Regression-LRM1

December 21, 2024

0.0.1 Choosing independant variables

1. *'Charge Amount'* might actually be part of how Customer Value is calculated, thus it'll be excluded from the independent variables.

2. *'Freq. of use'* This is most likely directly used to calculate Customer Value, the more the customer uses the service, the more he value he has.

3. *'Status'* (active/non-active), this actually would've been much more helpful if I had the duration of being active, as Status might help define a Customer's value, for example: "Active for 2 years, high value", "Active for 2 weeks, low value", hence, it'll be removed, along with those two reasons as well:

   - It might create noise in our model since all high-value customers must be active, but not all active customers have high value
   - The relationship might be too deterministic - being non-active almost automatically means lower Customer Value.

4. *'Age'* is clearly independent of Cutomer value.

5. *'Plan_pre-paid'* is an independent predictor with Charge Amount.

6. *'Call Failure'* isn't related to how to define a Customer value, so independent.

7. *'Complains'* is an independent predictor compared with Charge Amount

8. *'Distinct Called Numbers'* this cannot define a Customer Value so it's independent.

0.0.2 Final chosen independent variables

1. *Age*
2. *Plan_pre-paid*
3. *Call Failure*
4. *Complains*
5. *Distinct Called Numbers*