



Sentiment Analysis for Understanding Economic Trends

In collaboration with Better Dwelling

Aaron Tian: aarontian426@gmail.com

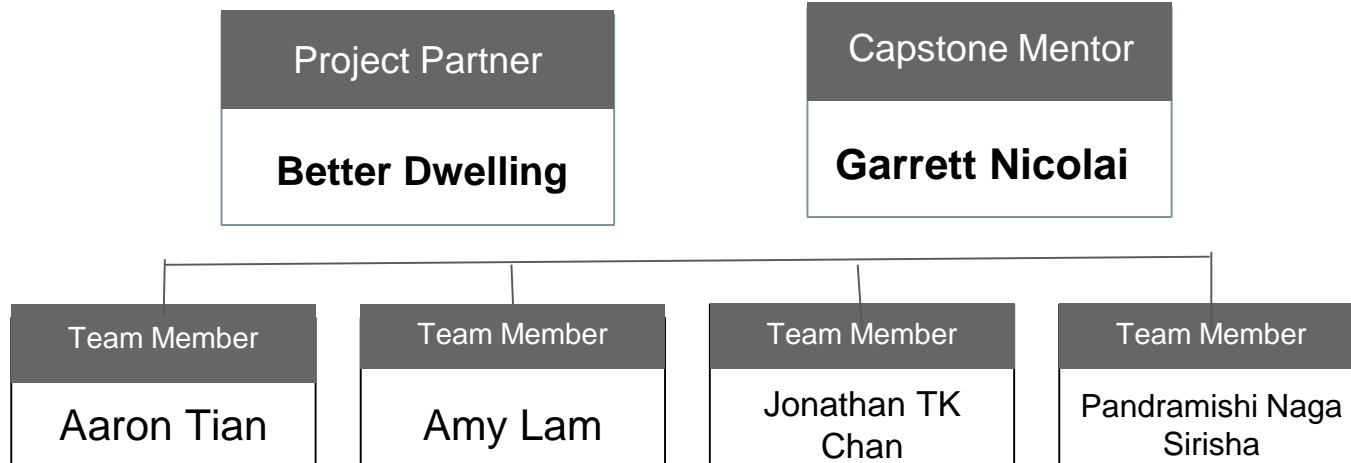
Amy Lam: kaman.lam@alumni.ubc.ca

Naga Sirisha: nagasirisha999@gmail.com

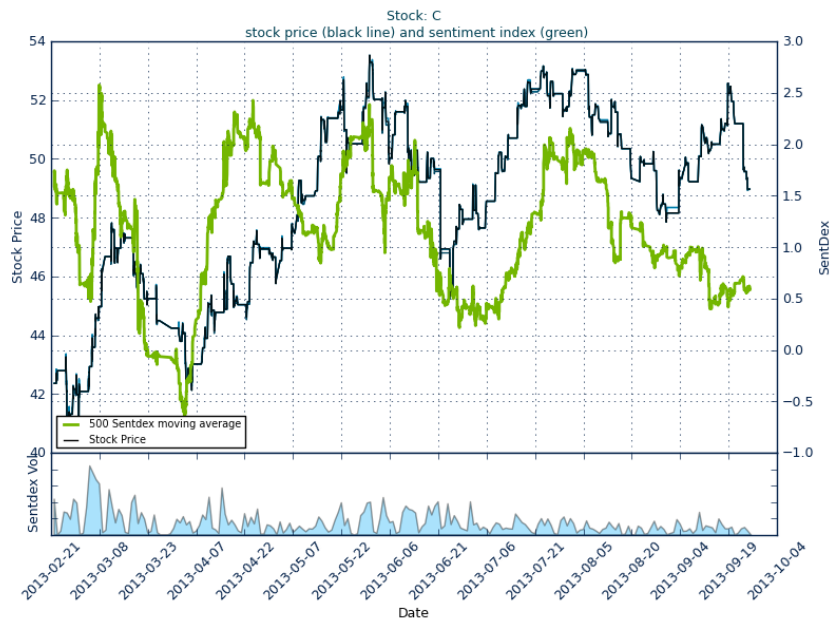
Jonathan T.K. Chan: jonathan.tk.chan@alumni.ubc.ca



