

# A-Z Machine Learning using Azure Machine Learning (AzureML)

Hands on AzureML: From Azure Machine Learning Introduction to Advance Machine Learning Algorithms. No Coding Required.

**BEST SELLER** ★★★★★ 4.3 (215 ratings) 1,597 students enrolled

Created by Jitesh Khurkhuriya Last updated 3/2018 English English

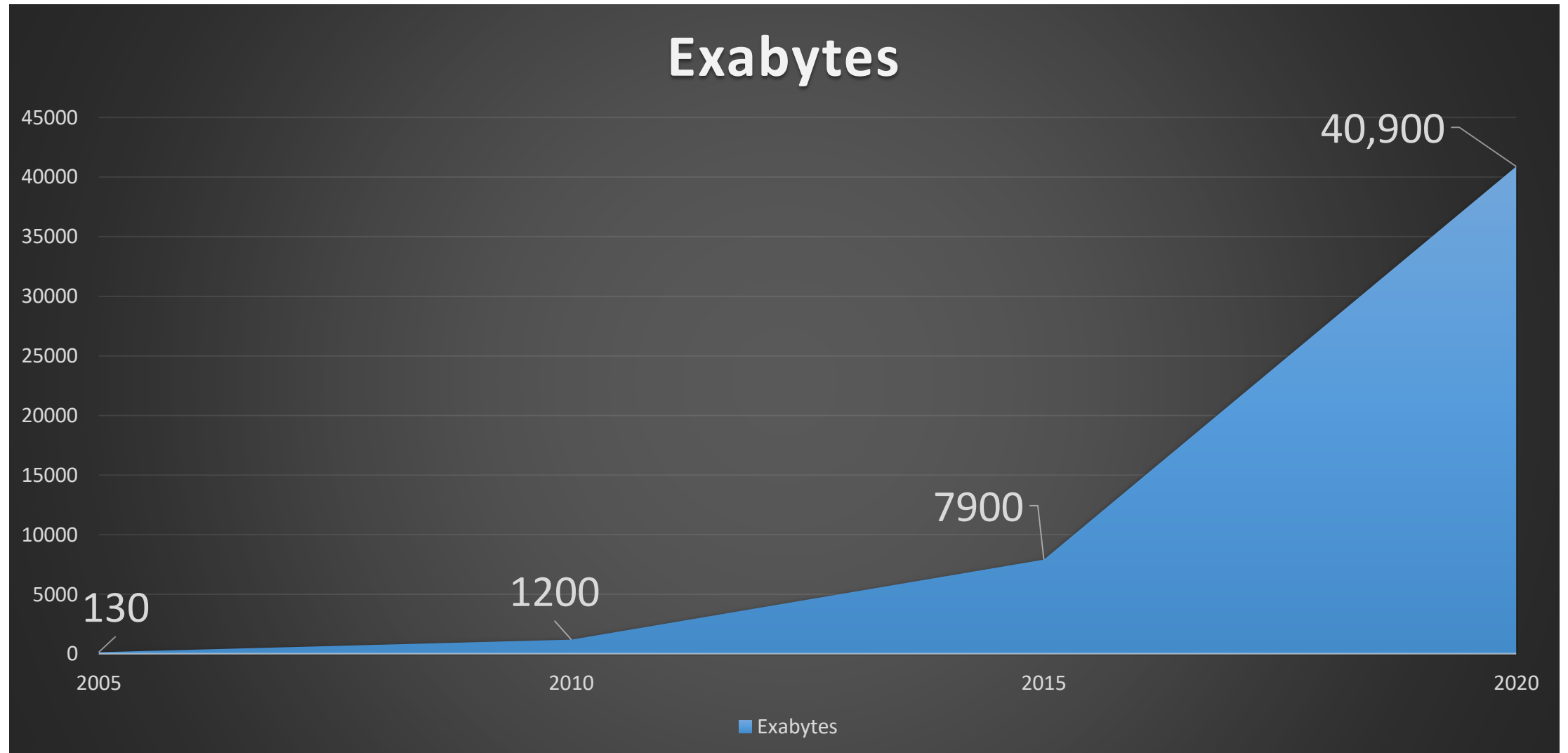
Gift This Course



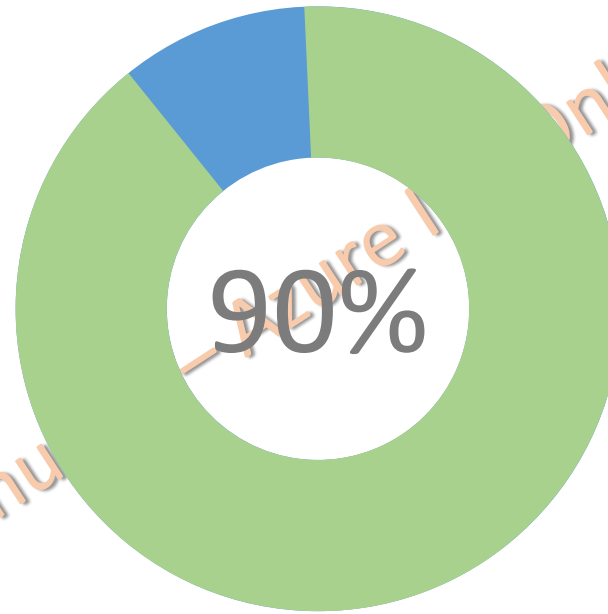
# Basics of Machine Learning

# Why Machine Learning is the Future?

# Data Growth



# Data Growth



90% of today's data has been created in last two years alone

# Benefits of Machine Learning

- Faster decisions
- Develop insights that are beyond human capabilities
- Act at the right time and take advantage of opportunities, converting them into closed deals.

# Why Azure ML?

# Why Azure ML?

- Drag and Drop interface and no Programming required
- Large variety of algorithm as modules
- From experiment to production API in minutes
- Supports R and Python to bring in your existing code
- Flexibility of data storage; supports variety of data storage options
- Large number of pre-built APIs available as a service

