The risk prediction works by using a simple but powerful type of AI called a **Bayesian Network**.

Think of it like a doctor diagnosing a patient. The doctor takes a few key factors (symptoms) to predict the likelihood of a condition. In our app, the "symptoms" are your income and spending habits, and the "condition" is your risk of going over budget.

Here's the step-by-step breakdown:

**1. The Model: A Simple Cause-and-Effect Map**

Our Bayesian Network has three variables (nodes):

* **Parent Nodes (The Causes):** Income Level and Spending Rate. These are calculated from your data.
* **Child Node (The Effect):** Risk of Overbudget. This is what we want to predict. Its value *depends* on the parents.

The relationship is straightforward: your income and spending habits directly influence your financial risk.

**2. Discretization: Turning Rupee Amounts into Categories**

The model doesn't work with exact numbers like ₹50,000. Instead, it simplifies your financial situation into broad categories: **Low**, **Medium**, and **High**. This process is called **discretization**.

This happens in your **app.py** file inside the /api/risk endpoint:

* **Income Level:** Your total income is categorized based on set thresholds.

Python

# Example from your code

if total\_income < 2000: income\_level = 0 # Low

elif total\_income < 5000: income\_level = 1 # Medium

else: income\_level = 2 # High

* **Spending Rate:** This is more clever. It's not about the total amount spent, but the **ratio of expenses to income**. Spending 85% of your income is high risk, whether you earn ₹20,000 or ₹2,00,000.

Python

# Example from your code

spending\_ratio = total\_expense / total\_income

if spending\_ratio < 0.5: spending\_level = 0 # Low (spending less than 50% of income)

elif spending\_ratio < 0.8: spending\_level = 1 # Medium (spending 50-80%)

else: spending\_level = 2 # High (spending more than 80%)

**3. The Brain: The Conditional Probability Table (CPT)**

This is the core of the model, located in your **pgm\_model.py** file. The CPT is a lookup table that stores all the "if-then" rules as probabilities. It answers the question: "Given a certain income level and spending rate, what is the probability of the risk being low, medium, or high?"

For instance, the table contains rules like this:

* **Rule 1:** IF Income is Low AND Spending is Low, THEN the probability of High Risk is 0.05 (5%).
* **Rule 2:** IF Income is Low AND Spending is High, THEN the probability of High Risk is 0.8 (80%).

This is represented in the code by the cpd\_risk variable, which you defined with pre-set, common-sense financial rules.

**4. The Final Prediction (Inference)**

When you load your dashboard, the following happens instantly:

1. **Calculate & Categorize:** The backend calculates your current income\_level (e.g., Medium) and spending\_level (e.g., High).
2. **Consult the PGM:** It feeds these two pieces of evidence—Income=Medium, Spending=High—into the pgmpy inference engine.
3. **Calculate Probabilities:** The engine uses the CPT rules to calculate the final probabilities for the Risk node. It might determine:
   * P(Risk = Low) = 15%
   * P(Risk = Medium) = 35%
   * P(Risk = High) = 50%
4. **Deliver Verdict:** The backend sees that "High" has the highest probability (50%) and sends the string "High" to your frontend, which then displays the warning in red.