XLNET-GRU Sentiment Regression Model for Cryptocurrency News in English and Malay

Nur Azmina Mohamad Zamani[1,2], Jasy Liew Suet Yan[1], Ahmad Muhyiddin Yusof[3]
[1]School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia.
[3]Academy of Language Studies, Universiti Teknologi MARA Perak Branch, Seri Iskandar Campus, 32610 Seri Iskandar, Perak, Malaysia.

namz.ina@gmail.com, jasyliew@usm.my, ahmadmuhyiddin4@uitm.edu.my

Abstract

Contextual word embeddings such as the transformer language models are gaining popularity in text classification and analytics but have rarely been explored for sentiment analysis on cryptocurrency news particularly on languages other than English. Various state-of-the-art (SOTA) pre-trained language models have been introduced recently such as BERT, ALBERT, ELECTRA, RoBERTa, and XLNet for text representation. Hence, this study aims to investigate the performance of using Gated Recurrent Unit (GRU) with Generalized Autoregressive Pretraining for Language (XLNet) contextual word embedding for sentiment analysis on English and Malay cryptocurrency news (Bitcoin and Ethereum). We also compare the performance of our XLNet-GRU model against other SOTA pre-trained language models. Manually labelled corpora of English and Malay news are utilized to learn the context of text specifically in the cryptocurrency domain. Based on our experiments, we found that our XLNet-GRU sentiment regression model outperformed the lexicon-based baseline with mean adjusted $R2 = 0.631$ across Bitcoin and Ethereum for English and mean adjusted $R2 = 0.514$ for Malay.

Keywords: sentiment analysis, deep learning, pre-trained language models, cryptocurrency news

1. Introduction

Cryptocurrency is a new form of digital currency designed to achieve transparency, decentralization, and immutability by utilising blockchain technology and cryptographic functions (Pintelas et al., 2020). As cryptocurrencies can be traded in coin exchanges, cryptocurrency price prediction models assist cryptocurrency investors in making investment decisions either to buy, sell or hold to maximise earnings, as well as policymakers and financial scholars in analysing the behaviour of cryptocurrency markets. Sentiment has shown to play a crucial role in cryptocurrency price prediction since the changing aspects of the cryptocurrency market is determined by sentiment from various sources such as online news and social media (Karalevicius et al., 2018; Rognone et al., 2020). Therefore, developing models that can accurately detect sentiment signals from text sources and measure the sentiment strength is an important first step to ensure reliability of the cryptocurrency price prediction in the downstream task.

The lexicon-based (Gurdgiev & O'Loughlin, 2020; Karalevicius et al., 2018; Loughran & Mcdonald, 2014; Mai et al., 2015) and machine learning methods utilising classic representations such as bag-of-words (BoW) or TF-IDF (Georgoula et al., 2015; Lamon et al., 2017) remain the most popular computational methods used to extract sentiment features for cryptocurrency price prediction. However, the sentiment prediction models that are often used as an intermediate component to generate sentiment features within the cryptocurrency price prediction pipeline are often not evaluated thoroughly and this has caused the effect of leveraging sentiment features in cryptocurrency price prediction models to be mixed in existing studies. Also, these sentiment analysis methods most often do not take context or relationship between words into account, and thus do not provide the best results when classifying or scoring text for sentiment.

Recently, word embedding has gained traction due to its enhanced functionality and performance (Li et al., 2017). From the vectors produced by the word embeddings, machine learning or deep learning methods such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have shown to yield good performance (Attila, 2017; Devika et al., 2016). One of the most popular word embeddings used in prior studies is Word2Vec coupled with LSTM to generate the sentiment score from tweets and news as sentiment features (Mohanty et al., 2018; Vo et al., 2019). Sentiment features produced by the Word2Vec deep learning models have yet to show any significant effect on the cryptocurrency price prediction models. Furthermore, dynamic embeddings or pre-trained language models have also been scarcely explored in the realm of sentiment analysis for cryptocurrency news especially in more than one language.

