Leveraging Latent Economic Concepts and Sentiments in the News for Market Prediction

Saeede Anbaee Farimani
Islamic Azad University
Mashhad, Iran
anbaee@mashdiau.ac.ir

Majid Vafaei Jahan
Islamic Azad University
Mashhad, Iran
VafaeiJahan@mshdiau.ac.ir

Amin Milani Fard
New York Institute of Technology
Vancouver, Canada
amilanif@nyit.edu

Gholamreza Haffari
Monash University
Melbourne, Australia
gholamreza.haffari@monash.edu

***Abstract*—Most of the existing news-based market prediction techniques disregard conceptual and emotional relations in the news stream. In this work, we consider the conceptual relationship between news documents using contextualized latent concept modeling as well as leveraging news sentiment and technical indicators. We present our approach as an open-source RESTFul API. We build a corpus of financial news related to currency pairs in the Foreign Exchange and Cryptocurrencies markets. Next, we apply BERT-based embedding to generate word vectors, cluster the vectors to create latent economic concepts, and propose a document representation based on the distribution of words on these concepts as well as news sentiment. We use a recurrent convolutional neural network to jointly use BERT-based text representation and technical indicators embedding for market time series prediction. We further augment our model with technical indicators using another recurrent layer. The experimental results show the superiority of our method compared to the baselines. Our *MarketNews* dataset, news crawler, and *MarketPredict* APIs are available for public use.**

***Index Terms*—Market prediction, latent concept modeling, news representation, BERT, convolutional neural network, sentiment analysis.**

I. Introduction

Applying deep learning for financial time series forecasting is a challenging task since financial data can exhibit non-stationary behavior over time due to a number of behavioral issues released in newsgroups and social media especially in enormous cryptocurrencies and Foreign Exchange (Forex) markets [31]. Leveraging human expert generated data in financial newsgroups and historical market information can help with better market decision making for investors [9], [13], [25], [27], [29], [37]. Most of previous methods only used information available in news title for sentiment analysis or directly embedding of news title [9], [29], [34], [35]. Recent methods [4], [21], [38] use word embedding representation of the news *title* only, through pre-trained models, while it has been shown that the information in the news *body* [24] is important as well. An approach for using all information in the news content is modeling concepts based on the probabilistic features of the occurrence of contextual words that are close to each other with word embedding. Latent concept modeling aims at creating low dimensional vectors to capture context proximity in the embedded space [16]. A latent concept refers to a cluster of semantically related words that occur in similar contexts [44]. For instance, the word 'Brexit' often occurs with 'political', 'UK', 'exit' and 'parliament' in one concept.

In this work, we propose a market prediction method that leverages latent concepts in news, called BERT-based Bag-of-Economic-Concepts (BERT-BoEC). It captures temporal characteristics in distribution of latent concepts in news stream as well as sentiment score of news titles by applying a recurrent convolutional neural network (RCNN) and also extract informative features from technical indicators. For concept modeling, we first generate vector representations of words based on semantic relationships in news using the Bidirectional Encoder Representations from Transformer (BERT) based [5] contextualized word embedding. We then cluster the vectors and consider each cluster as a latent concept. Then, document vectorization is done based on the words distribution, over the latent concepts. For sentiment analysis of news titles, we use pre-train DistilBERT [33]. We formulated our work as a regression problem for market price prediction. Our contributions include:

- We present a contextualized concept modeling for news document representation based on Bag-of-Concepts (BoC) [16], called Bag-of-Economic-Concepts (BoEC) for fundamental financial news analysis.
- We propose a deep learning approach that leverages information in news and historical market data, and use a pre-train model for sentiment analysis of news titles.
- We implemented our technique in a tool called *MarketPredict*[1] as an open-source RESTFul API that includes a scheduling module for news scraping services and prediction services that run hourly to POST the predicted price value on MongoDB storage services.

[1]https://github.com/MarketPredict-BoEC/MarketPredict-RESTFul-API

- We build a *MarketNews* dataset related to Forex and cryptocurrencies from specialized financial newsgroups.
- Our Github repository includes datasets, implementation of our crawler, *MarketPredict* services, and our service-oriented architecture description in Postman.

Apart from the application of *MarketPredict* in decision support, the research community can benefit from our APIs to develop and test their predictive models for specific currency pair as well as using our data collecting services through our scrapper tool.

II. Related Work

Text mining and machine learning have been used for knowledge extraction from unstructured text data to improve market prediction [11], [19], [23], [29]. Some researchers proposed sentiment analysis methods for analysis of financial news titles [3], [10], [26] and examined social media mood influence on investors' decisions [14], [30], [40], [42]. However, such methods mainly focus on the news title sentiment analysis and disregard the relationship between the news. Also, some researchers used news analysis via word embedding or jointly use of news representation and sentiment analysis [22], [45]. Huynh et al. [13] consider the semantic relationships among the words and use word embedding [28] within Bidirectional Gated Recurrent Unit (BGRU) [12] to forecast S&P500 stock market. This method uses bidirectional relations between words in news titles. Lutz et al. [26] used the Doc2vec embedding technique [20] for predicting the sentence sentiment in their financial decision support system. The generated low dimensional vectors capture the proximity in the news title and description; however, the meaning of each feature is not explainable.

To extract explainable features from the news, [15] used BERT [5] to capture deep semantic information. Chen et al. [4] proposed a news aggregating method for Forex market prediction based on BERT. They use *[CLS]* token as the vector representation of news title, while loosing important information in the news content. Other news representation works, such as [2], [8] use a transformer layer [39] for feature extraction from news title and use other information of user preferences or tags and images available in news content. In [6], [7] a knowledge graph embedding is used for feature extraction from new titles and [36] use a mixture of embedding of news words and knowledge graph of news title entities, while word2vec and Glove [20] embedding do not accurately reflect relevance among news because there exists polysemy problem in these methods [2].

Unlike previous work, in this paper, we proposed a context-aware conceptual document representation to model the relevance between the news based on all information in financial news titles and bodies. However, by applying a RCNN model, we utilize the temporal changes in latent concepts and sentiment from news and temporal changes in more informative public market data including technical indicators during news publishing time.

III. The Proposed Approach

We propose BERT-BoEC for financial market decision support. Our approach consists of three phases, including *latent concept modeling*, *feature extraction and fusion*, and *market prediction*, as depicted in Figure 1.

