INTRODUCTION TO MACHINE LEARNING & RELATED CONCEPTS

MACHINE LEARNING

- Machine learning is a tool **for turning information into knowledge**
- In the past 50 years, there has been an explosion of data
- This mass of data is useless unless **we analyze it and find the patterns hidden within**
- Machine learning techniques are used to automatically **find the valuable underlying patterns within complex data** that we would otherwise struggle to discover
- The hidden patterns and knowledge about a problem can be **used to predict future events and perform all kinds of complex decision making**

MACHINE LEARNING

- Traditionally, software engineering combined human created rules with data to create answers to a problem. Instead, machine learning uses data and answers to discover the rules behind a problem.

- Machines have to go through a learning process, trying different rules and learning from how well they perform. Hence, why it’s known as Machine Learning

IMPORTANT TERMS

DATASET

- Source Dataset that contain features important to solving the problem

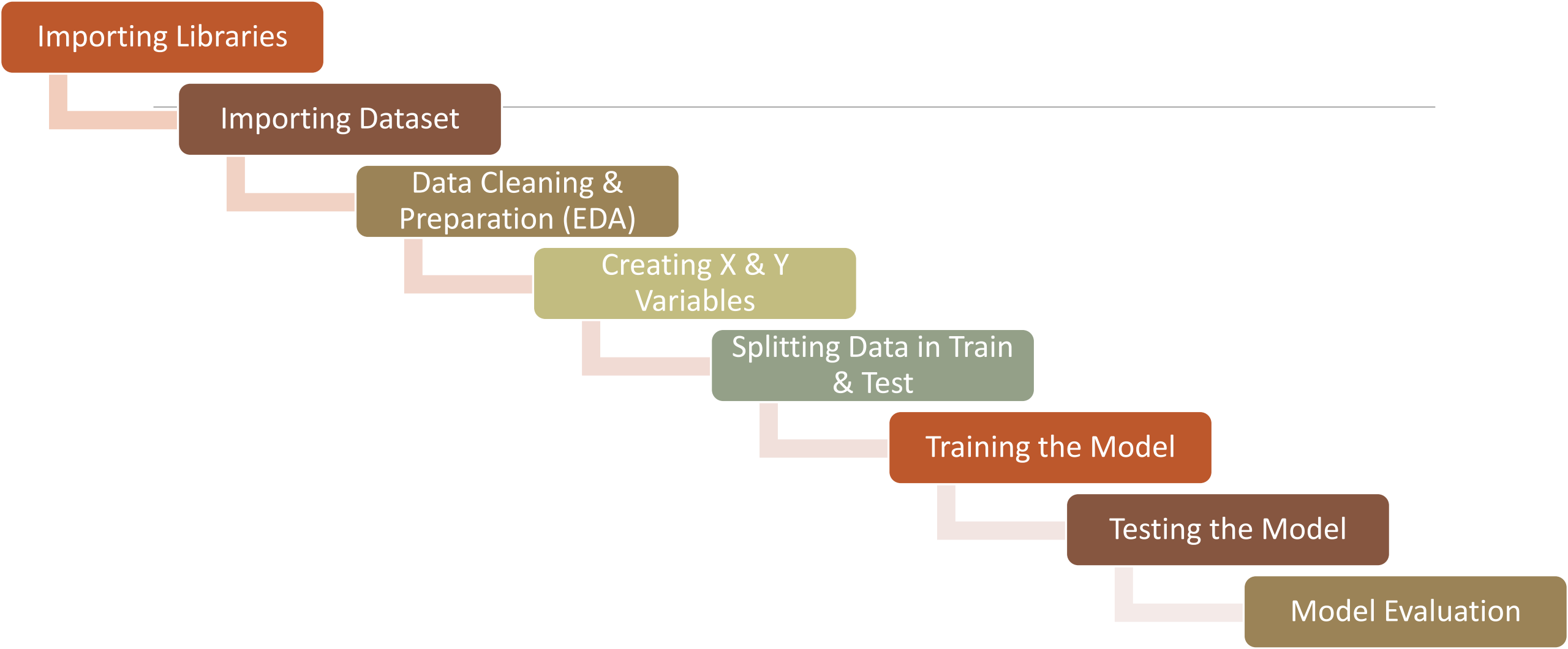
FEATURES

- Important pieces of data that help us understand a problem.
- These are fed in to a Machine Learning algorithm to help it learn

