Probability for Data Science

Definition of Probability

Probability is the ***likelihood*** or ***chance*** of an **event** occuring.

Analytics Vidhya
Learn everything about analytics

Relevance of Probability

5-20%

Probability of selling to a **new prospect**

60-70%

Probability of selling to an **existing customer**

Relevance of Probability

- Improve Business Sense

Ex. Bank providing better services to the customers who are likely to churn.

Churn Prediction

Problem : Identify customer who will churn?

	gender	age	occupation	churn
0	Male	young	salaried	0
1	Male	young	self_employed	0
2	Male	old	self_employed	0
3	Male	young	self_employed	0
4	Female	young	salaried	1
5	Male	old	salaried	0
6	Female	young	self_employed	1
7	Male	young	self_employed	0
8	Male	young	salaried	1
9	Male	young	salaried	0
10	Male	young	self_employed	1
11	Female	young	self_employed	1
12	Male	young	retired	0
13	Female	young	self_employed	0
14	Male	old	self_employed	0

Random Experiment

- *Random experiment* is a process with a number of possible outcomes.




- Those outcomes are **not necessarily** certain

Random Experiment

- can be repeated numerous times under the same conditions
- Generally the outcome of an individual random experiment is independent and identically distributed

Random Experiment

Ex. The profession of a customer?

Self Employed

Salaried

Retired

Sample Space

Sample Space associated with a random experiment is a set of all possible outcomes

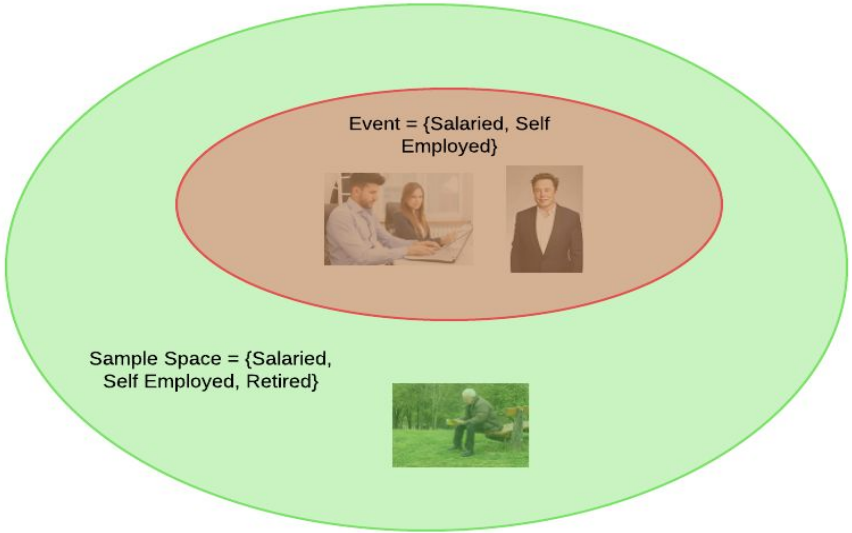
Ex. The profession of a customer?

Analytics Vidhya
Learn everything about analytics

Event

An event is a subset of Sample Space

Ex. If customer is currently working ?



Event

An event is a subset of Sample Space

Event 1: Getting 3 on a dice

Single Outcome

Event 2: Getting odd number on a dice

Multiple Outcomes

Probability

Probability is the ***likelihood*** or ***chance*** of an **event** occuring.