In the latent concept modeling phase, we first build embedding vectors for words in the news corpus and then cluster these vectors to construct latent economic concepts. We construct economic concepts based on contextualized word embedding of news to model the relevance between news distribution over the latent concepts. The idea of concept modeling is the same as Bag-of-Concepts (BoC) for document classification [16], which we apply to BERT word embedding for economic concept modeling. To the best of our knowledge, BoC that is based on word2vec [20] has not been studied or used for fundamental financial news analysis.

In the feature extraction and fusion phase, we apply a Recurrent Convolutional Neural Network (RCNN) model for feature extraction from the news that vectorizes based on our BERT-BoEC method over latent concepts and sentiment score of news. In addition, we select more informative market data and use a Recurrent Neural Network (RNN) layer for feature extraction from market data and technical indicators.

Finally, for the fusion part we concatenate extracted features from news and market data, which we previously aligned them based on timestamps in chronological order and finally predict the price values.

A. Latent Concept Modeling

The latent concept modeling phase has 2 steps. In the first step, we construct the contextualize embedded space $\mathcal{R}^m$ using BERT [5] (block 1). For contextualized word embedding, we choose the Tensorflow pre-trained BERT implementation [2] and use the BERT-base-uncased [3] version with 768 hidden state and 12 attention. We sum the output of top four layers of BERT and truncate each sentence in news documents to 128 tokens. We set batch size to 128. In the second step, we build latent concepts (block 2). Let us denote the context-dependent embedding vectors for the i-th token in the j-th sentence by $\mathcal{R}^m$. We construct context-aware concepts by clustering all embeddings $E[w_{ij}]$ in the corpus into k clusters using the k-means algorithm. Thus, each token in a sentence is assigned to the cluster with the closest cluster center. Let us denote the center of the l-th cluster by u_l. Then, the token w_{ij} is assigned to the cluster $c = \arg\min_l(E[w_{ij}].u_l)$. In our experiments we empirically determined the best value for k to be 210. We later explain in section V-B1 how we determine k.

B. Feature Extraction

1) News Based Feature Extraction: In our BoEC vectorization method (block 3), first the title and content of a news document is preprocessed by removing numbers, tags, URLs and stop words (lines 1-2). Next is title expansion in

[2]https://github.com/google-research/bert

[3]https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip

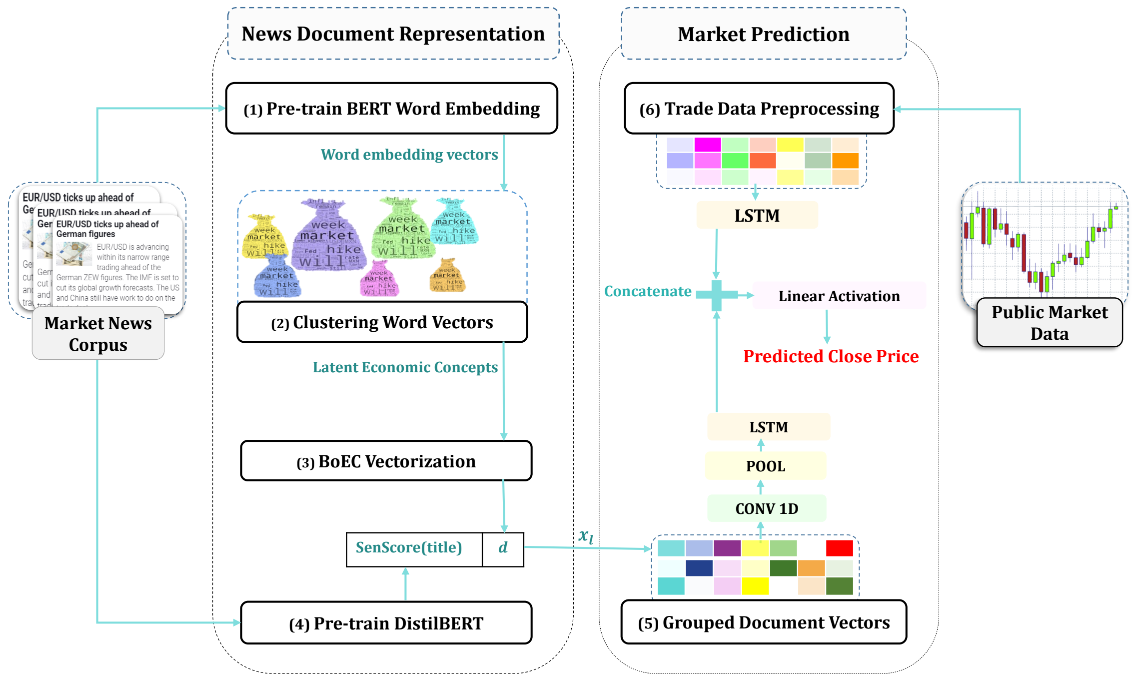


Fig. 1: Overview of the system.

Algorithm 1: Bag-of-Economic-Concept (BoEC)

input : News document D = {*title*, *content*, *timestamp*}, embedded latent concepts space $C^k = \{c_1, \ldots, c_k\}$, m dimensional embedding space $\mathcal{R}^m$, number of the most similar words to add n

output: Vector $\boldsymbol{d} \in C^k$ corresponding to the news document D

```
/* Step 1: feature selection */
1 titleKeywords ← PREPROCESS(title)
2 descKeywords ← PREPROCESS(content)
3 titleVectors ← titleKeywords corresponding vectors from R^m
4 extWords ← TOP_N_SIMILAR(titleVectors, n, R^m)
5 totalExtWords ← APPEND(titleKeywords, descKeywords,
    extWords)
/* Step 2: vectorization */
6 for i = 1 to k do
7   d[i] ← number of tokens in totalExtWords ∈ c_i
8 return d
```

---

order to strength the news topic (line 3-4). In line 3, vectors from $\mathcal{R}^m$ that correspond to words of the title are extracted and the subroutine TOP_N_SIMILAR() in line 4 returns top n most similar words to those vectors from $\mathcal{R}^m$ based on Cosine similarity. $totalExtWords$ holds all keywords in title, content and extended words ($extWords$). Finally, keywords distribution through latent concepts would be calculated (line 6). Hence, the dimension of the document vector $\boldsymbol{d}$ is identical to the number of latent concepts in embedding concepts space i.e., k.

In order to study the temporal characteristics of concepts distribution as well as emotions in news stream, we augment sentiment score and conceptual representation. For a news l we concatenate its BoEC representation $\boldsymbol{d}$, and the sentiment score of its title $SenScore(title_l) \in [-1, 1]$ into $x_l = [\boldsymbol{d}, SenScore(title_l)]$. We calculate the sentiment score of news title using pre-trained DistilBERT-base-uncased [4] [33] fine-tuned on SST-2 for sentiment analysis.