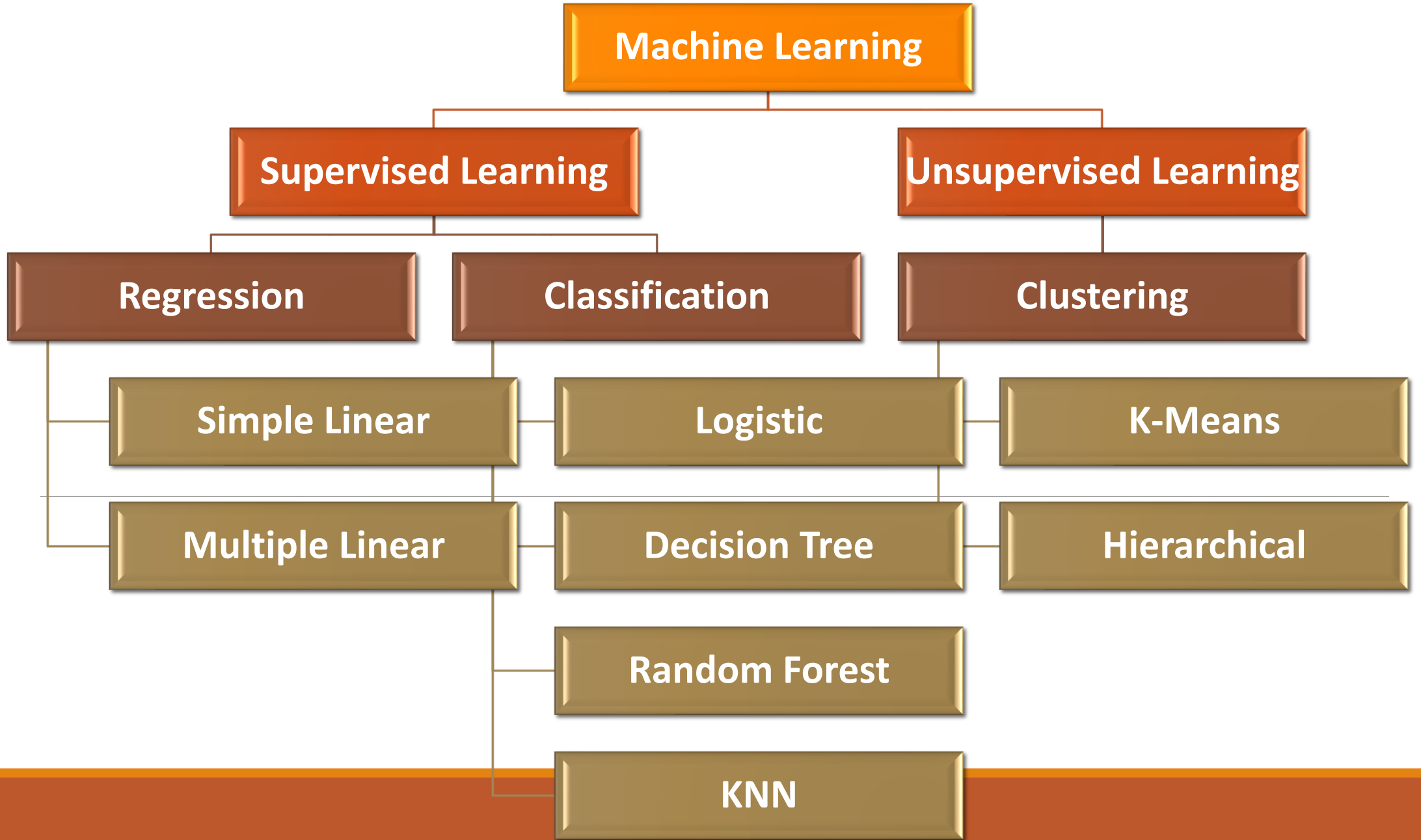
MODEL

- The model is the output you get after training an algorithm

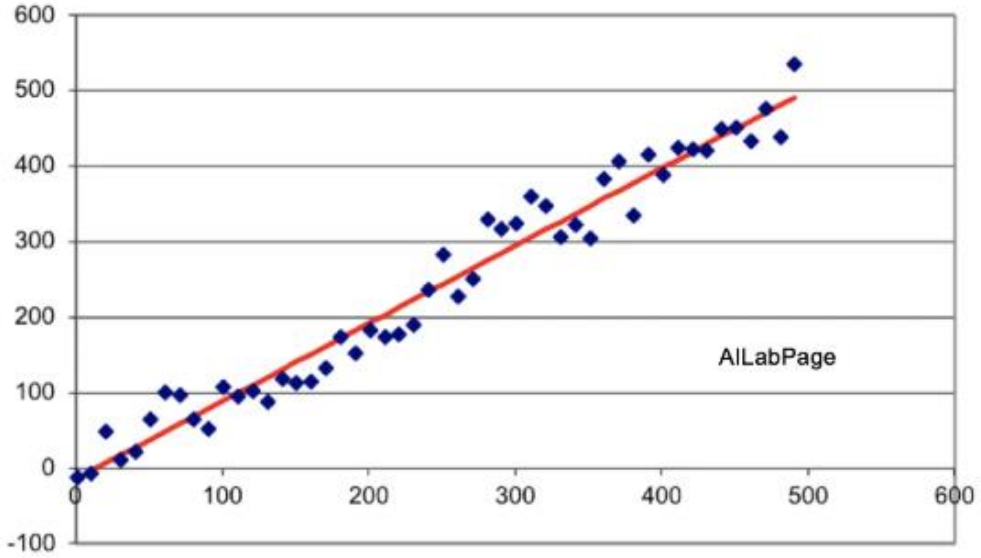
STEPS IN MACHINE LEARNING

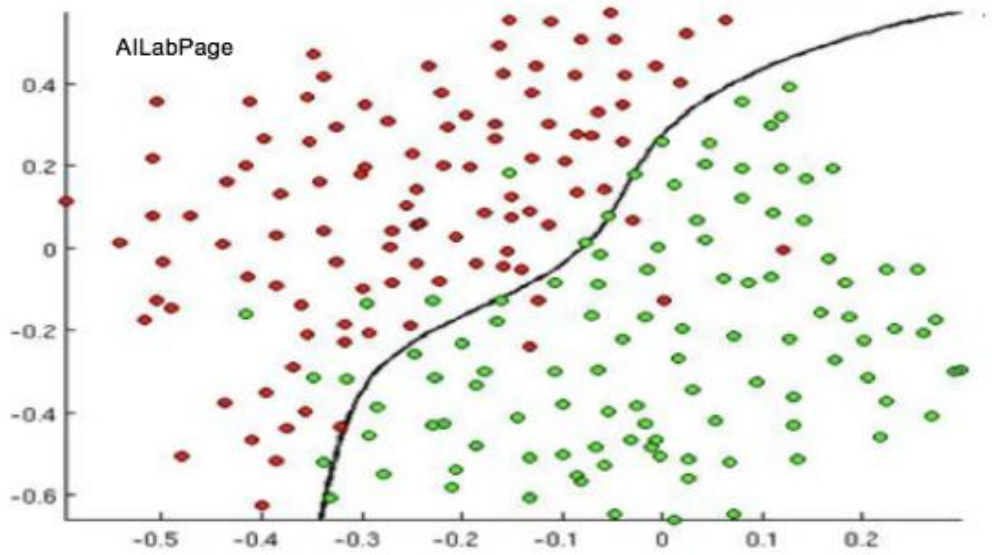


TYPES OF MACHINE LEARNING



REGRESSION VS CLASSIFICATION





Regression

The system attempts to predict a value for an input based on past data.

Example – 1. Temperature for tomorrow

Classification

In classification, predictions are made by classifying them into different categories.