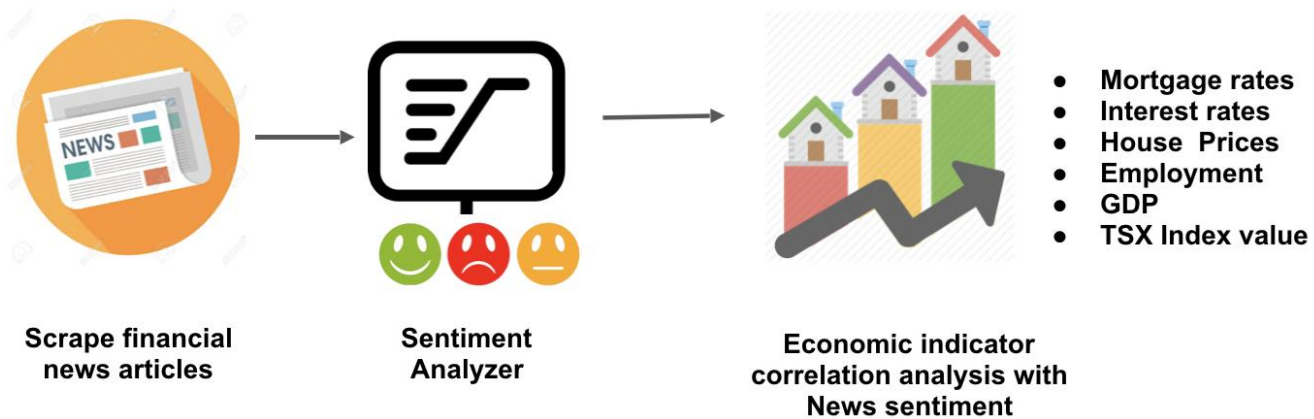
Project Team



Motivation: Sentiment index vs Stock price

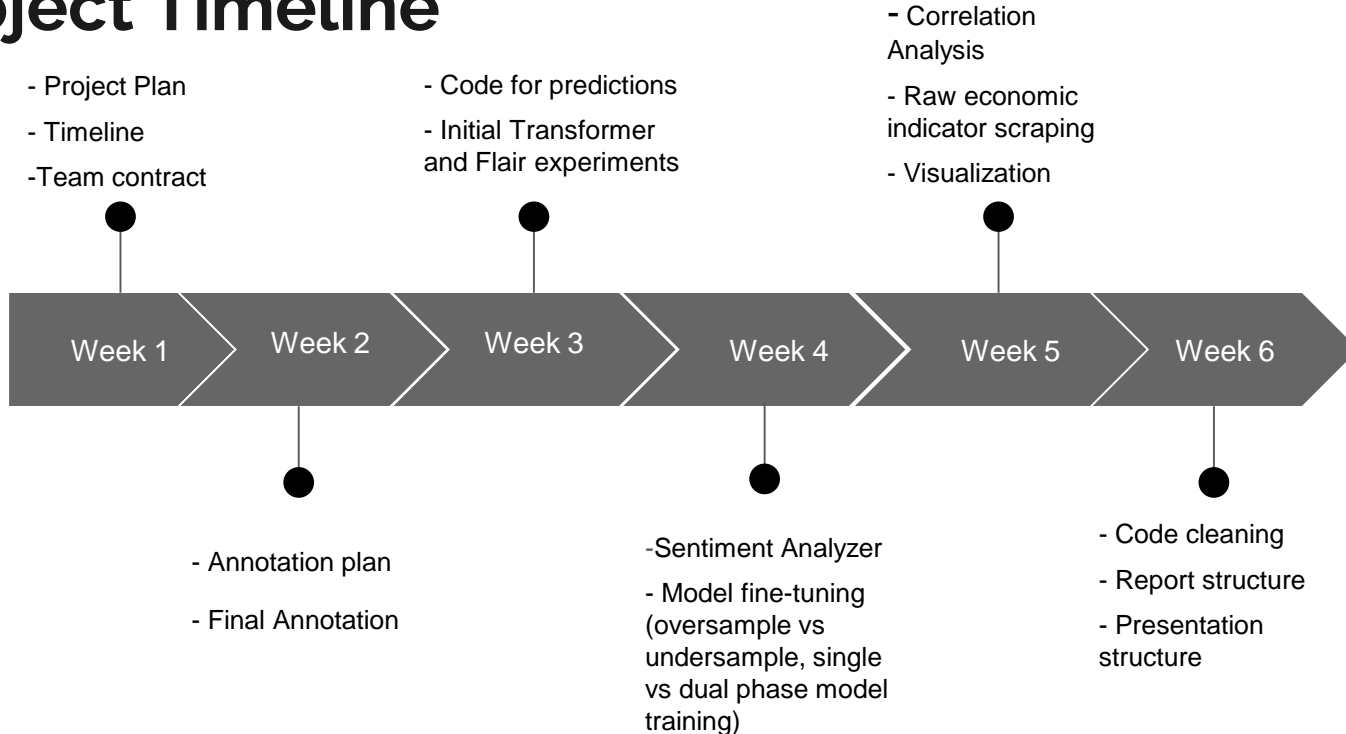


Motivation: Project Flow





Project Timeline



Data collection

Past one year data

Canadian news sources

The data is collected for the 2019-2020 in two phases:

- **News articles** for sentiment analyzer.
- **Raw economic indicator data** for the correlation analysis.