Let $X_N^{(t)}$ be a sequence $\{x_1, \ldots, x_L\}$ of BERT-BoEC representation of news published at time t, where $x_l \in C^{k+1}$ and L is the number of news in a delay window h arranged in chronological order (block 5). We later explain in section V-B1 how L is determined. We embed $X_N^{(t)}$ into a convolutional layer. Aim to capture conceptual and emotional relationships in a sequence of news, we use temporal convolution layer with a set of learnable filters called convolution kernels. Let us denote $W = \{w_1, \ldots, w_P\}$ to be a set of filters. For a news sequence $X_N^{(t)} = \{x_1, \ldots, x_L\}$, where L is the number of news, $w_P \in C^k$, and P is the length of the filter window, a one dimensional CNN performs $q_j = W.x_{j:j+P-1} + b$ at each possible window of the news vector to produce a feature map $Q = \{q_1, \ldots, q_{L-P+1}\}$ that is fed into a temporal max pooling function $h(q_j) = max\,\{0, q_p\}$, which captures the most important information in news sequence. Finally, to introduce non-linearity we use ReLU as the activation function for this layer output.

2) Market Based Feature Extraction: So far researchers use both raw data and technical indicators of the market, as such indicators can help in better decision makings [18], [43]. We select the initial set of market features including raw market data and technical indicators based on the amount of the information gains [32] of those features from the target variable. Then we monitor the model accuracy by recursive eliminating of features from low to high degree of

[4] https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english

informativeness of features. Let space S be our candidate set of market data, following [38], [43], [45] we use the close price, volume, Exponential Moving Average (EMA), Moving Average Convergence-Divergence (MACD), On Balance Volume (OBV), Bollinger Bands (BBs), Stochastic, Momentum, Williams, Relative Strength Index (RSI) and Average True Range (ATR) in z-score normalize forms.

Let $X_M^{(t)}$ be a matrix of prepossessed trading data corresponding to time t, $X_M^{(t)} \in S^{q*h}$, form q market data items of space S, during delay window of h (block 6). We later explain in section V-B2 how we determine q. We fed market data into RNN layer. In the recurrent layer, we use two separate LSTM units, one after the convolutional layer to learn time dependencies between a news sequence of $L-P+1$ time steps, and the other one for technical indicators inputs $X_M^{(t)} \in S^{q*h}$. For the RNN layer we use LSTM to store and access a longer range of contextual information in the sequential input, and also to handle the vanishing gradient problem. A single cell in LSTM has a cell state and three gates: an input gate i, a forget gate f, and an output gate o. Formally, the LSTM can be formulated as:

$$i_t = \sigma(W_i x_t + U_i \overrightarrow{h}_{t-1} + b_i)$$

$$f_t = \sigma(W_f x_t + U_f \overrightarrow{h}_{t-1} + b_f)$$

$$C_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t$$

$$O_t = \sigma(W_o x_t + U_o \overrightarrow{h}_{t-1} + b_o)$$

$$\hat{C}_t = W_c x_t + U_c h_{t-1} + b_c$$

$$h_t = O_t \odot \tanh(C_t)$$

where $X = x_1, x_2, \ldots, x_m$ is the input vector, W_i, W_f, W_c, W_o, b_i, b_f, b_c, b_o are the parameters as weight matrices and biases, $\odot$ is the sigmoid function, and $h_1, h_2, \ldots, h_m$ represent a sequence of semantic features.

C. Fusion and prediction

After news representation and market data prepossessing, we align news with market data based on timestamps in chronological order. Regarding extracted feature from $X_M^{(t)}$ via an RNN layer and $X_N^{(t)}$ via a recurrent convolution layer, we concatenate hidden states of two LSTM units as final features. Note that we align news and market data based on the timestamp in chronological order, therefore extracted features are based on news and market data during delay window of L arranged in the temporal order. Previous work [29], [35] have studied the investors' reactions to the Forex market within an hour from the news release. We also consider hourly time frame and put a simple dense layer at the last layer of our model with Linear activation function to predict the close price value of next hour. Since our task is formulated as a regression problem, we consider mean percentage absolute error (MAPE) loss function and use adaptive moment estimation (Adam) for minimization the loss value over all instance in training set.

IV. RESTful API and Dashboard

Figure 2 depicts our *MarketPredict* microservices architecture. All of our services work on RESTful API and use our MongoDB engine services for news, market data, concept clusters and prediction data. The current implementation of our tool supports the following three major components:

- *Data collection services* including news and public market information that is currently available for some currency pairs in Forex and Cryptocurrencies markets.
- *Model training services* that leverage latent relationships in the news stream along with market trading data through a Recurrent Convolutional Neural Network.
- *Prediction services* that provide forecasting of price for a given currency pair.

A. Data collection services

1) News Data Preparation: Our scraping microservices obtain news from well-known specialized financial newsgroups[5] that publish news related to currency pairs in the Foreign Exchange and Cryptocurrency markets. Every hour the scheduler module runs scraping microservices and POST the collected news to MongoDB storage services. Each news item contains title, content, timestamp (in Unix UTC timestamp).We remove records with duplicate news titles. The target currency pair is manually determined by human experts in the newsgroup. In our training services for training a currency pair predictive model, we *extract* news corpus by sending a GET request to our MongoDB storage services through related keywords parameter to that currency pair.

2) Market Data Preparation: We extract our market data from Finnhub REST API [6] that fetches Forex data from FXCM broker [7] and Cryptocurrencies data from Binance crypto Exchange [8].In our training services, we *extract* the candlestick data by sending a GET request to Finnhub. We then calculate technical indicators with TA python package[9] and used the default values.

B. Model training services

In our training service, for data *extraction*, we send a GET request to two microservices of news and market data provider. Next, we apply document vectorization and indicator calculation as data *transformation* and align the news vectors and market data based on their publishing timestamps in chronological order. Then *loads* them into training process. Finally, we train our BoEC model and save it for future prediction.