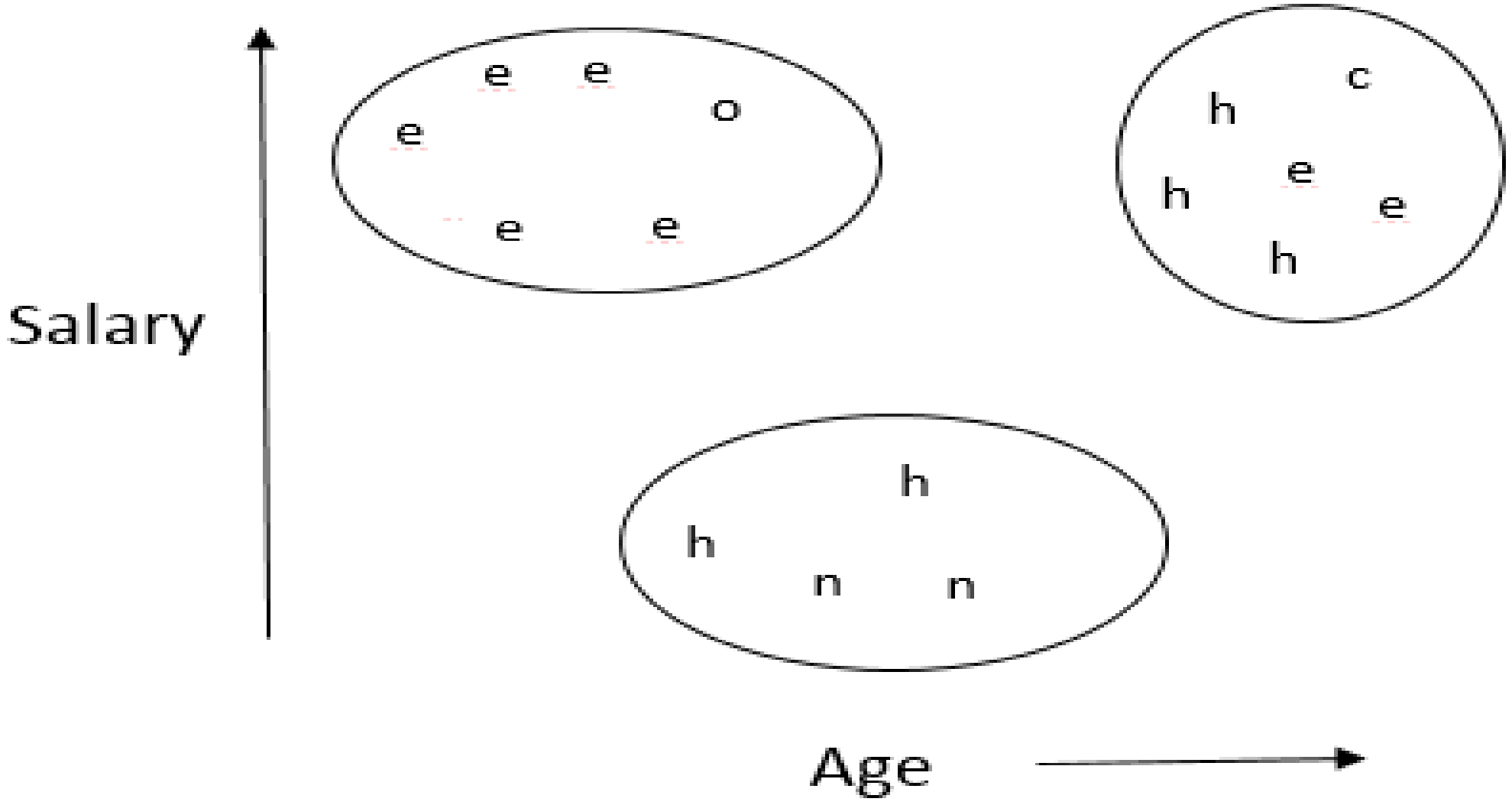
Example – 1. Type of cancer 2. Cancer Y/N

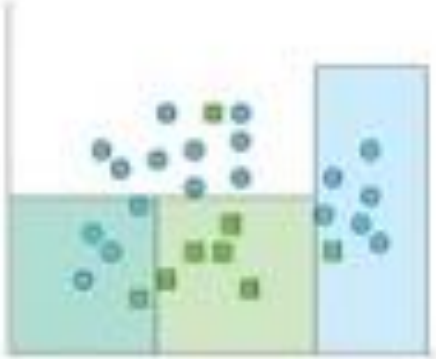
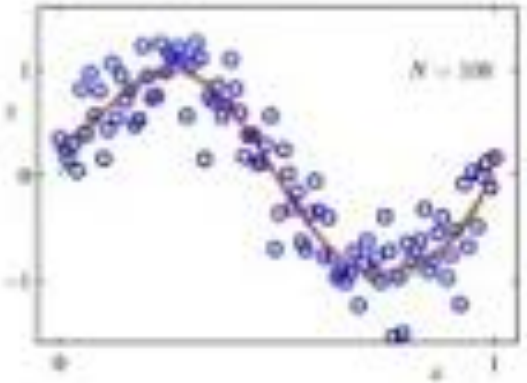
CLASSIFICATION VERSUS REGRESSION

2 KEY DIFFERENCES

CLASSIFICATION	REGRESSION
A tree model where the target variable can take a discrete set of values.	A tree model where the target variable can take continuous values typically real numbers.
The dependent variables are categorical.	The dependent variables are numerical.

CLUSTERING



Predictive methods	Descriptive methods
Classification  Learns a method for predicting the instance class from pre-labeled (classified) instances	**Clustering**  Finds "natural" grouping of instances given un-labeled data
Regression  An attempt to predict a continuous attribute	**Association Rules**  Method for discovering interesting relations between variables in large DBs

PROJECT DESCRIPTION

Problem Statement

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing numerous passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class. In this, we ask you to complete the analysis of what sorts of people were likely to survive.

In particular, we ask you to apply the tools of machine learning to predict which passengers survived the tragedy.

DATASET FEATURES

Variable	Definition	Key
survived	Survival	0 = No, 1 = Yes
Pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
Sex	Sex	
Age	Age in years	
Sibsp	# of siblings / spouses aboard the Titanic	
Parch	# of parents / children aboard the Titanic	
Ticket	Ticket number	
Fare	Passenger fare	
Cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

TO – DO TASKS

PART ONE

Complete EDA for Titanic Dataset. Try and draw as many insights about the passengers in the TITANIC.

PART TWO

Build Various Classification Models and Suggest the best Model based on Accuracy of Results.