Ethics Statement

Our data analysis will provide implications to financial institutions ranging from banks, to institutional investors such as insurance funds and hedge funds (amongst other asset managers) to create an implied default probability.

This prediction is based on various variables such as age, income amount, loan intent, loan amount, type of homeownership, and loan grade. With more than 28,000 entries in our data set, this amount of data under the financial institution reflects a heavy responsibility on keeping it safe. Being responsible for the data means there has to be consent from individuals on collecting data, storing it in a safe cloud space to prevent cyberattacks, and being cautious who can access the data. The data accuracy and representativeness are important variables for a fair unbiased model for investors. Data collection should have been equal and unbiased, giving data scientists well rounded sets of data to predict on. Any data biases could lead to unequal predictions, impacting the end result model. With investors relying on the model to make their investment decision, the predictions should be as accurate as possible. It is optimal to actively update the model and mitigate biases, allowing all investors to invest into projects with the greatest return. It will be unfair for individuals who need a loan, and are unable to get it because the model has biased historical data.

When looking at privacy within the data, financial institutions are sharing their client information to third parties. This incluses for both business purposes or marketing purposes [Figure 1]. Referring to the US Government Accountability Office article, individuals must share their information to open an account, and are able to opt out of some sharing of information, while some are not as flexible [1]. An example of this would be financial institutions sharing information to the service provider, such as banks outsourcing some of their projects to an overseas IT company. Understanding that there is a non-disclosure agreement signed between the banks and the outsourced company, individuals have no say whether their information can be shared or not. With such a large geographical distance and a different legal system for the outsourced service company providers, it is hard to trace who has access to the data, and what they would do with it. Worse, if a breach occurs, it is crucial to understand the liability and enforcing legal actions will be very complicated. At the end of the day, the victims are the individuals, who allowed financial institutions to collect their data when opening an account. No one can predict what will happen if financial institution information were to breach.

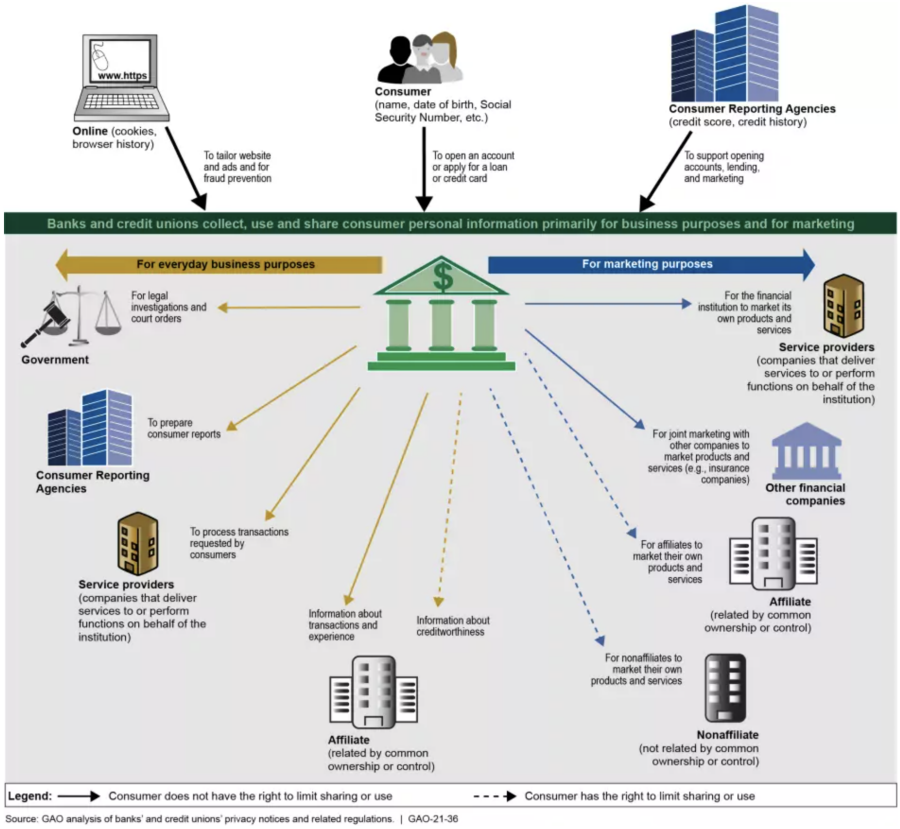

Figure 1: Financial Institution Client Information Sharing [1]

Overall, the credit risk fault prediction model is complicated, involving more than technical capabilities. Ethical concerns of accurate data, privacy challenges, and impact towards individuals needs to be considered. As data scientists, we need to understand the consequences our models bring us.

References

[1] U. S. G. A. Office, “Why do banks share your financial information and are they allowed to?,” U.S. GAO, https://www.gao.gov/blog/why-do-banks-share-your-financial-information-and-are-they-allowed#:~:text=Again%2C%20the%20answer%20is%20yes,and%20share%20with%20third%20parties (accessed Dec. 1, 2023).

Notes:

500 word commentary on **ethical aspects of the project**

- Possible consequences of work for stakeholders
- Address privacy matter and fairness

Ethical Implications

i was thinking a portion would be the privacy of the information used, such as the age, income, employment years, loan interest all that

we can also talk about the concept part and the storage

and then the bank selling it as a bundle

- This personal information - is it shared beyond the initial lender? Can it lead to discrimination?
- If the information is not shared, in situations where loans are collateralized (essentially packaged up and sold off) does this incentivize predatory practices, wherein the person pawning off the risk of the loan has an information advantage and can understate the risk, to their benefit