

## Part 1: Probability and Statistics

1. Probability is a stat which tells about the chances of a particular occurrence, helping to make predictions for a particular outcome. 0 Probability of something means that an event will not occur, that is it is not possible. Probability of 0.5 means that an event will happen once every twice turns (equally likely to occur and not occur). Probability of 1 signifies that an event will occur always in each and every single turn.
2.  $\frac{1}{6}$ (every once out of 6 times).  
Favorable Outcomes=1{3}. Possible Outcomes=6{1,2,3,4,5,6}.
3. Mean, Median and Mode.
4. To make analysis of the data given and make inferences on the basis of these central tendencies. It describes the essential features of a dataset.
5. Range is the interval within which the given data lies. It is the difference of the highest value of the data and the lowest value of the data. In the given example, the range is 40 (100-60).
6. The unit of variance is the original unit<sup>2</sup>. It tells the degree of spread in the given data (average of the squared differences from the mean). The unit of standard deviation is the original unit itself. It indicates the average or typical distance of the data points from the mean.
7. Real-world data is often noisy and uncertain. Machine

Learning models can be used to give outputs from the probabilities of the occurrence, hence making it more accurate. One example of this is the spam detection and filter found in mails.

8. Median is the central value of the dataset, whereas mean is the arithmetic average, or sum of values/number of values. In a skewed data, mean is often pulled towards the higher values, whereas mode always remains centrally aligned. Mean is not useful for data with outliers, where the use of medians is best, like incomes of the employees in a company.
9. Data Exploration refers to the understanding of the data patterns between the elements. It also refers to the analysis of their distribution and relations.
10. It checks the quality and validity of data and rigor of methods, with its collection and analysis. It directly impacts the decisions regarding medical approval.
11. It might be due to the fact that the prices are spread out and are not close to each other. There is a huge difference between the mean and other values of the data set. It would make me think that the values are far apart and the mean lies as it is almost at the center of the dataset, and the house prices are spread apart.
12. It tells us how the two axes are related to each other, and they might have an increasing change, as the graph is almost a straight line. The value is adjusted or change with reference to the gene, whereas fold value is telling the frequency in this plot. Up regulation

is showing the increase in activity, whereas down regulation is showing a decreased activity.

## **Part 2: Machine Learning Fundamentals**

13. The ability of computers to learn without being explicitly programmed.
14. Supervised Learning, Unsupervised Learning, Reinforcement Learning
15. Classification involves predicting a category or class, whose output will always be discrete. An example of this is spam classification. On the other hand, regression involves predicting a number, whose output is a number. An example of this is finding the price of a house.
16. The aim of unsupervised learning is identifying and finding the relations, structures and patterns in the data.
17. PCA stands for Principal Component Analysis. It is used to simplify data and speed up learning, and to help with visualization as well; mainly to preserve important information.
18. In traditional programming, the input provided is data with the set of rules in the form of a program; and the output received is the result required. On the other hand Machine learning involves data and result as input in the form of examples and gives the program as an output in the form of models and rules.

19. A machine learning algorithm is given lots of inputs in order to recognize patterns between them in order to build a model. This model is similar to identifying new, unseen cats; as in when a child learns to recognize them by seeing their pictures
20. An agent is the Machine Learning model in this context, like a robot or a recommendation system. It learns by taking actions in the environment to maximize its rewards and reduce its punishments.
21. Linear Regression and Logistic regression for Supervised Learning. Principal Componental Analysis for Unsupervised Learning.
22. The outliers and presence of null-values might alter the data, thus presenting inaccurate predictions and classifications. Right input features are required to present the correct output which is required.
23. False Positives as it has predicted a something which is not actually present.. It is because important emails might get missed, leading to inconvenience in the real world.

### **Part 3: Artificial Intelligence Concepts**

24. A broad field encompassing a list of tasks done by a computer typically requiring human intelligence, like reasoning, problem solving, learning and perception.

25. AI is a broad field which includes machine learning and deep learning which itself is inside machine learning.
26. Artificial Narrow Intelligence(Current)  
Artificial General Intelligence  
Artificial Superintelligence
27. Robotics and Knowledge Representation and Reasoning.
28. The system thinks like human as in the case of cognitive modelling. (GPS Tracking). On the other hand, the system tries to achieve the best possible outcome, as is dominant today.
29. It enables computers to understand and generate human language, as in Machine and Google Translations.
30. Generative AI can generate human-like text, images, code, translate and do more like answering questions. On the other hand, AI used in data analysis only checks for the data and makes predictions and algorithms based on these examples, like ChatGPT and LLMs.
31. AI can learn biases in the data, which can lead to unfair outcomes, as in the case with hiring, loan applications and face recognition. These might be due to the presence of large number of outliers, and the presence of null or impossible values.

32. It is to avoid any discrepancies due to error generated by it in some cases, which might affect the world globally in these critical areas, and to learn about its accountability as well.