Assignment Title: Statistics & Probability

Project Title: Fake News Detection

1. Reading Summary:

This week, I studied the concepts of Statistics and Probability which are essential for understanding data relationships.

I learned how to calculate mean, median, mode, variance, and correlation, and how these statistical measures help describe data behavior and relationships between variables.

I also understood how correlation helps identify the strength and direction of relationships between numerical features and the target variable.

2. Task Performed:

- Imported and explored the **Fake News dataset (Fake.csv)**.

- Added a **target column ('label')** to distinguish between fake (1) and real (0) news.

- Created numeric features such as **text length, title length, and word count**.

- Calculated **basic statistical measures** – mean, median, mode, variance.

- Performed **correlation analysis** to find how strongly features relate to the target variable.

- Visualized correlations using a **heatmap (Seaborn)** for better understanding.

3.Calculations and Analysis:

- Calculated mean, median, and mode for numeric columns (text_length, title_length, word_count).

- Found variance to measure the spread of data.

- Generated a correlation matrix and compared all numeric features with the target variable label.

4. Learning Outcome:

From this task, I learned :

- How to calculate and interpret **basic statistics** and **correlations**.

- How to identify which features are most related to the target variable.

- How **heatmaps** visually represent feature relationships.

- Why correlation is important before applying machine learning algorithms.

5. Challenges Faced:

- Initially, faced a **KeyError** due to missing label column.

- Needed to create a proper target variable to perform correlation.

- Correlation values appeared small because of limited numeric features in text dataset.

- Adjusted heatmap color schemes and size for better readability in Google Colab.

6. GitHub Repository Link:

GitHub repository link here:
https://github.com/AlmasMalik66/DataScience-AI-Assignment