Capstone Project Technical Report
Christopher Ng, Jake Thompson, Mikhaela Anderson, Nick Kartschoke, Ryan Lippe
5/16/2022

Introduction

We found datasets about Small Business Administration (SBA) disaster loans, the S&P 500 index, unemployment, and local government finances. We hypothesize that unemployment can be predicted by a combination of disaster loan data, local government finances, and market data. We will analyze data between 2008 and 2015 on a county level. We are planning to capture these relationships in an elastic net regression model using cross validation. We will see which factors were the best predictors of unemployment by looking at coefficient weights of the finished model. In addition to predicting unemployment, our analysis will be aimed at answering these questions:

1. Are disaster loans a significant predictor of unemployment?
2. Does the amount of the loan affect mitigating unemployment?
3. Which disasters saw the most loan money awarded?
4. Were local government revenues correlated with increased unemployment rate?
5. What states received the most disaster loans?
6. Is the S&P 500 a significant predictor for unemployment?

Machine Learning Model

For our machine learning model we used an elastic net and applied it to our filtered dataset from our SQL server, which eliminated all zero entries for the disaster number column. For our predictors we used only the numeric columns, except for “Unemployment”, which was our target. We then created our training and testing sets, with 75% being for training and 25% being for testing. We created our first model with the default parameters for elastic net and applied it to our dataset. We then found the R^2 score and MSE to obtain a baseline that will be used for comparisons with our tuned models. We then tried to adjust the hyper-parameters “alphas” and “l1_ratio” to increase the performance of the model. For the second model we tried to expand the ranges of the alpha and l1_ratio values, then did a gridsearchCV to see which combination of those two parameters gave the best results. For our third model we tried expanding the alpha value options by using a logspace to obtain more values between 0.1 and 2. While the third model performed better than the previous models, we then applied a logspace to the l1_ratio as well and created our fourth model. The fourth model performed the best so far with a R^2 score of 0.4028 and a MSE of 1.989.

Another thing we checked was if removing a column would lead to a better performing model. We found the descriptive statistics for the numeric predictors for general information for each of them. Then, we ran a correlation matrix to see the relationships between each numerical predictor and the “Unemployment” value. From the correlation matrix we concluded that the “Total Salaries & Wages" variable had the lowest correlation. So we then created a new elastic net model on the dataset, dropping the “Total Salaries & Wages" column, with log spaces for both the alpha and

l1_ratio parameters. This model had a MSE of 1.987 and a R^2 of 0.3711. When comparing this model to the fourth model, we can conclude that the fourth model had a higher R^2 and slightly higher MSE. Even though the fourth model is more complex, we believe that it is the more optimal model for predicting employment.

Results

1. Are disaster loans a significant predictor of unemployment?

We saw that disaster loans were overall very weak predictors of unemployment, both in general and relative to other factors we investigated. Our model minimized the weights of these factors, but did not improve when they were removed. We also saw that the approved amount of economic injury disaster loans (EIDL) was the disaster loan factor with the strongest correlation to unemployment, with a correlation of -0.116. This was interesting, because the negative correlation implies that places which received greater amounts in loans saw slightly lower unemployment.

