IC22007 Team summary

Members: Jieqian Xiao, Rui Jin, Tianli Ding, Zihan Zhang

Mentor: Jeff Lang

We are the team 07 in the data analysis track, working on the project of PPP Removed Applications project.

The formation of the team was suggested by Zihan and we are all first year students of MSBA. Although we have different background of undergrad, we share the same interest and passion into data analysis. Therefore, we decide to corporate and form a team to compete in the info challenge.

We have a team with great diversity in terms of skill set. Finance, Chemistry, Computer Science, Data Science, Business, the various skill sets allow us to corporate very well and really work as a team.

In this competition, we have deployed a logistic regression model and KNN algorithms to describe and predict which factors are significant when we are applying the loan. We have measured the applying pass rate from different perspectives and discover the most significant roles in deciding whether an apply would be removed from paycheck protect program database. We also take political, pandemic factors into consideration and get pretty good story.

What we found in the competition:

There is a lag of the effect of PPP program on SMEs. And

Banks do ensure the program carry on efficiently, but more banks or financial institutions involved will help the loan approvement to be fairer.

Business owners should be aware for following characteristics if they intend to apply loans: amount applied, job retained, date applied, lender, hubzone, Low- and Moderate-Income Communities.

During this challenge info competition, we have developed models and applied these to questions. It allows us to have an opportunity to apply what we learned from classes to real world questions. We also get an opportunity to dive deep into the data scientists’ world to have a look at how powerful the data analysis is.

We would like to thank our mentor Jeff and the all the IC staff. We really appreciate this opportunity. If we can take this project further, we could try some deep learning algorithms to develop a better model. We could also utilize tableau to develop some interactive diagram.