[5]www.fxstreet.com and www.newsbtc.com and www.cointelegraph.com and www.investing.com
[6]https://finnhub.io
[7]www.fxcm.com
[8]https://www.binance.com
[9]https://technical-analysis-library-in-python.readthedocs.io/en/latest/

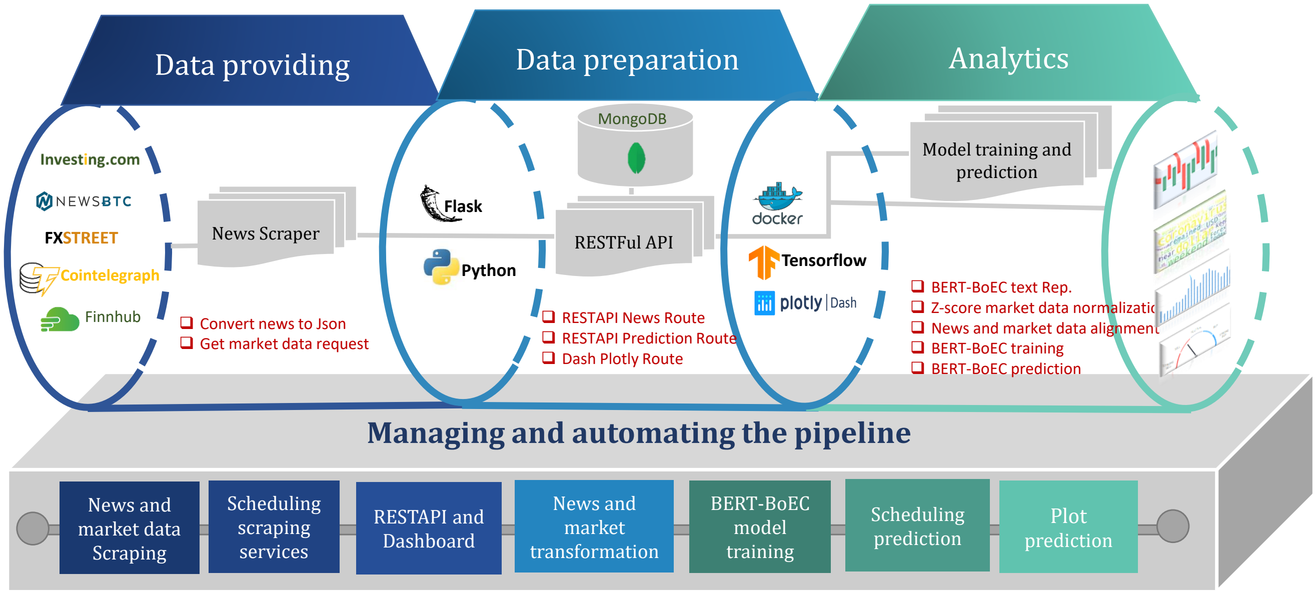


Fig. 2: Data pipeline from news and market data scraping to prediction plot.

C. Prediction Services

Our scheduler module calls prediction services every hour, which in turn provide the data from our data providing micro services, transformed the data same as our training services, load the trained model and send the predicted result by POST request to MongoDB engine services. Once the model is trained for a particular target currency pair, the user can send a GET request for prediction. We send the predicted price values as a user response based on the timestamps of the user request.

D. Front-end Dashboard

In order to visualize the market data and plot predictions as well as market candlestick chart for all currency pairs, we use Dash [10] python package as a tool for front-end development.

V. Experiments

A. Dataset

News dataset. The *MarketNews* dataset contains over 100,000 unique records of 33-month of data from September 2018 to May 2021. We explore subsets of news in our *MarketNews* dataset with keyword EUR/USD, GBP/USD, USD/JPY and BTC/USD (indicated by human expert in newsgroups) for training our predictive model as shown in Table I. Note that the distribution of news among days of the week is not uniform. For currency pairs in the Forex market, few news are released in holidays.

Market dataset. Each record is time-stamped in the GMT time zone every 60 minutes and contains Low, High, Open, Close price as well as the trading Volume information as shown in Table I. We use Binance:BTC/USDT for bitcoin price prediction.

[10]https://dash.plotly.com

B. Experiment Setting

As for our model parameters, we set the kernel size in CNN to 3 and we use 64 filter units with stride 1 and same padding. In the recurrent layer, we use two separate LSTMs with 128 units. For optimizer, we used Adam [17] with initial learning rate of 0.001. The batch size is set to 32. The minimum number of training epoches is 30 with early stopping by monitoring validation set loss. For the training objective, we use Mean Absolute Percentage Error with L2 regularization. The weight of L2 regularization is 0.015. We use Keras in Tensorflow 2 for the implementation. In our experiment we set delay window to later 7 time steps. We work with *hourly* trading data and use 60% as training samples, 20% as validation, and 20% as test set.

We build a corpus consisting of all news related to the target currency pair in Forex or cryptocurrency markets and then apply preprocessing steps on the corpus to removes numbers, URLs, tags and stopwords. We then insert [*CLS*] token before each sentence and [*SEP*] token between adjacent sentences. Next, we apply embedding and clustering. For contextualized word embedding, we choose the pre-trained BERT implementation [11] and use the BERT-base-uncased [12] version with 768 hidden state and 12 attention. We sum the output of top four layers of BERT and truncate each sentence in news documents to 128 tokens. For building embedded latent concepts space, we exploit k-means++ in Sklearn package. Regarding our largest news subset *EUR/USD news*, We set L as max number of news during delay window h up to the current hour in the *EUR/USD News* dataset to 15 since the max number of news

[11]https://github.com/google-research/bert

[12]https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip

TABLE I: Dataset statistics.

Dataset		Time Interval	# Instance	# News
EUR/USD	Training	2018-09-24 to 2020-05-26	10,380	3,731
	Validation	2020-05-27 to 2020-10-23	2,594	1,322
	Test	2020-10-24 to 2021-05-04	3,243	2,359
USD/JPY	Training	2018-09-23 to 2020-05-26	10,357	2,621
	Validation	2020-05-27 to 2020-10-23	2,595	692
	Test	2020-10-24 to 2021-05-04	3,245	1,156
GBP/USD	Training	2018-09-23 to 2020-05-26	10,381	4,912
	Validation	2020-05-27 to 2021-02-25	2,594	410
	Test	2020-02-26 to 2021-05-04	3,243	507
BTC/USDT	Training	2020-06-24 to 2021-01-11	4,576	2,329
	Validation	2021-01-12 to 2021-03-03	1,143	542
	Test	2021-03-04 to 2021-05-05	1,429	1,288

for this currency-pair during the training phase was 15. The newsgroup publishes an average of 12.03 news with standard deviation 7.06 per day for EUR/USD currency-pair and we have max number of 5 news per hour in *EUR/USD news*. For sentiment analysis of news titles, we use pre-train DistilBERT-base-uncase[13] which is fine tuned for GLUE benchmark [41] on SST-2 for sentiment analysis task. We leave sentiment analysis of both news title and content for a future work.

We calculate technical indicators with TA python package[14] and used the default values. We report experiment setting for EUR/USD currency pair. We do the experiment setting in the same way for other currency pairs and only present the results.