2. Does the amount of the loan affect mitigating unemployment?

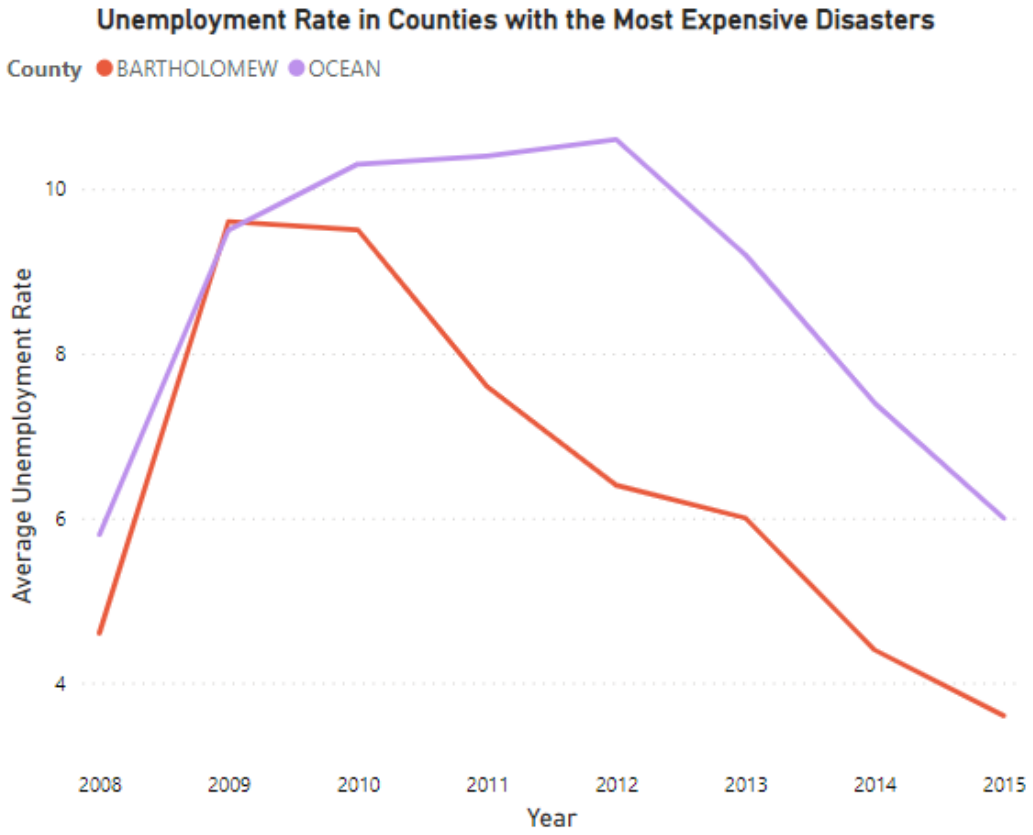

Figure 1.

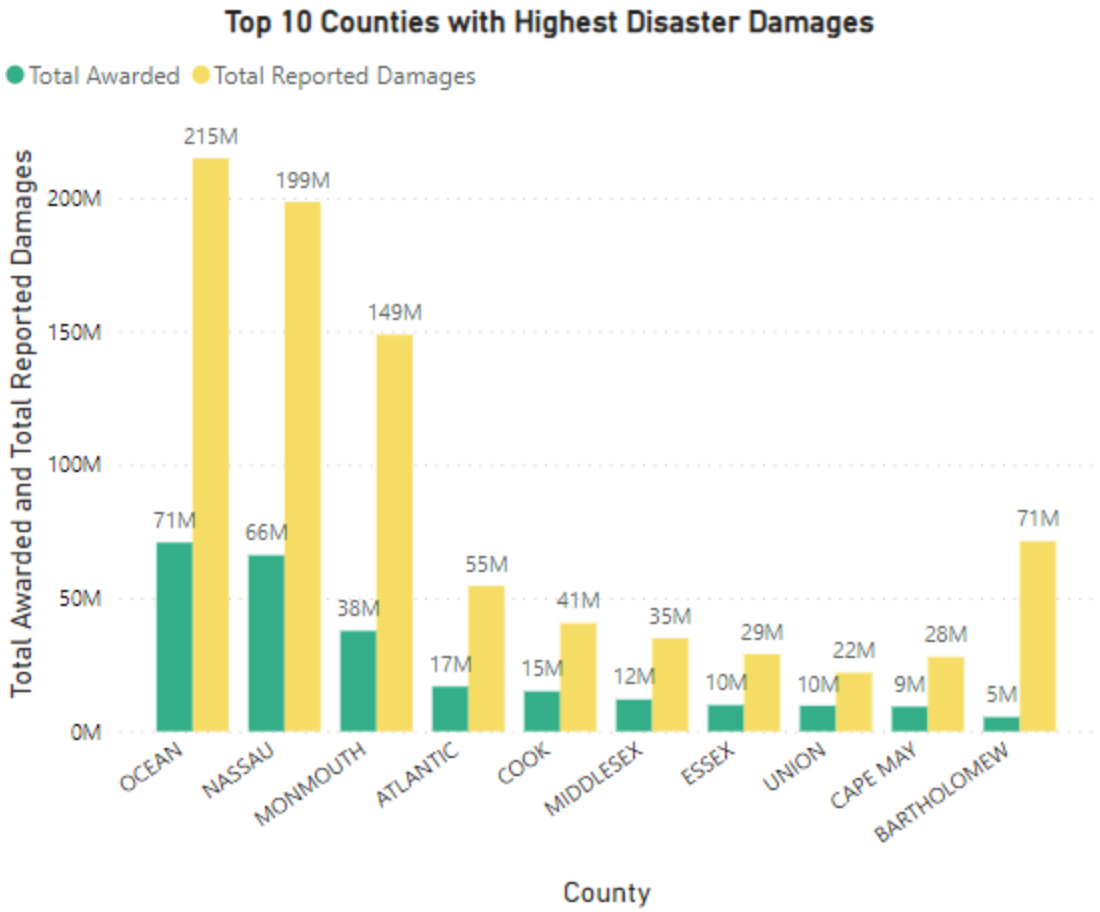

Figure 2.

