Hidden Markov Model: Student Mood Analysis

Task Description

Given Dataset:

The dataset contains observations of 5 students over 20 days each:

- Each day records:
  - Student's mood (Hidden state): Happy(H) or Sad(S)
  - Color of shirt worn (Observation): Red(R), Green(G), Blue(B)
- Data Format: StudentID, Day, Mood, ShirtColor

Part 1: Parameter Learning

Calculate following matrices:

- Initial Probability Distribution:
  - $P(M_1)$ = Count(M_1 on first day) / Total students where $M_1 \in \{H,S\}$
- Transition Matrix:
  - $P(M_2|M_1)$ = Count($M_1 \rightarrow M_2$) / Count(M_1) where $M_1, M_2 \in \{H,S\}$
- Emission Matrix:
  - $P(C|M)$ = Count(C,M) / Count(M) where $M \in \{H,S\}$, $C \in \{R,G,B\}$

Part 2: Testing Model

Given: New sequence of observed shirt colors: R, B, G

Find: Most likely sequence of moods (M_1,M_2,M_3) by calculating:

$$P(M_1, M_2, M_3, R, B, G) = P(M_1) \times P(R|M_1) \times P(M_2|M_1) \times P(B|M_2) \times P(M_3|M_2) \times P(G|M_3)$$

Calculate for all possible sequences:
{H,H,H},
{H,H,S},

{H,S,H},
{H,S,S},
{S,H,H},
{S,H,S},
{S,S,H},
{S,S,S}

Find: argmax P(M_1,M_2,M_3,R,B,G)

Required Deliverables:

1. Three probability matrices
2. Probability calculations for all possible mood sequences
3. Identification of most likely sequence