ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING TRAINING TR-102 REPORT DAY 9 3 JULY 2025

Overview:

The ninth day of training focused on statistical concepts that are essential in data analysis and Machine Learning — including Mean, Median, Mode, Standard Deviation, Percentile, and Scatter Plot visualization.We learned how to calculate these measures using both manual methods and Python libraries, and how they help in understanding data distributions.

The session also included hands-on practice using NumPy and Matplotlib to visualize data through scatter plots.This day built the foundation for analyzing datasets and preparing data for AI model training.

Learning Objectives:

- Understand and compute measures of central tendency — mean, median, and mode.
- Learn about measures of dispersion — standard deviation and percentiles.
- Apply these concepts using Python libraries like NumPy and statistics.
- Visualize data distribution using scatter plots.
- Interpret the statistical insights in relation to Machine Learning applications

Mean (Average):

The Mean is the average of all numerical values in a dataset. It provides a central value around which the data points are distributed.

Example in Python:

```
import numpy as np

data = [10, 20, 30, 40, 50]

mean_value = np.mean(data)
```

```
print("Mean:", mean_value)
```

Median:

The Median is the middle value when data is arranged in ascending or descending order. It divides the dataset into two equal halves.

Example in Python:

```
data = [5, 10, 15, 20, 25]
import numpy as np
median_value = np.median(data)
print("Median:", median_value)
```

Mode:

The Mode is the value that appears most frequently in a dataset.

Example in Python:

```
from statistics import mode

data = [10, 20, 20, 30, 30, 30, 40]

mode_value = mode(data)

print("Mode:", mode_value)
```

Standard Deviation:

The Standard Deviation (σ) measures how spread out the data values are around the mean. A low standard deviation indicates that the data points are close to the mean, while a high standard deviation shows greater variability.

Example in Python:

```
import numpy as np

data = [10, 12, 23, 23, 16, 23, 21, 16]
```

```
std_dev = np.std(data)

print("Standard Deviation:", std_dev)
```

Percentile:

A Percentile indicates the value below which a given percentage of data points fall. For example, the 50th percentile is the median.

Example in Python:

```
import numpy as np
data = [15, 20, 35, 40, 50]
percentile_25 = np.percentile(data, 25)
percentile_75 = np.percentile(data, 75)
print("25th Percentile:", percentile_25)
print("75th Percentile:", percentile_75)
```

Scatter Plot:

A **Scatter Plot** is a visual representation of the relationship between two numerical variables. It helps identify **patterns, correlations, and trends** in data.

Example in Python:

```
import matplotlib.pyplot as plt

x = [10, 20, 30, 40, 50]
y = [15, 25, 35, 30, 45]

plt.scatter(x, y)
plt.title("Scatter Plot Example")
plt.xlabel("X Values")
plt.ylabel("Y Values")
plt.show()
```

Applications in AI and Machine Learning:

- Mean, Median, and Mode help summarize datasets before model training.
- Standard Deviation indicates how consistent the data is — useful in normalization.
- Percentiles help detect outliers in datasets.
- Scatter Plots are used for visualizing data distributions and relationships between features.

Conclusion:

Day 9 helped us develop a solid grasp of descriptive statistics and data visualization.
We learned to compute and interpret mean, median, mode, standard deviation, percentiles, and scatter plots, which are essential tools for any data scientist or AI developer.Understanding these statistical measures enhances our ability to analyze datasets accurately and prepare them effectively for AI and Machine Learning models.This session strengthened the analytical foundation needed to transition from data understanding to data modeling in upcoming training modules.