It appears that the amount of disaster loans doesn't significantly mitigate unemployment. In the cases of Ocean and Bartholomew County, we see that despite reporting much higher awarded loans and a higher proportion of damages awarded as loans (**Figure 2**), Ocean County saw higher unemployment (**Figure 1**). Looking at the counties with the highest disaster loan awards, unemployment rates vary widely, but they all appear to move with a similar trend despite very different award amounts. This doesn't necessarily rule out a causal relationship between loan amounts and unemployment rates, but our analysis did not find any evidence for one. The disaster loan factor which our model weighed the most heavily was the approved amount of EIDL, but this was still a very weak predictor relative to other factors.

3. Which disasters saw the most loan money awarded?

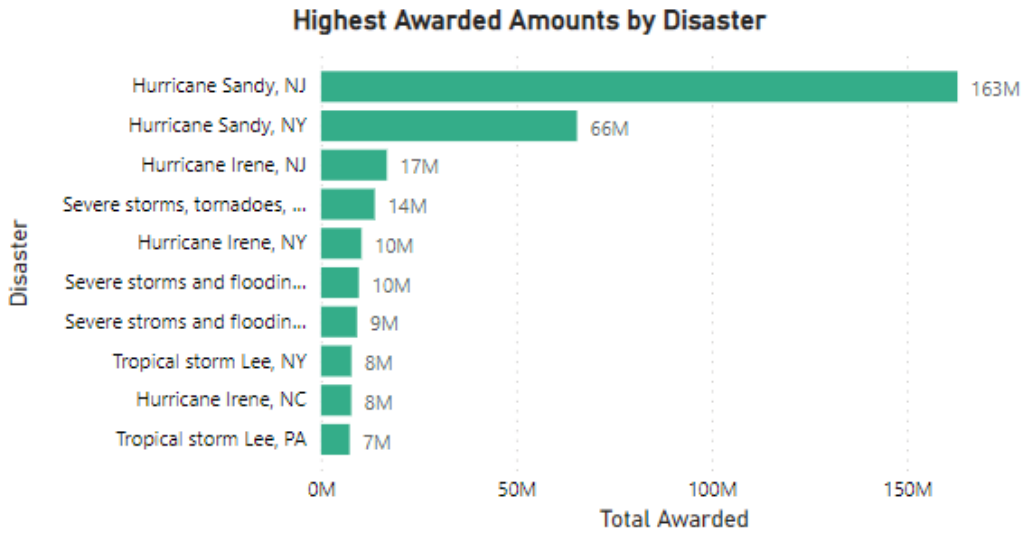

Figure 3.

By far the most destructive single disaster in terms of loans awarded was Hurricane Sandy, which caused $229 million in disaster loans just in New York and New Jersey (**Figure 3**) Hurricane Irene was the next largest cause of disaster loans. In general, severe storms and flooding was the disaster category which incurred the most loans.

4. Were local government revenues correlated with increased unemployment rate?

We saw that local government revenues, specifically from fines and forfeits as well as revenue from interest on debts held by local governments, were both weighed more heavily by our model than any SBA disaster loan factors. Revenue from fines and forfeits was our second best predictor of unemployment overall, though the relationship was still weak. Fines and forfeits had a correlation with unemployment of -0.161, and interest revenue had a correlation of -0.143. This implies that governments which saw greater revenue from these sources saw slightly lower unemployment.

5. What states received the most disaster loans?

The top 3 recipients of disaster loans between 2008-2015 were New Jersey, with $180,667,100, New York, with $88,670,300, and Illinois, with $17,681,200. New Jersey and New York received significantly more loans than any other state, and we saw that they were both affected by Hurricane Sandy (**Figure 3**) which was the most destructive single disaster we found.