The probability of an Event E, is a number P, between 0 and 1 that measures the likelihood that Event will occur.

$$P(event) = \frac{count\ of\ outcomes\ in\ Event}{count\ of\ outcomes\ in\ Sample\ Space}$$

P = 0 → Impossible event

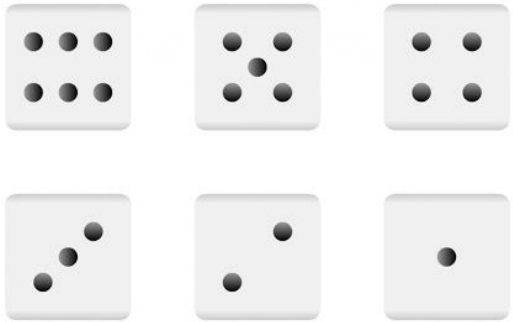
P = 1 → Certain event

Probability

P = 0 → Impossible Event

P = 1 → Certain Event

Ex. Getting 8 on rolling a six faced dice

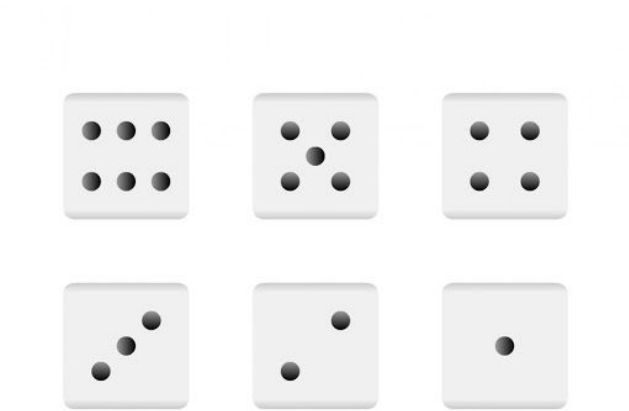


Probability

P = 0 → Impossible Event

P = 1 → Certain Event

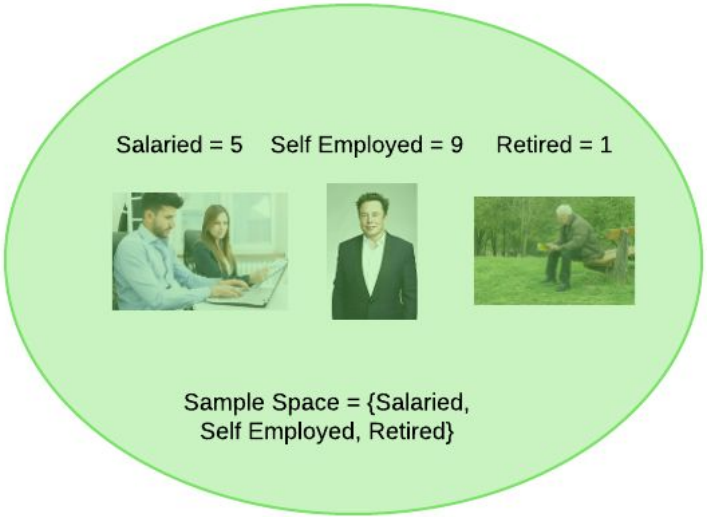
Ex. Getting a number less than 7, on rolling a six faced dice



Example: Probability of a Random Experiment

	gender	age	occupation	churn
0	Male	young	salaried	0
1	Male	young	self_employed	0
2	Male	old	self_employed	0
3	Male	young	self_employed	0
4	Female	young	salaried	1
5	Male	old	salaried	0
6	Female	young	self_employed	1
7	Male	young	self_employed	0
8	Male	young	salaried	1
9	Male	young	salaried	0
10	Male	young	self_employed	1
11	Female	young	self_employed	1
12	Male	young	retired	0
13	Female	young	self_employed	0
14	Male	old	self_employed	0

Ex. The profession of a customer?



