Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Predicting economic growth using machine learning techniques and sentiment analysis[1]

Berkay Akisoglu,

Central Bank of the Republic of Turkey

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

TÜRKİYE CUMHURİYET MERKEZ BANKASI

IFC- BNM- ECB

Predicting Economic Growth Using Machine Learning Techniques and Sentiment Analysis

Berkay Akışoğlu
Statistics Department , Central Bank of Turkey

16 August 2019 | Malaysia

Outline

I. Motivation

II. Methodology

III. Data

IV. Machine Learning Models

V. Sentiment Index

VI. Future Works

I. Motivation

- Time lag in macro-economic statistics
- The problems with regular surveys
- New data sources to monitor economic developments
- Financial and economic news

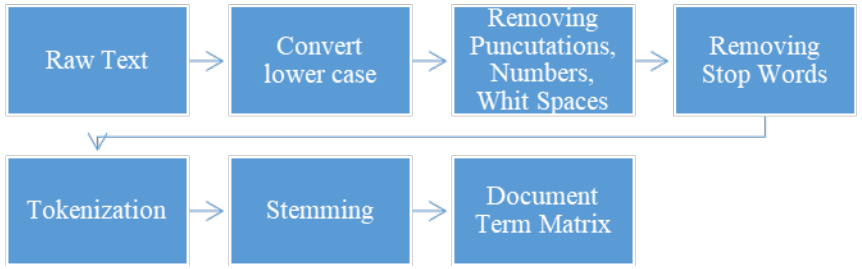
II. Methodology

- Sentiment Analysis: A field of research aiming to determine subjective information such as emotion, opinion, attitude stated in the text by using methods and techniques from fields such as natural language processing (NLP), statistics and computer science.

- Methods used in sentiment analysis
  - Lexical Approach
  - Machine Learning Models

II. Methodology

- Preprocessing Steps:
  - Convert lower case
  - Removing punctuations, number and white spaces
  - Removing stop words
  - Tokenization
  - Stemming



- Document Term Matrix

Doc 1: «**TÜRKİYE genelinde konut metrekare fiyatları nisan sonu itibarıyla son bir yılda…..»**

Vector-space representation

Docs	artış	aya	Fiyat	konut	liraya	metrek are	nisan	son
1	2	1	2	3	2	3	2	3

III. Data

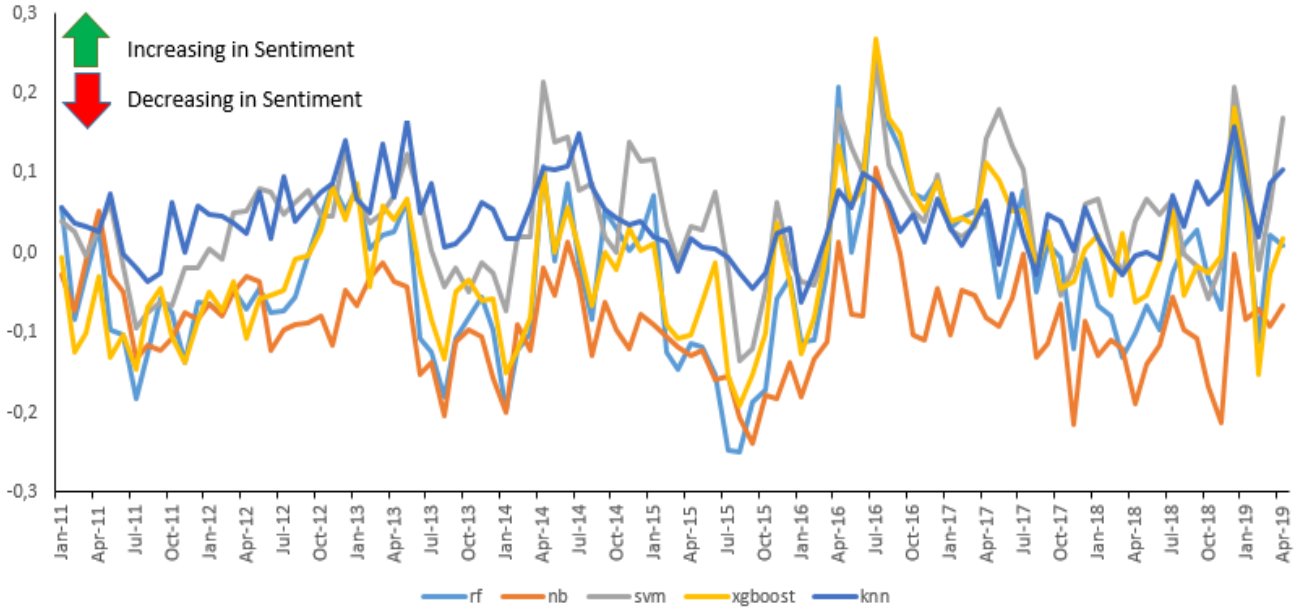
- Total 131601 financial and economic news
- Each includes name and phrases specified by Central Bank
- Mainstream and local media

