AI for Finance

Financial sentiment of cryptocurrency microblog content

Student: Phoe Chuan Bin   Supervisor: A/P Erik Cambria

Abstract:

Over the last decades, the increasing capability of NLP makes it possible to capture market sentiments more accurately and the semantics of financial corpus in a more nuanced way. Such opinion mining allows sentiment data from media to be fused with other structured data coming from the stock market for investing decisions.

Project Objectives:

To leverage on the use of NLP techniques to better predict the financial sentiment of social media cryptocurrency content.

Symbolic Approach:

Using current literature to create and utilize best-in-class lexicons for financial sentiment prediction. Explore use of combinations for improvements.

Individual Results:

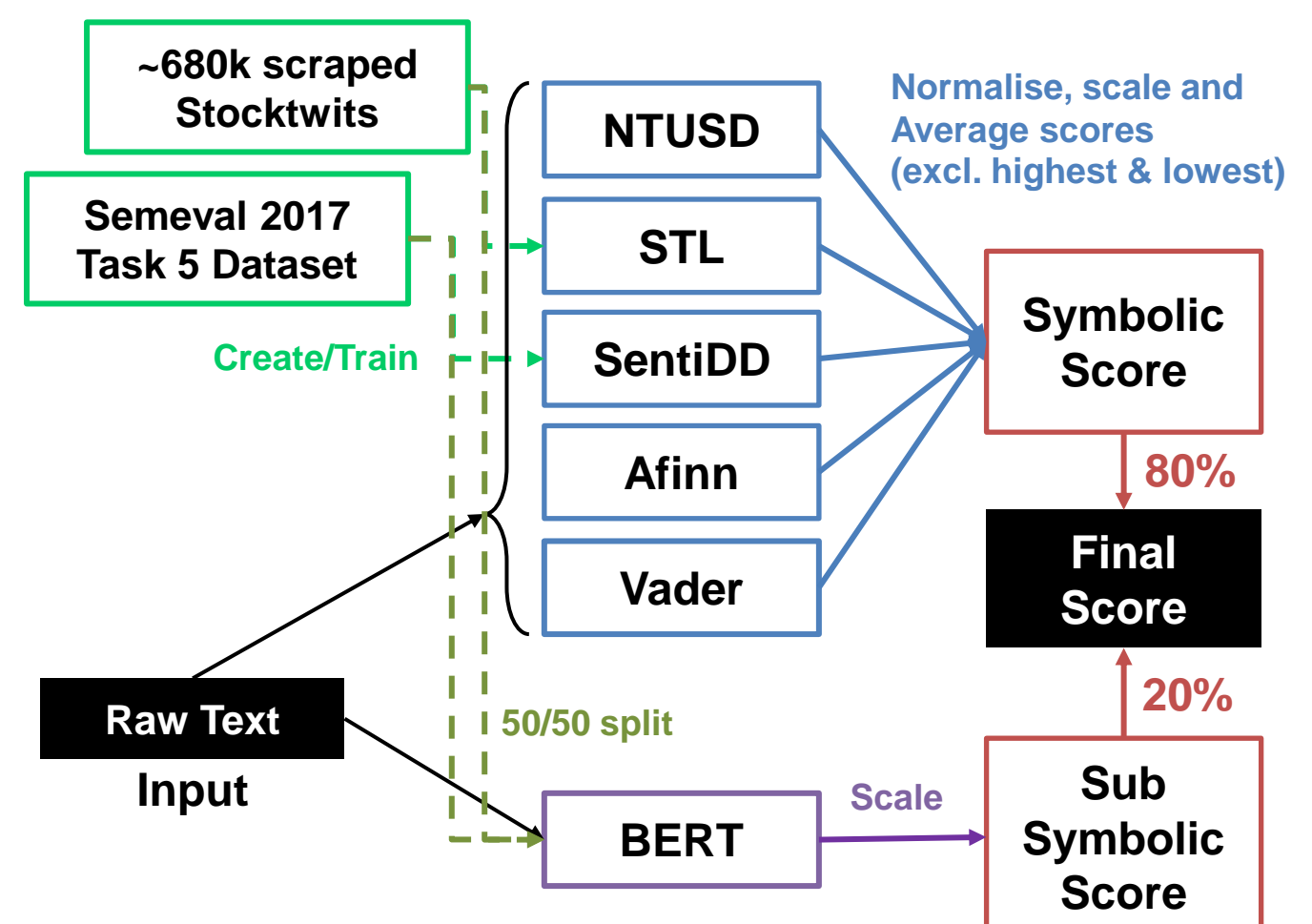| Lexicon	Average of 2 Test Datasets	
	Accuracy	F1 Score
NTUSD	0.712135	0.765286
STL*	0.629407	0.755706
Senticnet	0.626706	0.714841
Vader	0.648328	0.657089
SWN	0.596784	0.638872
Afinn	0.628022	0.624538
SentiDD*	0.543308	0.422547

*Created lexicons

Combination Results (Top 3):

Combination Technique	Average of 2 Test Datasets	
	Accuracy	F1 Score
Soft voting & Leave 2 out - NTUSD, STL, Afinn, Vader, SentiDD	**0.756192**	**0.807582**
Soft voting & Leave 2 out - NTUSD, SWN, STL, Vader, SentiDD	0.750179	0.805732
Soft voting only– NTUSD, SWN, STL, Vader	0.751049	0.804324

Sub symbolic Approach:

Engage Deep Learning techniques such as transformer architectures (BERT) to capture nuances in financial text.

BERT trained on	Average of 2 Test Datasets	
	Accuracy	F1 Score
Semeval 2017 Task 5 Dataset	0.7505	0.8014
Scraped Stocktwits Dataset	0.6782	0.7730
Combine 50% each	**0.7412**	**0.8019**

Hybrid Approach:

Combine both scores for the final hybrid score.

Symbolic Weightage	Average of 2 Test Datasets	
	Accuracy	F1 Score
20%	0.74336	0.88334
40%	0.74922	0.88881
60%	0.75895	0.89789
80%	**0.77526**	**0.90432**

Future Works:

Use the final sentiment score as a data input for upstream financial prediction tasks such as volatility forecasting, asset allocation, and market trend predictions.