

## **Week 2: Ashoka Horizons Achievers Program Faculty Lecture Homework.**

### Part 1: Probability and Statistics

#### *Easy*

1. Probability basically signifies the likelihood of some event to occur. A probability of 0 means that event is impossible whereas the probability of 1 suggests that the event will definitely occur. A probability of 0.5 means that there is equal likelihood of the event happening or not happening.
2. Total possible outcomes: 6, Favourable outcomes: 1, Therefore probability:  $1/6$
3. Mean, Median and Mode
4. The purpose of descriptive statistics is to show us the main features of our dataset.
5. Range is the maximum value – minimum value. For the given scores, the range shall be  $100 - 60 = 40$ .
6. Variance is the average of squared differences from the mean, whilst the standard deviation is the square root of variance. Variance is in squared units whereas standard deviation is in the original units.

#### *Medium*

7. ML models use probability to quantify the likelihood of something occurring. In the slides, the example of a model quantifying “80% chance this email is spam” is one such example of a ML model using probability.
8. We would probably use median over mean when the dataset has anomalies/outliers. For example, in a company, when giving an average salary, the mean may be an inaccurate representation due to extremely high salaries of CEOs or larger postholders. Therefore, a median salary average would be preferred as it portrays the entire dataset while being unaffected by a few extreme values.
9. Data exploration basically helps us understand patterns, trends or relationships in data before we model the data. Thus, we can include better features in the model and gain better insights.
10. The Friedrich’s Ataxia Case Study displayed how a combination of good quality data with clever ML methods can help in identifying certain patterns (biomarkers) in diseases. Not only does this improve the diagnosis standard, but it also aids in improving the treatment techniques.

#### *Hard*

11. House prices vary a lot, depending on size, location and the interiors – like the difference between apartments or mansions. If there is such a large standard

deviation, that means that the mean average on its own is highly misleading, as it doesn't accurately reflect the true house price paid by the large majority of people.

12. The Volcano Plot gives a visual representation of the gene changes. "Up regulated" signifies increased activity, whereas "Down regulation" signifies decreased activity.

## Part 2: Machine Learning Fundamentals

### *Easy*

13. In 1959, Arthur Samuels defined Machine Learning as "A field of study that gives computers the ability to learn without being explicitly programmed."
14. Supervised Learning, Unsupervised Learning, Reinforcement Learning
15. Classification predicts categories as a discrete output whereas Regression predicts continuous values as a numerical output. An example of Classification is whether a tumor is Malignant/Benign. An example of Regression is predicting the price of a house.
16. The main goal of unsupervised learning is to discover hidden structures, patterns or relationships in a particular data set.
17. PCA stands for Principal Component Analysis and its primary purpose is to reduce data dimensions while preserving the important information.

### *Medium*

18. In traditional programming, data + rules are inputted, and results are outputted. However, in Machine Learning, data + results are inputted, and a model(the rules) is outputted.
19. Like a child learns about cats by viewing many cat pictures, ML algorithms learn from labelled data example to identify new ones.
20. An agent interacts with an environment, learns through trial and error and makes decision to maximise a reward.
21. Supervised: Logistics Regression, Decisions Trees + Unsupervised: K-Means, PCA

### *Hard*

22. Bad, inaccurate data results in a confused model which produces mediocre results. If these steps are performed incorrectly, the model may learn incorrect patterns or may not learn the patterns at all.
23. This is a False Positive. It is a bad result because you may miss important messages.

## Part 3: Artificial Intelligence Concepts

### *Easy*

- 24. Artificial Intelligence is programming computers to perform tasks that would normally require human intelligence.
- 25. Artificial Intelligence (AI) is the broad field, Machine Learning (ML) is a subset of AI and Deep Learning (DL) is a subset of ML.
- 26. Artificial Narrow Intelligence (today's level), Artificial General Intelligence, Artificial Superintelligence.
- 27. Natural Language Processing (NLP) and Robotics.

#### *Medium*

- 28. "Thinking Humanly" is to emulate human thought processes, whereas "Acting Rationally" is to logically make the best possible decision, regardless of human thought.
- 29. NLP is the way in which computers understand and using human language. Examples would be chatbots like Siri or Google Translate.
- 30. Generative AI can create new, original content, that didn't exist earlier – such as text or images. For example, ChatGPT can write its own code or generate an original image.

#### *Hard*

- 31. AI may reinforce or perpetuate certain prejudices if it is trained on biased data. For example, a predictive policing AI could reinforce racial biases by sending excess patrols to regions that were previously over-policed. This causes some communities to be unfairly targeted, over and over again.
- 32. When AI is used in sensitive areas like healthcare – where literal lives are at stake – it is essential to understand what is going on under the hood. For example, we need to completely and thoroughly understand the AI's reasoning behind suggesting a patient diagnosis before trusting it. Otherwise, it may result in risky and irresponsible outcomes.