THE UNIVERSITY OF CHICAGO

PREDICTING STOCK RETURNS WITH NEWS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES DIVISION

IN CANDIDACY FOR THE DEGREE OF

PH.D.

DEPARTMENT OF DEPARTMENT OF COMPUTER SCIENCE

BY

JINGWEN JIANG

CHICAGO, ILLINOIS

FEBRUARY 15, 2022

Copyright © 2022 by Jingwen Jiang

All Rights Reserved

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

With the rise of deep learning in image analysis and natural language processing, we have witnessed a growing interest in the analysis of alternative data including images, text, and audio. Those alternative data are most commonly unstructured data, which does not have a pre-defined data model or even unorganized, posting challenges of extracting useful information from them. In fact, research shows that unstructured information accounts for more than 90% of the digital universe.

It has been a while since the financial industry led the application in unstructured text data analysis. The most commonly researched area in text is to study the sentiment. For example, RavenPack, one of the industry-leading commercial vendor of financial news sentiment scores, provides a data analysis service to hedge funds and large financial services organizations. Aside from the data vendors, some financial institutions themselves spent efforts on news analysis. State Street provides a indicator called MediaStats which covers 100,000 digital media sources and offers in-depth sentiment analytical statistics.

Led by the success in industry, academic research has been growing attention on textual data and gradually consolidating the value behind. One of the first foundational work is Tetlock (2007), which validates the predictive power of news stories in term of both earnings and stock returns. It further shows that the short-term predicted return will soon decay to nothing. Loughran and McDonald (2011) and Jegadeesh and Wu (2013) both conducts textual analysis to quantify document tone from 10-Ks. Chew et al. (2017) parses the headlines of financial news and use Name Entity Recognition Finkel et al. (2005) to match them to each company. Based on that, they apply Global Vector for Word Representation (GloVe) model Pennington et al. (2014) and k-means clustering to identity significant events and the post-event stock returns. Cong et al. (2019) applies a data-driven approach to form textual factors by clustering on vector word embedding. Recently, Ke et al. (2019) propose a text-mining approach that learns sentiment information from news for return prediction. Bybee et al. (2020) introduce a method of unsupervised clustering model that group news by topic.

However, methodologies of most literature are limited to static features such as bag of words, and word embedding. The reasons are two-folded: first, it has just been recent years that we have witnessed the great success of Transformer-based NLP models; second, though powerful, those deep learning models suffer from the complexity of interpretation due to the the blackbox nature. Still it has been a growing interest in the financial literature. Kölbel et al. (2020) uses BERT to analyze 10-K reports filed by companies to extract climate risk information from text data and shows that it substantially outperforms other classical algorithms. Jha et al. (2020) measures popular sentiment towards finance from BERT embedding of finance related 5-grams extracted from published books.

In this paper, we apply a novel way of predicting stock returns from textual news based on a BERT-based contextual model. In addition, we compare the context-based model with the other two canonical models, namely bag-of-words and word embeddings. The advantage of the BERT-based model not only lies in its better prediction performance but also in its interpretation of news based on the context.

Table 2.1: International Countries Model Specifications

	BOW/W2V Tokenizer	BOW Alpha	BERT Tokenizer
US	en_core_web_sm	100/200	bert-base-uncased
China	jieba	100/200	bert-base-chinese
UK	en_core_web_sm	100/200	bert-base-uncased
Australia	en_core_web_sm	100/200	bert-base-uncased
Canada	en_core_web_sm	100/200	bert-base-uncased
Japan	nagisa	100/200	cl-tohoku/bert-base-japanese
Germany	de_core_news_sm	100/200	bert-base-german-cased
Italy	it_core_news_sm	100/200	dbmdz/bert-base-italian-cased
France	fr_core_news_sm	100/200	bert-base-multilingual-cased
Sweden	xx_ent_wiki_sm	100/200	KB/bert-base-swedish-cased
Denmark	da_core_news_sm	100/200	bert-base-multilingual-cased
Spain	es_core_news_sm	100/200	dccuchile/bert-base-spanish-wwm-uncased
Finland	xx_ent_wiki_sm	100/200	TurkuNLP/bert-base-finnish-cased-v1
Portugal	pt_core_news_sm	50	neuralmind/bert-base-portuguese-cased
Greece	el_core_news_sm	50	nlpaueb/bert-base-greek-uncased-v1
Netherlands	nl_core_news_sm	50	wietsedv/bert-base-dutch-cased

Note: This table reports the model specifications for each country. BOW/W2V Tokenizer shows the language pipeline used from spacy to tokenize news into words for Bag-of-words and Word2vec. BOW Alpha is the number of positive/negative words used in the Bag-of-words model. BERT Tokenizer corresponds to the pretrained model from Hugging Face.

CHAPTER 2

METHODOLOGY

2.1 Overview

For simplicity, we only include news that are tagged to a single stock. We use the cross-sectionally rank-normalized three-day return, beginning the day before the article is published and ending the day after, as the label y_i for news article of stock i.

We compare the performance of 6 different models including 2 variants of bag-of-words models, 2 variants of word embedding models, and 2 variants of BERT models, namely BOW100, BOW200, W2V(LOGIT), W2V(OLS), BERT(LOGIT), and BERT(OLS). Table. 2.1 shows the model specifications for each country.

The labels used by each model are all variants of the rank-normed 3-day returns ranging from -1 to 1. For BERT(OLS), and W2V(OLS), we directly use the ranked-normed returns. For BERT(LOGIT) and W2V(LOGIT), the labels are 0 if the ranked-normed returns are less than 0 and 1 otherwise. For BOW100 and BOW200, we rescale the returns from 0 to 1 to be consistent with the MLE methodology in Ke et al. (2019).

2.2 Bag-of-Words

The bag-of-words (BOW) model is first proposed by Harris (1954) as a tool of feature generation where an article is represented as a multiset (bag) of words in vector form. The BOW representation is able to extract the information of word occurrence and frequency from the article but loses track of grammar and word order.

2.2.1 Model Setup

We follow the approach from Ke et al. (2019) to implement a bag-of-word model using maximum likelihood estimation (MLE). Given a dictionary of m words, we generate a word count for each article i using a vector $d_i \in \mathbb{R}_+^m$, where $d_{i,j}$ is the number of times word j appears in article i. To isolate sentiment-neutral words and tease out words that carry sentiment information, we count the frequency with which word j co-occurs with a positive return as

$$f_j = \frac{\#\text{articles including word j AND having } sgn(y) = 1}{\#\text{articles including word j}}$$

for each $j = 1, \ldots, m$. Then, with an hyperparameter α, we pick the top α words with the highest f_j as positive sentiment words and the bottom α words with the lowest f_j as negative sentiment words. For statistical accuracy, we restrict the above word selection process to words which appear at least κ times in-sample, where $\kappa = 3000$ is another hyperparameter. We then follow the MLE described in Ke et al. (2019) with regularizer coefficient $\lambda = 10$ to obtain the predicted sentiment of news articles.

BOW100: MLE using $\alpha = 100$ positive sentiment words and negative sentiment words each. Portugual, Greece and Netherlands use $\alpha = 50$ due to limited size of vocabularies.

BOW200: MLE using $\alpha = 200$ positive sentiment words and negative sentiment words each.

Portugal, Greece and Netherlands use $\alpha = 50$ due to limited size of vocabularies.

2.3 Word Embeddings

A common approach in Natrual Language Processing is to learn word embeddings under a high-dimensional vector space for each unique word in the vocabulary of size 1 million to 3 million. The two popular word embedding models are *word2vec* (Mikolov et al. (2013)) created by Google and *fast-Text* (Bojanowski et al. (2017)) created by Facebook's AI Research (FAIR) lab. Both models are trained using a method called the skipgram architecture, where the existence of nearby words are predicted given a source word. The pre-trained vectors of words provide distributional information that captures relationships in the vector space. A famous example from in Mikolov et al. (2013) shows that the resulting vector from the operation vector("King") - vector("Man") + vector("Woman") is closest to the vector representation of the word Queen.

2.3.1 Model Setup

We downloaded pre-trained word vectors for English and foreign languages from fastText. For English word vectors, we pick the model *wiki-news-300d-1M* (Mikolov et al. (2018)) which contains 1 million word vectors trained on Wikipedia 2017, UMBC webbase corpus and statmt.org news dataset with 16 billion tokens in total. For foreign languages, the word vectors are trained on Common Crawl and Wikipedia (Grave et al. (2018)).

For each news, we sum up the vectors corresponding to each word and use the average vector as the embedding.

W2V(OLS): OLS regression with the average word vector as regressors.

W2V(LOGIT): Logistic regression with the average word vector as regressors.

2.4 BERT Embedding

BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al. (2018)) is a deep neural network-based natural language processing model. Unlike the traditional unidirectional language models, BERT is a bidirectional model designed to learn contextual relations.

BERT is derived from a sequence-to-sequence model that takes in a sequence (series of words) and outputs another sequence. A sequence-to-sequence model consists of an encoder, which compiles the information into vectors, and a decoder, which in turn produce the output sequence from the vectors. Encoders and decoders are often recurrent neural networks (RNN). Previously, the sequence-to-sequence model process each word from a input sentence one at a time, which is challenging to process long sentences. Bahdanau et al. (2014) and Luong et al. (2015) proposed a technique called ”Attention” that allows the decoder to focus on relevant parts of the input sequence.

The main architecture of BERT is called the Transformer proposed by Vaswani et al. (2017), which is based on attention mechanisms. The Transformer is composed of multiple stacked encoders and decoders. Each encoder and decoder contains a self-attention layer and a fully connected layer. The self-attention layer is composed of a Query matrix, a Key matrix, and a Value matrix, which are trainable parameters and are used to learn interactive relations between tokens.

BERT is originally trained with two unsupervised tasks simultaneously: Masked Language Model (MLM) and Next Sentence Prediction (NSP). MLM simply masks 15% of tokens at random from the input sequence and then predicts the masked tokens. The MLM objective enables the representation of a token to incorporate information from both the left and the right contexts. NSP, as the name suggests, aims to predict whether a sentence pair is adjacent or not. In this paper, we only pre-train the model using MLM to improve the bidirectional representation since the main tasks that benefit from NSP, e.g. Question Answering and Natural Language Inference, are out of scope.

There are two versions of the pre-trained BERT model shared by Google: $BERT_{BASE}$ with

12 layers and 768 hidden features and $BERT_{LARGE}$ with 24 layers and 1024 hidden features. Both models have a cased and uncased variant. Both models are pre-trained on the BooksCorpus with 800 million words and English Wikipedia with 2,500 million words. Each model is trained for 40 epochs which takes 4 days to complete on Google Cloud TPUs (4 TPUs for the $BERT_{BASE}$ and 16 TPUS for $BERT_{LARGE}$). One advantage of BERT over the other classic language models is transfer learning. Text sequences can be directly fed into the pre-trained BERT model and the output feature embedding already employs the benefits from the training data of over 3.3 billion words. In addition, the pre-trained BERT model can be used as a starting point for future fine-tuning on different tasks.