Year	Number of News (as of April 2019)
2011	18.474
2012	21.042
2013	14.978
2014	15.400
2015	16.606
2016	15.227
2017	13.736
2018	13.397
2019	2.741
Total	**131.601**

III. Data

- No ready-to-use dataset whose sentiment orientation is known.
- Randomly generated news bundles were evaluated by Central Bank employees, academicians, graduates and students of TED University.
- Question: «What is your expectations of future economic activity after reading this news.»
- 1357 news labelled (three labels: «-1», «0», «1»)

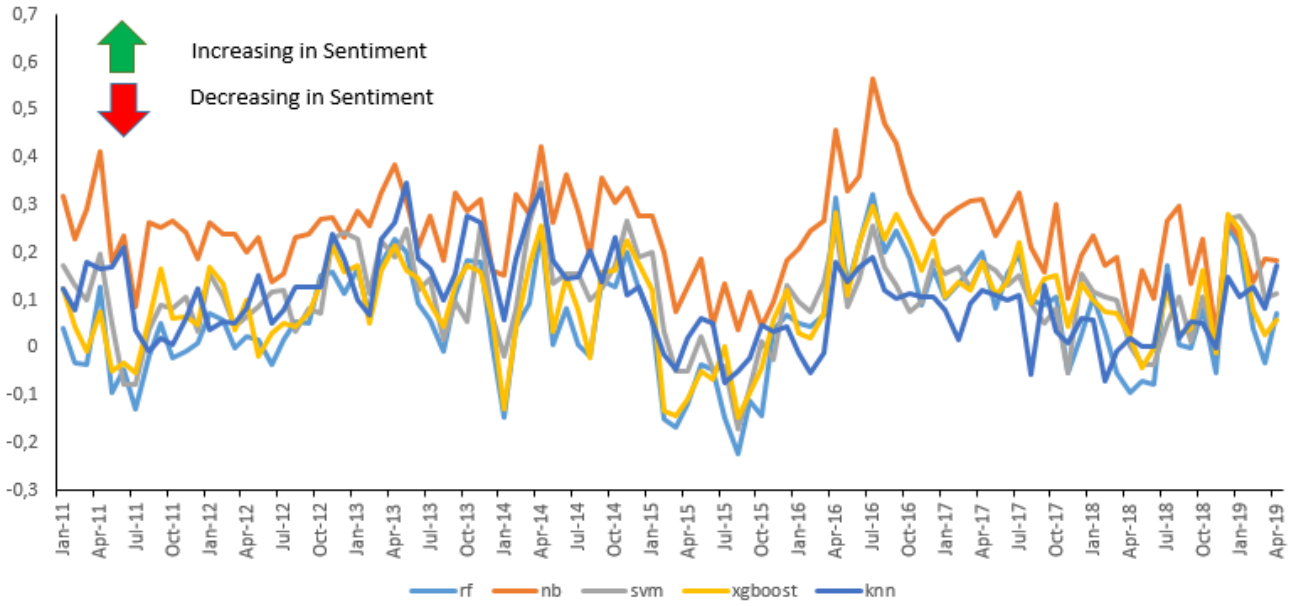
IV. Machine Learning Models and Sentiment Indexes (3 labels; -1,0,1)



Algorithm	Accuracy (%)
Random Forest	0.51
SVM	0.44
Naive-Bayes	0.47
KNN	0.43
XGBoost	0.45

	rf	nb	svm	xgboost	knn
rf	1.00	0.72	0.70	0.87	0.60
nb	0.72	1.00	0.60	0.55	0.50
svm	0.70	0.60	1.00	0.76	0.59
xgboost	0.87	0.55	0.76	1.00	0.51
knn	0.60	0.50	0.59	0.51	1.00