Thus, the goal of this paper is to explore the development and evaluation of a deep learning sentiment model, GRU with the latest transformer language model, Generalized Autoregressive Pretraining for Language (XLNet) for sentiment detection and scoring in English and Malay cryptocurrency news. XLNet is chosen because it has shown to yield better performance than BERT as XLNet considers all possible permutations of the factorisation order instead of the fixed forward-backward factorisation order in BERT and thus can avoid the pretrain-finetune discrepancy faced in BERT (Devlin et al., 2019; Yang et al., 2020).

First, we present the architecture of our XLNet-GRU model for sentiment analysis in English and Malay news respectively. Second, we evaluate the performance of our XLNet-GRU model against VADER (i.e., the most commonly used lexicon-based sentiment scoring method) and currently the most popular transformer language model

(i.e., Bidirectional Encoder Representations or BERT). In addition, we also compare our results with other three state-of-the-art (SOTA) pre-trained language models including A Lite BERT (ALBERT) (Lan et al., 2020), Pre-training Text Encoders as Discriminators Rather Than Generators (ELECTRA) (Clark et al., 2020), and Robustly Optimized BERT Pre-training Approach (RoBERTa) (Liu et al., 2019). To our knowledge, this is the first study investigating sentiment analysis from Malay news sources (resource poor language) in addition to English news. Second, we review and compare the overall performance of the XLNet-GRU model across English and Malay languages.

We also contribute by creating two news sentiment corpora specifically on the topic of Bitcoin and Ethereum, one corpus in English and the second corpus in Malay. Both news sentiment corpora are manually labeled and carefully curated to serve as training and test sets for the XLNet-GRU sentiment model to perform sentiment regression. The cryptocurrency domain has given birth to a rich set of terminologies and jargons, making its news vocabulary to be significantly different from general, financial and economic news. As such, sentiment models trained using general or financial news may not generalize well on cryptocurrency news. Therefore, our cryptocurrency news sentiment corpora make an important contribution to advance the development of sentiment models particularly in the new but growing cryptocurrency domain.

2.   Related Work

A majority of prior studies applying the lexicon-based approach focused on VADER and Textblob for sentiment scoring and classification. Stenqvist & Lönnö (2017) applied VADER to first score the sentiment of each tweet and then determined the sentiment class (positive or negative). The sentiment labels were then aggregated based on time series intervals to be used as features in cryptocurrency price prediction. Similarly, Valencia et al. (2019) also used VADER to extract positive, negative, neutral and compound scores from tweets. Prajapati (2020) applied VADER, Textblob and Flair to obtain multiple sentiment views on English news for cryptocurrency price prediction while Chin & Omar (2020) utilized the NTUSD dictionary (Chen et al., 2018). However, these studies did not directly evaluate the performance of the sentiment scoring method before feeding the sentiment feature into the cryptocurrency price prediction model. Instead, the cryptocurrency price prediction results (i.e., the downstream task) is used as a proxy to assess the effectiveness of the sentiment features.

More recent studies have also ventured into applying static word embeddings such as Word2Vec and FastText as the representation on cryptocurrency-related texts, but mainly in English language. Mohanty et al. (2018) and Vo et al. (2019) first labeled news with sentiment based on the rise and fall of the cryptocurrency prices (i.e., increasing price represented positive sentiment and decreasing price represented negative sentiment). The news with the sentiment labels were then transformed into sentiment feature vectors using the Word2Vec embeddings in both

studies. These studies also did not evaluate the accuracy of the sentiment feature extraction method and used the cryptocurrency price prediction results as an indirect measure to assess the effectiveness of the sentiment features.