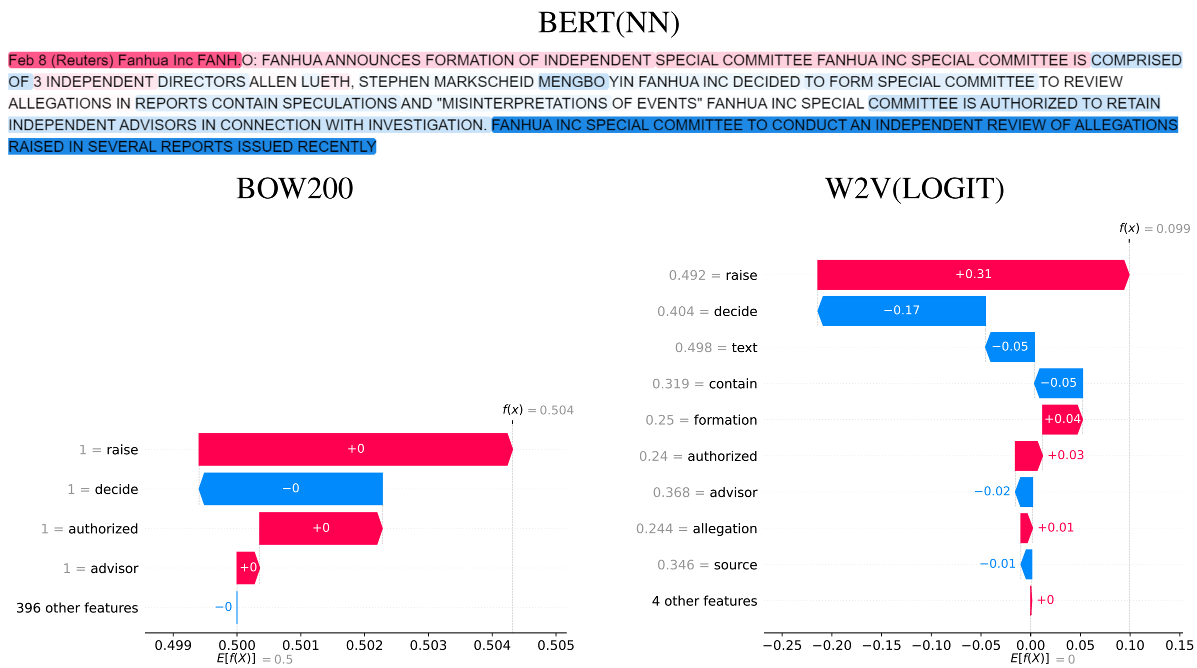
Due to limited computing power, we choose uncased $BERT_{BASE}$ for all English texts. We use language specific pre-trained models for foreign languages, and cased multilingual model for French and Danish as no language specific models are available. The third column in Table. 2.1 shows the pretrained BERT models from Hugging Face used for each country.

Compared with the classical bag-of-words and word embedding models, BERT is able to distinguish the contextualized interpretation of a word from a news instead of relying on a fixed, unique word feature. Figure. 2.1 shows the model interpretations of BERT(NN), BOW200, and W2V(LOGIT) on the following news with negative sentiment:

> *Feb 8 (Reuters) Fanhua Inc FANH.O: Fanhua announces formation of independent special committee Fanhua Inc special committee is comprised of 3 independent directors Allen Lueth, Stephen Markscheid Mengbo Yin Fanhua Inc decided to form special committee to review allegations in reports contain speculations and "misinterpretations of events" Fanhua Inc special committee is authorized to retain independent advisors in connection with investigation. Fanhua Inc special committee to conduct an independent review of allegations raised in several reports issued recently.*

We can see that all three models rely heavily on the last sentence "Fanhua Inc special committee to conduct an independent review of allegations **raised** in several reports issued recently"

Figure 2.1: Contextualized Interpretation of BERT



Note: Interpreted by SHAP (SHapley Additive exPlanations) from Lundberg and Lee (2017). When interpreting BERT, segments with positive SHAP values are highlighted in red and segments with negative SHAP values are highlighted in blue. The darker the color is, the larger magnitude the SHAP value is. When interpreting Bag of Words (BOW) and Words to Vectors (W2V), we use a waterfall plot which shows all features contributing to the prediction. The blue features push the prediction to the negative side while the red features push the prediction to the positive side.

for the final prediction. However, BOW and W2V mistakenly assume the positive sentiment from the word "raise" based on the prior even though the context here is "allegations raised". BERT manages to infer from the context and assign strong negative sentiment to the last sentence. For more news examples where BERT outperforms bag-of words models and word embedding models, please see Section. A.3 in the Appendix.

One limitation of BERT is that it is only able to encode a sequence of maximum 512 tokens. For US news, there are over 60% of news that can fit into BERT's length limitation. For the remaining longer news, we follow the canonical setting of the NLP community by truncating the first 512 tokens. Experiments have shown that the first 512 tokens have enough information to predict future returns.

2.4.1 Model Setup

Based on the pre-trained BERT model from Hugging Face, we fine-tune the model using news from Thomson Reuters in a masked language modeling setting. All BERT models use embedding of the first 512 tokens generated by the fine-tuned model. The output feature embedding is a vector of length 768 for each text sequence.

BERT(OLS): OLS regression with the feature embedding as regressors.

BERT(LOGIT): Logistic regression with the feature embedding as regressors.

2.5 Novelty of News

As indicated by Ke et al. (2019), a significant amount of news is old news and the information is already reflected in prices when the news is published. In order to avoid this issue and predict returns using "fresh news" only, we follow the same approach from Ke et al. (2019) to calculate the novelty of news. For all models, we use the same measure of article novelty based on cosine similarity of the BOW representations, described in the previous section of bag-of-words. For each article of stock i at time t, we calculate its cosine similarity with all articles about stock i within 5 trading days of t (denoted by the set $\chi_{i,t}$). Let $d_{i,t}$ be the vector of word count of the news article of stock i at time t, we define the novelty as

$$\text{Novelty}_{i,t} = 1 - \max_{j \in \chi_{i,t}} \left(\frac{d_{i,t} \cdot d_j}{||d_{i,t}|| \cdot ||d_j||} \right)$$

.

2.6 Interpretation using SHAP Values

We use a method called SHAP (SHapley Additive exPlanations) proposed by Lundberg and Lee (2017) to interpret individual predictions. SHAP is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. The SHAP values can be viewed as expected change in predictions conditional on specific features. When interpreting BERT, segments with positive SHAP values are highlighted in red and segments with negative SHAP values are highlighted in blue. The darker the color is, the larger magnitude the SHAP value is. When interpreting bag-of-words models and word embedding models, we use a waterfall plot which shows all features contributing to the prediction. The blue features push the prediction to the negative side while the red features push the prediction to the positive side.

Intuitively, the model itself best explains itself but often it is too complicated to understand. Thus, SHAP value explains a model using a simplified *explanation model*. Let f be the original prediction model to be explained and g the explanation model. Meanwhile, let x be a single input to f and x' the corresponding simplified input to g that map to the original inputs through a mapping function $x = h_x(x')$ such that $g(x') \approx f(h_x(x'))$. Furthermore, the method of SHAP value assumes the original model to be an *additive feature attribution method*, which has an explanation model that is a linear function of binary variables, i.e.

$$g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i,$$

where $x' \in \{0,1\}^M$ and $\phi_i \in \mathbb{R}$ is the impact attributed to each feature.

A significant attribute of the class of additive feature attribution methods is the unique existence of a solution that satisfies the following three desirable properties:

1. **Local Accuracy:** $f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i$
2. **Missingness:** Features with $x'_i = 0$ have no attributed impact, i.e. if $x'_i = 0$, then $\phi_i = 0$.
3. **Consistency:** Let $f_x(z') = f(h_x(z'))$ and $z' \backslash i$ denote setting $z'_i = 0$. For any two models f and f', if

$$f'_x(z') - f'_x(z' \backslash i) \geq f_x(z') - f_x(z' \backslash i)$$

for all inputs $z' \in \{0, 1\}^M$, then $\phi_i(f', x) \geq \phi_i(f, x)$.

SHAP values are the estimated solution that is a unified measure of feature importance.

CHAPTER 3

NEWS PREDICTION FOR US STOCK RETURNS

3.1 Data and Pre-processing

We obtain news of US from two sources: Thomson Reuters Real-time News Feed (RTRS) and Archive (3PTY) from 1996 to 2019. There are two types of news, namely **articles** and **alerts**. News articles are the common news we see that include both a headline and a body article for each news. They often provide a detailed and thorough report on the story it covers. On the other hand, a news alert provides immediate coverage for emergent news. In order to deliver the real-time event, a news alert only has a headline with no body.

Of all news articles and news alerts, we only use those that are tagged to a single stock with available three-day close-to-close returns. In addition, we filter out news articles of length less than 30 for China and 100 for all other countries (see Table. A.1 for the summary of article length before filtering). Following Ke et al. (2019), we filter out half news below the median novelty and only use the top half "fresh news" for training and testing.

Table 3.1: News Summary Statistics

News Article

	Raw Articles			Articles Tagged with Single Stock			Articles With	After Filtering	After Filtering
	RTRS	3PTY	Total	RTRS	3PTY	Total	Available Returns	Short Articles	Old News
US	6,366,019	4,843,867	11,209,886	2,863,166	4,123,823	6,986,989	4,755,247	4,123,279	2,061,640 (605,232)

News Alerts

	Raw Alerts			Alerts Tagged with Single Stock			Alerts With	After Filtering	First In
	RTRS	3PTY	Total	RTRS	3PTY	Total	Available Returns	Old News	Take Sequence
US	4,976,374	4	4,976,378	4,054,683	4	4,054,687	3,286,003	1,682,883 (451,938)	930,349 (263,025)

Note: In this table, we report the impact of each filter we apply on the number of news articles and news alerts in the top and bottom panels respectively. Raw Articles/Alerts shows the number of articles/alerts from Thomson Reuters Real-time News Feed (RTRS) and Archive (3PTY). Articles/alerts Tagged with Single Stock shows the number of articles/alerts that are tagged with a single stock. In the following three columns, we show the number of remaining articles/alerts after filtering out those with unavailable returns, those with short length of news body (less than 100), those with larger than median cosine similarities. The parenthesized number of US corresponds to the number of articles/alerts of companies belonging to the S&P 500.