Economic indicators considered:

- GDP
- Housing Price
- Mortgage rates
- Interest Rates
- TSX/Stock Market
- Employment

News Sentiment Data

Annotation Statistics:

Sentiment	CBC	Bloomberg	Total
Positive	61	70	131
Neutral	222	147	369
Negative	77	149	226

We scraped the financial news data of the Canadian market in order to analyze sentiment:

The news sources are:

- **CBC**
- **Bloomberg**
- Benchmark dataset of financial news data set.

Note: Considered only title and description as headlines generally tend to capture the sentiment of the article.

News Sentiment Data

For the economic indicator under scrutiny, the annotation scale:

0 : Irrelevant or no impact

-1 : Influence in a negative way.

1 : Influence in a positive way.

Economic Indicator	Factors that influence the economic indicator
GDP	Employment, Economy, Investment in a company, Layoffs, Province-level increase/decrease.
Employment	Recession, Layoffs, Investment in a large company.
Stock Market	Interest rates, Employment, General economy
Interest rate	Economy, Employment, Inflation target
Mortgage rate	Interest rate as a cost of funding, Bank competition(price war to gain market share), Lay offs, Economy, Housing price
Housing price	Economy, Insolvency/bankruptcy (leads to force sale),investment by a big company, Surge in investors, Money laundering(push up certain luxury housing price), Layoffs(if it increases insolvency rate of home owners, ...)

Raw Economic Indicator Data

- GDP

Percent change from last year same month to this year same month.

- Housing Price

Percentage change in the composite HPI from month to month.

-Mortgage rates

Mortgage for Total, funds advanced, residential mortgages, insured

-Interest Rates

Monthly aggregated Interest rates

-TSX/Stock Market

Monthly aggregated closing tsx price

-Employment

Employment numbers for both sexes over the age 15



Sentiment Analyzer

- State-of-the-art performance
- Stacked word embeddings
- Fast training speed

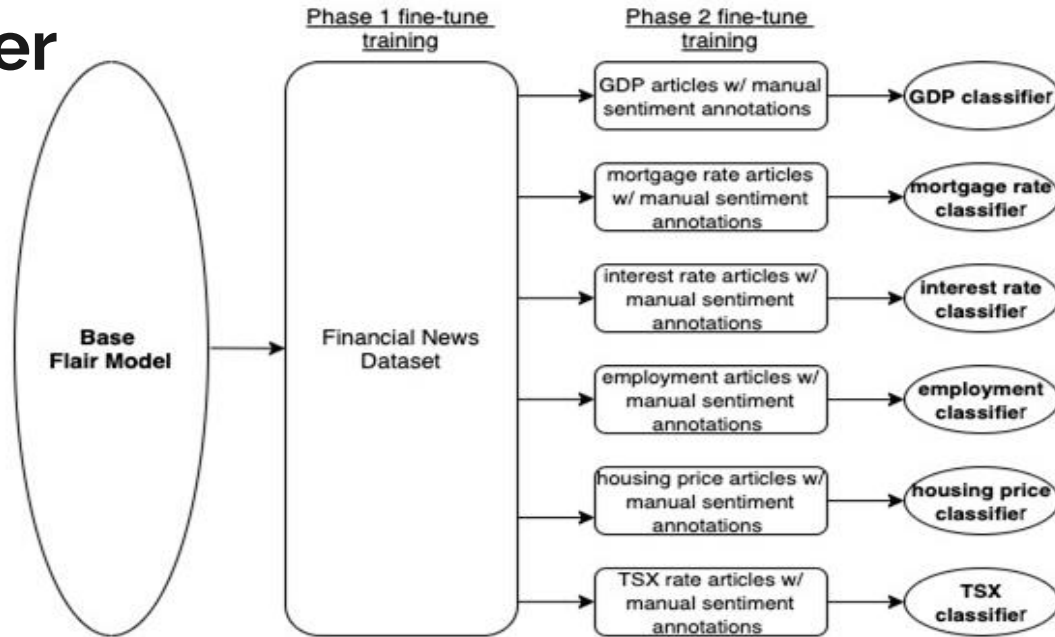
flair



Challenges



Classifier



- Fine tuning: Take existing model and optimize for particular task



Performance

Classifier	Best Accuracy	Best F1-score
GDP	0.82	0.72
Mortgage Rate	0.72	0.56
Interest Rate	0.83	0.70
Employment	0.75	0.64
Housing Price	0.74	0.62
TSX	0.83	0.75

Visualization

Canadian News Sentiment Swings As Economy Evolves

Pick an economic indicator and a news source:

Interest rates

×

▼

CBC

×

▼

Pick a chart type:

☒ Monthly average ☐ Daily datapoints

Correlation coefficient between interest rates and selected source sentiment is: 0.69

in the period

01/01/2019

→

06/01/2020

interest rates VS news sentiment



Notes:



