

Some possible edge cases / issues that might occur

- How to extract information from password encrypted bank statement?
 - What if salary is paid in cash? How will system detect that?
 - What if salary is paid in cheque for 2 months out of 6 months, then how will my system detect a cheque salary?
 - Since this is not a typical model, how will we evaluate success/failure?
 - Freelancer's 6-month bank statement vs An employee's 6-month bank statement. How will my system correctly identify and distinguish?
 - Salary can be credited early/late. How will the system detect misalignment in credit time?
 - Several other outlier scenarios (side income, annual bonus/incentives)
-

When can this system break?

- The system relies on **fixed thresholds** & those values are set globally. Might not be optimal for all banking systems because different employers have different payment patterns that fall outside of the rules.
- Sudden, **significant salary increase** or **receiving large bonuses**, these will be flagged as inconsistent until enough new consistent data is accumulated
- **Limited Data Window**; this system needs a certain amount of historical data. For new accounts or shorter bank statements, it might struggle to make an exact decision due to insufficient data.
- **Purely Rule based system, no learning** ; it's not learning from new data or adapting to evaluating patterns. If there's a new pattern /new edge case, then we'll need to manually readjust thresholds, unlike ML models can be retained.
- **Single source-consistency**; if the initial description is atypical, the system might skew the similarity scores for the entire cluster, because it compares the descriptions to the first transactions description in the cluster