The use of contextual word embeddings such as BERT and ULMFiT language models to perform sentiment analysis for cryptocurrency price prediction is currently still limited. Cerda (2021) performed two types of experiments, one for sentiment analysis and the second for stance detection. For sentiment analysis, SentiWordNet was used to generate the sentiment score (continuous label ranging from -1 to 1) while for stance detection, the pre-trained Universal Language Model Fine-tuning (ULMFiT) and BERT models were fine-tuned individually as the features to perform the sentiment classification task. Then, the features were fed into an Extreme Gradient Boosting (XGBoost) model. Both ULMFiT and BERT applied with the XGBoost model produced positive results with F1-scores of 0.62 and 0.67 respectively in classifying sentiment. ULMFiT was good in identifying positive and neutral tweets but performed badly in classifying negative tweets as it had the tendency to misclassify the negative ones as positive. On the contrary, BERT was reported to show better classification performance for positive, neutral and negative tweets compared to ULMFiT.

Newer pre-trained language models have been introduced since BERT but have yet to be explored in sentiment analysis for cryptocurrency news. Prior studies are even rarer in the Malay language. In this paper, we propose a news sentiment regression model applying XLNet as the text representation to be fed into the GRU deep learning model for sentiment regression of English and Malay news. As BERT has shown good performance in sentiment classification, BERT is used as the comparison with our sentiment model together with ALBERT, ELECTRA and RoBERTa.

3.   Data

3.1   Data Collection

Cryptocurrency news on Bitcoin and Ethereum in English and Malay were collected for a duration of one year from 1 January 2021 until 31 December 2021.

NewsAPI[1] was used to extract English cryptocurrency news from various sources such as Reuters, Forbes, Yahoo, and New York Times. Using NewsAPI, we extracted English online news based on the queries: "(Bitcoin OR BTC)" and "(Ethereum OR ETH)". Parsehub was utilised to extract Malay news from various Malay online news sites such as Intraday.my[2], Berita Harian[3], Utusan Malaysia[4], and Harian Metro[5] (i.e., local Malay news sites). Parsehub extracted Malay news from Intraday using the web page URLs. The data collection process is illustrated in Figure 1.

The raw data extracted was filtered by splitting into two different corpora: English news corpus and Malay news corpus. Table 1 shows the number of English and Malay news for Bitcoin and Ethereum. The purpose for the split

---

was to allow sentiment analysis to be performed separately on each language.

Only the news headlines were included in our analysis as news headlines sufficiently capture the main points of the topic as opposed to the news content that may introduce unnecessary noise into the sentiment model. Any non-English and non-Malay instances were removed. Tokenization, lemmatization, and special characters removal were applied to each corpus using natural language processing tools appropriate for each language. Only the pre-processed text would then proceed with sentiment scoring.
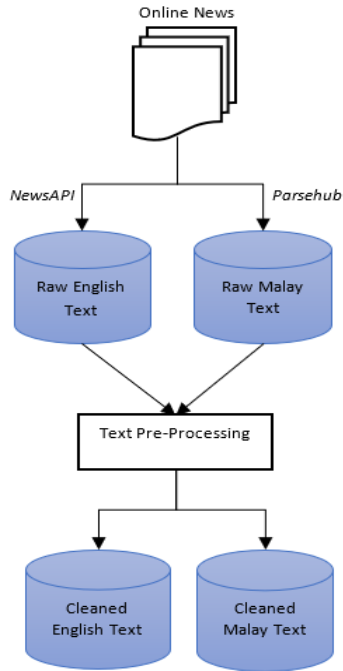

Figure 1: Data extraction and pre-processing.

News-Topic	English	Malay
Bitcoin	1518	1521
Ethereum	1205	1453

Table 1: Number of news documents by language and topic.

3.2 Data Annotation

To obtain training and test data for the sentiment model, manual annotation was performed by three annotators including the primary researcher. All three annotators worked on annotating both English and Malay news. The annotators were required to have basic understanding on the cryptocurrency topic and must be able to read and comprehend English and Malay languages. The annotation team was made up of one female and two male university students and lecturer between the age of 24 – 31. All annotators were well-versed in both English and Malay.

Each news headline was labelled with sentiment polarity scores ranging from -1 to +1 with one decimal place (-1: very negative, +1: very positive and 0: neutral).

