Hypothesis

DS Bang - Chen Wei (cwei24), Chenxi Wu (cwu101), Minfeixue Zong(mzong1), Zhenhao Sun(zsun32)

Hypothesis

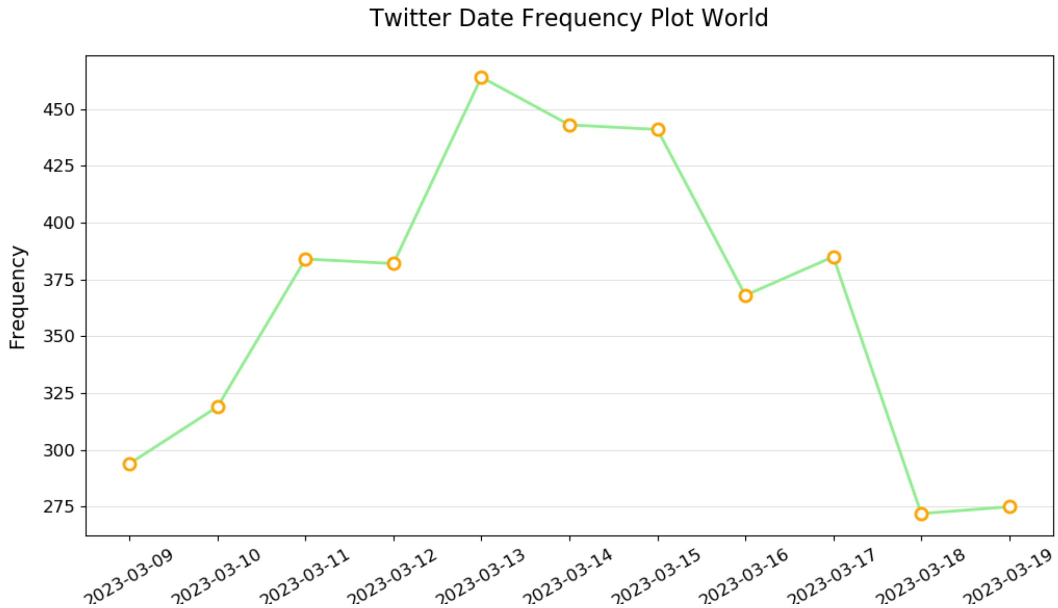
Sentiment analysis is one of the most common NLP applications. We sought out to investigate people's attitude towards SVB, Credit Sussie and their relationship using data scraped from Twitter. Such an investigation is meaningful as it helps us better grasp people's satisfaction, loyalty and potential risks or opportunities for the bank and the world economy as a whole. Also, it helps us to better study the trend and see if there is a relationship between people's attitudes and the bank's collapse and whether the collapse of banks necessarily lead to a negative impact on people's perception of the world economy. Specifically, we are testing on the following 3 hypothesis:

1. Before the collapse, 50% of people had a negative attitude towards SVB.
2. Before collapse or being bought, people's attitude towards SVB and Sussie Credit are identical.
3. The collapse has a close relationship with people's attitude towards the world economy.

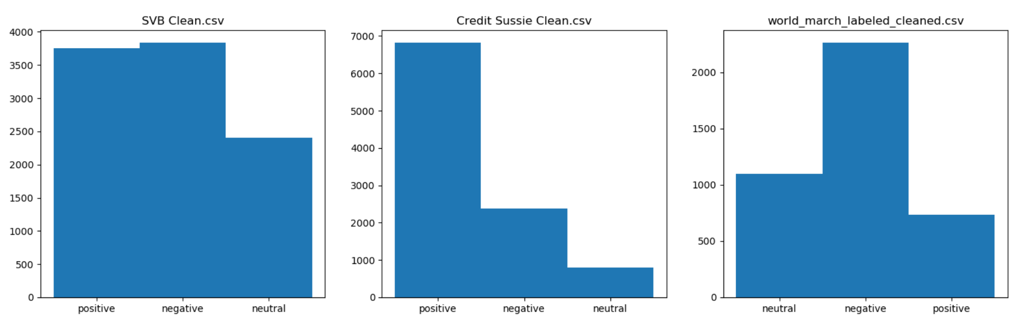
And in the findings part, we will illustrate our findings from our hypothesis testing.

Data

We scraped data from twitter which contains Tweet_id, User, Content, Like_Count, Quote_Count, Reply_Count, Url, Date. After carefully handling sensitive information like User, we want to perform the sentiment analysis task and hypothesis testing on the Content of the twitter. We have 3 datasets, which correspond separately to the twitter world_economy data, twitter credit_sussie data, and twitter SVB data. Due to the scraping technical constraint, the world_economy data ranges from the date 2023-03-09 to 2023-03-19 and the credit suisse (2023.3.18 - 2023.3.19) and SVB (2023.3.9 - 3.10) data is just the twitter at the time of the collapse.