1) News Based Setting: We evaluate the sensitivity of two hyper-parameters k (document vector dimension that is equal to the number of concept clusters) and n (the most similar words for title expansion) in our BERT-BoEC vectorization. We conduct experiments to find the best values for parameters k and n. We test k values from 10 to 300 on the *EUR/SD News* corpus and monitor model accuracy on validation set versus different values of n. Figure 3 depicts the MAPE results for testing both k and n settings. When we increase the concept numbers to 210 the model performance also improves and we observe a small decrease in the performance for $k > 210$. In the *EUR/USD News* corpus, after text preprocessing (removing numbers, tags, URLs and stopwords), we have an average of 8.24 keywords in the news title with standard deviation of 2.49, and an average of 161 keywords in the news body with standard deviation of 111. Given the distribution of keyword counts in *EUR/USD News*, the model achieves its best performance for document vectorization with 210 concept clusters and $n = 7$ most similar words.

As an example, consider the news title *"EUR/USD price analysis: A series move below 1.10 loses traction"* published in *2/6/2020 12:06:33 PM GMT, By Pablo Piovano in FXStreet*. Table II depicts its tokens and top 7 most similar words for each token. For each token in the news title, we pick up the top 7 most similar words from $\mathcal{R}^m$ and add those to the set of candidate tokens from the news title and body. Given the contextualized word embedding from BERT, we have more than one representation for each word and thus we have

[13]https://huggingface.co/transformers/model_doc/distilbert.html

[14]https://technical-analysis-library-in-python.readthedocs.io/en/latest/

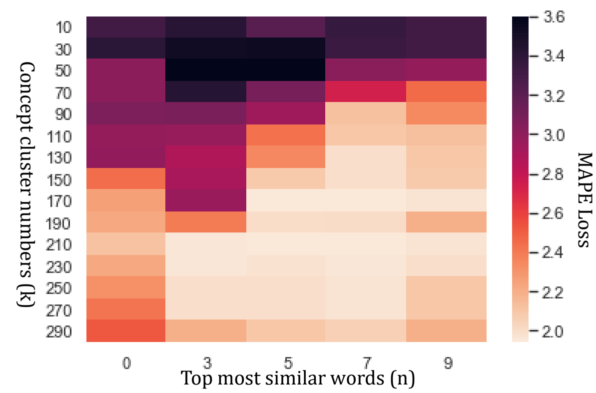


Fig. 3: Sensitivity analysis for different values of number of concept clusters k and the most similar words for title expansion n for EUR/USD with respect to MAPE loss on the validation set for hourly market prediction. Brighter areas represent more accurate predictions.

TABLE II: Example of news title and tokens with 7 most similar words.

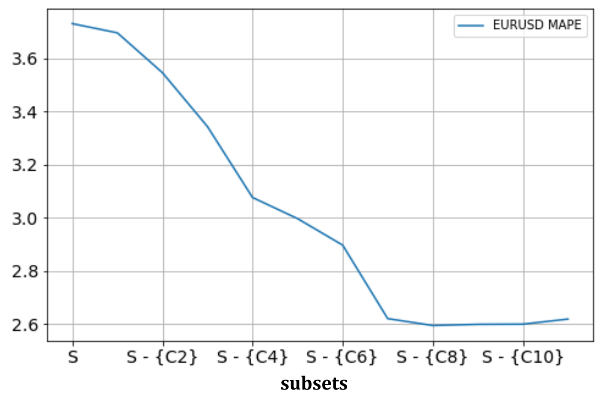
Token	**Top 7 most similar words**
eur	*eur, eur, eur, eur, eur, eur, eur*
usd	*usd, usd, usd, usd, usd, usd, usd*
price	*price, price, price, price, price, price, price*
analysis	*analysis, analysis, analysis, analysis, analysis, analysis, analysis*
series	*significant, sharp, sharp, sharp, already, strong, more*
move	*move, move, move, move, move, move, move*
below	*below, below, below, below, below, below, below*
loses	*loses, loses, loses, loses, loses, loses, loses*
traction	*traction, traction, traction, impetus, traction, momentum, ground*

duplicate terms at each row in table II. Then for vectorization, we compute the corresponding cluster number frequencies of tokens. The results in Figure 3 suggests that leveraging title and content together in the process of document vectorization helps in increasing the accuracy of prediction.

2) Market Based Setting: To evaluate the sensitivity of hyper-parameter q as number of market based features of space S, we train our model with candidate set of features. Table III shows mutual information of each feature in $X_M^{(t)}$ (trading data

TABLE III: Information gain and L^2 norm of LSTM layer weights (block 6) corresponding to each market features (W_o). Cells with * mark, negatively correlated with target close price of the corresponding market.

	EUR/USD		USD/JPY		GBP/USD		BTC/USDT	
Feature	**IG**	**Weight**	**IG**	**Weight**	**IG**	**Weight**	**IG**	**Weight**
Close	3.6594	1.8347	3.4958	3.9258	3.5043	2.5246	3.9462	83.8695
EMA	2.9630	1.6276	2.8130	3.7880	2.8076	2.2409	3.3755	83.2995
BB-Mean	2.7330	1.6685	2.5587	3.7263	2.5448	2.1705	3.1825	83.3435
OBV	1.5036	1.3054	1.6658*	2.5355	1.4185	1.3175	2.6562	88.3308
ADI	1.8476	1.2718	1.9415*	2.6007	1.8238	1.2011	3.2460	76.7559
Volume	0.0890	1.1176	0.04246	2.5745	0.0638*	1.0519	0.2102	61.7795
ATR	0.3293	1.1779	0.2456	2.9986	0.2009*	1.3233	1.5258	58.2721
MACD	0.2332	1.2358	0.2292	2.45255	0.2265	1.2636	0.7403	56.1560
Momentum	0.1607	1.2521	0.1496	2.5442	0.1798	1.3641	0.5865	55.1003
RSI	0.0957	1.2436	0.0778	2.6187	0.1177	1.3725	0.1020	55.8891
Stochastic	0.0364	1.2392	0.0289	2.5165	0.0310	1.2863	0.0446	54.4809
Williams	0.0365	1.2713	0.0289*	2.4765	0.0310	1.2694	0.0447	54.5761



S	Close, EMA, BB-Mean, OBV, ADI, Volume, ATR, MACD, Momentum, RSI, Stochastic, Williams
C2	Stochastic, Williams
C4	Momentum, RSI, Stochastic, Williams
C6	ATR, MACD, Momentum, RSI, Stochastic, Williams
C8	ADI, Volume, ATR, MACD, Momentum, RSI, Stochastic, Williams
C10	BB-Mean, OBV, ADI, Volume, ATR, MACD, Momentum, RSI, Stochastic, Williams

Fig. 4: MAPE loss on validation set of EUR/USD predictive model on different subsets of features.

corresponding to time t and target close price of time t+1) and also the L^2 norm of weights corresponding to each feature in $X_M^{(t)}$ learned through LSTM layer (block 6). Results indicates that more informative features gave more weights during end-to-end learning of our models.