Sentiment scores within the range of 0.1 – 0.3 would be considered low positive, 0.4 – 0.6 to be moderately positive and 0.7 – 1 very positive. Similar ranges were applied to the negative sentiment scores. We chose a numeric scale as the sentiment output instead of discrete polarity classes to capture more nuanced sentiment expressions and features from the news headlines. The following shows examples of news headlines for Bitcoin being assigned with the positive and negative sentiment scores.

Example 1 – Positive Sentiment
Bitcoin: Tesla invests about $1.50 billion in bitcoin – Reuters.com
[Sentiment Score: 1]

Example 2 – Negative Sentiment
Bitcoin: Chinese local government auctions seized bitcoin mining machines
[Sentiment Score: -0.8]

A codebook describing the annotation task and examples was provided to the annotators. All annotators were trained using the same set of samples and disagreements were resolved through discussion to obtain the final sentiment score. Training was conducted for several rounds until the percentage agreement reached greater than 50%. Once the expected percentage agreement was achieved, each remaining news headline would be annotated by at least two annotators and the final sentiment score would be computed by taking the mean score across all annotators.

Inter-annotator reliability is computed using Krippendorff's alpha (Krippendorff, 2018) to measure the agreement between annotators for the continuous labels. Table 2 depicts the inter-annotator reliability measures achieved for English and Malay news headlines.

Inter-rater-Metric	English	Malay
Agreement %	63%	61%
Krippendorf's	0.58	0.56

Table 2: Inter-annotator reliability measures.

The alpha within the range of 0.56 to 0.58 indicates acceptable agreement between annotators given the subjectivity and complexity of the annotation task. We observed that the sentiment scores assigned by annotators most often only vary less than ±0.2.

4. Methodology

4.1 Deep Learning Architecture for Sentiment Regression

Figure 2 illustrates the XLNet encoding architecture to produce the vector representation. The news headlines were first tokenized using the XLNet pre-trained word embedding ('xlnet-base-cased'[6] for English text) and 'xlnet-base-bahasa-cased'[7] for Malay text) with 12 layers of transformer blocks, 768 hidden layers (dimensions), and 12 self-attention heads (Gong et al., 2019). Next, the encoding was performed with the use of attention mask where the permutation language modelling took place. This

---

[6] English XLNet: https://huggingface.co/xlnet-base-cased

[7] Malay XLNet: https://huggingface.co/malay-huggingface/xlnet-base-bahasa-cased

allowed bidirectional contexts to be captured for the positional encoding based on factorization order instead of the sequence order. Then, the context vector produced was utilised for the XLNet fine-tuning (training) process.

Fine-tuning was required to learn the context of the word with the newly assigned sentiment weights based on cryptocurrency-related sentiment from the labeled data. The fine-tuned XLNet model was then incorporated into the GRU sentiment deep learning model. The GRU sentiment deep learning architecture is shown in Figure 3.
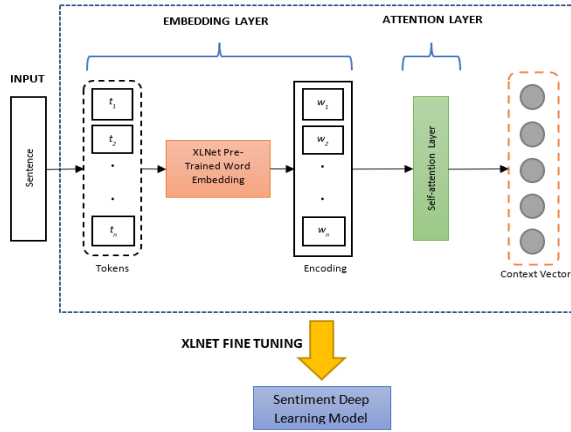

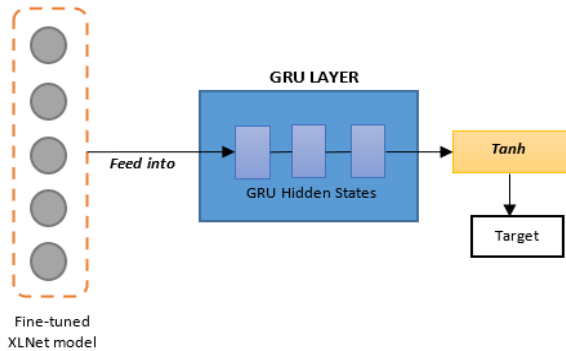Figure 2: XLNet encoding architecture.