As it can be seen, our world_economy data is not screwed that much by the date and is normally distributed. The sentiment distribution is like the following:



Although it is a little bit imbalanced, it is still acceptable if we mainly use the sentiment to do the hypothesis testing and perform some extra steps (sampling) when we did the machine learning part.

Findings

Claim #1: Before the collapse, 50% of people had a negative attitude towards SVB.

Support for Claim #1:

We have a p-value of 1.00 and a z-statistic of -22.981 for testing the hypothesis that the proportion of negative attitudes towards SVB is larger than 50%. The results suggest that we fail to reject the null hypothesis. Less than 50% of the tweets in your sample have a negative attitude towards SVB. Since the z-statistic is negative, it means that the sample proportion of negative attitudes is less than the hypothesized proportion of 50%.

When we change the hypothesized proportion to 30% and obtain a z-statistic of 17.153 and a p-value of 0.000, it means there is strong evidence to reject the null hypothesis. We can conclude that the proportion of negative attitudes towards SVB in your sample is significantly larger than 30%. More than 30% of the tweets in your sample have a negative attitude towards SVB.

Claim #2: Before collapse or being bought, people's attitude towards SVB and Sussie Credit are identical.

Support for Claim #2:

We choose to use the Chi-Square test to determine if there is a significant difference in the distribution of the sentiment labels between the two groups. After testing, we get the p-value 0.9629995589325446 > 0.05. So we fail to reject the null hypothesis.

Claim #3: People have the same attitudes towards the world economy before and after the collapse.

Support for Claim #3:

We use two sample t-tests for this hypothesis. The sentiment scores before and after the collapse are taken at two different times. The null hypothesis is that the average difference between sentiment scores before and after the collapse is zero. The p-value is 0.0400179627, which is smaller than the significance level 0.05. We reject the null hypothesis that people have the same attitudes towards the world economy before and after the collapse. The average sentiment scores from Twitter towards the world economy are statistically different before and after the collapse.

ML (Sentiment prediction and LDA)

DS Bang - Chen Wei (cwei24), Chenxi Wu (cwu101), Minfeixue Zong(mzong1), Zhenhao Sun(zsun32)

Goal

Sentiment analysis prediction is one of the most common applications of NLP. We first implemented an unsupervised Latent Dirichlet Allocation, which can identify the topics most debated in the context of the world economy. We then proceeded to train two supervised sentiment prediction models, the Multinomial Naive Bayes Model and the Support Vector Machine, to predict people's attitudes based on their language. Our training and validation data were sourced from posts we scraped from Twitter. This type of investigation is meaningful as it aids us in better generalizing people's thoughts and understanding their satisfaction, loyalty, and potential risks or opportunities for both the bank and the world economy. Furthermore, our predictions of people's overall sentiment towards the world economy could provide valuable insights for policy makers, economists, and social scientists.

Data

We scraped data from twitter which contains Tweet_id, User, Content, Like_Count, Quote_Count, Reply_Count, Url, Date. After carefully handling sensitive information like User, we want to perform the sentiment analysis task and hypothesis testing on the Content of the twitter. We have 3 datasets, which correspond separately to the twitter world_economy data, twitter credit_sussie data, and twitter SVB data. Due to the scraping technical constraint, the world_economy data ranges from the date 2023-03-09 to 2023-03-19 and the credit suisse (2023.3.18 - 2023.3.19) and SVB (2023.3.9 - 3.10) data is just the twitter at the time of the collapse. The visualization is the same as the hypothesis testing part.

We manually labeled the world_economy sentiment (and only used this dataset) for our model training. We ask 3 of us to label the data and one to organize our votes, use the mode as the overall sentiment and manually deal with the outliers (i.e. some illegal values, the situation where 3 of us all think differently and there is not really a mode). This means if 2 out of 3 think that a twitter is negative, its overall sentiment label is negative. For negative, positive, and neutral sentiments, we use 0, 1, and 2 respectively to represent them in our data.