**TATA  
MOTORS**

**UBER**



**Rolls-Royce**



**XBOX**

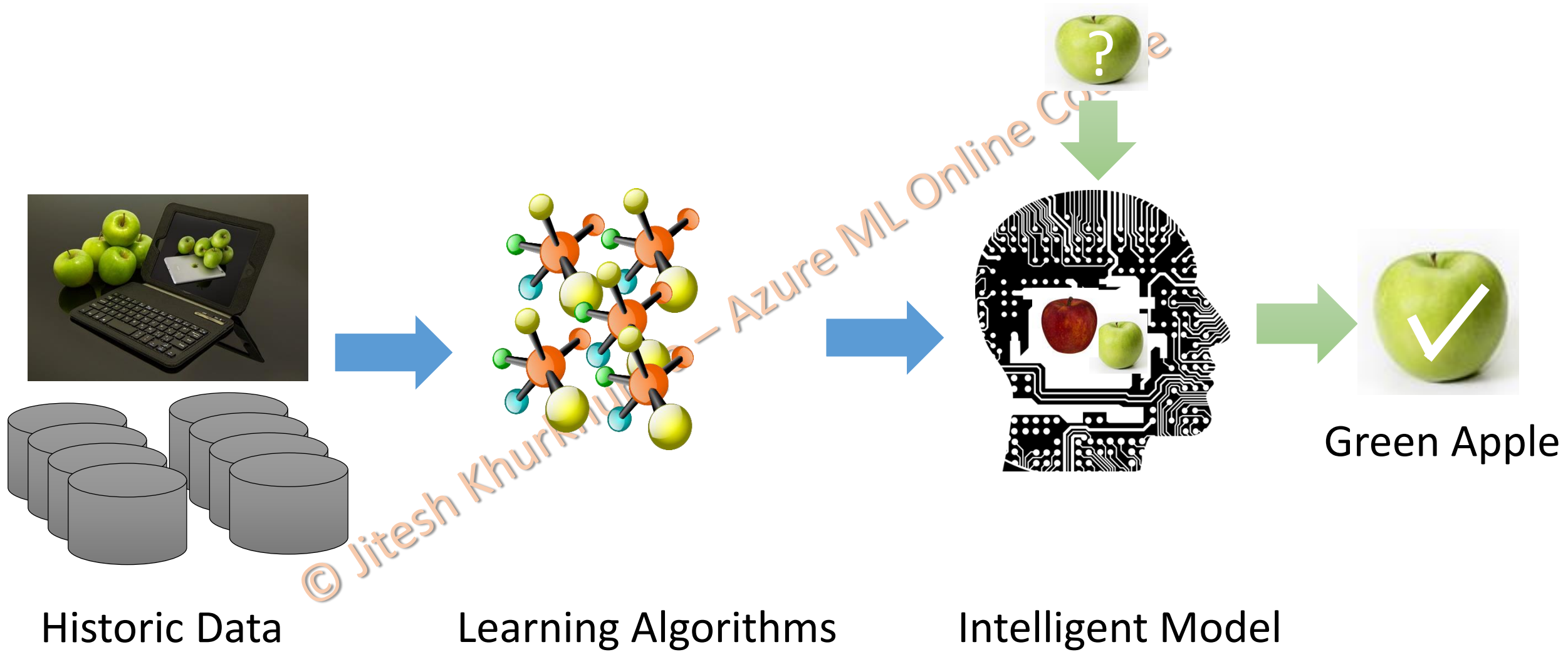


# What is Machine Learning?

# What Is Machine Learning?

- **Machine learning** is the subfield of computer science that gives computers the ability to learn without being explicitly programmed.
  - Arthur Samuel, 1959
- Extraction of knowledge from data
- Learns from past behaviour and make predictions or decisions

# How Machines Learn?



# Supervised, Unsupervised and Reinforcement Learning

# Supervised Machine Learning

- Data is labelled
- There is an Input variable “X” or set of input variables and an output variable “Y”

$$Y = f(X)$$

- The function is approximated to predict new values of Y given X
- Examples
  - **Regression** – Output variable is a real value such as Amount, Height, Weight etc
  - **Classification** – Output variable is a category, such as Yes, No, Red, Blue, Yellow etc

Loan_ID	Gender	Married	Dependents	Self_Employed	Income	LoanAmt	Term	CreditHistory	Property_Area	Status
LP001002	Male	No	0	No	\$5,849.00		60	1	Urban	Y
LP001003	Male	Yes	1	No	\$4,583.00	\$128.00	120	1	Rural	N
LP001005	Male	Yes	0	Yes	\$3,000.00	\$66.00	60	1	Urban	Y
LP001006	Male	Yes	2	No	\$2,583.00	\$120.00	60	1	Urban	Y

# Unsupervised Machine Learning

- Only X or input variable is known
- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.
- There is no correct answers and there is no teacher.
- Algorithms are left on their own to discover and present the interesting structure in the data.
- Examples
  - Clustering – Customer behaviour grouping
  - Association – Recommendation model