Figure 3: GRU Sentiment deep learning architecture.

From Figure 3, the pre-trained XLNet word embedding that had gone through finetuning was fed into the GRU layer consisting of 768 hidden nodes, one output layer and a dropout rate of 0.1. Tanh was used as the activation function as it is suitable for the sentiment regression output values of between -1 to +1.

Grid search with 5-fold cross validation was used to find the most optimized hyperparameter settings for the XLNet-GRU sentiment regression model. The range of hyperparameter values included in the grid search during the optimization process are listed as follows:

- Batch size: {8, 16, 32, 64}
- Learning rate: {2e-5, 3e-5, 3e-4}
- Number of epochs: {30, 40, 50}

Finally, a batch size of 8 was utilized for the AdamW optimizer with the learning rate of 2e-5 and 30 epochs for training.

Each labeled corpus was first split into 80% as the training (finetuning) set and 20% as the test set. From each corpus, 200 samples were reserved for the test set for each coin and the remaining samples were used for training and validation (~1500 instances). The training set was fed into the XLNet-GRU sentiment regression model for training.

Adjusted R^2 was used as the primary performance metric to measure how well the predicted values fit to the original values. The higher the value of adjusted R^2, the better the model performance. In addition, we also report RMSE and MAE as error measures (i.e., lower error means better model performance).

5. Results and Analysis

5.1 English Sentiment Models

The VADER compound score (lexicon-based approach) is used as a simple baseline for comparison in English news. VADER is chosen as the baseline for lexicon-based approach because it is the most common lexicon-based method encountered in existing studies.

We also set up the BERT, ALBERT, ELECTRA and RoBERTa pre-trained language models in our experiments for the English model as they served as SOTA language models found in existing studies (Farha & Magdy, 2021; Pranesh et al., 2020). BERT ('bert-base-cased')[8] was fine-tuned to learn word contexts from the English news sentiment corpus with the GRU deep learning model (BERT-GRU) We also fine-tune ALBERT ('albert-base-v2')[9], ELECTRA ('google/electra-small-discriminator')[10], and RoBERTa ('roberta-base')[11] to be incorporated into the GRU deep learning model. Table 3 shows the results obtained for English VADER, BERT-GRU, ALBERT-GRU, ELECTRA-GRU, RoBERTa-GRU, and XLNet-GRU.

ENGLISH CRYPTOCURRENCY NEWS			
Bitcoin			
Model	RMSE	MAE	Adjusted R^2
VADER	0.483	0.388	0.081
BERT-GRU	0.346	0.219	0.527
ALBERT-GRU	0.527	0.428	-0.095
ELECTRA-GRU	0.356	0.229	0.499
RoBERTa-GRU	0.325	0.209	0.583
XLNet-GRU	**0.296**	**0.185**	**0.654**
Ethereum			
Model	RMSE	MAE	Adjusted R^2
VADER	0.423	0.324	-0.139
BERT-GRU	0.257	0.154	0.582
ALBERT-GRU	0.395	0.333	0.007
ELECTRA-GRU	0.292	0.162	0.459
RoBERTa-GRU	0.277	0.149	0.512
XLNet-GRU	**0.249**	**0.131**	**0.607**

Table 3: Model performance for English cryptocurrency news.