Analytics Vidhya
Learn everything about analytics

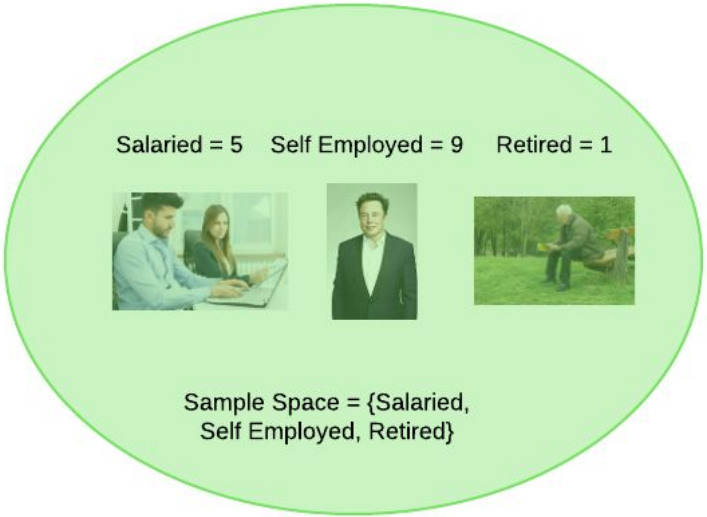
Example: Probability of a Random Experiment

Ex. The profession of a customer?

P(Customer = Salaried) = 5/15 = 0.333

P(Customer = Self Employed) = 9/15 = 0.6

P(Customer = Retired) = 1/15 = 0.067



Example: Definite and Impossible Outcome

Ex. If a salaried female customer is going to churn ?

Example: Definite and Impossible Outcome

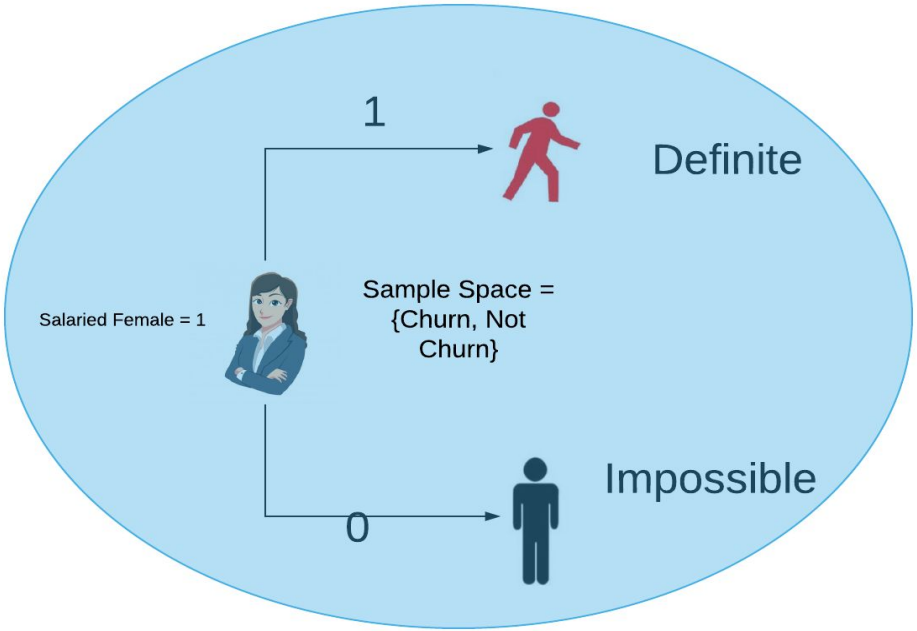
Ex. If a salaried female customer is going to churn ?

Churn = 1

Churn = 0

Example: Definite and Impossible Outcome

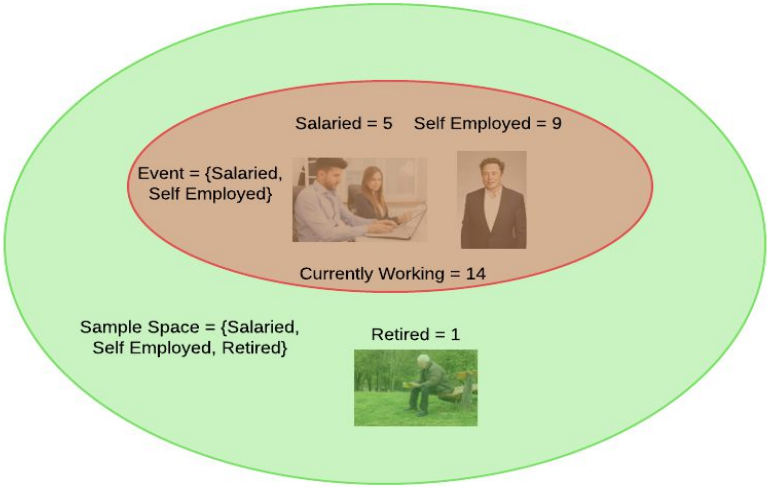
Ex. If a salaried female customer is going to churn ?