Customers who viewed this item also viewed these products



Dualit Food XL1500  
Processor

\$560

Add to cart



Kenwood kMix Manual  
Espresso Machine

★★★★★  
\$250

Select options



Weber One Touch Gold  
Premium Charcoal  
Grill-57cm

\$225

Add to cart



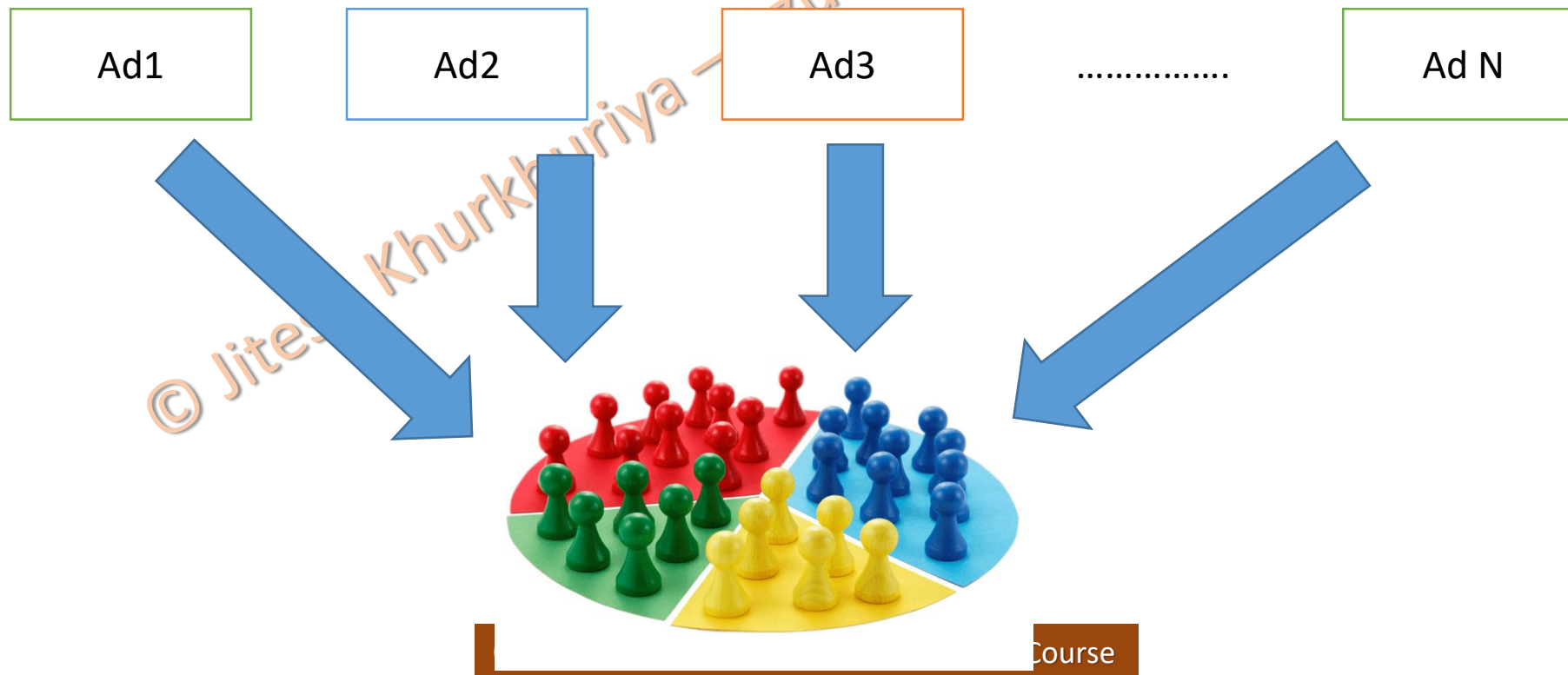
NoMU Salt Pepper and  
Spice Grinders

\$3

View options

# Reinforcement Learning

- Reinforcement learning rewards good behaviour and penalizes bad ones
- The idea is to maximise the gain or reward



# Understanding Data, Variables/Features



# Understanding The Variables Using a Dataset

Loan_ID	Gender	Married	Dependents	Self_Employed	Income	LoanAmt	Term	CreditHistory	Property_Area	Status
LP001002	Male	No	0	No	\$5,849.00		60	1	Urban	Y
LP001003	Male	Yes	1	No	\$4,583.00	\$128.00	120	1	Rural	N
LP001005	Male	Yes	0	Yes	\$3,000.00	\$66.00	60	1	Urban	Y
LP001006	Male	Yes	2	No	\$2,583.00	\$120.00	60	1	Urban	Y

## Types of Variables

- **Predictor/Independent**

- Gender
- Married
- Dependents
- Self\_Employed
- Income
- LoanAmt
- Term
- CreditHistory
- PropertyArea