To find the best value of q features, we recursively eliminate features from low to high degrees of informativeness and perceive the model accuracy. Figure 4 depicts the EUR/USD MAPE loss during eliminating low informative features. In this figure, at each step of the experiment, we remove the low informative feature based on the reverse order shown in table III and report the MAPE loss value on the validation set until we have only the close price value as the market feature. We repeat this process for all currency pairs and notice that the accuracy increases by eliminating low informative features and decreases by eliminating high informative features. Finally from all of the features in candidate sets, we select *Close*, *EMA*, *BB-Mean*, and *OBV* as final market features. Models with these $q = 4$ features produced better results and by eliminating each of these features the model loss on the validation set increased, especially for BTC/USDT model which gave the heaviest weight to the OBV indicator in presence of news.

C. Comparison with the Baselines

To assess the effect of data sources and text representation in our BERT-BoEC model, we consider five variations of our proposed model as shown in Table IV. Moreover, we compare the effectiveness of our method in price prediction against two time series-based financial prediction methods that directly use market price and group news in the training process:

- **SI-RCNN**: As presented in [38], after performing an average Word2vec embedding of all the words in the news title, a hybrid model is composed by a RCNN for the financial news and a LSTM for technical indicators. We adopted this model to hourly market price prediction and set model parameters according to [38]. We extract word embedding through Google news pre-train model with embedding dimension of 300 and set the delay window to 7 hours ago same as our method.
- **BHAM**: The BERT-based Hierarchical Aggregation Model (BHAM) [4], first performs [*CLS*] token through BERT pre-trained model for news title and then groups news based on timestamp and applies summarization. They used a SOTA based summerization method. We used top 5 news for group aggregation. This model uses a multi-layer perceptron for feature extraction from trading data and then predict the market trend based on jointly using text information and trading data. We adopted this model to hourly market price prediction and for labeling, we use next hour close price. For trade data preprocessing, following [4], we use min-max normalization. We use three dense layer for feature extraction from trade data followed by a dense node with linear activation function for price regression.

The results in Table V show the superiority of our BERT-BoEC technique compared to the baselines. The difference prediction error of BERT-BoEC and BERT-NoNews shows the

TABLE IV: Variations of our proposed BERT-BoEC model.

Model Name	Training Data	News Vectorization	Sentiment Analysis
BERT-NoNews	Only trade data	–	–
BERT-BoEC-title	News and trade data	BoEC for only title	DistilBERT
BERT-BoEC-NoSentiment	News and trade data	BoEC (Algorithm 1)	–
BERT-Doc2vec	News and trade data	Doc2vec	DistilBERT
BERT-BoEC	News and trade data	BoEC (Algorithm 1)	DistilBERT

TABLE V: Comparison of the mean absolute percentage error (MAPE) loss.

	EUR/USD			**USD/JPY**			**GBP/USD**			**BTC/USDT**		
Model	**Training Loss**	**Validation Loss**	**Test Loss**	**Training Loss**	**Validation Loss**	**Test Loss**	**Training Loss**	**Validation Loss**	**Test Loss**	**Training Loss**	**Validation Loss**	**Test Loss**
BERT-NoNews	0.1076	3.6233	6.9652	3.4815	2.6950	3.0853	0.1529	0.6915	6.4349	85.9197	95.4747	96.7576
BERT-BoEC-title	0.2156	3.3473	6.6911	2.5218	2.7268	3.1422	0.1529	0.7015	6.5235	7.2233	70.0934	75.6124
BERT-BoEC-NoSentiment	0.185	2.9854	5.5416	2.2218	**2.0281**	3.0092	0.1823	0.6735	6.4867	6.0983	65.4708	73.6804
BERT-Doc2vec	0.3511	3.8434	7.1660	3.0352	2.7490	3.2843	0.2745	0.7615	6.8232	9.9273	69.0854	77.6674
BERT-BoEC	**0.185**	**2.6941**	**5.1476**	**2.1478**	2.2248	**2.9422**	**0.1615**	**0.6113**	**6.3831**	**5.5122**	**65.1852**	**73.2147**
SI-RCNN	0.2156	3.4321	6.8432	2.5218	2.7008	3.2128	0.2083	0.6805	6.4951	7.9623	71.3084	75.9024
BHAM	0.3017	3.2434	6.1660	2.4798	2.6894	3.0829	0.2814	0.6913	6.4221	6.8963	67.5364	75.0879

importance of integrating news in financial decision support. In BoEC, we group news based on delay window length together with technical market data for prediction. This enables decision making during the trading time even when there is no news in special cases. The superiority of BERT-BoEC against BERT-BoEC-title, BERT-BoEC-NoSentiment indicates the importance of using all information in news *title*, *body* and *sentiment*.

We also notice that BERT-BoEC compared to BERT-doc2vec, SI-RCNN, and BHAM, has smaller prediction error which indicates the ability of BoEC text representation to model context proximity between news. The results clearly show that incorporating relevant news reduces the prediction error. However, the non-stationary and multimodal nature of financial time series severely affect the performance of our model in training and validation sets. The differences between loss values in training and validation pertains to the change in means and variances of market data during the phases of normalization. This specially affects BTC/USDT prediction in validation and test set due to a large difference between price value in training and validation set. To assess the ability of our model for BTC/USDT trend behavior prediction, we add a bias value to all of the predicted prices and noticed at least 60% improvement in BERT-BoEC results while when we did the same increase for BERT-NoNews, we did not notice a major improvement. We plan to resolve this problem in future by end-to-end training of our neural network normalization same as [31].

D. Effects of Positive and Negative News

To better understand how BERT-BoEC can be useful in practice for financial market decision support, we study two cases for GBP/USD and BTC/USDT and look into prediction plots in Figures 5 and 6 that show the prediction against the target close price values corresponding to same time steps.

We noticed a major drop for GBP/USD in March 2020 (falling 900 pips from March 9 to March 18). We plot the first time steps of this fall from March 9 to March 12 in Figure 5, where BoEC predicts falling of the close price earlier than *NoNews* (160 pips falling during 9th 12:00 to 10th 10:00).