	gender	age	occupation	churn
0	Male	young	salaried	0
1	Male	young	self_employed	0
2	Male	old	self_employed	0
3	Male	young	self_employed	0
4	Female	young	salaried	1
5	Male	old	salaried	0
6	Female	young	self_employed	1
7	Male	young	self_employed	0
8	Male	young	salaried	1
9	Male	young	salaried	0
10	Male	young	self_employed	1
11	Female	young	self_employed	1
12	Male	young	retired	0
13	Female	young	self_employed	0
14	Male	old	self_employed	0

Example: Probability of an Event

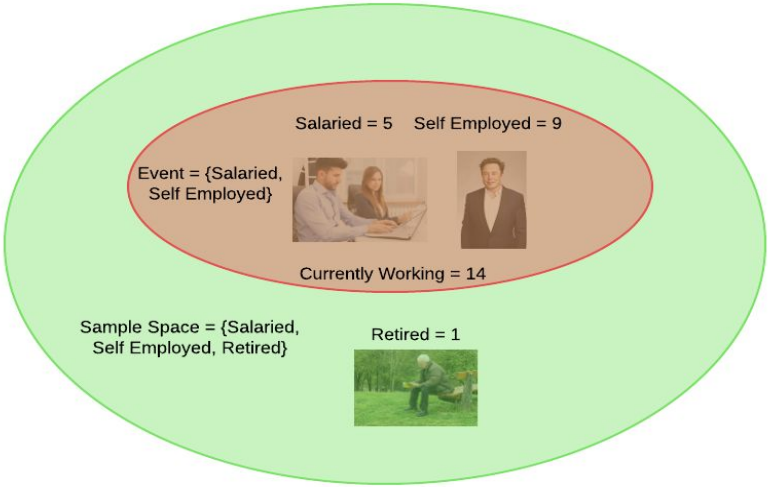
Ex. If a customer is currently working ?



	gender	age	occupation	churn
0	Male	young	salaried	0
1	Male	young	self_employed	0
2	Male	old	self_employed	0
3	Male	young	self_employed	0
4	Female	young	salaried	1
5	Male	old	salaried	0
6	Female	young	self_employed	1
7	Male	young	self_employed	0
8	Male	young	salaried	1
9	Male	young	salaried	0
10	Male	young	self_employed	1
11	Female	young	self_employed	1
12	Male	young	retired	0
13	Female	young	self_employed	0
14	Male	old	self_employed	0

Example: Probability of an Event

Ex. If a customer is currently working ?