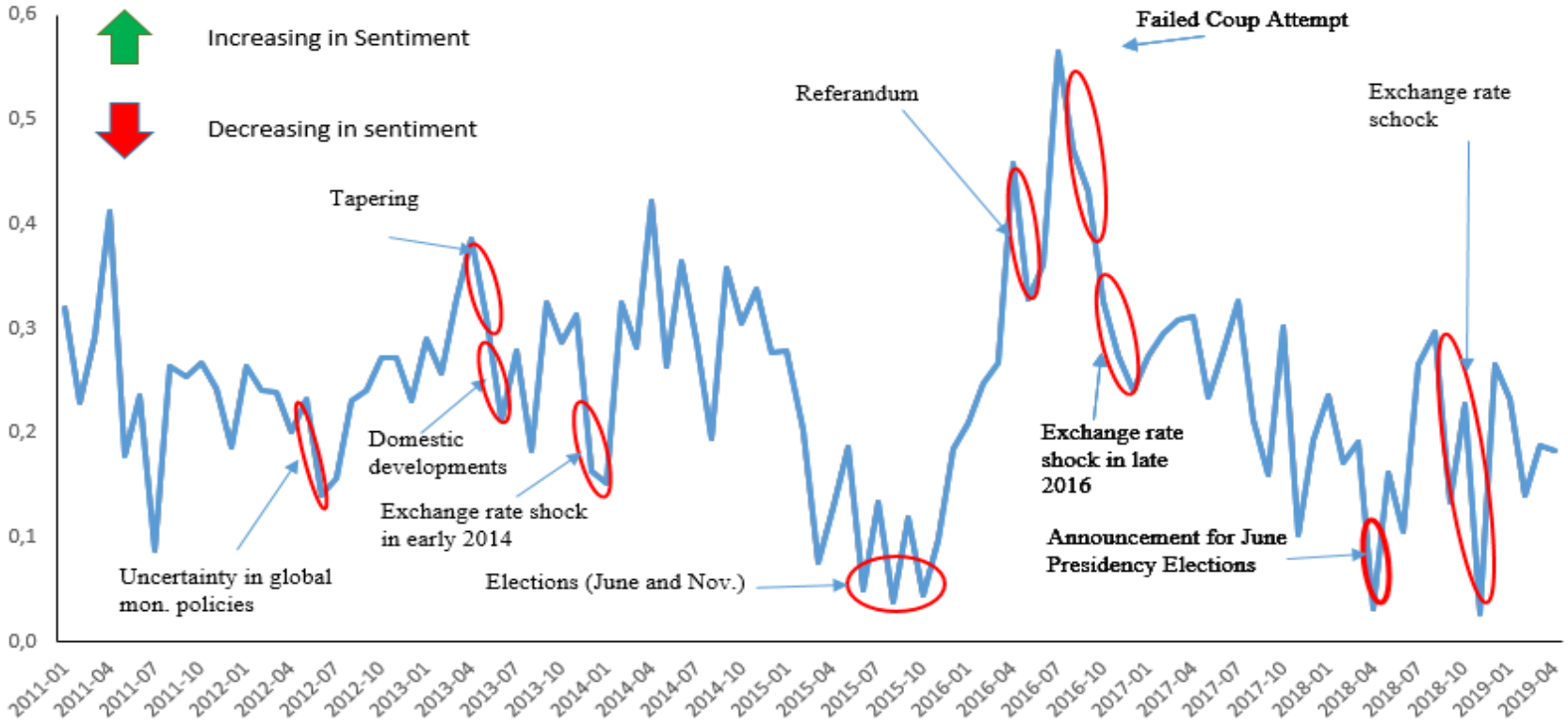
IV. Machine Learning Models and Sentiment Indexes (2 labels;-1,1)



	Accuracy	Precision	Recall
Random Forest	0.65	0.65	0.67
SVM	0.61	0.62	0.58
Naive-Bayes	0.61	0.59	0.66
KNN	0.57	0.55	0.66
XGBoost	0.61	0.60	0.66

	rf	nb	svm	xgboost	knn
rf	1.00	0.80	0.78	0.98	0.57
nb	0.80	1.00	0.69	0.73	0.54
svm	0.78	0.69	1.00	0.78	0.57
xgboost	0.93	0.73	0.78	1.00	0.49
knn	0.57	0.54	0.57	0.49	1.00

VI. Interpretation of Sentiment Index for Turkish Economy



Conclusion and Future Works

Conclusion

- First study using this techniques on Turkish financial and economic news
- Captures important economic and political events

Future Works

- Improving stemming
- Differantations between news
- Conducting survey with larger participation
- Running different algorithms on the same data set
- Trying econometric models to get most benefit from sentiment index
- Investigate the relations with other economic indicators

THANK YOU