Viabill

Saturday, 17 April 2021 11:46

Notes:
attach lending club assignments

Look at all available materials to get an idea

Look at the big picture - aim of the task and business context

Objectives:
- Explore data
- Discover risk factors → late payments → default
- predict → late payments → default

Focus on customers, not shops

Frame the problem

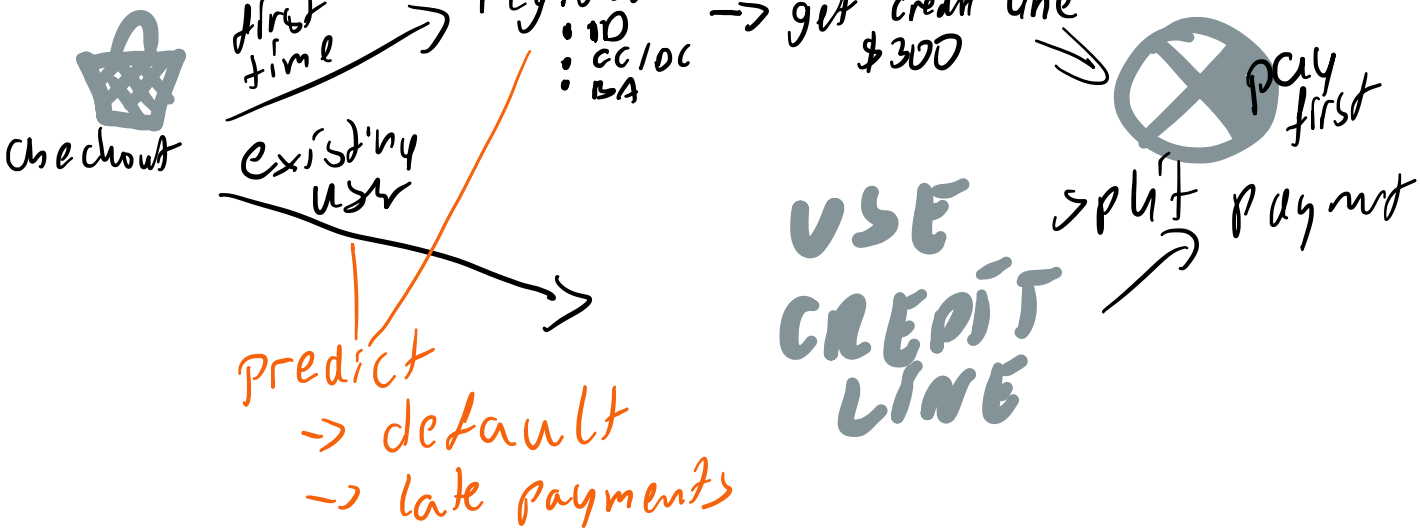
① Buy Now Pay Later → monthly installments

④ At the time of checkout you are charged for first installment
⑤ email → payment schedule
⑥ installment date is the last billing day of the month → email reminders
⑦ Can repay earlier → MyVia Bill dashboard
⑧ No charges for a customer
⑨ Via bill doesn't → report to Credit Bureau
→ pull data from Credit Bureau
⑩ Initial sign up → ID details
→ credit / debit card
→ bank account connection → PSD2 ?
→ credit limit $300

⑨ Requirements for being approved:
- valid email
- min. 18 years old
- valid mobile number
- valid ID
- credit / debit card (no prepaid)
- sufficient funds

⑩ Connect bank account

PROCESS



GET CREDIT LINE

USE CREDIT LINE

How the results will be used?

- at this point just to check what we can get from data,

- no model deployment at this point

Get the data,

Have a look at data structure and plan kind of transformations/algorithms you will use

"**Customers.csv**" has 500k rows and six columns.

CustomerID	Integer	1..500000
Sex:	Integer	1 (male), 2 (female) 0 (other)
Age (years):	Integer	1..100
Residential Address:	String	free text
Postal Address	String	free text
Income (dollars per year):	Integer	100..100000

Transactions.csv

This table has 2 million rows and 8 columns.

Transaction ID	Integer	1..2000000
ShopID	Integer	100.999
CustomerID2	Integer	1..500000
Price (dollars)	Integer	3..200
PaymentStatus1	Integer	0 paid on time 1 paid but late 2 never paid
PaymentStatus2	Integer	
PaymentStatus3	Integer	
PaymentStatus4	Integer	

Note that **transactions.CustomerID2** maps to **customers.CustomerID**

- relational data (simple joins)
- not demanding aggregations
- possible window function

↳ pandas

Choose a framework / analytical platform

↳ PYTHON

↳ kedro for project structure
↳ mlflow for experiment tracking

↳ laptop will be sufficient for calculation

Discover and visualize the data to gain insight

→ jupyter

→ One shop (114 ?) has default rate 100% ⇒ FRAUD

↓

delete from further analysis

→ may be some relation between sex and income

→

To check

↳ no income -?

↳ client segments → NEW customer

→ Existing customer

Look for an opportunity to enrich the data

→ geocoder

→ address list

Prepare data for Machine Learning model

Use transaction count

Select a model and train it

Fine-tune your model

Present your solution

Launch, monitor, and maintain your system