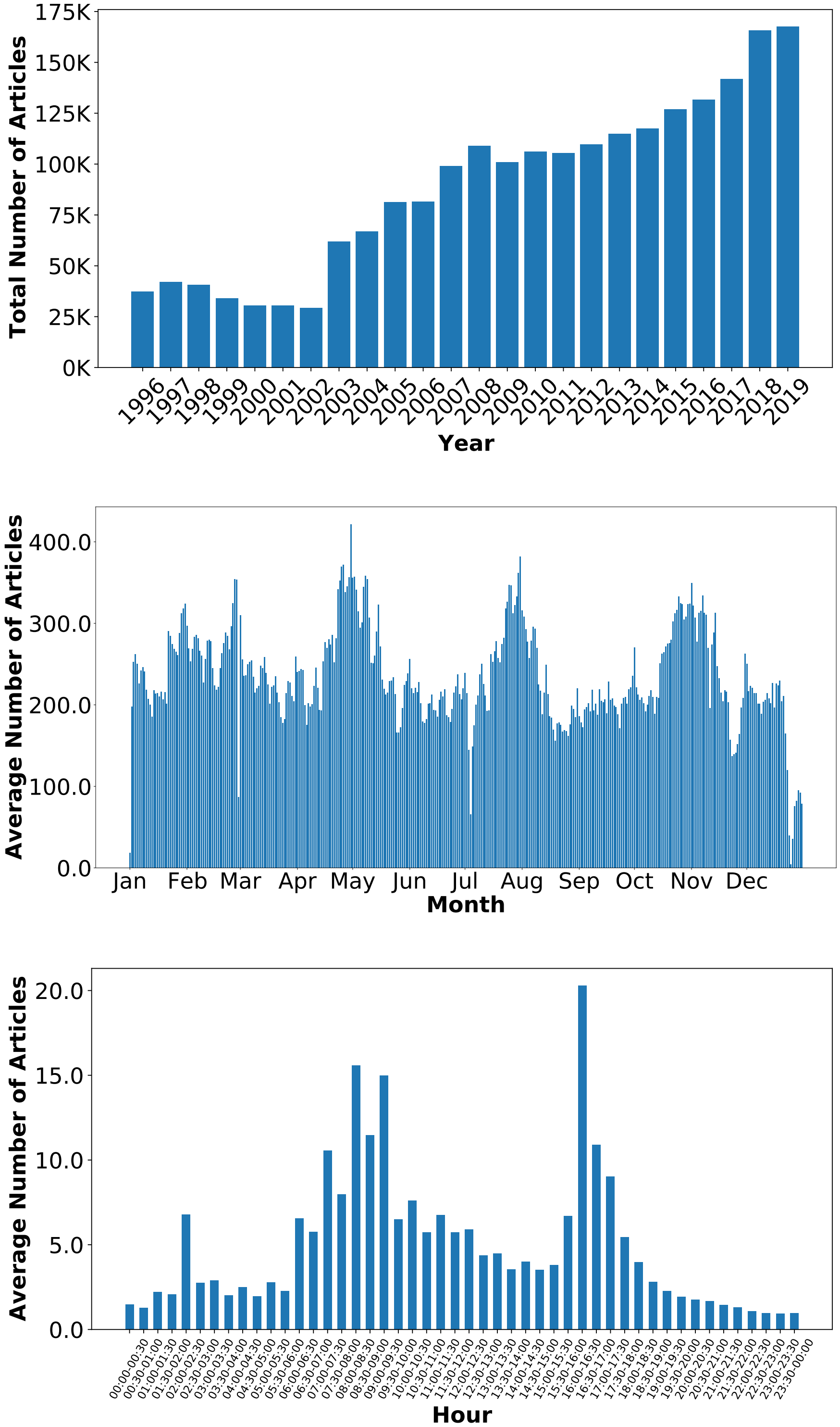
Table. 3.1 shows the summary statistics of news articles and alerts. After filtering, there are around 2 million news articles (of which 600k on S& P 500 companies) and around 1 million news alerts (of which 263k on S& P 500 companies).

Figure. 3.1 and Figure. 3.2 show the yearly, monthly, and hourly news count for articles and alerts separately. Both news articles and news alerts follow the same trend. From the top panel of yearly news count, there is a continuous increase in the number of news from 2003 to 2019. From the middle panel of the monthly count, the quarterly earnings season effects are quite notable, especially for news alerts, around February, May, August, and November. From the bottom panel of the hourly count, News arrive much more frequently around the market open at 9:30 am and the market close at 16:00 pm.

Figure. 3.3 plots the average daily number of stocks with active news and alerts, respectively. There is also an increase in the number of stocks that are covered by news since 2003. Starting from 2015, on average more than 300 stocks are covered by at least a news article each day and more than 250 stocks are covered by at least a news alert.

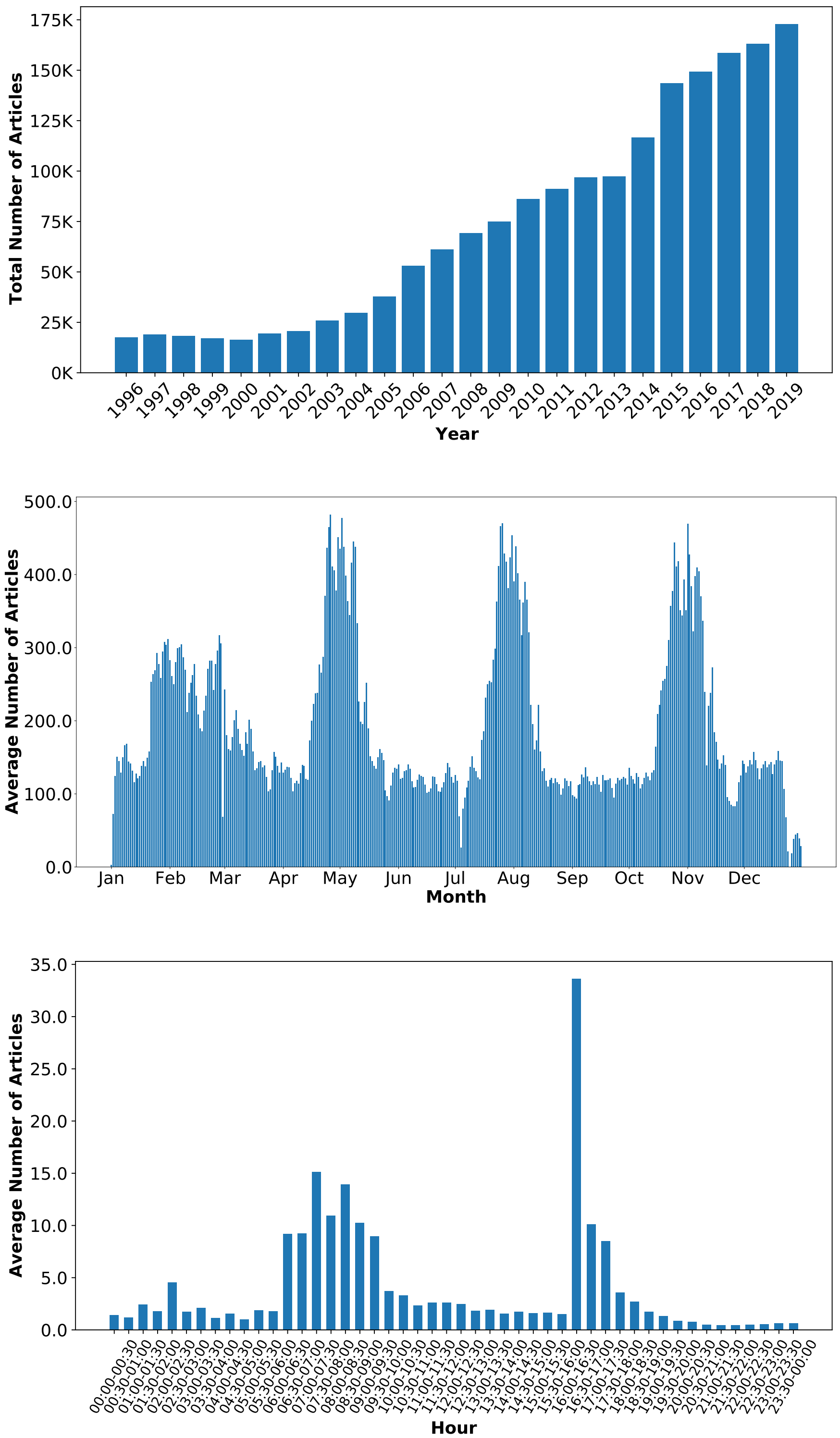
For BOW and W2V models, we apply the same pre-processing approach described in Ke et al. (2019). We use the natural language processing package spaCy to preprocess the data. First, we normalize the text including converting the whole article to lower case and expanding contractions such as “haven’t” to “have n’t”. Then, we lemmatize all words to their base forms, e.g., “was” to “be”, and “n’t” to ”not”. In the thirds step, we tokenize the article into a list of words. The fourth step removes pronouns, proper nouns, punctuations, special symbols, numbers, non-English words, and stop words such as “and”, “the”, and “is”. Finally, each article is transformed into a vector of word counts, which can then be used as the “bag of words” representation or to calculate the average word embedding. For BERT models, we do not apply pre-processing and feed the raw article to the model since BERT is pre-trained on unprocessed raw corpus.

Figure 3.1: US News Article Counts



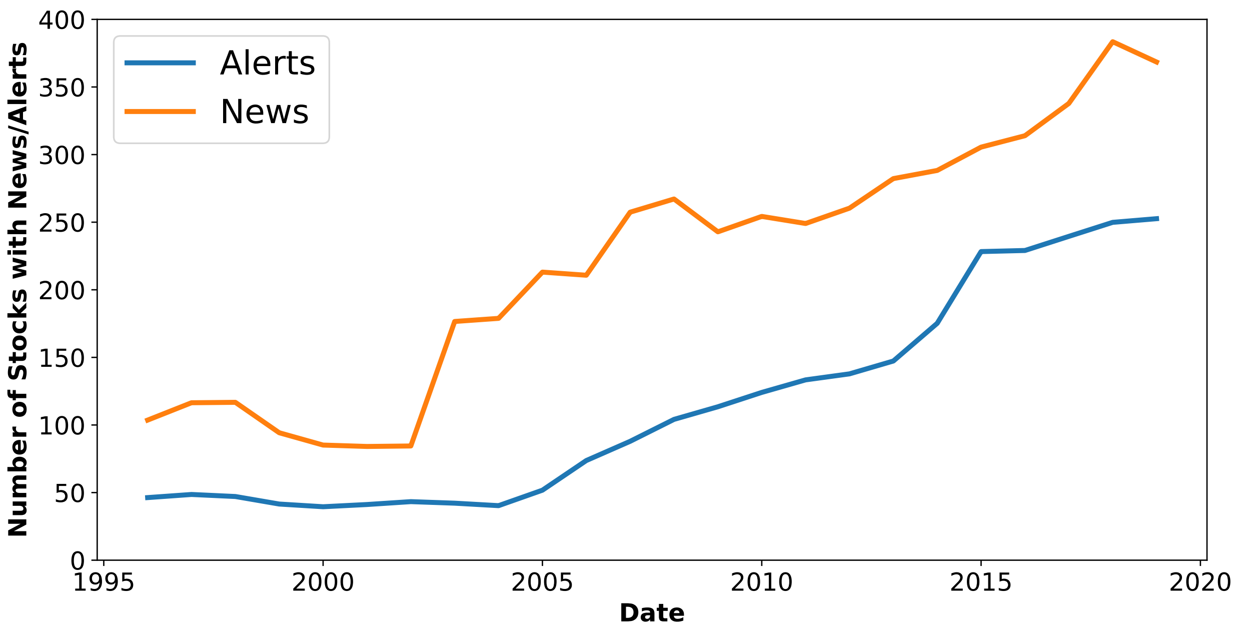
Note: The top figure plots the annual time series of the total number of news articles per year, the middle figure plots the average numbers of news articles per half an hour (24 hour local time), and the bottom figure plots the average numbers of news articles per calendar day.

Figure 3.2: US News Alerts Counts



Note: The top figure plots the annual time series of the total number of news alerts per year, the middle figure plots the average numbers of news alerts per half an hour (24 hour local time), and the bottom figure plots the average numbers of news alerts per calendar day.

Figure 3.3: US Average Daily Number of Stocks with News/Alerts



Note: This figure plots the average daily number of stocks with news/alerts.