- **Target/Dependent**

- Status

## Data Type

- **Character/String**

- Gender
- Married
- Self\_Employed
- Property\_Area
- Status

- **Numeric**

- Dependents
- Income
- LoanAmt
- Term
- CreditHistory

## Category

- **Categorical**

- Gender
- Married
- Self-Employed
- CreditHistory
- Property\_Area
- Status

- **Continuous**

- Dependents
- Income
- LoanAmt
- Term

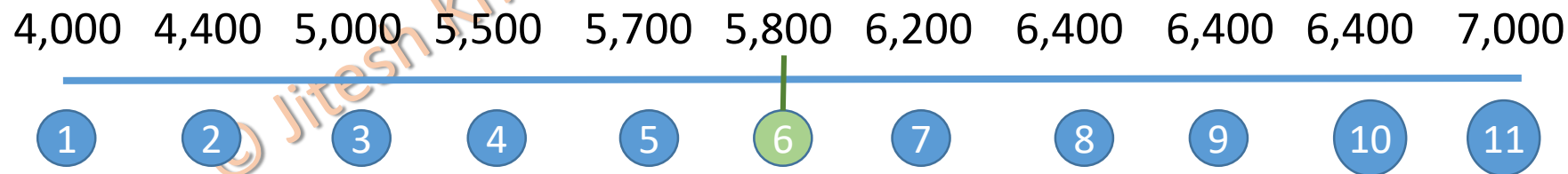
# Recap of Common Terms

# Mean and Median

- **Mean** – Average of all the values

$$\begin{aligned}\text{Mean} &= \text{Sum of Salary} / \text{Number of observations} \\ &= 62,800 / 11 \\ &= \$ 5709.09\end{aligned}$$

- **Median** – Numerical Middle value of the sorted observations with equal number of observations on both sides,



Salary
\$ 4,000
\$ 4,400
\$ 5,000
\$ 5,500
\$ 5,700
\$ 5,800
\$ 6,200
\$ 6,400
\$ 6,400
\$ 6,400
\$ 7,000
Sum
\$ 62,800

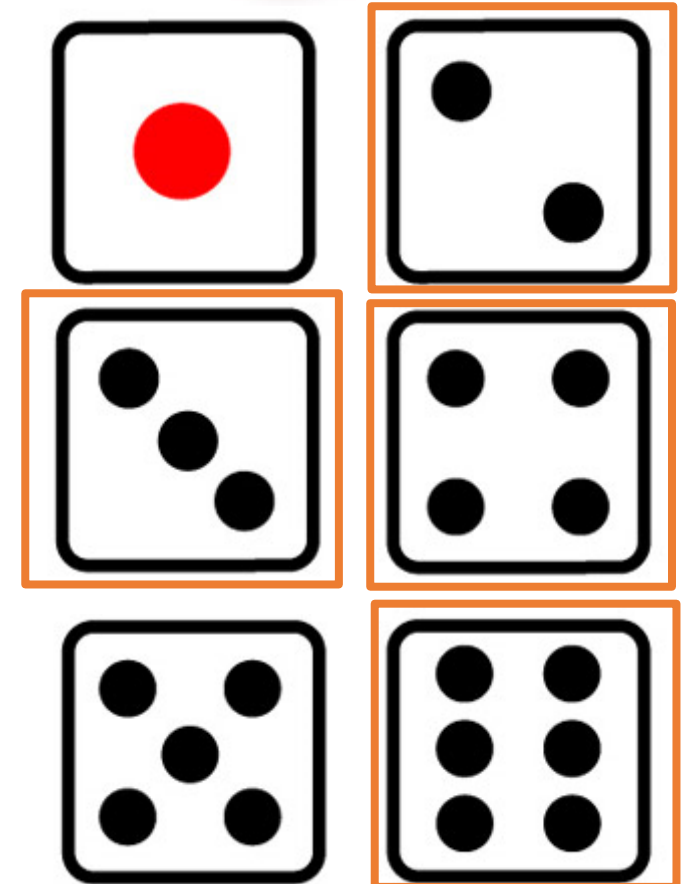
# Mode and Range

- **Mode** - The value that appears most often in a set of data  
**6,400**
- **Range** – The difference of highest and lowest values in a sample of observations  
**7000 – 4000 = 3,000**

Salary
\$ 4,000
\$ 4,400
\$ 5,000
\$ 5,500
\$ 5,700
\$ 5,800
\$ 6,200
\$ 6,400
\$ 6,400
\$ 6,400
\$ 7,000
Sum
\$ 62,800