---

[8] English BERT: https://huggingface.co/bert-base-cased
[9] English ALBERT: https://huggingface.co/albert-base-v2

[10] English ELECTRA: https://huggingface.co/google/electra-small-discriminator
[11] English RoBERTa: https://huggingface.co/roberta-base

Based on Table 3, VADER achieved adjusted R^2 of 0.081 and -0.139 for Bitcoin and Ethereum respectively. The baseline results utilizing purely a lexicon-based approach indicate very low accuracy. The negative adjusted R^2 value is treated as 0, which signifies poor fit as sentiment words in the VADER dictionary is made up of common sentiment words in general and not catered specifically to handle cryptocurrency-related sentiment words. The same negative adjusted R^2 value is also observed for Bitcoin using ALBERT-GRU whereas, an adjusted R^2 of 0.007 is observed for Ethereum. Although a positive value was obtained for Ethereum, the performance is extremely low.

On the contrary, the adjusted R^2 in BERT-GRU, RoBERTa-GRU and XLNet-GRU models demonstrate more promising results. For Bitcoin, the RoBERTa-GRU model obtained an adjusted R^2 of 0.583, while our XLNet-GRU model achieved a higher score of 0.654. Similar observation applies to our XLNet-GRU model for Ethereum with an adjusted R^2 of 0.607. Surprisingly, BERT-GRU (adjusted R^2 = 0.582) manages to surpass RoBERTa-GRU for Ethereum.

The results clearly show that our XLNet-GRU model consistently yields the best performance in sentiment regression on English cryptocurrency news for both Bitcoin and Ethereum in comparison to VADER and the other SOTA pre-trained language models.

5.2 Malay Sentiment Models

VADER cannot be directly applied to Malay text as it only supports English. For Malay, we use the Open Multilingual WordNet[12] to first retrieve the English synonym to each Malay word in a news headline before feeding the English synonyms into VADER for sentiment scoring. WordNet-Bahasa[13] from the Open Multilingual WordNet is used to recognize the Malay words.

To set up the SOTA models for Malay, BERT ('bert-base-bahasa-cased')[14] and ALBERT ('malay-huggingface/albert-base-bahasa-cased')[15] pre-trained on Malay text were fine-tuned to learn the context words from the Malay sentiment corpus and then fed into a separate GRU deep learning model. VADER, BERT-GRU, and ALBERT-GRU were evaluated against our XLNet-GRU model for Malay cryptocurrency news. As there is no pre-trained language model available for ELECTRA and RoBERTa in Malay, both are excluded for comparison. The results on Malay cryptocurrency news are shown in Table 4.

From Table 4, VADER shows the poorest performance as indicated by the negative adjusted R^2 scores for both Bitcoin and Ethereum. Such low performance is attributed to the lack of reliable Malay sentiment lexicons, especially those containing words that are related to cryptocurrency. On the other hand, ALBERT-GRU shows improved performance than VADER but still yields a negative adjusted R^2 of -0.119 for Bitcoin.

BERT-GRU and XLNet-GRU fare better in scoring sentiment for Malay cryptocurrency news. The performance scores between BERT-GRU and XLNet-

GRU for Malay cryptocurrency news show only a small difference, particularly for Ethereum. For Bitcoin, the XLNet-GRU model achieved better performance in RMSE (reduction in error) and adjusted R^2 (increase in fit between the predicted and actual scores) compared to BERT-GRU. While both BERT-GRU and XLNet-GRU reported MAE scores with a very slight difference, the larger RMSE score observed in BERT-GRU compared to XLNet-GRU indicate a greater variance in error. Therefore, we can conclude that XLNet-GRU still yields performance advantages in Bitcoin news.

MALAY CRYPTOCURRENCY NEWS			
Bitcoin			
Model	RMSE	MAE	Adjusted R^2
VADER	0.584	0.471	-0.590
BERT-GRU	0.364	**0.252**	0.383
ALBERT-GRU	0.490	0.458	-0.119
XLNet-GRU	**0.351**	0.255	**0.428**
Ethereum			
Model	RMSE	MAE	Adjusted R^2
VADER	0.598	0.485	-0.956
BERT-GRU	**0.271**	**0.144**	0.597
ALBERT-GRU	0.327	0.209	0.414
XLNet-GRU	**0.271**	0.168	**0.599**

Table 4: Model performance for Malay cryptocurrency news.