3.2 Model Training

We train each model using annually updated rolling windows. Each rolling window consists of a 15 year interval for in-sample training with the first 10 years used as training set and the next 5 years as validation set. The subsequent one-year data is then set aside for out-of-sample testing. The out-of-sample data range from 2011 to 2019, totaling 9 years. Following Ke et al. (2019), each model is trained to predict the cross-sectionally rank-normalized three-day return, beginning the day before the article is published and ending the day after. Note that the three-day return is only used for training sentiment of news.

A vocabulary is constructed for each out-of-sample year based on the trailing in-sample data. The vocabulary is used to filter words for BOW and W2V models as well as to calculate news novelty to filter out "old news". For all in-sample news articles, we only include words that appear at least 1000 times and exclude stop words and pronouns. Table. 3.2 shows the size of in-sample vocabulary for US.

Table 3.2: Size of US In-Sample Vocabulary for Out-of-Sample Years

	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Vocab
US	8843	8909	8969	9025	9064	9085	9120	9146	9178	9539

Note: The table reports the size of vocabulary for each year built from the trailing 15 years of news. The vocabulary universe is built with words that occur more than 1000 times in the in-sample data.

3.3 Predicting US Returns with News Articles

The BERT(LOGIT) and W2V(LOGIT) models both predicts an estimate for the probability of a positive sentiment. The other four models each predicts a raw sentiment score with positive values indicating positive sentiments and negative values vice versa. For all 6 models, the larger the predicted outcome the more positive sentiment they assign to a news article. By setting the threshold at 50% for BERT(LOGIT) and W2V(LOGIT) and 0.0 neutral sentiment for the rest 4 models, we can calculate the prediction accuracy that matches that realized three-day returns.

Table 3.3: Out-of-Sample Prediction Accuracy and Correlation

	BERT(LOGIT)		BERT(OLS)		W2V(LOGIT)		W2V(OLS)		BOW100		BOW200	
	Acc.	Corr.	Acc.	Corr.	Acc.	Corr.	Acc.	Corr.	Acc.	Corr.	Acc.	Corr.
2011	52.4%	9.7%	52.4%	9.9%	52.6%	10.8%	52.5%	10.5%	53.3%	11.4%	53.2%	11.8%
2012	53.2%	10.4%	53.3%	10.8%	53.3%	11.9%	53.0%	12.1%	53.7%	12.6%	53.3%	12.5%
2013	51.6%	9.2%	51.9%	9.7%	53.9%	12.9%	54.0%	13.0%	54.6%	13.6%	54.2%	13.2%
2014	53.1%	11.7%	53.2%	11.8%	52.9%	10.2%	52.9%	10.0%	53.1%	10.2%	52.9%	10.4%
2015	53.1%	12.6%	53.3%	12.9%	51.6%	7.1%	51.7%	7.2%	52.1%	8.0%	52.1%	8.4%
2016	53.1%	11.6%	53.1%	11.7%	52.0%	7.0%	51.9%	7.6%	52.3%	8.0%	52.5%	8.6%
2017	53.5%	12.4%	53.4%	12.6%	52.5%	8.1%	52.5%	8.8%	52.4%	8.1%	52.5%	8.9%
2018	52.6%	9.9%	52.4%	10.1%	51.6%	6.9%	51.7%	7.1%	51.3%	6.4%	51.5%	7.4%
2019	52.8%	11.8%	52.8%	11.9%	52.8%	8.0%	52.8%	8.2%	52.1%	8.5%	53.1%	9.1%
Total	52.8%	11.0%	52.9%	11.3%	52.6%	9.2%	52.5%	9.4%	52.8%	9.6%	52.8%	10.0%

Note: The table reports out-of-sample prediction performance for the models. We calculate classification accuracy and correlation cross-sectionally each period then report time series averages over each period in the test sample.

Table. 3.3 reports the out-of-sample prediction accuracy and correlation (all statistics are significant, so we suppress p-values). All 6 models exceed the accuracy of a random guess (50%) and the two BERT models have the highest accuracy and correlation compared with other 4. In particular, the average cross-sectional correlation of BERT(LOGIT) and BERT(OLS) are 11.0% and 11.3%, respectively, significantly higher than the best of the classical models BOW200 at 10.0%. The other three models all have correlation less than 10%.

3.3.1 News Articles Portfolio Performance

To further exploit the advantage of BERT models, we conduct portfolio analysis in this section. We follow the approach of Ke et al. (2019) to form long-short portfolios. Specifically, we exclude news from 30 minutes before market open to market open (9:00 to 9:30 EST for US) from trading, although these news are still used for training and validation purposes. We form a zero-investment portfolio that longs the top 50 stocks with the highest predicted sentiment and shorts the bottom 50 stocks with the lowest predicted sentiment. When a stock is mentioned by multiple news on the same day, we predict the sentiment for the next day using average sentiments among multiple news. For news that occur on day 0, we build positions at the market opening on day 1, and rebalance at the next market opening, holding the positions of the portfolio within the day. We call this portfolio Day+1 portfolio. Similarly, we can define Day 0 and Day$\pm$1 portfolios. We report portfolio performance in term of both equal-weighted and value-weighted.

Table 3.4: US News Articles Day+1 Portfolio Performance

BERT(LOGIT)

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.41	-0.18	0.59	0.27	0.08	0.19
Std	0.18	0.18	0.13	0.16	0.18	0.12
SR	2.29	-1.00	4.60	1.66	0.47	1.60
Turnover			22.51			25.45

BERT(OLS)

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.41	-0.20	0.61	0.26	0.08	0.18
Std	0.18	0.18	0.13	0.16	0.17	0.12
SR	2.24	-1.08	4.58	1.60	0.47	1.52
Turnover			22.49			25.50

W2V(LOGIT)

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.34	-0.06	0.40	0.22	0.14	0.08
Std	0.17	0.19	0.13	0.16	0.18	0.11
SR	1.95	-0.35	3.17	1.33	0.79	0.69
Turnover			22.37			25.10

W2V(OLS)

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.31	-0.10	0.41	0.21	0.16	0.05
Std	0.18	0.19	0.13	0.16	0.18	0.12
SR	1.73	-0.54	3.19	1.28	0.87	0.42
Turnover			22.33			25.14

BOW100

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.33	-0.01	0.34	0.25	0.14	0.11
Std	0.17	0.19	0.12	0.17	0.17	0.11
SR	1.91	-0.05	2.80	1.51	0.82	0.95
Turnover			21.20			22.86

BOW200

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.35	-0.09	0.44	0.26	0.13	0.13
Std	0.18	0.18	0.12	0.17	0.17	0.12
SR	1.98	-0.50	3.81	1.53	0.76	1.10
Turnover			21.73			24.09

PAST 1-DAY TREND

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.05	-0.22	0.27	-0.15	-0.29	0.15
Std	-0.22	-0.24	-0.20	-0.21	-0.24	-0.23
SR	0.21	-0.93	1.34	-0.68	-1.22	0.66
Turnover			21.34			25.09

PAST 5-DAY TREND

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.02	-0.29	0.31	-0.25	-0.24	-0.00
Std	-0.21	-0.25	-0.21	-0.20	-0.25	-0.23
SR	0.09	-1.16	1.45	-1.26	-0.97	-0.02
Turnover			20.12			21.42

PAST 20-DAY TREND

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.01	-0.40	0.40	-0.22	-0.30	0.08
Std	-0.20	-0.26	-0.22	-0.19	-0.27	-0.25
SR	0.03	-1.50	1.81	-1.16	-1.12	0.34
Turnover			19.41			19.90

PAST 60-DAY TREND

	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	-0.06	-0.32	0.26	-0.25	-0.31	0.06
Std	-0.20	-0.27	-0.23	-0.20	-0.28	-0.26
SR	-0.30	-1.18	1.14	-1.23	-1.10	0.25
Turnover			19.03			18.93

Note: Performance of equal-weighted and value-weighted day+1 long/short portfolios formed by top/bottom 50 out-of-sample predicted returns. Each panel reports annualized returns, annualized standard deviations, and annualized Sharpe ratios for long, short, and long-short portfolios. We also report monthly turnover of each strategy. We compare 10 different models, namely BERT(LOGIT), BERT(OLS), W2V(LOGIT), W2V(OLS), BOW100, BOW200, and 4 reversal indicators of length 1/5/20/60 days. All BERT models uses embedding of the first 512 tokens generated by the pretrained BERT model from Hugging Face. BERT(LOGIT) and BERT(OLS) apply logistic regression and OLS regression, respectively, on the feature embedding. W2V(LOGIT) and W2V(OLS) apply logistic regression and OLS regression, respectively, on the average of vectors of all words from each news. BOW100 and BOW200 are Bag-of-words model from Ke et al. (2019) with $\alpha_+ = \alpha_- = 100$ and $\alpha_+ = \alpha_- = 200$, respectively.

Table. 3.4 reports the performance of Day+1 long/short portfolios in terms of annualized average returns and Sharpe ratios. For comparison we include traditional trend indicators based on past 1/5/20/60-day returns. For each model, we report the equal-weighted portfolio on the left and value-weighted portfolio on the right, each including statistics of the long/short leg and the long-short portfolio. Following Gu et al. (2020), we calculate monthly turnover as:

$$\text{Turnover} = \frac{1}{M}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i}\left|w_{i,t+1} - \frac{w_{i,t}(1+r_{i,t+1})}{1+\sum_j w_{j,t}r_{j,t+1}}\right|\right),$$

where M is the number of months in the holding period, T is the number of trading periods, $r_{i,t+1}$ is the return of stock i at time $t+1$, and $w_{i,t}$ is the portfolio weight of stock i at time t. Since the trading frequency is daily, we multiply the turnover by 21 to approximate the average monthly turnover.

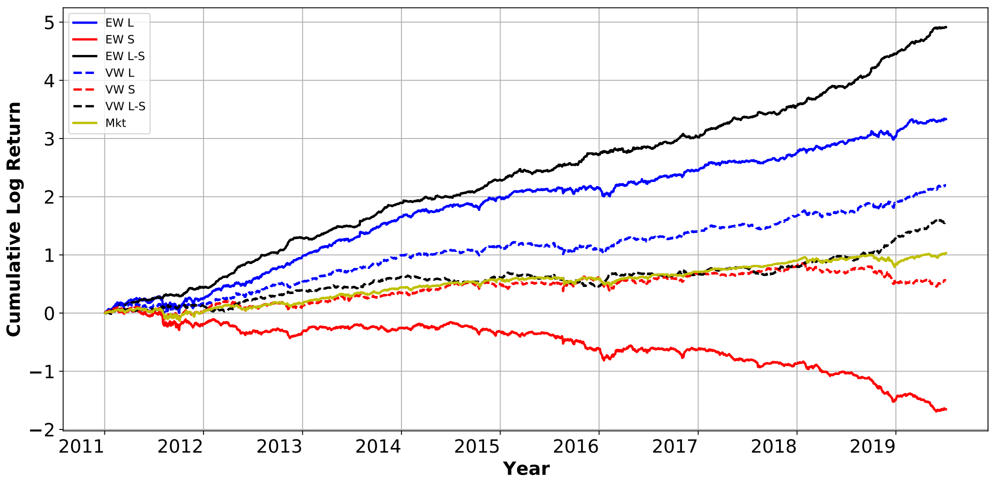
The turnovers of the Past 1/5/20/60-Day TREND strategies can be used as a benchmark for turnover, which is around 20 (except for 25.09 for the value weighted portfolio of PAST 1-DAY RET). The turnover of the two BERT models are generally the highest with equal-weight at 22.5 and value-weight at 25.5. The W2V and BOW models have turnover slightly higher than TREND strategies but lower than BERT models.

