Assignment #2 [HW] – Applied Data Science with ML and AI

Name: Yuvraj Khanna

Course: Ashoka Horizons, Summer 2025

Instructor: Rintu Kutum, Gautam Ahuja

## 1. Probability

Probability is the measure of how likely an event is to occur. A probability of 0 means the event is impossible, 0.5 implies a 50% chance, and 1 means certainty.

## 2. Rolling a 3

A six-sided die has 6 outcomes. Rolling a 3 is one favorable outcome. So, probability = 1/6.

## 3. Measures of Central Tendency

Mean, Median, and Mode.

## 4. Purpose of Descriptive Statistics

To summarize, describe, and understand the key features of a dataset.

## 5. Range Definition

Range = Max - Min. For [60, 70, 80, 90, 100], Range = 100 - 60 = 40.

## 6. Variance vs Std Deviation

Variance is the average squared deviation from the mean; SD is the square root of variance, in original units for easier interpretation.

## 7. Probability in ML

ML models often predict probabilities (e.g., email being spam). Understanding these is key to interpreting output and risk.

## 8. Median over Mean

When the data is skewed or has outliers (e.g., housing prices), the median better represents the center.

## 9. Data Exploration

It means examining data to find patterns, trends, anomalies, and guide preprocessing before modeling.

## 10. FRDA Case Study

It showed that data and methods together enable insights — without both, critical patterns like gene-disease links would be missed.

## 11. Large Std Deviation

It suggests prices vary widely, so the mean might not reflect typical values. This could mislead interpretations.

## 12. Volcano Plot

It visualizes which genes are significantly and strongly affected. Up-regulated = higher activity, Down-regulated = lower activity.

## 13. Definition of ML

Arthur Samuel: ML is the field of study that gives computers the ability to learn without being explicitly programmed.

## 14. Three Types of ML

Supervised, Unsupervised, and Reinforcement Learning.

## 15. Classification vs Regression

Classification predicts categories (e.g., spam detection), regression predicts numbers (e.g., house prices).

## 16. Goal of Unsupervised Learning

To find hidden patterns or groupings in data without labels.

## 17. PCA

Principal Component Analysis reduces data dimensionality while preserving variance.

## 18. Traditional vs ML

Traditional: Rules + Data → Output. ML: Data + Output → learns Rules.

## 19. Learning from Examples

ML learns by observing many examples. For cats, it learns by seeing many labeled images.

## 20. Agent in RL

An agent interacts with an environment and learns via rewards and punishments.

## 21. Common Algorithms

Supervised: Linear Regression, Decision Trees; Unsupervised: K-Means.

## 22. Preprocessing Importance

Without proper preprocessing, models can learn from noise, leading to poor performance and bias.

## 23. Spam Filter Error

Marking real mail as spam is a False Positive. It’s bad because it hides important messages.

## 24. Definition of AI

AI is the simulation of human intelligence in machines.

## 25. AI vs ML vs DL

AI is the broadest, ML is a subset of AI, and DL is a subset of ML.

## 26. Three Types of AI

Narrow AI, General AI, and Super AI. Today we have only Narrow AI.

## 27. AI Foundations

Logic and Search; Probability and Statistics.

## 28. AI Thinking vs Acting

Thinking humanly = mimicking thought; Acting rationally = optimizing outcomes.

## 29. NLP

Natural Language Processing helps machines understand human language. Example: Chatbots.

## 30. Generative AI

It creates new content (text, images). Unlike analytics AI, it generates original outputs.

## 31. Bias in AI

AI learns bias from biased data. E.g., facial recognition performing poorly on minorities.

## 32. Explainability in AI

Understanding AI decisions builds trust, especially in healthcare, where wrong predictions can cost lives.