSION



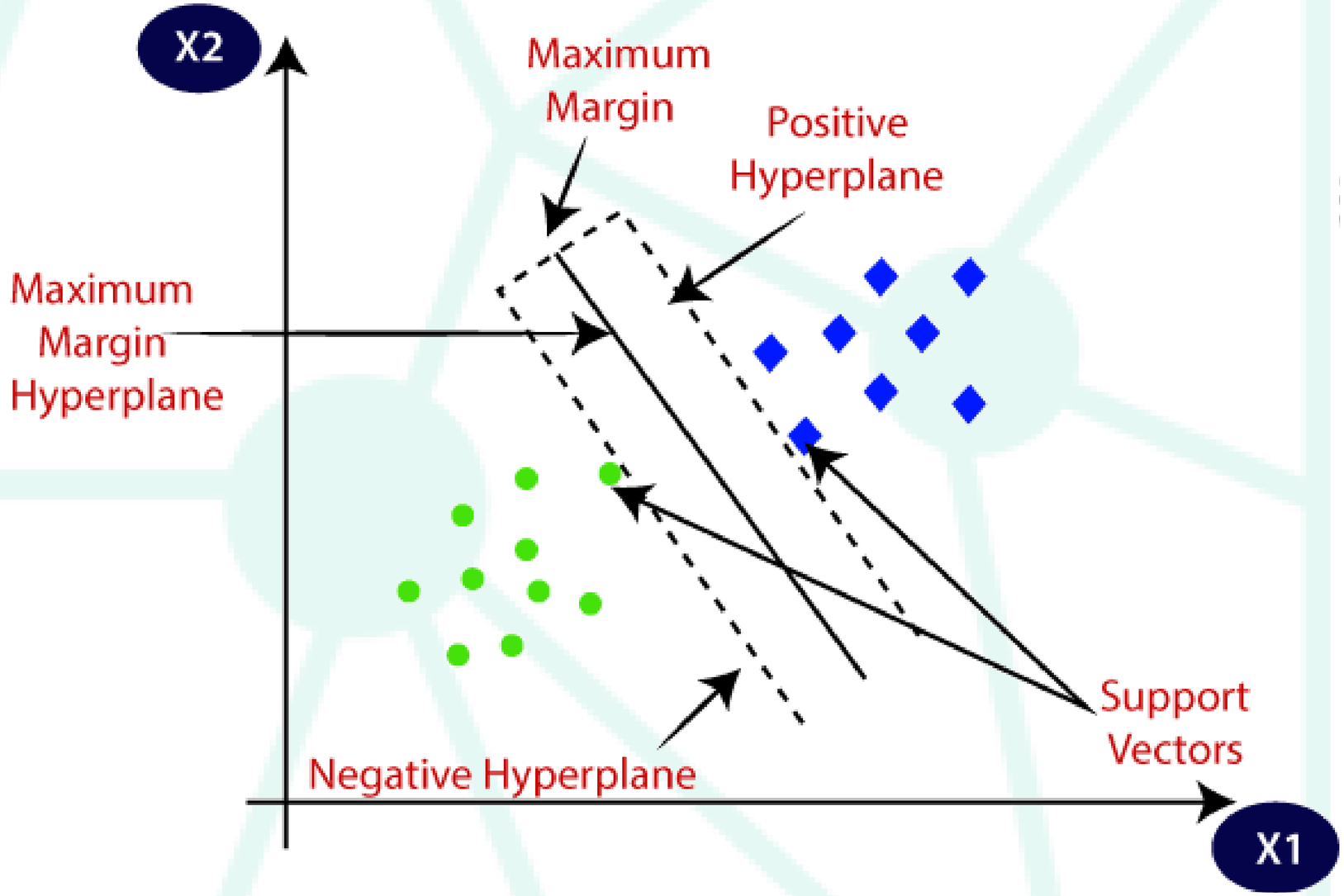
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are a set of supervised learning methods used for classification, regression, and outliers detection. The basic principle behind SVM is to find the best boundary (or hyperplane) that separates data points of different classes. This boundary is chosen to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