In terms of Sharpe ratios, both BERT models dominate the rest in both equal-weighted portfolio and value-weighted portfolio. BERT(LOGIT) produces a L-S equal-weighted Sharpe ratio of 4.6, which is more than 2 times the highest TREND strategies (1.81 for PAST 20-DAY TREND), and a L-S value-weighted Sharpe ratio of 1.6, roughly 3 times the highest TREND strategies (0.66 for PAST 1-DAY TREND). BERT(OLS) almost has the same great performance as BERT(LOGIT). Except for BOW200, none of the rest models has value-weighted Sharpe ratio greater than 1. Aside from the two BERT models, BOW200 has the highest Sharpe ratio both equal-weighted and value-weighted.

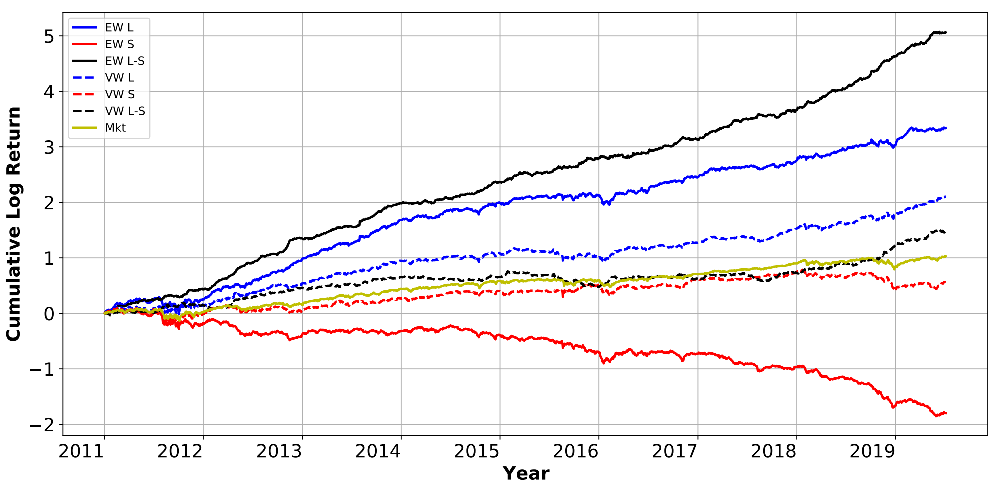
Figure. 3.5 plots the cumulative one-day trading strategy returns (calculated from open-to-open) based on out-of-sample BERT models predictions. We report the long and short legs separately, as well as the overall long-short strategy performance. In addition, we compare the performance of equal-weighted and value-weighted

Figure 3.4: US One-day-ahead Performance Comparison (BERT)

BERT(LOGIT)



BERT(OLS)



Note: This figure compares the out-of-sample cumulative log returns of portfolios sorted on sentiment scores. The black, blue, and red colors represent the long-short (L-S), long (L), and short (S) portfolios, respectively. The solid and dashed lines represent equal-weighted (EW) and value-weighted (VW) portfolios, respectively. The yellow solid line is the market return.

Table 3.5: Exposure to Aggregate Risk Factors

	Sharpe Ratio	Average Return	FF3		FF5		FF5+MOM	
			α	R^2	α	R^2	α	R^2
	BERT(LOGIT)							
EW L-S	4.59	58	58	0.5%	57	0.6%	57	2.0%
EW L	2.29	40	34	34.2%	34	34.6%	35	34.7%
EW S	0.99	17	23	34.3%	22	34.9%	22	36.1%
VW L-S	1.59	19	19	0.8%	18	1.1%	18	1.5%
VW L	1.67	27	19	27.9%	19	28.1%	19	28.2%
VW S	-0.47	-8	0	26.1%	0	26.1%	-1	26.2%
	BERT(OLS)							
EW L-S	4.57	60	59	0.7%	59	0.7%	59	2.1%
EW L	2.24	41	34	32.5%	35	32.8%	35	32.9%
EW S	1.07	19	24	33.8%	24	34.3%	24	35.6%
VW L-S	1.52	17	18	1.2%	17	1.4%	17	1.6%
VW L	1.60	26	18	27.3%	18	27.5%	18	27.5%
VW S	-0.47	-8	0	26.9%	0	27.1%	0	27.1%

Note: The table reports the performance of equal-weighted (EW) and value-weighted (VW) long-short (L-S) portfolios and their long (L) and short (S) legs. The performance measures include (annualized) annual Sharpe ratio, annualized expected returns, risk-adjusted alphas, and R^2s with respect to the Fama-French three-factor model ("FF3"), the Fama-French five-factor model ("FF5"), and the Fama-French five-factor model augmented to include the momentum factor ("FF5+MOM").

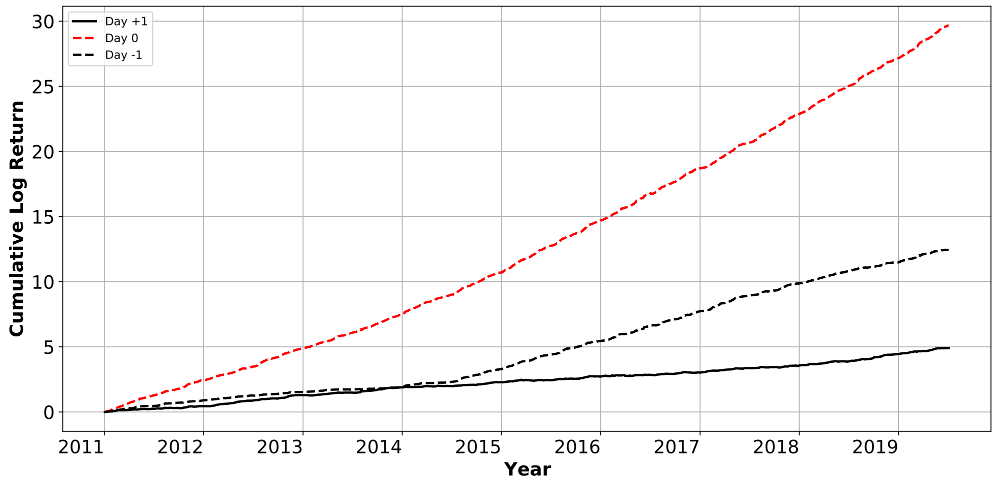
Table. 3.6 shows that the BERT models have little exposure to the aggregate risk factors. The individual long and short legs of the portfolio have at most a 36.1% daily R^2 when regressed on Fama-French factors. As for the long-short portfolio, the R^2 is at most 2.1%. The average returns of the spread portfolio are almost all alpha for both models.

3.3.2 Lead-lag Relationship Among News and Prices

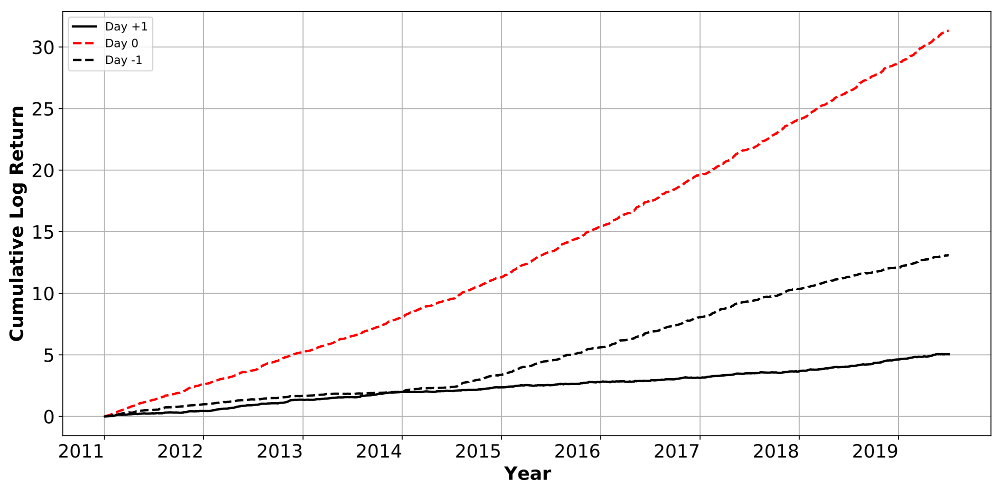
We use three-day return, starting from one day before the news and end at the next day after the news is published, when training the model. In Figure. 3.5, we separately investigate the subsequent out-of-sample association between news sentiment on day t and returns on day t-1 (from open t-1 to open t), day t, and day t+1. As each year's models are trained using rolling windows of preceding years, the portfolio formations of all three portfolios are complete out-of-sample. Those two strategies Day -1 and Day 0 are not implementable since traders don't have time take positions based on the real-time signals. However, the portfolio performance of the above two

Figure 3.5: US Price Response On Days -1, 0, and +1

BERT(LOGIT)



BERT(OLS)



Note: This figure compares the out-of-sample cumulative log returns of long-short portfolios sorted on sentiment scores. All strategies are equal-weighted. The Day -1 strategy (dashed black line) shows the association between news and returns one day prior to the news; the Day 0 strategy (dashed red line) shows the association between news and returns on the same day; and the Day +1 strategy (solid black line) shows the association between news and returns one day later.

strategies can be interpreted as out-of-sample correlation between the predicted article sentiments and realized returns. Furthermore, the Day +1 strategy is the implementable trading strategy with significant returns.

Table. 3.6 reports the summary statistics for these portfolios, including their annualized Sharpe

Table 3.6: Price Response On Days -1, 0, and +1

	Sharpe Ratio	Average Return	FF3 α	FF3 R^2	FF5 α	FF5 R^2	FF5+MOM α	FF5+MOM R^2
				BERT(LOGIT)				
				Day -1				
L-S	10.02	148	147	0.3%	147	0.5%	147	0.6%
L	5.41	100	94	33.8%	94	33.9%	94	34.2%
S	2.54	47	53	33.6%	52	33.9%	52	34.6%
				Day 0				
L-S	18.60	353	352	0.4%	352	0.5%	352	0.8%
L	10.04	207	201	28.3%	201	28.6%	201	28.6%
S	7.26	145	151	27.2%	151	27.4%	150	27.9%
				Day +1				
L-S	4.59	58	58	0.5%	57	0.6%	57	2.0%
L	2.29	40	34	34.2%	34	34.6%	35	34.7%
S	0.99	17	23	34.3%	22	34.9%	22	36.1%
				BERT(OLS)				
				Day -1				
L-S	10.41	155	154	0.5%	154	0.7%	154	0.9%
L	5.46	102	95	34.5%	95	34.7%	96	34.9%
S	2.86	53	59	33.0%	58	33.3%	58	34.0%
				Day 0				
L-S	19.49	373	372	0.5%	372	0.6%	372	1.0%
L	10.24	213	206	29.1%	206	29.4%	206	29.4%
S	7.94	160	165	26.6%	165	26.9%	165	27.5%
				Day +1				
L-S	4.57	60	59	0.7%	59	0.7%	59	2.1%
L	2.24	41	34	32.5%	35	32.8%	35	32.9%
S	1.07	19	24	33.8%	24	34.3%	24	35.6%

Note: The table repeats the analysis of Table 2 for the equal-weighted long-short (L-S) portfolios, as well as their long (L) and short (S) legs. Sharpe ratios are annualized, while returns and alphas are in basis points per day.

ratios, average returns, alphas, and turnover. For this analysis, we specialize on equally weighted portfolios. The Day-1 strategy proxies the correlation between today's news and last days return. The Sharpe ratio of around 10 for both BERT(LOGIT) and BERT(OLS) provides significant evidence that prices move ahead of news to some extent. We found out in same manually selected cases that this behavior is led by return-led news. In other words, stocks with extreme returns tend to be reported in next day's news. The Day 0 strategy with high Sharpe ratio of 18.6 for BERT(LOGIT) and 19.5 for BERT(OLS) indicates a strong contemporary correlation between returns and news. Although impossible to trade, the Day 0 strategy shows that our BERT models indeed extract new information from the same day's fresh news that affects the same day's return.

