

Questions

- 1) In the dataset the column credit limit what does it indicates whether it is monthly or average credit limit?

Answer: Monthly credit limit.

- 2) What does it mean when the data have an Avg_Utilizatio_Ratio of 0?
- 3) Is it 0 because they can no longer use the credit card since they are already churned or
- 4) is it 0 because they havent used it for a few months?

Answer: It is the ratio of $(\text{credit card spent} + \text{money withdrawal}) / (\text{Total available limit for credit card spends} + \text{Total money withdrawal limit})$

- 5) How would you treat 'Unknown' label?

Answer: replace them with nulls first, since they don't provide any information.

Thereafter imputed them with values based on K nearest neighbor to remove all nulls.

However, all three have minimal effect on the dependent variable, so they can be left alone as well

- 6) Could you please elaborate this column? What do "Total_Trans_Amt" and "Total_Revolving_Bal " mean?

- 7) What is the difference?

Answer: Total_Trans_amt is the sum of transactions one has done in the last 12 months.

This basically tells us the total usage of a credit card by the user.

Revolving balance is the unpaid amount that carries off on your next credit card's cycle.

Total_revolving_bal would be its sum. This tells us the amount that they don't pay on time.

8) However, I'm not sure what the attributes "Total_Relationship_Count", "Contacts_count", "Total_Amt_Chng_Q4_Q1", "Total_Ct_Chng_Q4_Q1" represent for the costumers.

9) Furthermore, how was "Avg_Utilization_Ratio" calculated? Do you think it should be included in the training of a machine learning model?

Answer:

Total_Relationship_Count: total number of products held by the customers (cards, accounts, etc.)

Contacts_count_12_mont: I guess it holds the number of times the bank contacted the customer and/or viceversa. There doesn't seem to have any relationship with other fields (e.g. contact clients that left the bank, customers with revolving balance, etc). I guess it has to do with ad campaigns.

Total_Amt_Chng_Q4_Q1: represents how much the customer increased their expenditure when comparing the 4th quarter against the 1st.

Total_Ct_Chng_Q4_Q1: similar to the previous but in number of transactions.

10) Is this dataset authenticated or somebody made it for visualization purpose?

Answer: I don't know the authenticity of the dataset. But personally I feel the most of the dataset available on Kaggle are good for learning purpose.