SUPPORT VECTOR MACHINES



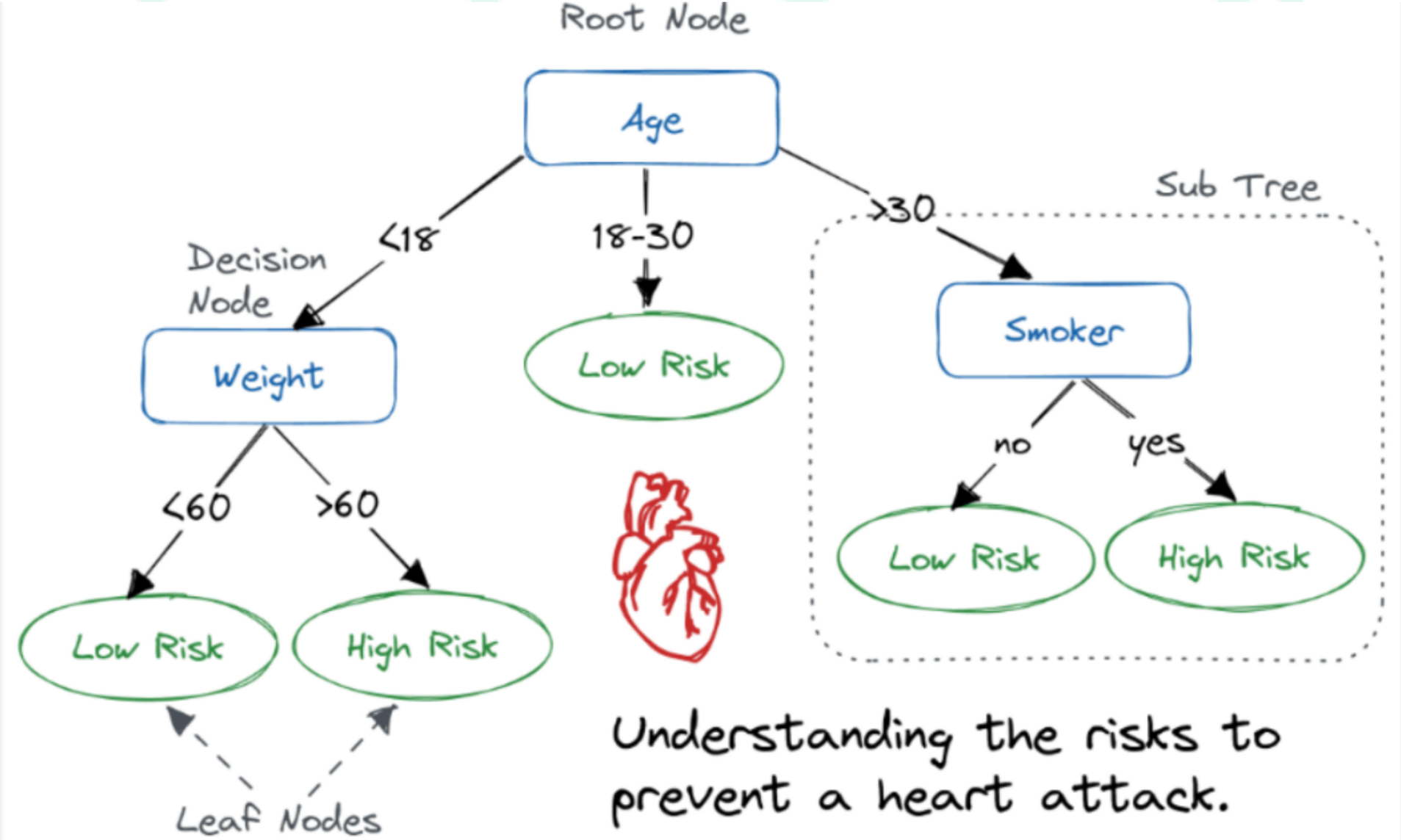
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

DECISION TREE

Decision Tree algorithms are a type of probability tree-like structural model that continuously separates data to categorize or make predictions depending on the results of the previous set of questions. The model analyzes the data and responds to the questions to assist you in making more informed choices.