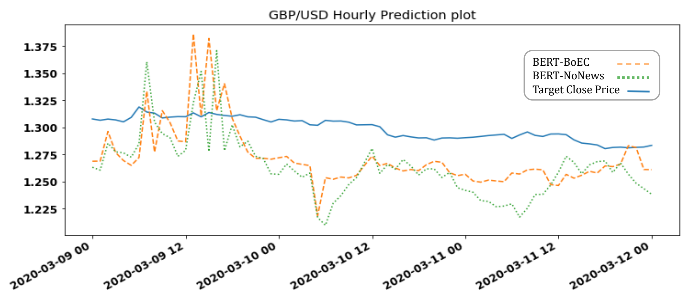


Fig. 5: GBP/USD hourly predicted close price. BoEC predicts falling of the close price during 9th 12:00 to 10th 10:00 earlier than NoNews.

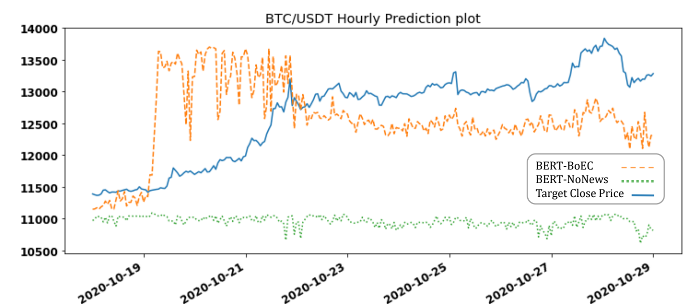


Fig. 6: BTC/USDT hourly predicted close price.

Table VI depicts some *negative news* during this period, which indicates the efficacy of our BoEC model. In another observation shown in Figure 6 we study the efficacy of our model under *positive news* (Table VI) about the bitcoin market. Figure 6 clearly shows the importance of incorporating news for early identification of bitcoin market rising that results in BoEC to predict the increase of BITCOIN/TETHERUS in October 2020 earlier than NoNews.

TABLE VI: Examples of negative news for GBP/USD and positive news for BTC/USDT.

GBP/USD Negative News Title	**GBP/USD Negative News Body**	**Timestamp**
GBP/USD: Failing to break critical resistance	The Relative Strength Index on the four-hour chart is above 70 – pointing to overbought conditions that imply a downside correction. GBP/USD was rejected at 1.32, which now turns into a triple top after holding the currency pair down in January and in February. It remains critical resistance	Mar 09 2020, 10:31 GMT
GBP/USD: Brexit more dangerous than coronavirus for the pound	GBP/USD was above 1.32 as growing coronavirus fears weigh on the US dollar while worries about Brexit have limited the upside move, Yohay Elam from FXStreet briefs	Mar 09 2020, 10:42 GMT
GBP/USD Price Analysis: Snaps five-day winning streak as MACD teases bears on H4	GBP/USD drops from five-week high, slips below 61.8% Fibonacci retracement. 200-bar SMA acts as the key support, buyers will look for entry beyond two-month-old horizontal resistance.	Mar 10 2020, 01:10 GMT
BTC/USDT Positive News Title	**BTC/USDT Positive News Body**	**Timestamp**
Here's Why Bitcoin Breaking Higher Now Would Trigger a Massive Move	While Bitcoin has seen strong volatility on a day-to-day and week-to-week basis, it is flat on a macro scale. The crypto-asset has basically traded in the same $3,000 range for five months now.	Sun, 18 Oct 2020 10:00:30
Options Trends Makes it Hard for This Analyst to Imagine a Bitcoin "Mega Pump"	Bitcoin has been consolidating within a $1,500 range over the past seven weeks. With BTC currently pushing higher and the fundamentals aligning in favor of bulls, analysts have begun to expect a strong rally into the end of a year.	Mon, 19 Oct 2020 00:00:23
3 Reasons Why Bitcoin Rebounded, Dollar Slumped Ahead of Powell's Speech	It is Monday. Bitcoin is holding above crucial technical support at $11,400. Meanwhile, its safe-haven rival, the US dollar, is experiencing a sharp decline.	Mon, 19 Oct 2020 13:00:56

E. Generalizability

An external validity threat is regarding the results generalization and to mitigate that we conducted our experiments on four currency pairs EUR/USD, USD/JPY, GBP/USD and BTC/USDT as shown in Table V. Moreover, we scraped news from two specialized financial newsgroups of Foreign Exchange and Cryptocurrency markets. With regards to the generalizability on other financial markets, such as commodities like gold, or stock market, we believe that our approach can similarly utilize trading data and technical indicators that are commonly used in technical analysis of such markets and leave it as a future work. Following [4], [29] we considered news effects on three major currency pairs prediction in the Forex market. We analyzed the multifractal correlation behavior of non-major currency pairs such as USD/CHF and USD/CAD with the major currency pairs in Forex in a previous work [1] and plan to study news effect for such non-major currency pairs in our future work.

VI. Conclusion and Future Work

News-based financial market prediction requires using relevant news to the target market and the semantic relationships between the news. Most of the previous work relies on news title embedding representation, which does not completely reflect the relationships in news. We argued that discovering the distribution of news over context-aware latent concepts and monitoring the temporal changes in the distribution of such concepts and news emotion, together with using more informative market technical indicators, can enhance financial market forecasting.

In this work, we proposed a generalizable market prediction tool as well as an open reliable news dataset scraped from specialized financial newsgroups. Our predictive model leverages emotional and conceptual relationships in the documents in BERT-based latent embedding space. In the proposed BERT-BoEC vectorization scheme, in addition to embedding news sentiment and the conceptual relationship in news document vectors, deep information about syntactic and semantic of words in title and body of news are extracted using BERT. Finally, each news document is vectorized based on the frequency of repetition of expanded title and news content terms in latent concepts. This results in a document representation that reflects sentiment and different latent economic concepts in documents. We then utilize a recurrent convolution network to jointly use emotional and conceptual relationships in news as well as market data and technical indicators.

We evaluated the generalizability of our tool in the Forex and Cryptocurrency markets and studied the effectiveness of using both news *sentiment*, news *title* and *content* in document representation. Experimental results show the effectiveness of incorporating our BERT-based concept modeling in market prediction. The improved prediction accuracy suggests that investors should also pay attention to the content in addition to the title of financial news. We plan to investigate other markets such as commodities like gold and oil, and stock market. Also, we would like to use other information in news such as images of technical analysis on market charts in the news and also the news author embedding in our model.

VII. Acknowledgment

The authors would like to acknowledge the financial support of Iran's Ministry of Science, Research and Technology for this project under the grant number *12-99-02-000049*.