# Probability

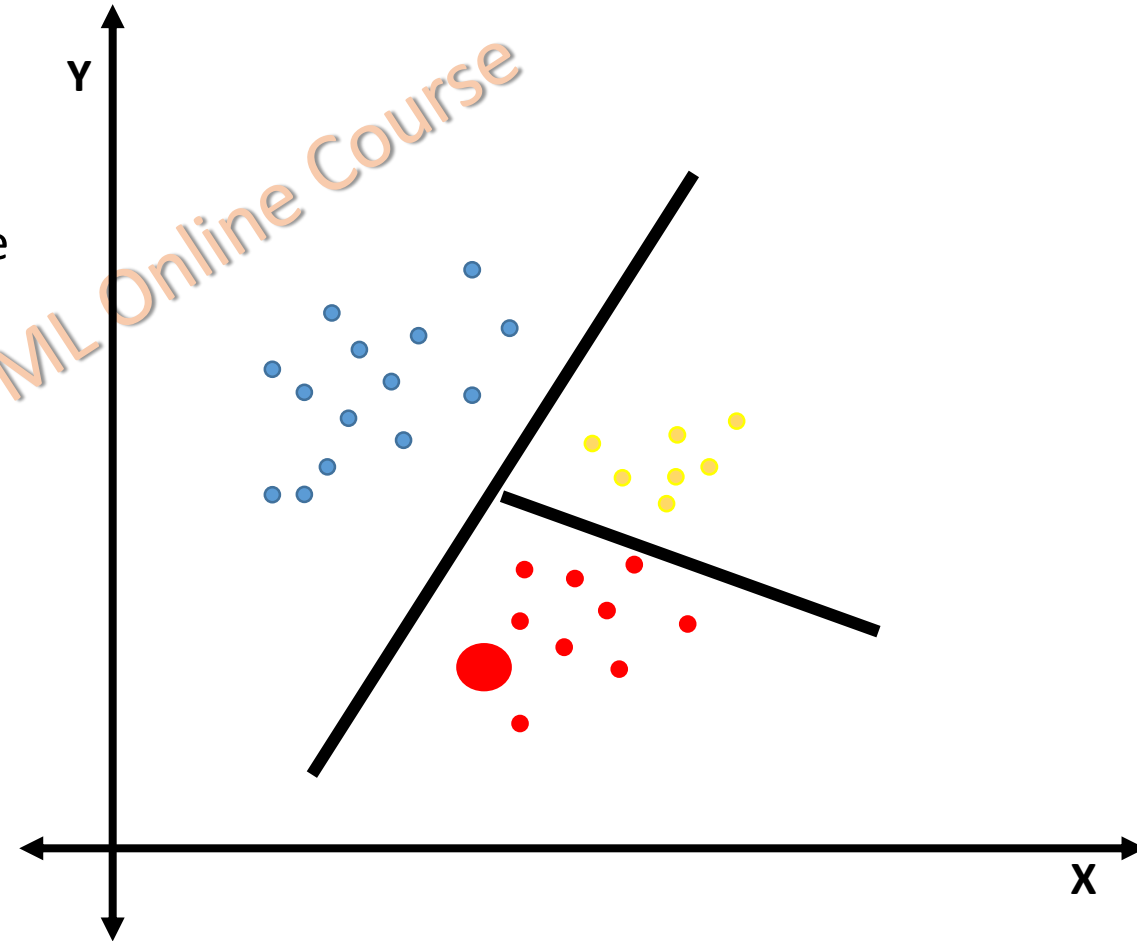
- Probability is a numerical way of describing how likely something is going to happen.
- Sample Space (S) – Set of possible outcomes that might be observed for an event
  - Dice Sample Space (S) = {1, 2, 3, 4, 5, 6}
- Probability of 3
  - $P(A) = 1/6 = 0.1667$
- Probability of getting an even number from the given sample space
- How many even numbers are there? 2,4,6
- So number of even occurrences = 3
- Probability of getting an even number is  $P(A) = 3/6 = 0.5$  or 50%



