A white robot hand with red and blue stripes

Description automatically generated

**Module-5**

**AI & ML**



**LAB MANUAL**

**Disclaimer: The content is curated from online/offline resources and used for educational purpose only**

**Computing and Visualizing Skewness and Kurtosis for Literacy Rate Distribution Across Indian States.**

**Objective:**

* Compute **skewness** and **kurtosis** for literacy rate distribution across Indian states.
* Interpret **whether the distribution is symmetric, left-skewed, or right-skewed**.
* Understand if the literacy rate distribution has **heavy tails (outliers)** or follows a normal pattern.
* Visualize the distribution using a **histogram with a density curve**.

**Duration** 2 Hrs

**Problem Statement:**

Literacy rate is a crucial socio-economic indicator. Understanding its distribution across Indian states can help policymakers address educational disparities. **Skewness** measures whether the data is **symmetrically distributed**, while **kurtosis** determines if the data has **heavy or light tails** compared to a normal distribution. In this activity, we will compute and visualize **skewness and kurtosis** of literacy rates across different Indian states.

**Requirement:**

Python with libraries**:**

* **NumPy** (for numerical operations)
* **Pandas** (for data handling)
* **SciPy** (for statistical computations)
* **Matplotlib & Seaborn** (for visualization)

**Dataset:** <https://www.kaggle.com/datasets/doncorleone92/govt-of-india-literacy-rate>

**Procedure:**

**Step 1: Install Required Libraries**

**Step 2: Import Necessary Libraries**

**A screen shot of a computer

AI-generated content may be incorrect.**

**Step 3: Load the Literacy Rate Dataset**

A screenshot of a computer

AI-generated content may be incorrect.

**Step 4: Extract literacy rate from the dataset.**

**A black screen with white text

AI-generated content may be incorrect.**

**Step 5: Compute Skewness**

A screen shot of a computer code

AI-generated content may be incorrect.

**Skewness Explanation:**

* skew () calculates **skewness**, which measures **asymmetry** in data distribution.
* **Interpretation of Skewness:**
  + **Skewness = 0** → Perfectly symmetrical distribution.
  + **Skewness > 0** → Right-skewed (more low values).
  + **Skewness < 0** → Left-skewed (more high values).

**Kurtosis Explanation:**

* kurtosis () computes **kurtosis**, which indicates whether the distribution has heavy or light tails.
* **Interpretation of Kurtosis:**
  + **Kurtosis ≈ 0** → Normal distribution.
  + **Kurtosis > 0** → Leptokurtic (heavy tails, more outliers).
  + **Kurtosis < 0** → Platykurtic (light tails, fewer outliers).

**Step 6: Visualizing the Distribution and print the result.**

**A screen shot of a computer program

AI-generated content may be incorrect.**

A screen shot of a computer code

AI-generated content may be incorrect.

**Explanation:**

* **sns.histplot()** plots a **histogram** to visualize literacy rate distribution.
* **kde=True** overlays a smooth density curve.
* **Red dashed line** represents the **mean literacy rate**.

A graph with a blue line

AI-generated content may be incorrect.

**Conclusion**

Analysing the skewness and kurtosis of literacy rates helps us understand the shape and distribution of literacy levels across Indian states. Skewness reveals whether literacy rates are symmetrically distributed or skewed toward higher or lower values. Kurtosis measures how peaked or flat the distribution is. A high positive kurtosis suggests that most states have literacy rates clustered around the mean, while negative kurtosis indicates a wider spread. This statistical evaluation assists education policymakers in identifying regions requiring more attention and resource allocation.