6. Is the S&P 500 a significant predictor for unemployment?

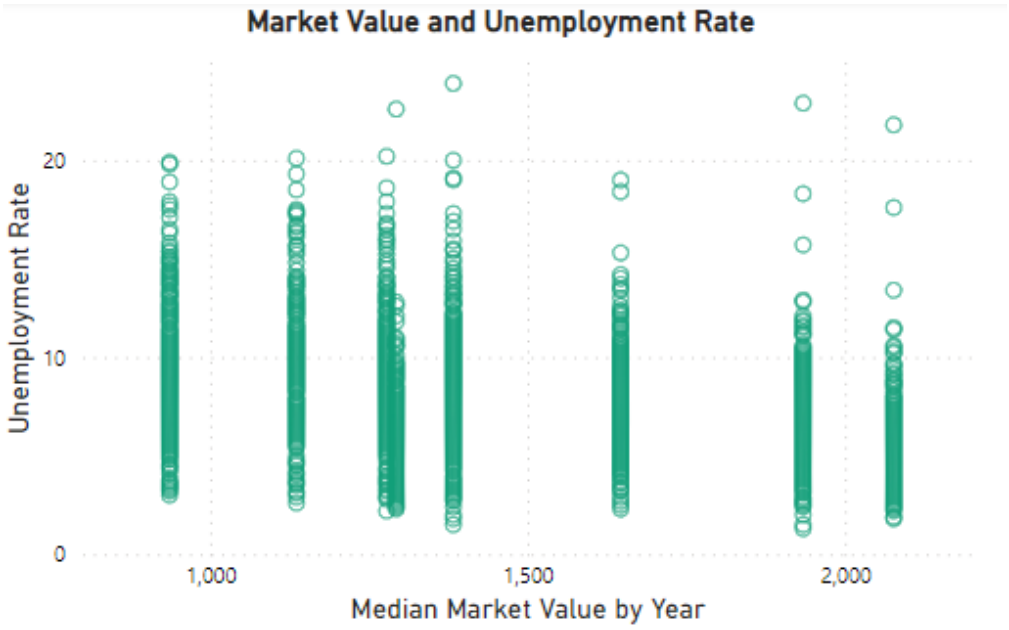

Figure 4.

Market value from the S&P 500 (**Figure 4**) was the strongest predictor of unemployment from all the features we included in our model by a wide margin, with a coefficient weight over 10x greater than government revenue from fines and forfeits, which was our next strongest predictor. Interestingly, it was our strongest predictor despite having a fairly weak correlation with unemployment of -0.1132. Fines and forfeits, interest revenue, and approved amount of EIDL all had stronger correlations despite being weaker predictors.

Conclusion

We successfully made an elastic net model to predict unemployment, and our final R^2 score was 0.4028. We saw that the factors we chose were overall weak predictors of unemployment, with S&P 500 market value being the strongest predictor, and with local government revenue from fines and forfeits being the next strongest predictor. Unemployment is complex and is influenced by many factors, so for additional progress we recommend trying to improve the model by considering more features. The scope of our research included correlation and using features as predictors, but for further analysis to investigate causation, we recommend the features be compared with unemployment in later years. For example, we could consider whether disaster loans received impacted the next year's unemployment rate, or unemployment over the next 5 years.

We were interested in seeing whether disaster loans were a significant predictor of current unemployment, and we saw that they were not. Further research is needed to see whether disaster loans are predictors of future unemployment as mentioned above. We also researched whether the amount of loans affected unemployment, and we saw that unemployment appeared insensitive to the amount of disaster loans.

Next, we looked into which disasters saw the most loan money awarded, and we found that the most destructive disasters from 2008-2015 in terms of disaster loans were Hurricane Sandy and Hurricane Irene. In general, severe storms and flooding were the largest category of disasters.

We then considered whether local government revenues correlated with unemployment. We saw that government revenue from fines and forfeits was the second strongest predictor of unemployment overall, and had a weak negative correlation with unemployment of -0.161.

The next question we looked into was which states received the most in disaster loans, and we saw that the top recipients were New Jersey, New York, and Illinois. Each of these states saw significant loans due to storms, with the top 2 being caused by Hurricane Sandy.

The final question we researched was whether the S&P 500 was a significant predictor of unemployment. We saw that it was our strongest predictor by a wide margin despite being less correlated with unemployment than several other features we considered. For future research, we recommend considering more market factors as predictors of unemployment due to this being much better than other features we included.