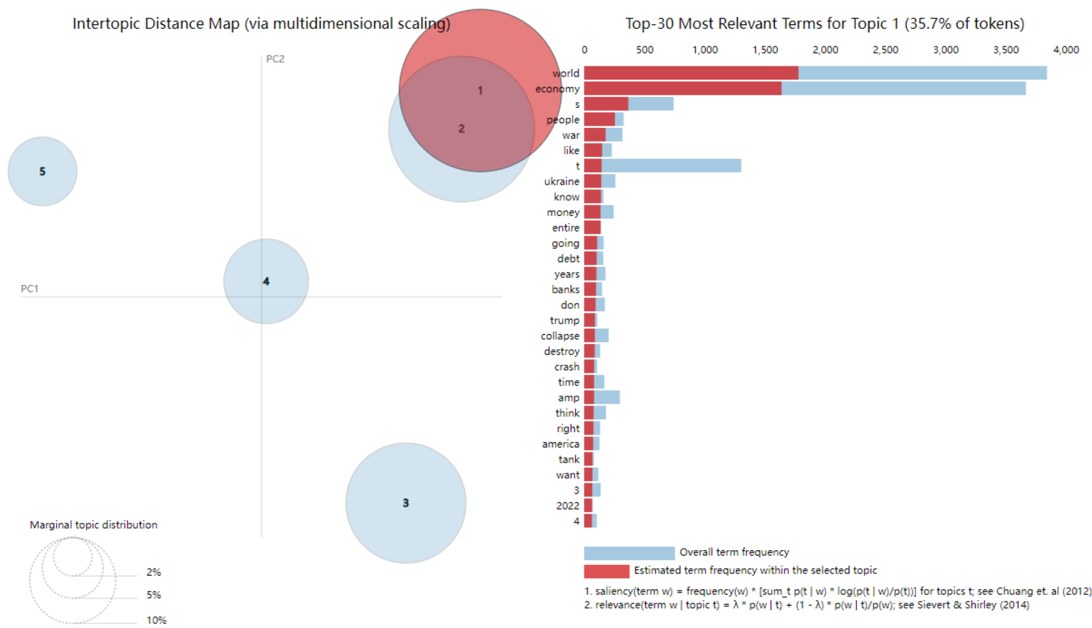
Model+Evaluation Setup

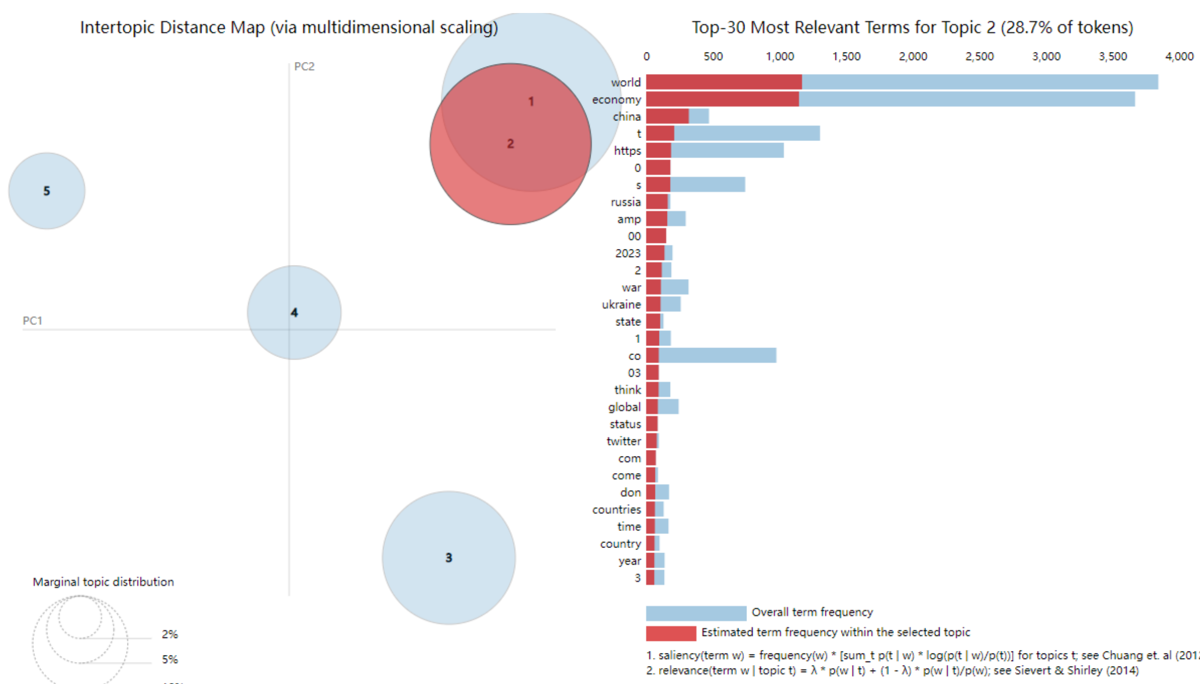
We care about forecasting the sentiment of twitter especially related to SVB, Credit Suisse and the world economy. For supervised learning models, we do the 80%: 20% train - validation split on our data and take user's input (unseen data) as test data. And we are using k-fold validation to make our model more robust.

