xepelin

Financial model

2022-05-03

Andrés Cervantes

Goal

Deliver a model that:

a. Predicts the amount of money used with our product
b. Predicts the amount of money financed by Xepelin

Results

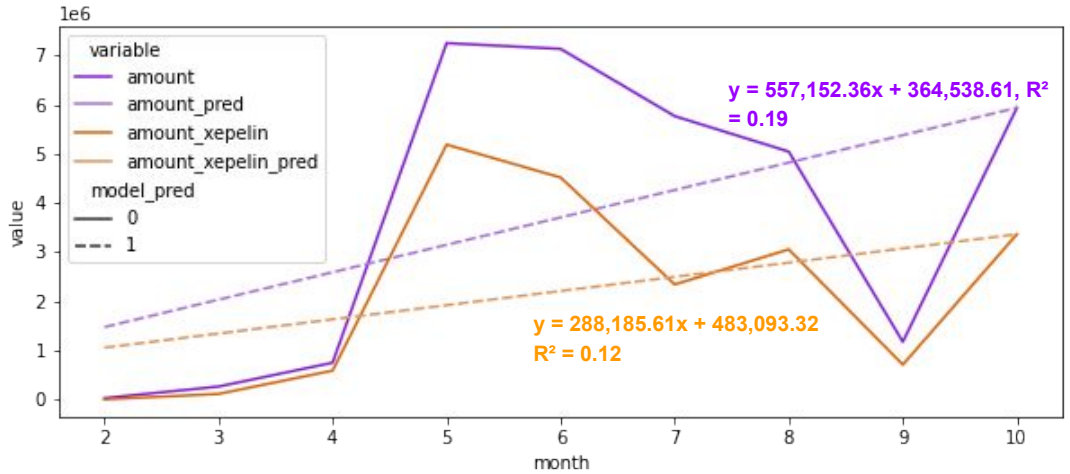

Fig 1: historical transactional amounts

In October, the expected amount to be paid using our product is **$5.9M.**

Of which **58% ($3.4M)** will be financed by Xepelin.

Methodology

- Model: Ordinary least squares Linear Regression.
- This model was selected since the data provided is scarce and doesn't show seasonality.
- Only transactions with status 'PAID' were considered since other status have no date and represent a smaller fraction of the dataset. Furthermore, the other features (IDs) don't provide useful information.

	PayerId	ReceiverId	invoiceId	paidAt	amount	amountfinancedByXepelin	status
0	53	10	18660729	2021-08-23	1490.46	0.00	PAID
1	93	11	18660730	2021-03-22	6418.28	624.48	PAID
2	122	12	18660731	2021-02-18	27979.20	10520.15	PAID
3	85	13	18660732	2021-03-15	183070.77	79421.63	PAID
4	87	14	18660733	2021-07-23	20532.00	20532.00	PAID

Table 1: Sample of the dataset. The first three features don't provide useful information.

Considerations

- The benefits of this model are that:
  - It's simple.
  - It has high interpretability.
  - It's computationally efficient.

- The caveats of this model is that:
  - It's over-simplistic.
  - It's severely affected by outliers and noise.
  - It's not taking into account other features except time.
  - It assumes the amounts are dependent to time.
  - We might be overfitting the data since there was no train-test split.

Conclusions

The predicted output of $3.4M to be financed by Xepelin should be considered as a rough estimate. For Xepelin to have enough financial muscle for next month the RMSE should be added ($1.8M), having a total of $5.2M to finance SMEs.

The model is weak (R^2 is 0.18 and 0.12 for ‘amount’ and ‘amount financed by Xepelin’, respectively) and needs more data to build the next versions of it.

As more data is collected, next versions of the model could be built using ARIMA, SARIMA, and even use deep-learning with LSTM.