# Types of Models

# Classification

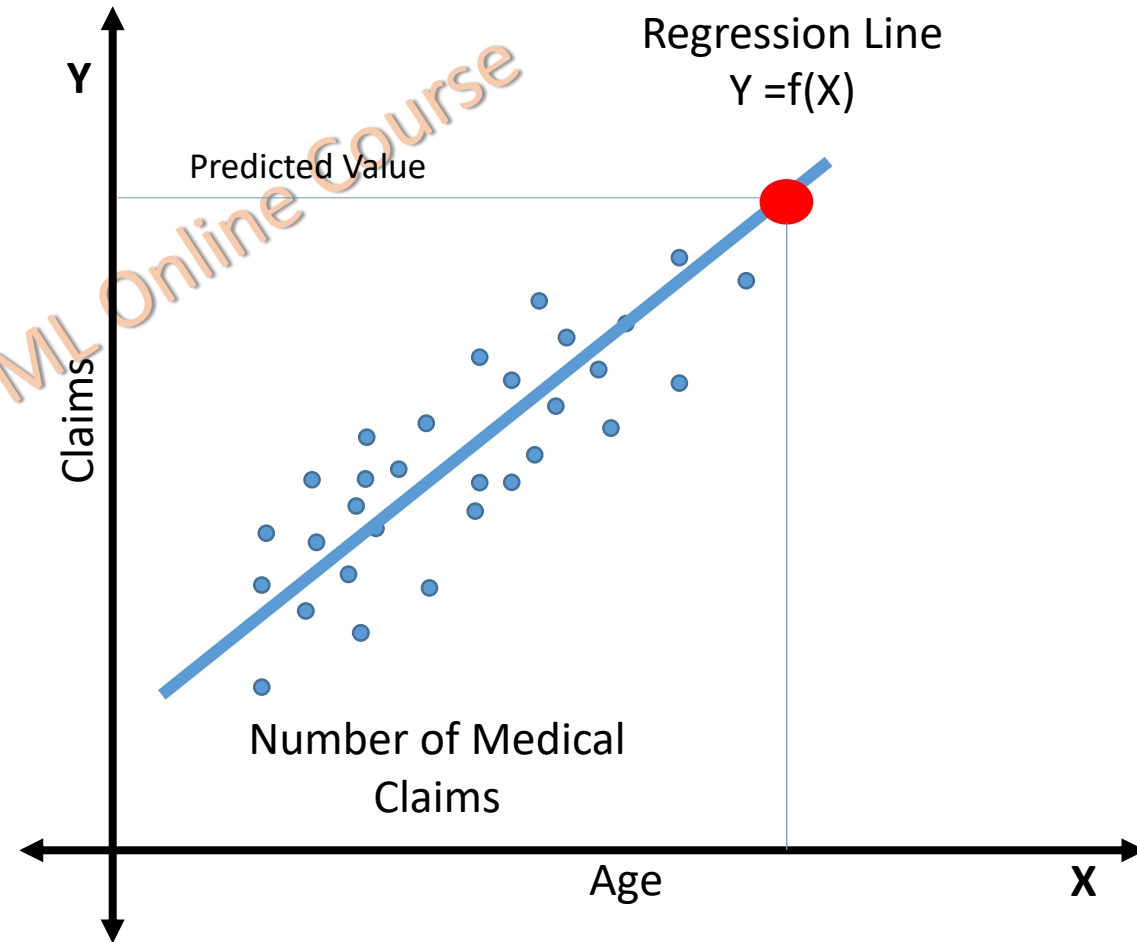
- Identification of category of data
- Binary/Two-Class Classification – Either/Or, Yes or No type
- Multi-Class Classification – One of the many alternatives
- Examples
  - Assigning a given email into "spam" or "non-spam" classes Or Primary, Social or Promotional emails
  - Will this customer default on loan repayment?
  - Will this customer buy my product?



Predicting the value for categorical variable.

# Regression Analysis

- Estimating the relationships among variables
- Predictor is a continuous variable
- Examples
  - Predicting the future sale of products
  - Computing fair price of the product or service
- One of the most common methods used in Machine Learning
- Infer causal relationships between dependent and independent variables.