As for Ethereum, XLNet-GRU shows slightly higher MAE than BERT-GRU but the greater difference between RMSE and MAE of BERT-GRU indicates that BERT-GRU still suffers from a greater variance in error, which is consistent to our findings from Bitcoin. XLNet-GRU still achieved adjusted R^2 of 0.599, which is slightly better than BERT-GRU with adjusted R^2 of 0.597.

From the overall results, XLNet-GRU is still deemed the winner for Malay cryptocurrency news as it achieved the best RMSE and adjusted R^2 scores.

5.3 Comparing English and Malay Sentiment Models

To compare the XLNet-GRU in both Malay and English languages, the mean RMSE, MAE and adjusted R^2 are calculated across Bitcoin and Ethereum and shown in Table 5.

English XLNet-GRU			
	RMSE	**MAE**	**Adjusted R^2**
Bitcoin	0.296	0.185	0.654
Ethereum	0.249	0.131	0.607
Mean	**0.273**	**0.158**	**0.631**
Malay XLNet-GRU			
Bitcoin	0.351	0.255	0.428
Ethereum	0.271	0.168	0.599
Mean	**0.311**	**0.212**	**0.514**

Table 5: Comparison of results achieved for English and Malay texts using XLNet-GRU.

---

[12] Open Multilingual WordNet:
http://globalwordnet.org/resources/wordnets-in-the-world/
[13] WordNet Bahasa: http://wn-msa.sourceforge.net/

[14] Malay BERT: https://huggingface.co/malay-huggingface/bert-base-bahasa-cased
[15] Malay ALBERT: https://huggingface.co/malay-huggingface/albert-base-bahasa-cased

Based on the comparable mean error and adjusted R^2 scores, we can conclude that XLNet-GRU shows fairly consistent performance across both English (mean adjusted $R^2 = 0.631$) and Malay (mean adjusted $R^2 = 0.514$) with only slightly better performance observed in the English model mainly due to the availability of more training data in the English news sentiment corpus. The results could also imply that the pre-trained English language model contains vocabulary that is more relevant to cryptocurrency terms in comparison to Malay. Thus, our method proves that it is possible to create a Malay sentiment model that is comparable in terms of performance to an English sentiment model despite the more limited language resources in Malay.

6. Conclusion

To conclude, we presented a XLNet-GRU model to perform sentiment regression for cryptocurrency news in English and Malay. Our XLNet-GRU sentiment regression model applies the latest XLNet transformer-based contextual word embedding for both English and Malay cryptocurrency news (Bitcoin and Ethereum). XLNet is a new pre-trained language model, which has not been explored in sentiment analysis of cryptocurrency news. Our experiment results show that XLNet-GRU outperforms BERT-GRU and other SOTA baselines as well as the naïve lexicon-based baseline, VADER. The performance of XL-Net-GRU is comparable in both English and Malay news.

To the best of our knowledge, this is the first study experimenting with a deep learning sentiment model for Malay cryptocurrency news. In addition, we also curated an English sentiment corpus and a Malay sentiment corpus specifically in the cryptocurrency news domain, which can serve as the benchmark to evaluate the quality of sentiment features extracted in the intermediate step within the cryptocurrency price prediction pipeline. We hope to release and share the cryptocurrency news sentiment corpora for the benefit of future research in the financial domain.

For future work, we hope that this cryptocurrency sentiment corpora will motivate further research through experimentation with other contextual word embeddings and deep learning methods particularly for Malay (i.e., the poor language resource). The XLNet-GRU model can be applied and evaluated on other domains and text sources such as tweets.

7. Acknowledgement

8. References

Attila, S. D. (2017). Impact of social media on cryptocurrency trading with deep learning. *Scientific Students' Conference*, 47.

Cerda, G. N. C. (2021). *Bitcoin price prediction through stimulus analysis: On the footprints of Twitter's crypto-influencers* [Master's Thesis, Pontificia Universidad Católica de Chile].