References

[1] Z. Alamatian, M. Vafaei Jahan, and A. Milani Fard. Using market indicators to eliminate local trends for financial time series cross-correlation analysis. In *Proceedings of the 34th Canadian Conference on Artificial Intelligence (Canadian AI)*, 2021.

[2] X. Bu, B. Li, Y. Wang, J. Zhu, X. Qian, and M. Zhao. Semantic gated network for efficient news representation. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, ICMR '20, page 251–255, New York, NY, USA, 2020. ACM.

[3] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco. Affective computing and sentiment analysis. In *A practical guide to sentiment analysis*, pages 1–10. Springer, 2017.

[4] D. Chen, S. Ma, K. Harimoto, R. Bao, Q. Su, and X. Sun. Group, extract and aggregate: Summarizing a large amount of finance news for forex movement prediction. *ArXiv*, abs/1910.05032, 2019.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
[6] X. Ding, Y. Zhang, T. Liu, and J. Duan. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.
[7] X. Ding, Y. Zhang, T. Liu, and J. Duan. Knowledge-driven event embedding for stock prediction. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*, pages 2133–2142, 2016.
[8] S. Ge, C. Wu, F. Wu, T. Qi, and Y. Huang. Graph enhanced representation learning for news recommendation. In *Proceedings of The Web Conference 2020*, WWW '20, page 2863–2869, New York, NY, USA, 2020. ACM.
[9] M. Hagenau, M. Liebmann, and D. Neumann. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, 55(3):685–697, 2013.
[10] J. Z. G. Hiew, X. Huang, H. Mou, D. Li, Q. Wu, and Y. Xu. Bert-based financial sentiment index and lstm-based stock return predictability. *arXiv preprint arXiv:1906.09024*, 2019.
[11] C.-S. Ho, P. Damien, B. Gu, and P. Konana. The time-varying nature of social media sentiments in modeling stock returns. *Decision Support Systems*, 101:69–81, 2017.
[12] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
[13] H. D. Huynh, L. M. Dang, and D. Duong. A new model for stock price movements prediction using deep neural network. In *Proceedings of the Eighth International Symposium on Information and Communication Technology*, pages 57–62, 2017.
[14] Johan Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
[15] Joshua Zoen Git Hiew, X. Huang, H. Mou, D. Li, Q. Wu, and Y. Xu. Bert-based financial sentiment index and lstm-based stock return predictability. *arXiv preprint arXiv:1906.09024*, 2019.
[16] H. K. Kim, H. Kim, and S. Cho. Bag-of-concepts: Comprehending document representation through clustering words in distributed representation. *Neurocomputing*, 266:336–352, 2017.
[17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017.
[18] C. D. Kirkpatrick II and J. A. Dahlquist. *Technical analysis: the complete resource for financial market technicians*. FT press, 2010.
[19] S. Krishnamoorthy. Sentiment analysis of financial news articles using performance indicators. *Knowledge and Information Systems*, 56(2):373–394, 2018.
[20] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, page II–1188–II–1196. JMLR.org, 2014.
[21] C.-Y. Lee and V.-W. Soo. Predict stock price with financial news based on recurrent convolutional neural networks. In *2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 160–165. IEEE, 2017.
[22] Q. Li, Y. Chen, L. L. Jiang, P. Li, and H. Chen. A tensor-based information framework for predicting the stock market. *ACM Transactions on Information Systems (TOIS)*, 34(2):1–30, 2016.
[23] X. Li, C. Wang, J. Dong, F. Wang, X. Deng, and S. Zhu. Improving stock market prediction by integrating both market news and stock prices. In *International Conference on Database and Expert Systems Applications*, pages 279–293. Springer, 2011.
[24] Q. Liu, X. Cheng, S. Su, and S. Zhu. Hierarchical complementary attention network for predicting stock price movements with news. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1603–1606, 2018.
[25] W. Long, L. Song, and Y. Tian. A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity. *Expert Systems with Applications*, 118:411–424, 2019.
[26] B. Lutz, N. Pröllochs, and D. Neumann. Predicting sentence-level polarity labels of financial news using abnormal stock returns. *Expert Systems with Applications*, 148:113223, 2020.
[27] A. C. MacKinlay. Event studies in economics and finance. *Journal of economic literature*, 35(1):13–39, 1997.
[28] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
[29] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo. Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications*, 42(1):306–324, 2015.
[30] T. M. Nisar and M. Yeung. Twitter as a tool for forecasting stock market movements: A short-window event study. *The journal of finance and data science*, 4(2):101–119, 2018.
[31] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis. Deep adaptive input normalization for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 2019.
[32] B. C. Ross. Mutual information between discrete and continuous data sets. *PloS one*, 9(2):e87357, 2014.
[33] V. Sanh, L. Debut, J. Chaumond, and T. Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.
[34] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen. Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3):458–464, 2012.
[35] S. Seifollahi and M. Shajari. Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to forex market prediction. *Journal of Intelligent Information Systems*, 52(1):57–83, 2019.
[36] H.-S. Sheu and S. Li. Context-aware graph embedding for session-based news recommendation. In *Fourteenth ACM Conference on Recommender Systems*, RecSys '20, page 657–662, New York, NY, USA, 2020. ACM.
[37] Y. Shi, K.-Y. Ho, and W.-M. Liu. Public information arrival and stock return volatility: Evidence from news sentiment and markov regime-switching approach. *International Review of Economics & Finance*, 42:291–312, 2016.
[38] M. R. Vargas, C. E. dos Anjos, G. L. Bichara, and A. G. Evsukoff. Deep leaming for stock market prediction using technical indicators and financial news articles. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.
[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.
[40] A. F. Vilas, R. P. D. Redondo, K. Crockett, M. Owda, and L. Evans. Twitter permeability to financial events: an experiment towards a model for sensing irregularities. *Multimedia Tools and Applications*, 78(7):9217–9245, 2019.
[41] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 2018.
[42] Xue Zhang and H. Fuehres and P. A. Gloor. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia - Social and Behavioral Sciences*, 26:55–62, 2011. The 2nd Collaborative Innovation Networks Conference - COINs2010.
[43] Y. Zhai, A. Hsu, and S. K. Halgamuge. Combining news and technical indicators in daily stock price trends prediction. In *International symposium on neural networks*, pages 1087–1096. Springer, 2007.
[44] W. Zhang, Y. Li, and S. Wang. Learning document representation via topic-enhanced lstm model. *Knowledge-Based Systems*, 174:194–204, 2019.
[45] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and S. Y. Philip. Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems*, 143:236–247, 2018.