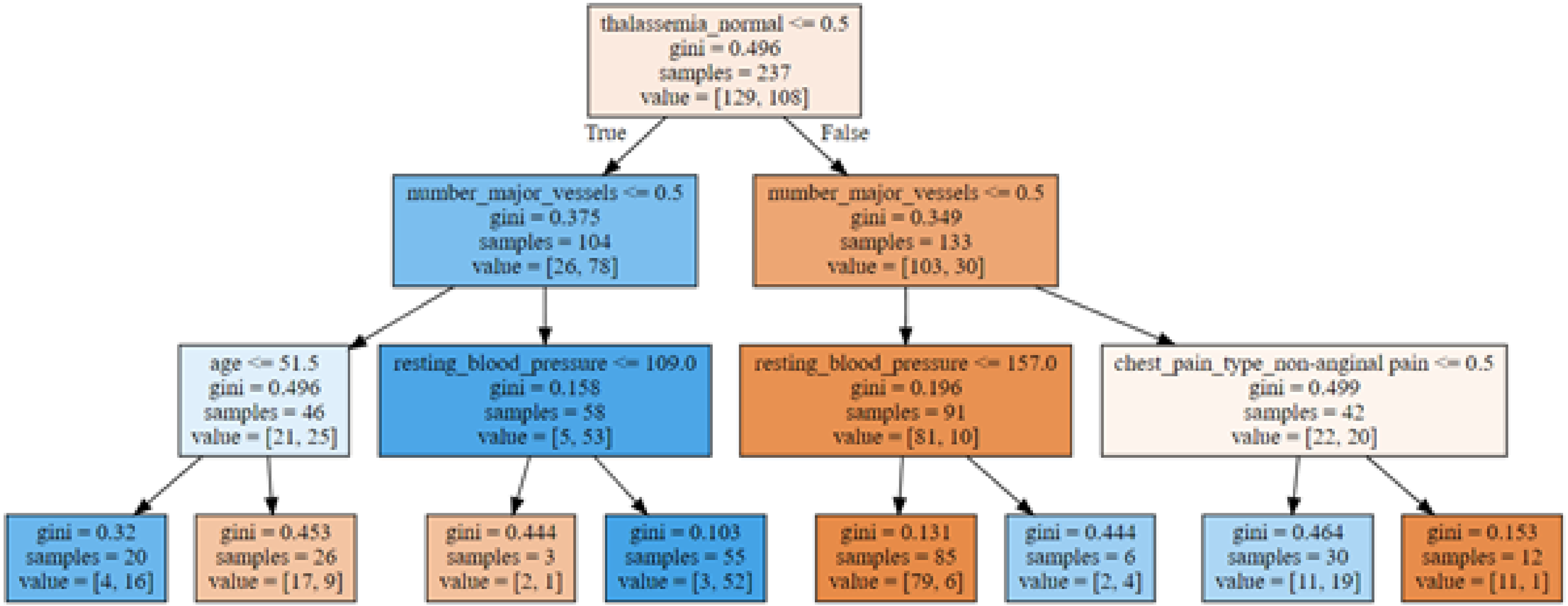
Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