Chen, C.-C., Huang, H.-H., and Chen, H.-H. (2018). NTUSD-Fin: A market sentiment dictionary for financial social media data applications. *Proceedings of the 1st Financial Narrative Processing Workshop (FNP 2018)*, 37–43.

Chin, C. K., and Omar, N. (2020). Bitcoin price prediction based on sentiment of news article and market data with LSTM model. *Asia-Pacific Journal of Information Technology and Multimedia*, *9*(1), 1–16.

Clark, K., Luong, M.-T., and Le, Q. V. (2020). *ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators*, 18.

Devika, M. D., Sunitha, C., and Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, *87*, 44–49.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.

Farha, I. A., and Magdy, W. (2021). Benchmarking Transformer-based Language Models for Arabic Sentiment and Sarcasm Detection. *Proceedings of the 6th Arabic Natural Language Processing Workshop*, 21–31.

Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D. N., and Giaglis, G. M. (2015). Using time-series and sentiment analysis to detect the determinants of Bitcoin prices. *SSRN Electronic Journal*.

Gong, X.-R., Jin, J.-X., and Zhang, T. (2019). Sentiment analysis using autoregressive language modeling and broad learning system. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1130–1134.

Gurdgiev, C., and O'Loughlin, D. (2020). Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty. *Journal of Behavioral and Experimental Finance*, *25*, 100271.

Karalevicius, V., Degrande, N., and De Weerdt, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance*, *19*(1), 56–75.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (Fourth). SAGE Publications.

Lamon, C., Nielsen, E., and Redondo, E. (2017). Cryptocurrency price prediction using news and social media sentiment. *SMU Data Science Review*, 1–22.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2020). *ALBERT: A Lite BERT for self-supervised learning of language representations*. 1–17.

Li, M., Lu, Q., Long, Y., and Gui, L. (2017). Inferring Affective Meanings of Words from Word Embedding. *IEEE Transactions on Affective Computing*, *8*(4), 443–456.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv:1907.11692 [Cs]*.

Loughran, T., and Mcdonald, B. (2014). Measuring Readability in financial disclosures. *The Journal of Finance*, *69*(4), 1643–1671.

Mai, F., Bai, Q., and Shan, J. (2015). The impacts of social media on Bitcoin performance. *International Conference on Information Systems (ICIS)*, 1–16.

Mohanty, P., Patel, D., Patel, P., and Roy, S. (2018). Predicting fluctuations in cryptocurrencies' price using users' comments and real-time prices. *7th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, 6.

Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., and Pintelas, P. (2020). Investigating the problem of cryptocurrency price prediction: A deep learning approach. In I. Maglogiannis, L. Iliadis, and E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (Vol. 584, pp. 99–110). Springer International Publishing.

Prajapati, P. (2020). Predictive analysis of Bitcoin price considering social sentiments. *ArXiv:2001.10343 [Cs]*.

Pranesh, R., Kumar, S., and Shekhar, A. (2020). CLPLM: Character Level Pretrained Language Model for ExtractingSupport Phrases for Sentiment Labels. *Proceedings of the 17th International Conference on Natural Language Processing*, 475–480.

Rognone, L., Hyde, S., and Zhang, S. S. (2020). News sentiment in the cryptocurrency market: An empirical comparison with Forex. *International Review of Financial Analysis*, *69*, 101462.

Stenqvist, E., and Lönnö, J. (2017). *Predicting Bitcoin price fluctuation with Twitter sentiment analysis* [Degree Project]. KTH Royal Institute of Technology School of Computer Science and Communication.

Valencia, F., Gómez-Espinosa, A., and Valdés-Aguirre, B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy*, *21*(6), 589.

Vo, A.-D., Nguyen, Q.-P., and Ock, C.-Y. (2019). Sentiment analysis of news for effective cryptocurrency price prediction. *International Journal of Knowledge Engineering*, *5*(2), 47–52.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., and Le, Q. V. (2020). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems (NeurIPS 2019)*, 1–11.