The Day+1 strategy, which is a tradable strategy, achieves Sharpe ratios of 4.6 for both models. Therefore, it is clear that the predicted signal from news is robust to a one-day delay.

3.4 Predicting US Returns with News Alerts

In the previous section, we have analyzed the return predictability of news articles with detailed bodies. Two natural question arises. First, does the detailed information from news articles really matter? Or in other words, for a more accurate prediction, do we want to trade the lack of detailed information for a faster releasing speed of news? Second, when the news is shorter with less context, do BERT models still have advantage over W2V and BOW that rely on static word features? The news from Thomson Reuters Real-time News Feed that is tagged as alerts serves as a great source of data to answer the above questions. The alerts consist of only a headline with no body texts. The headline is often a sentence briefly summarizing the real-time event. We train separate models with news alerts only following the same methodology describe in the previous sections and form portfolios based on out-of-sample news alerts only. As we see in Figure. 3.3, the average daily number of stocks with alerts is at least one half of that of stocks with news articles. Since we only trade the top and bottom 50 stocks, the comparison of portfolio performance will hardly be affected by the different sizes of stock universe.

3.4.1 News Alerts Portfolio Performance

Table. 3.7 reports the portfolio performance of Day+1 long-short based on news alerts. The Sharpe ratios under all settings are generally higher that those of portfolios based on news articles. Except for W2V(OLS) (with equal-weighted Sharpe ratio of 5.92), all models have equal-weighted Sharpe ratios greater than 6. BERT(LOGIT) still have the highest value-weighted Sharpe ratio of 3.51. The higher Sharpe ratios of news alerts confirm our hypothesis that the releasing speed of news over weighs the detailed news information since the fresh news is impounded into stock prices quickly. Another interesting finding is the great performance of BOW models. In particular, BOW200

Table 3.7: US News Alerts Day+1 Portfolio Performance

BERT(LOGIT)(ALERT)						
	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.51	-0.29	0.79	0.36	-0.02	0.38
Std	0.18	0.19	0.12	0.16	0.17	0.11
SR	2.81	-1.54	6.53	2.20	-0.12	3.51
Turnover			21.20			22.26

BERT(OLS)(ALERT)						
	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.49	-0.29	0.78	0.35	-0.00	0.35
Std	0.18	0.19	0.12	0.17	0.17	0.11
SR	2.69	-1.57	6.31	2.10	-0.03	3.16
Turnover			21.19			22.22

W2V(LOGIT)(ALERT)						
	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.51	-0.28	0.79	0.34	0.02	0.32
Std	0.17	0.19	0.13	0.16	0.17	0.11
SR	2.97	-1.45	6.06	2.09	0.11	2.84
Turnover			21.24			22.28

W2V(OLS)(ALERT)						
	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.50	-0.28	0.78	0.34	-0.01	0.35
Std	0.17	0.20	0.13	0.16	0.17	0.11
SR	2.89	-1.44	5.92	2.12	-0.03	3.15
Turnover			21.20			22.26

BOW100(ALERT)						
	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.56	-0.30	0.86	0.36	0.03	0.33
Std	0.18	0.20	0.13	0.17	0.17	0.12
SR	3.11	-1.52	6.59	2.16	0.19	2.68
Turnover			20.42			20.72

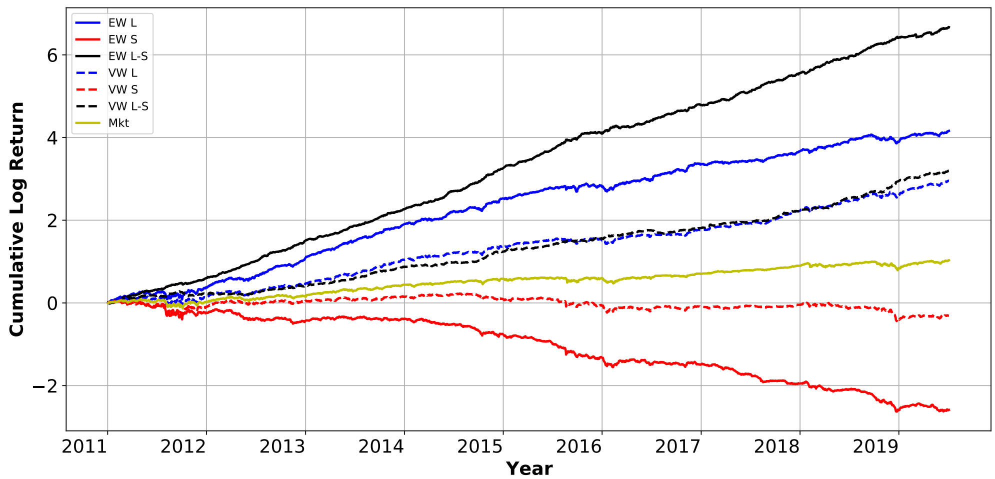
BOW200(ALERT)						
	EW			VW		
	Long	Short	L-S	Long	Short	L-S
Ret	0.55	-0.31	0.86	0.38	0.00	0.38
Std	0.18	0.19	0.12	0.16	0.17	0.11
SR	3.08	-1.64	6.92	2.33	0.01	3.31
Turnover			20.52			21.15

Note: Performance of equal-weighted and value-weighted day+1 long/short portfolios formed by top/bottom 50 out-of-sample predicted returns based on alerts with headlines only. Each panel reports annualized returns, annualized standard deviations, and annualized Sharpe ratios for long, short, and long-short portfolios. We also report monthly turnover of each strategy.

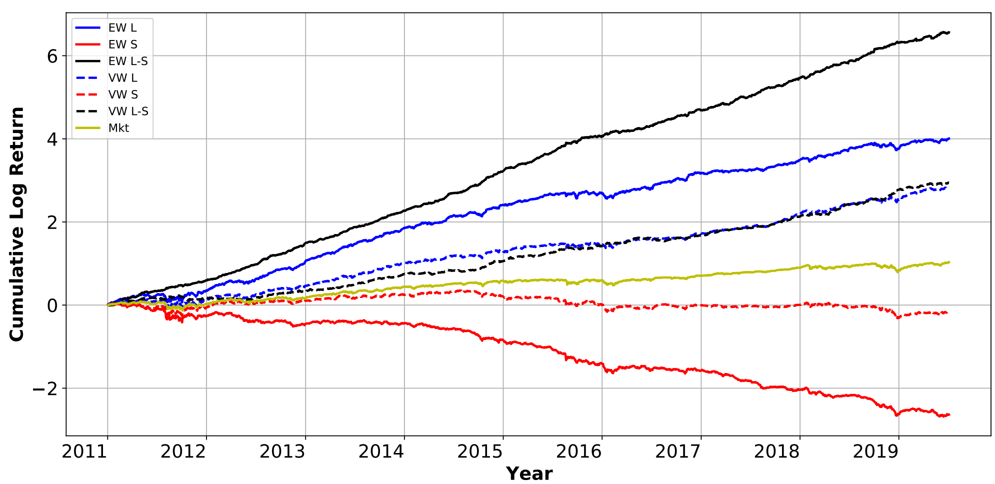
achieves equal-weighted Sharpe ratio of 6.92, higher than all other models including BERT models (with BERT(LOGIT) having equal-weighted Sharpe ratio of 6.53). It is not surprising that BERT models lose the edge of context interpretation when each news consist of only headlines. However, it is worth mentioning that BERT(LOGIT) still has the highest value-weighted Sharpe ratio of 3.51, which is more practical to trade. Figure. 3.6 plots the cumulative one-day trading strategy returns (calculated from open-to-open) based on out-of-sample BERT models predictions.

Figure 3.6: US Alert One-day-ahead Performance Comparison (BERT)

BERT(LOGIT)

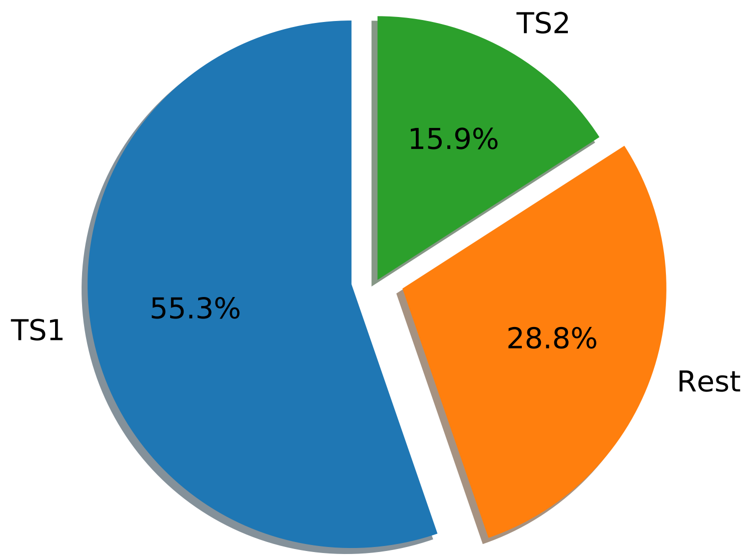


BERT(OLS)



Note: This figure compares the out-of-sample cumulative log returns of portfolios sorted on sentiment scores. The black, blue, and red colors represent the long-short (L-S), long (L), and short (S) portfolios, respectively. The solid and dashed lines represent equal-weighted (EW) and value-weighted (VW) portfolios, respectively. The yellow solid line is the market return.

Figure 3.7: Percentage of Alerts with Different Take Sequences



Note: TS1 corresponds to alerts of take sequence 1, TS2 corresponds to alerts of takes sequence 2, and Rest corresponds to all alerts with take sequence greater than 2.