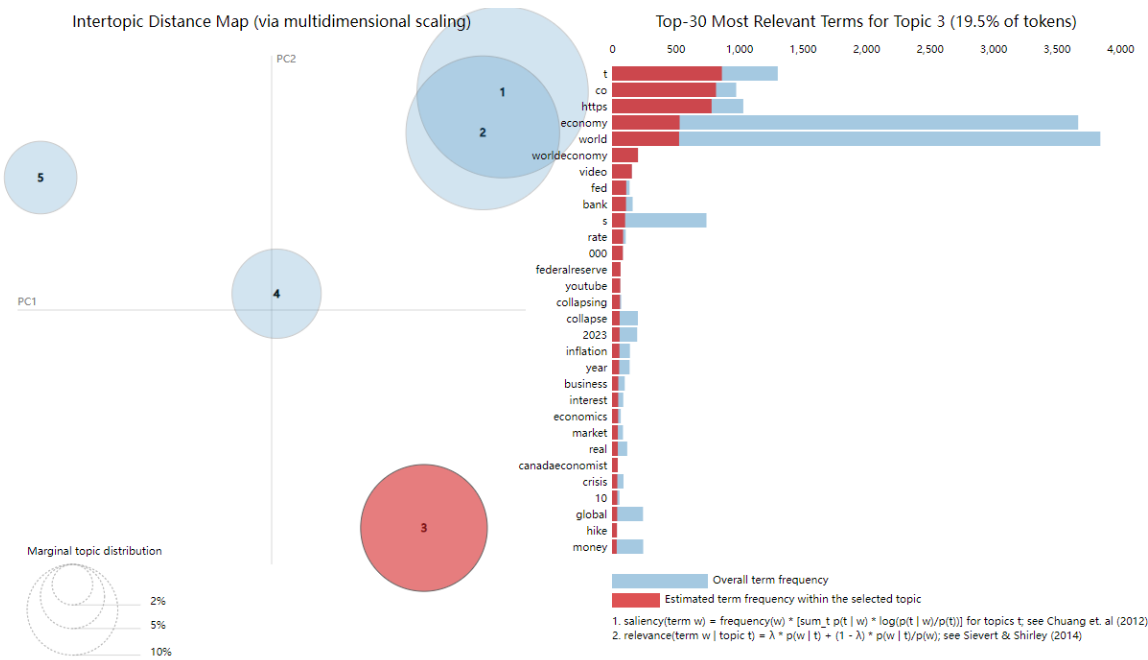
Results and Analysis

Claim #1: Through our LDA model, we have identified key subjects under global economic discourse, including the effect of war on the world economy, the interconnection between international relations and economic conditions, the impact of monetary policies on the global economy, and the differing viewpoints towards the world's economic state.

Support for Claim #1: The results of the topic modeling show that the top two common topics are related to the war's impact on the world economy and international relationships. The third topic is more focused on monetary policies and the world economy. The fourth and fifth topics are less established, but they give insights into people's attitudes towards the world economy, such as the need for "change", "control", and opinions about "failure". Overall, the model's predictions are satisfactory.







Claim #2: Multinomial Naive Bayes model moderately succeeded in predicting tweet sentiments with around 76% accuracy.

Support for Claim #2: The model performs best when identifying positive sentiments (class 1), with a high recall and F1 score. It struggles most with identifying negative sentiments (class 0), where it has a lower recall, suggesting that it's misclassifying a significant proportion of negative sentiments as either neutral or positive. The model's performance on neutral sentiments (class 2) is somewhere in the middle. Its overall accuracy is 76%.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.53      0.64       471
           1       0.75      0.90      0.82       430
           2       0.74      0.85      0.79       455

    accuracy                           0.76      1356
   macro avg       0.76      0.76      0.75      1356
weighted avg       0.77      0.76      0.75      1356
```

Claim #3: Support Vector Machine model demonstrates strong performance in identifying negative and positive sentiments with an overall accuracy of 72%, but it struggles significantly in recognizing neutral sentiments.

Support for Claim #3: The SVM model is performing well when identifying negative sentiments and neutral sentiments, with high recall and balanced F1 scores. However, it's struggling significantly with neutral sentiments, where it has a very low recall, suggesting that it's often misclassifying neutral sentiments. While the precision for neutral sentiment is high, the low recall pulls the F1-score down. The overall accuracy of 72%.

	precision	recall	f1-score	support
negative	0.73	0.87	0.79	386
neutral	0.75	0.20	0.32	103
positive	0.73	0.73	0.73	320
accuracy			0.73	809
macro avg	0.74	0.60	0.62	809
weighted avg	0.73	0.73	0.71	809