

What exactly is the problem?

Given a bank statement, build a system that:

- predicts whether a user is salaried
- If YES, then estimate the monthly salary

How does a sample bank statement look like?

Statement From : 01-10-2025 to 31-12-2025					
Date	Details	Ref No/Cheque No	Debit	Credit	Balance
01/10/2025	WDL TFR UPI/DR/527472660403/Kamdhenu/YESB/q604165733/UPI 0097693162093 AT 08077 PALBAZAR (KOLKATA)		100.00		1035.40
02/10/2025	WDL TFR UPI/DR/527541238410/BOUDIR M/BARB/sweet98305/UPI 0097694162092 AT 08077 PALBAZAR (KOLKATA)		15.00		1020.40
02/10/2025	WDL TFR UPI/DR/527504966304/85830533/UCBA/8583053312/UPI 0097694162092 AT 08077 PALBAZAR (KOLKATA)		795.00		225.40
02/10/2025	WDL TFR UPI/DR/527502384306/SPENCERS/HDFC/spencers re/Gene 0097694162092 AT 08077 PALBAZAR (KOLKATA)		18.50		206.90

What separates a salaried user from a non-salaried user?

And how does it reflect on bank statement?

Salaried Guy	Non-Salaried Guy
<div><ul style="list-style-type: none">. Credit (not debit). Fixed or near-fixed amount. Monthly occurence(28-31 day interval). Should come from same source/employer. Contain keywords like CREDIT, NEFT,...</div>	<div><ul style="list-style-type: none">. Multiple irregular credits. Cash Deposits. UPI transfers from individuals. Freelance payents with high variance. Contain keywords like UPI, ATM, WDL...</div>

Any constraints / Any underlying conditions?

- Raw Transactional Data
- No ground truth labels initially
- High Noise in Transaction 'Details'