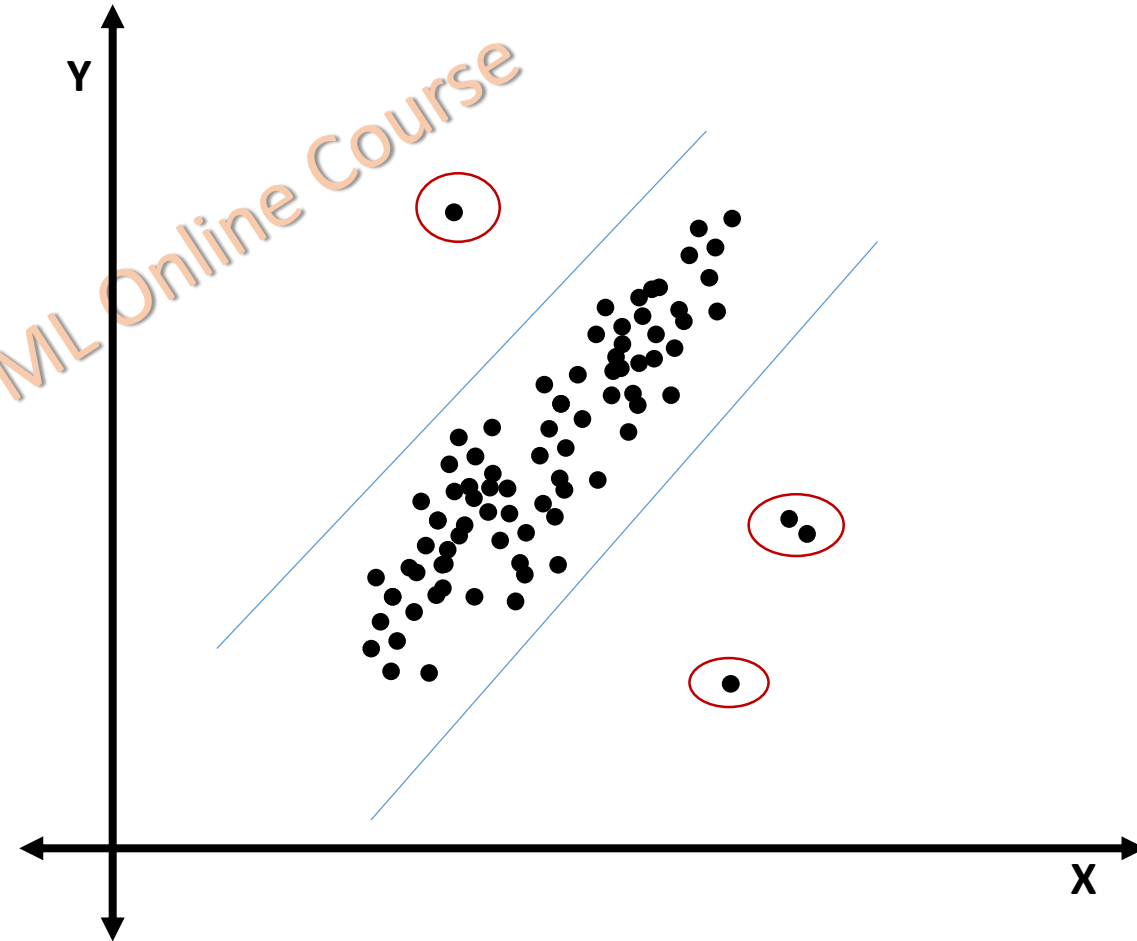
# Clustering or Cluster Analysis

- Clustering is the task of grouping a set of objects in such a way that
  - objects in the same group (called a cluster) are
  - more similar (in some sense or another)
  - to each other than
  - to those in other groups (clusters)
- Unsupervised Learning model
- Customers who make lot of long-distance calls and don't have a job. Who are they?



# Anomaly Detection

- Anomaly detection (also outlier detection) is the
- Identification of items, events or observations which
- Do not conform to an expected pattern or other items in a dataset.
- Typically the anomalous items will translate to some kind of problem such as
  - Bank fraud
  - Credit Card Fraud
  - Structural defect
  - Medical problems
- Anomalies are also referred to as outliers, novelties, noise, deviations and exceptions.



Thank You and Have a Great Time!