DECISION TREE



Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

DECISION TREE



Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

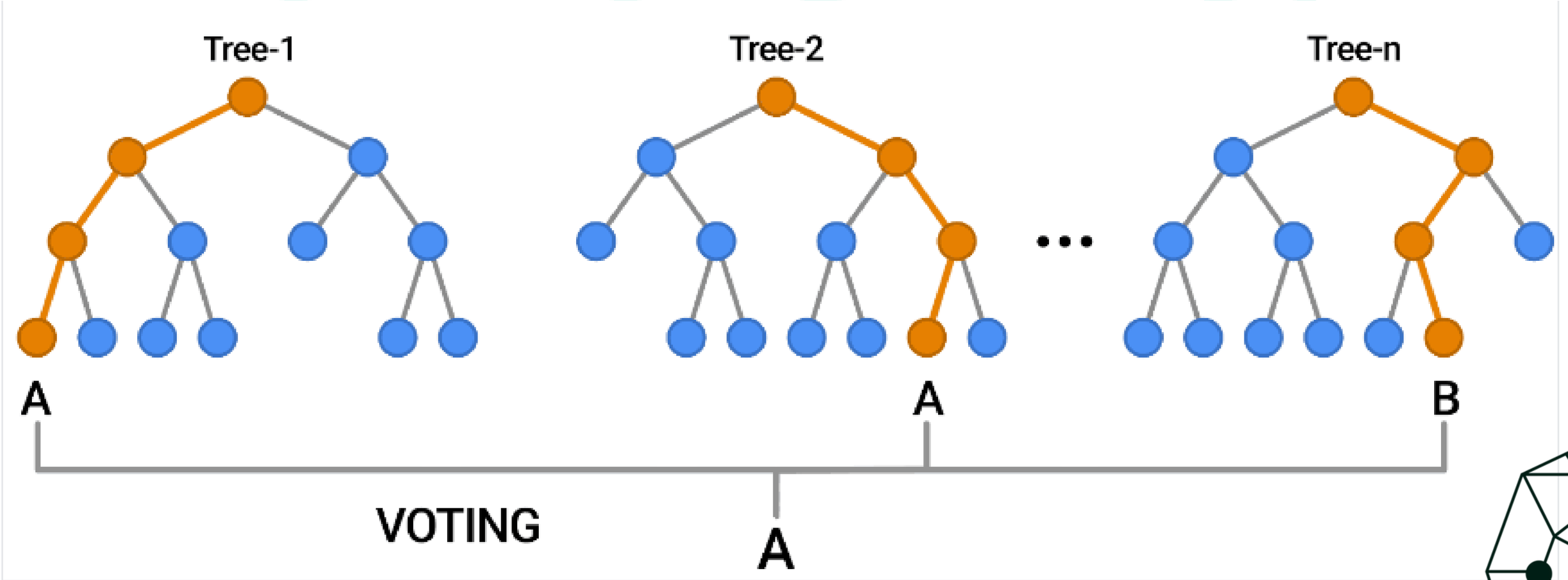
RANDOM FOREST

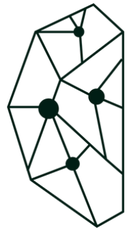
Random Forest is another example of an algorithm that is built on trees, just like Decision Trees. In contrast to Decision Tree, which only consists of a single tree, Random Forest employs a number of Decision Trees to make judgments, creating what is essentially a forest of trees.

It does this by combining a number of different models in order to produce predictions, and it may be used for both classification and regression.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

RANDOM FOREST

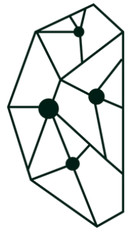




Data Overflow

INSAT ACM Student Chapter
IEEE Computer Society INSAT

LET'S PRACTICE



Data Overflow

INSAT ACM Student Chapter
IEEE Computer Society INSAT

GITHUB LINK

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

LET'S TEST OUR KNOWLEDGE!

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 1: What is supervised learning?

A) A type of learning where the model tries to predict the output from the input data without any labeled examples.
B) A type of learning where the model is trained on labeled data to predict the output for new inputs.
C) A type of learning where the model organizes the input data into clusters without labeled examples.
D) A type of learning where the model only uses text data for training.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 1: What is supervised learning?

A) A type of learning where the model tries to predict the output from the input data without any labeled examples.

B) A type of learning where the model is trained on labeled data to predict the output for new inputs.

C) A type of learning where the model organizes the input data into clusters without labeled examples.

D) A type of learning where the model only uses text data for training.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 2: What is a label in a supervised learning dataset?

A) A unique identifier for each data point.

B) The feature that is used to divide the dataset into training and testing sets.

C) The variable that is being predicted by the model.

D) The algorithm that is used to train the model.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 2: What is a label in a supervised learning dataset?

A) A unique identifier for each data point.

B) The feature that is used to divide the dataset into training and testing sets.

C) The variable that is being predicted by the model.

D) The algorithm that is used to train the model.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 3: In the context of supervised learning, what is regression used for?

A) To predict categorical outcomes.
B) To predict numerical continuous outcomes.
C) To classify images into different categories.
D) To cluster data into different groups.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 3: In the context of supervised learning, what is regression used for?

A) To predict categorical outcomes.

B) To predict numerical continuous outcomes.

C) To classify images into different categories.

D) To cluster data into different groups.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 4: Which of the following is a regression algorithm?

A) K-Means Clustering

B) Decision Tree

C) Naive Bayes Classifier

D) Linear Regression

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 4: Which of the following are regression algorithms?

A) K-Means Clustering

B) Decision Tree

C) Naive Bayes Classifier

D) Linear Regression

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 5: What is classification in supervised learning?

A) Predicting the exact numerical value for new data points.

B) Predicting the category or class to which new data points belong.

C) Dividing data into clusters that contain similar characteristics.

D) Estimating the probability of an event occurring.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 5: What is classification in supervised learning?

A) Predicting the exact numerical value for new data points.

B) Predicting the category or class to which new data points belong.

C) Dividing data into clusters that contain similar characteristics.

D) Estimating the probability of an event occurring.

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 6: Which of the following are classification algorithms?

A) Logistic Regression

B) Decision Tree

C) Linear Regression

D) Random Forest

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

QUIZ

Question 6: Which of the following are classification algorithms?

A) Logistic Regression

B) Decision Trees

C) Linear Regression

D) Random Forest

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

THANK YOU
FOR YOUR ATTENTION!

Data Overflow
INSAT ACM Student Chapter
IEEE Computer Society INSAT

CONTACT ME!

Eya Laouini

Eya Laouini

Eya-Laouini

eyai.laouini@gmail.com

Eya Laouini

Eya Laouini