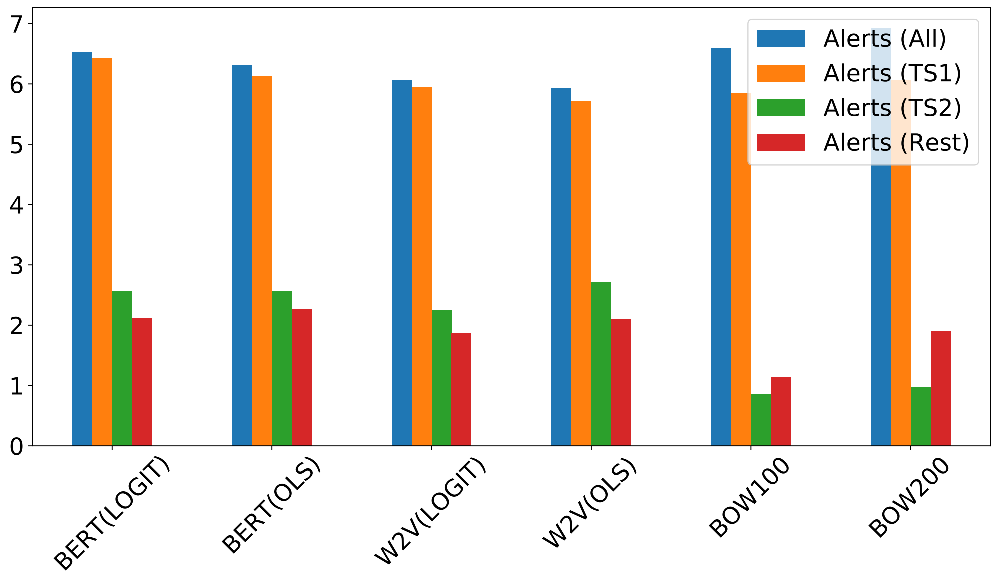
3.4.2 Fresh Alerts and Stale Alerts

In this section, we investigate the difference in price response to fresh versus stale alerts, where the novelty of alerts is defined by the take sequence of the news. For all news alerts covering the same story, each news alert is labeled by a take sequence number starting from 1 (the freshest). A take sequence greater than 1 reports the follow-up development. We split all alerts into three partitions: take sequence 1 (denoted as TS1), take sequence 2 (TS2), and take sequence greater than 2 (Rest). Figure. 3.7 shows the distribution of the tree partitions, where TS1 accounts for 55.3% of all alerts, TS2 for 15.9%, and the rest alerts for 28.8%.

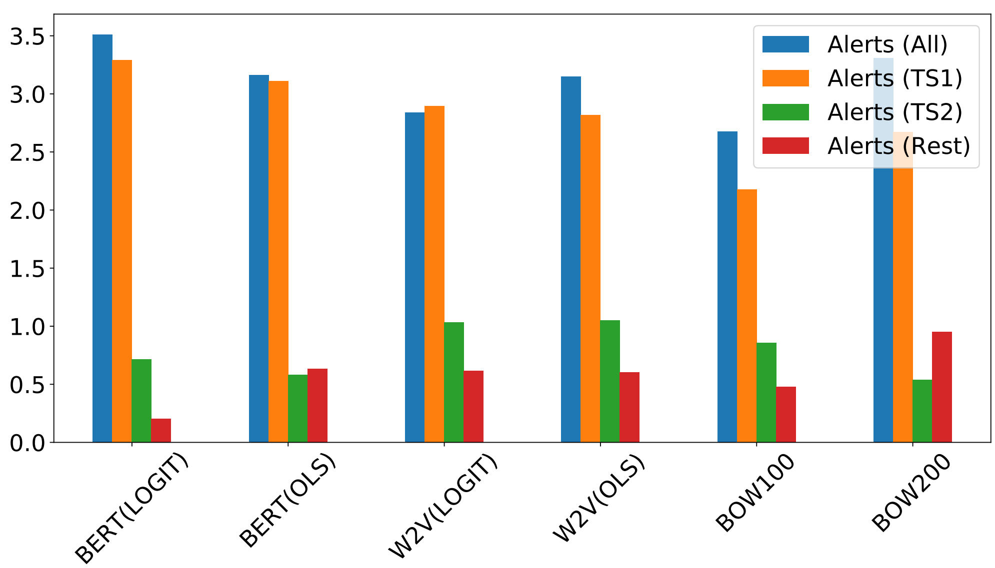
For each partition, we form a long-short portfolio using alerts under that partition only. For a fair comparison, the model used for all partitions is the same model trained with all alerts. Figure. 3.8 shows the SR decay versus the novelty of alerts for all models. We can see from the bar chart that portfolios using all alerts have the highest Sharpe ratios in terms of both equal-weighted

Figure 3.8: SR Decay of US Alerts with Different Take Sequences

Equal Weight



Value Weight



Note: This figure reports the SR decay of long-short portfolio of alerts with different take sequences under the universe of all three US stock exchanges. The blue bars correspond to portfolio using all alerts, the orange bars, green bars, and red bars correspond to alerts of take sequence 1, 2, and all after 2, respectively. Each model is trained using all alerts.

and value-weighted portfolios for all models (with an exception of W2V(LOGIT) value-weighted portfolio). Although the number of TS1 alerts is only around one half of all alerts, the Sharpe ratios of portfolios they generate are almost as high as those of portfolios formed by using all alerts. As the market has absorbed the new information from the first take sequence of stories, further reports from the later take sequences have little impact on the stock price. It is further supported by the low Sharpe ratios of TS2 and Rest. One may suspect that the lower performance of TS2 is due to the smaller amount of news (only less than one third of TS1). However, the portfolio performance of alerts from the Rest partition is generally worse than that of TS2, even though the Rest partition is twice as large as TS2.

The analysis of alerts with different take sequences provides further evidence that stock prices take in fresh news quite efficiently.

3.5 Portfolio Performance on Large Stocks

Table 3.8: SR of Long-Short Portfolio with News and Alerts

	EW					VW				
	Articles	Alerts (All)	Alerts (TS1)	Alerts (TS2)	Alerts (Rest)	Articles	Alerts (All)	Alerts (TS1)	Alerts (TS2)	Alerts (Rest)
All Stocks										
BERT(LOGIT)	4.60	6.53	6.42	2.57	2.12	1.60	3.51	3.29	0.71	0.21
BERT(OLS)	4.58	6.31	6.13	2.56	2.26	1.52	3.16	3.11	0.58	0.63
W2V(LOGIT)	3.17	6.06	5.94	2.25	1.87	0.69	2.84	2.90	1.03	0.62
W2V(OLS)	3.19	5.92	5.72	2.71	2.09	0.42	3.15	2.82	1.05	0.60
BOW100	2.80	6.59	5.85	0.85	1.14	0.95	2.68	2.18	0.86	0.48
BOW200	3.81	6.92	6.06	0.97	1.90	1.10	3.31	2.67	0.54	0.95
S&P 500										
BERT(LOGIT)	0.85	2.12	1.97	0.31	0.13	0.55	1.76	1.78	0.24	-0.05
BERT(OLS)	0.77	2.31	1.96	0.10	0.44	0.60	1.67	1.78	0.02	0.24
W2V(LOGIT)	0.88	2.66	2.50	1.25	0.24	-0.14	2.36	1.89	1.62	0.21
W2V(OLS)	0.56	2.34	2.28	1.26	0.52	-0.12	2.18	1.95	1.58	0.68
BOW100	0.93	2.19	2.09	0.19	0.96	0.26	1.58	1.56	0.17	0.82
BOW200	1.00	2.82	2.24	0.59	1.28	0.08	1.91	1.41	0.63	0.93

Note: The table shows the Sharpe ratios of long-short portfolio with news and alerts of different take sequences. The top panel shows the SR of portfolios formed by all stocks; the bottom panel shows the SR of portfolios formed by S&P 500 only. the columns correspond to news, all alerts, alerts of take sequence 1, 2, and all after 2, respectively. News models and alerts models are trained separately with news and alerts data.

Since it's more practical to trade only stocks with large market capitalization for higher liquidity and lower transaction cost, we form long-short portfolios on S& P 500 stocks under different settings. Table. 3.8 reports the portfolio performance on S&P 500 (top panel) compared with that on the whole universe (bottom panel). The Sharpe ratios decrease significantly when we limit to trading only large stocks. The BERT(LOGIT) model has equal-weighted Sharpe ratio of 0.85 using articles and 2.12 using alerts. In term of the value-weighted portfolios formed from news articles, both BERT models show significant Sharpe ratios of 0.6 while all other models' performance is insignificant.

CHAPTER 4

NEWS PREDICTION FOR INTERNATIONAL COUNTRIES' STOCK RETURNS

Table 4.1: International News Summary Statistics

International News Article

	Raw Articles			Articles Tagged with Single Stock			Articles With	After Filtering	After Filtering
	RTRS	3PTY	Total	RTRS	3PTY	Total	Available Returns	Short Articles	Old News
UK	707,288	1,050,467	1,757,755	196,573	773,266	969,839	906,705	901,838	450,920
Australia	261,020	1,203,784	1,464,804	100,444	1,113,347	1,213,791	388,585	382,114	191,057
Canada	255,933	473,686	729,619	126,281	431,401	557,682	481,891	478,205	239,103
China	3,537,487	7,287,688	10,825,175	1,140,542	5,558,763	6,699,305	2,086,045	305,335	152,668
Japan	3,259,103	38,860	3,297,963	1,210,077	16,850	1,226,927	405,341	399,185	221,181
Germany	2,423,671	1,751,231	4,174,902	480,264	880,650	1,360,914	238,577	229,265	114,633
Italy	1,022,204	337,322	1,359,526	194,650	227,599	422,249	173,250	168,410	84,205
France	2,422,338	1,587,490	4,009,828	298,886	670,469	969,355	174,917	174,784	87,392
Sweden	288,395	189,424	477,819	96,039	124,862	220,901	126,211	126,168	63,084
Denmark	261,146	124,209	385,355	93,596	57,768	151,364	53,056	52,381	26,191
Spain	2,748,601	165,468	2,914,069	257,739	46,829	304,568	47,541	45,597	22,801
Finland	108	125,595	125,703	47	87,266	87,313	38,163	38,123	19,062
Portugal	1,097,055	39,086	1,136,141	155,755	13,638	169,393	11,284	11,231	5,616
Greece	85,915	14	85,929	19,156	6	19,162	10,093	10,082	5,041
Netherlands	37,215	213,732	250,947	12,267	69,156	81,423	7,137	7,128	3,564

Note: In this table, we report the impact of each filter we apply on the number of news articles and news alerts in the top and bottom panels respectively. Raw Articles shows the number of articles from Thomson Reuters Real-time News Feed (RTRS) and Archive (3PTY). Articles Tagged with Single Stock shows the number of articles/alerts that are tagged with a single stock. In the following three columns, we show the number of remaining articles/alerts after filtering out those with unavailable returns, those with short length of news body (less than 30 for China and less than 100 for all other countries, see Table. A.1 for details), those with larger than median cosine similarities. The parenthesized number of US corresponds to the number of articles/alerts of companies belonging to the S&P 500.

In this chapter, we expand the methodology to international countries in order to test the robustness of BERT models and its advantage over classical models under different settings, i.e. different market and news of different languages. Same as US news, we obtain news of 15 international countries from two sources: Thomson Reuters Real-time News Feed (RTRS) and Archive (3PTY). International countries don't have enough alerts available in recent years and therefore we only show the performance of news articles. We follow the same filtering procedure and the numbers of news articles are reported in Table. 4.1.

Following the same approach for US, we build a yearly updated vocabulary for each country using the in-sample 15-year rolling window. Table. 4.2 reports the size of vocabulary by year for each country.