Visualization

Canadian News Sentiment Swings As Economy Evolves

Pick an economic indicator and a news source:

Employment × ▼

CBC × ▼

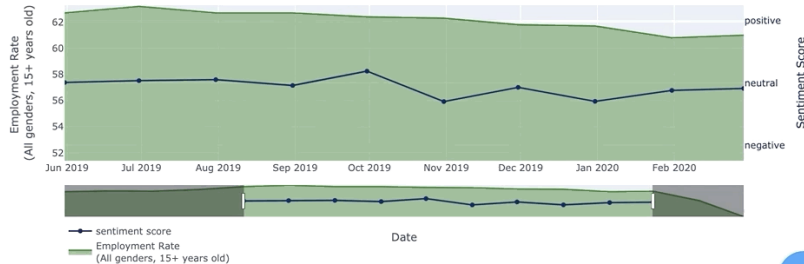
Pick a chart type:

☒ Monthly average ☐ Daily datapoints

Correlation coefficient between employment and selected source sentiment is: 0.43
in the period

01/01/2019 → 03/01/2020

employment VS news sentiment



Notes:



Canadian News Sentiment Swings As Economy Evolves

Pick an economic indicator and a news source:

Employment × ▼

CBC × ▼

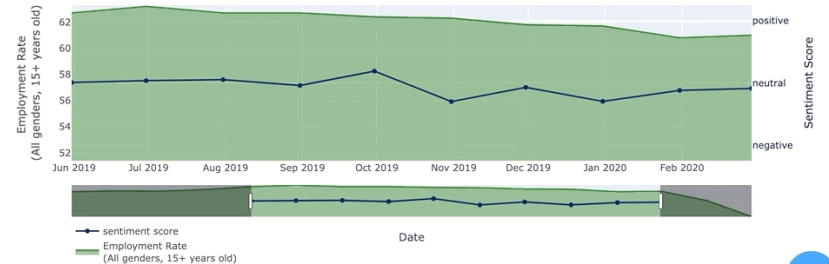
Pick a chart type:

☒ Monthly average ☐ Daily datapoints

Correlation coefficient between employment and selected source sentiment is: 0.43
in the period

01/01/2019 → 03/01/2020

employment VS news sentiment



Notes:





Analysis - Correlation more obvious during crisis

Sources	Full Observation period	Full Observation period correlation	2019 months correlation	2020 months correlation	2020 Feb-Mar correlation
GDP	Apr 2019 - Mar 2020	0.59	0.14	0.71	1.0
Mortgage Rates	Jan 2019 - Mar 2020	-0.45	-0.56	0.99	1.0
Interest Rates	May31 2019 - Jun1 2020	0.42	NA(target interest rate entirely flat)	0.43	1.0
Employment	Jan 2019 - Apr 2020	0.38	0.56	0.95	1.0
Housing prices	Jan 2019 - Apr 2020	0.32	-0.47	0.91	-1.0 (resumes to 0.92 in Feb-Apr)
TSX	Jan31 2019 - May31 2020	0.29	0.32	0.87	1.0



Error examples - reasonable

"Consumer worries cloud **best** Canadian output **gain** in two years . Canada 's economy recorded a **stronger-than-expected rebound** in the second quarter as exports recovered , but surprisingly **weak** consumption and business investment will **cast doubts** on the expansion 's sustainability ."

Predicted label	Annotated label
-1	0

**Possible source of error:
negative wordings**



Error examples - less reasonable wrongs

"**Weak** Canadian GDP data opens door for stimulus budget : Scotiabank . Canada 's **sluggish** economic data opens the door for Finance Minister Bill Morneau to **spend more** in his pre-election **budget** , according to Bank of Nova Scotia ."

Predicted label	Annotated label
1	-1

Possible source of error: Probable association of "stimulus" with positive sentiment from annotated examples.



Sources of Error

- Small phase 2 training data
 - Further hand annotations, use of Mturk
 - Identify which indicators are worth exploring
- Ambiguity in training data
 - Neutral tag = irrelevant OR relevant w/ neutral sentiment
 - Noisiness of training data varied for each model
- Classify articles based on full text
 - More computational power required



Future Work

- To include more annotations for training the model better to identify the intricacies of the factors influencing the economic indicator
- We would like to rebuild the classifier to classify the articles based on average paragraph sentiment instead of just considering the title and description
- To cherry pick the articles to train by using more scrupulous methods for data preprocessing to distinguish between irrelevant and neutral articles



Conclusion

- Data pipeline successfully built
 - Data extraction
 - Sentiment analysis
 - Visualization interface
- Correlation values were determined and visualized
- Future work
 - Expand training data
 - Improve classifier accuracy
 - Address sentiment ambiguity