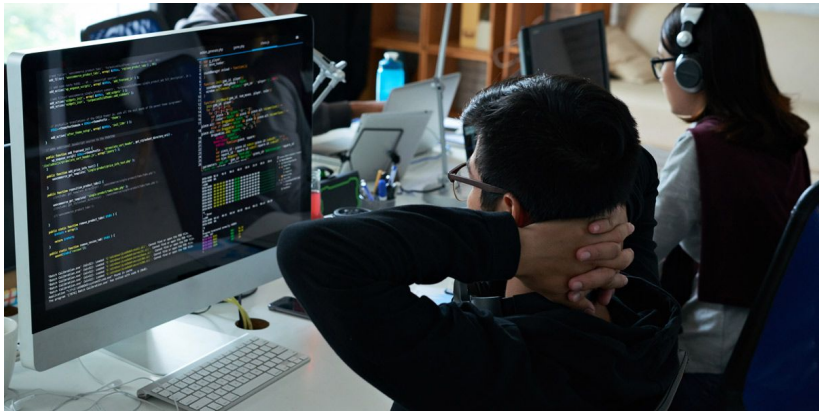
P(customer = working) = 14/15 = 0.933

	gender	age	occupation	churn
0	Male	young	salaried	0
1	Male	young	self_employed	0
2	Male	old	self_employed	0
3	Male	young	self_employed	0
4	Female	young	salaried	1
5	Male	old	salaried	0
6	Female	young	self_employed	1
7	Male	young	self_employed	0
8	Male	young	salaried	1
9	Male	young	salaried	0
10	Male	young	self_employed	1
11	Female	young	self_employed	1
12	Male	young	retired	0
13	Female	young	self_employed	0
14	Male	old	self_employed	0

Thank You!

Relevance of Probability

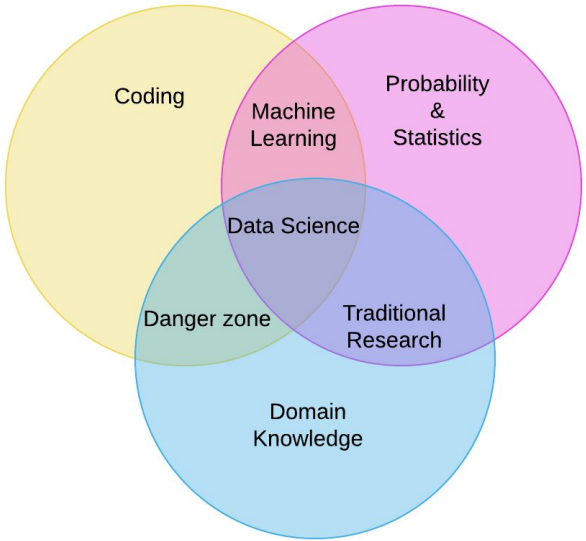
A Data Scientist without the knowledge of Probability and Statistics is like a Pilot without the knowledge of Aerodynamics



Analytics Vidhya
Learn everything about analytics

Relevance of Probability

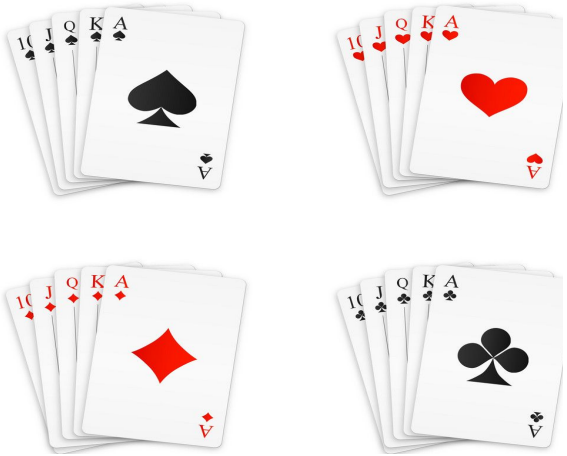
- Foundational language of Data Science

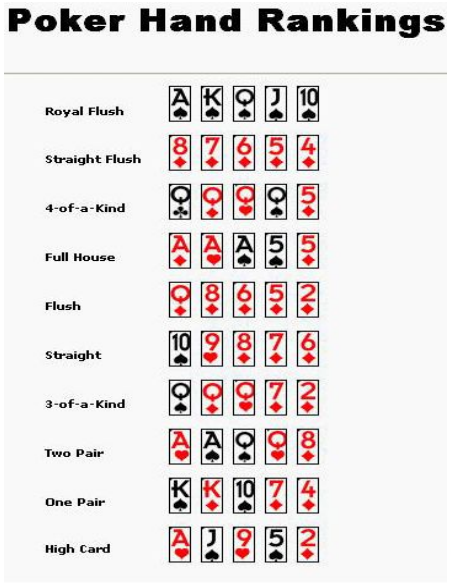


Probability

The probability of an outcome O is a number P, between 0 and 1 that measures the likelihood that O will occur.

P = 1 → Definite outcome
Ex. Winning a pot with royal flush





Analytics Vidhya
Learn everything about analytics