Table 4.2: Size of International In-Sample Vocabulary for Out-of-Sample Years

	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Vocab
China	1627	1628	1705	1713	1779	1844	1890	1968	2014	2162
UK	2973	2974	2977	2981	2981	2984	2984	2984	2984	2987
Australia	1930	1931	1931	1929	1929	1929	1929	1927	1928	1933
Canada	3588	3589	3592	3593	3593	3594	3594	3595	3598	3605
Japan	1194	1194	1196	1204	1207	1204	1204	1202	1199	1217
Germany	2551	2576	2643	2663	2673	2698	2711	2729	2737	2805
Italy	2154	2165	2172	2176	2177	2178	2178	2178	2187	2199
France	2764	2785	2847	2869	2944	2976	2983	2990	2996	3015
Sweden	/	/	/	/	/	3105	3135	3152	3169	3260
Denmark	811	819	838	842	844	849	854	856	856	857
Spain	580	580	580	580	580	580	580	580	579	581
Finland	/	/	/	/	/	/	991	994	1000	1060
Portugal	/	/	155	156	156	161	161	161	161	161
Greece	/	/	/	/	/	/	/	180	180	183
Netherlands	/	127	127	126	124	128	127	126	126	131

Note: The table reports the size of vocabulary for each year built from the trailing 15 years of news. The vocabulary universe is built with words that occur more than 1000 times in the in-sample data. For some countries with not enough years of in-sample data, the cell is marked by "/".

4.1 Portfolio Performance

Table 4.3: International Countries Statistics

	Language	Market Open	Market Close	Total News Articles	Start Day	End Day	Total Days	Average News
US	English	09:30	16:00	2061640	1996-01-02	2019-08-01	5933	347.49
China	Chinese	09:30	16:00	152668	1996-01-02	2019-08-01	5569	27.41
UK	English	08:00	16:30	450920	1996-01-02	2019-08-01	5983	75.37
Australia	English	10:00	16:00	191057	1996-01-02	2019-08-01	5926	32.24
Canada	English	09:30	16:00	239103	1996-01-03	2019-08-01	5917	40.41
Japan	Japanese	09:00	15:00	221181	1996-01-04	2019-08-01	5777	38.29
Germany	German	09:00	17:30	114633	1996-01-02	2019-08-01	5940	19.30
Italy	Italian	09:00	17:30	84205	1996-01-05	2019-08-01	5715	14.73
France	French	09:00	17:30	87392	1996-01-02	2019-08-01	5924	14.75
Sweden	Swedish	09:00	17:25	63084	2001-06-06	2019-08-01	4522	13.95
Denmark	Danish	09:00	16:55	26191	1996-01-19	2019-07-31	4422	5.92
Spain	Spanish	09:00	17:30	22801	1996-01-04	2019-08-01	5340	4.27
Finland	Finnish	10:00	18:25	19062	2002-04-24	2019-08-01	3850	4.95
Portugal	Portuguese	11:30	16:30	5616	1998-06-17	2019-08-01	2538	2.21
Greece	Greek	10:15	05:20	5041	2003-02-19	2019-07-31	2721	1.85
Netherlands	Dutch	09:00	17:30	3564	1996-01-04	2019-07-31	2410	1.48

Note: This table summarizes market information and final dataset for each country. The columns correspond to the language of news articles, local times for market open and market close, total number of news articles, the start day and end day of news articles available, total number of days with news articles, and the average number of news articles per day.

Table. 4.3 reports the market information of international countries, including languages of news articles, local times for market open and market close and the summary statistics of the news dataset we use. Almost all countries has news data starting from 1996, with a few exceptions, Sweden starting from 2001, Finland starting from 2002, Portugal starting from 1998, and Greece starting from 2003. The average daily number of news ranges from 1.5 to 75.4. There are 6 countries whose average daily number of news less than 10. China, UK, Australia, Canada, and Japan all have at least 25 news per day on average.

Table 4.4: International Day+1 Porfolio Performance

	BERT(LOGIT)		BERT(OLS)		W2V(LOGIT)		W2V(OLS)		BOW100		BOW200	
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW
US	4.60	1.60	4.58	1.52	3.17	0.69	3.19	0.42	2.80	0.95	3.81	1.10
China	1.33	0.68	1.38	0.94	0.65	0.53	1.05	0.66	1.15	0.91	0.66	0.61
UK	2.26	0.86	2.49	0.80	1.96	0.94	1.70	0.99	1.03	0.42	1.52	0.42
Australia	0.02	0.00	-0.17	-0.10	0.43	0.74	0.24	0.59	-0.17	-0.14	-0.17	-0.35
Canada	1.94	0.76	2.24	1.14	1.47	1.35	1.85	1.42	-0.02	-0.06	0.21	0.16
Japan	0.10	0.26	0.17	0.27	0.40	-0.15	0.36	-0.33	-1.23	-0.96	-0.87	-0.72
Germany	1.42	0.88	1.59	1.07	1.34	1.50	1.11	1.07	0.71	1.07	0.89	0.90
Italy	0.71	0.00	0.84	0.50	0.98	0.39	0.98	0.27	0.51	0.34	0.51	0.13
France	1.18	0.45	1.09	0.08	0.79	-0.43	0.67	0.17	-0.13	0.26	0.35	0.29
Sweden	1.01	0.79	0.48	0.33	1.42	0.15	1.55	0.49	-0.41	0.53	0.82	1.14
Denmark	0.65	0.81	0.90	1.08	0.68	1.00	-0.15	-0.02	0.34	0.19	0.24	0.42
Spain	0.87	0.56	0.62	0.45	0.66	0.84	-0.33	-0.36	-0.39	-0.58	0.15	-0.25
Finland	-1.15	-0.81	-0.99	-0.97	-0.08	-0.14	-0.97	0.07	-1.32	-0.14	-1.07	-0.09
Portugal	0.26	0.41	-0.46	-0.41	0.21	0.20	0.56	0.60	0.90	1.05	0.90	1.05
Greece	0.05	0.47	-0.29	0.46	2.71	3.09	3.02	3.01	1.16	1.12	1.16	1.12
Netherlands	1.01	1.04	0.67	0.97	0.09	-0.01	-0.60	-0.31	-0.35	-0.34	-0.35	-0.34

Note: This table reports the equal-weight and value-weight SR of long-short portfolio formed by different models. We compare 7 different models, namely BERT(LOGIT), BERT(OLS), W2V(LOGIT), W2V(OLS), BOW100, and BOW200. All BERT models uses embedding of the first 512 tokens generated by the pretrained BERT model from Hugging Face. BERT(LOGIT) and BERT(OLS) apply logistic regression and OLS regression, respectively, on the feature embedding. W2V(LOGIT) and W2V(OLS) apply logistic regression and OLS regression, respectively, on the average of vectors of all words from each news. BOW100 and BOW200 are Bag-of-words model from Ke et al. (2019) with $\alpha_+ = \alpha_- = 100$ and $\alpha_+ = \alpha_- = 200$, respectively, with exception of Portugual, Greece and Netherlands, which uses $\alpha_+ = \alpha_- = 50$ due to limited size of vocabularies.

Table. 4.4 shows the performance of day+1 long-short portfolios based on news articles for all international countries.

CHAPTER 5

WHAT'S LEARNT BY BERT?

Unlike classical models (W2V and BOW), BERT is a contextual model. Instead of relying on static feature representation of words, BERT represent each word as a function of the entire context. In other words, BERT manages to keep the word dependencies and sentence structures. Recent researches have successfully utilized BERT to understand certain text-based tasks (e.g. Kölbel et al. (2020)). In this section, we investigate the mechanics of how BERT model works and try to explain why BERT interprets news better than classical models.

5.1 BERT Embeddings

To extract sentiments from financial news requires understanding of specialized language with domain knowledge (Araci (2019)). As the BERT model is pre-trained from unlabeled data of BooksCorpus and English Wikipedia, we further fine-tune the model with our financial news dataset under the original setting of a masked language modeling objective. We use specific examples to show that our fine-tuned BERT model performs better in picking up finance-related sentiments compared with raw BERT model. Furthermore, we show that the BERT embedding of size 768 is a joint feature with no single value dominating the others.

5.1.1 Improvement through Pretraining with News

We form long-short portfolios with all news articles from 2011 to 2019 using the raw BERT embedding and our fine-tuned BERT embedding with news. Table. 5.1 reports the portfolio performance of raw BERT (top panel) vs news-pretrained BERT (bottom panel). Both BERT(LOGIT) and BERT(OLS) have equal-weighted Sharpe ratios of 4.6, compared with the 4.0 from RAW BERT models. What really differentiates the performance is the value-weighted Sharpe ratios, with 1.6 of BERT(LOGIT) almost twice as large as that of RAW BERT(LOGIT) (BERT(OLS) and RAW

BERT(OLS) are in the same situation). The huge increase in portfolio performance is a direct evidence suggesting that our pre-training process immerse the BERT model with finance-sensitive features that directly pick up return predictability from news.

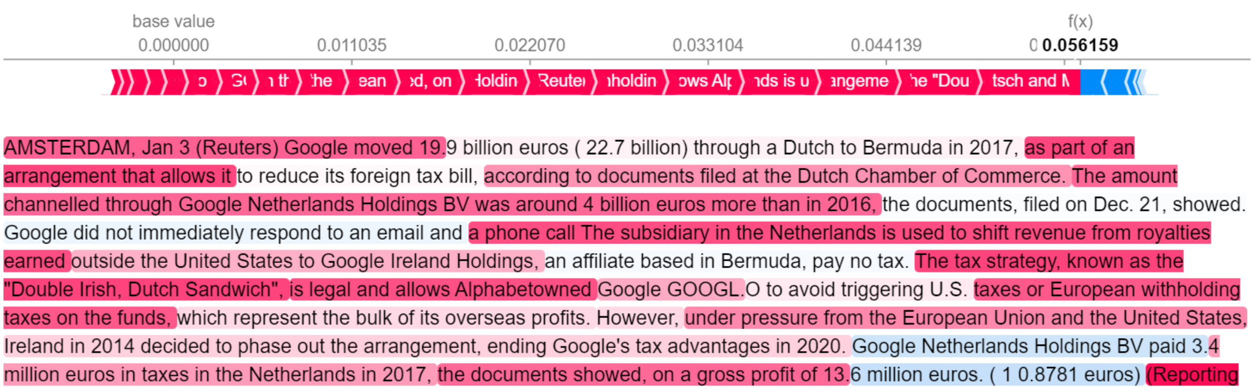
Table 5.1: Raw BERT vs News-Pretrained BERT

	RAW BERT(LOGIT)							RAW BERT(OLS)					
	EW			VW				EW			VW		
	Long	Short	L-S	Long	Short	L-S		Long	Short	L-S	Long	Short	L-S
Ret	0.38	-0.10	0.48	0.26	0.16	0.10	Ret	0.38	-0.12	0.50	0.26	0.15	0.10
Std	0.18	0.17	0.12	0.16	0.17	0.11	Std	0.18	0.17	0.12	0.16	0.17	0.11
SR	2.13	-0.60	4.05	1.62	0.92	0.91	SR	2.12	-0.68	4.02	1.57	0.90	0.90
Turnover			22.59			25.43	Turnover			22.61			25.57
	BERT(LOGIT)							BERT(OLS)					
	EW			VW				EW			VW		
	Long	Short	L-S	Long	Short	L-S		Long	Short	L-S	Long	Short	L-S
Ret	0.41	-0.18	0.59	0.27	0.08	0.19	Ret	0.41	-0.20	0.61	0.26	0.08	0.18
Std	0.18	0.18	0.13	0.16	0.18	0.12	Std	0.18	0.18	0.13	0.16	0.17	0.12
SR	2.29	-1.00	4.60	1.66	0.47	1.60	SR	2.24	-1.08	4.58	1.60	0.47	1.52
Turnover			22.51			25.45	Turnover			22.49			25.50

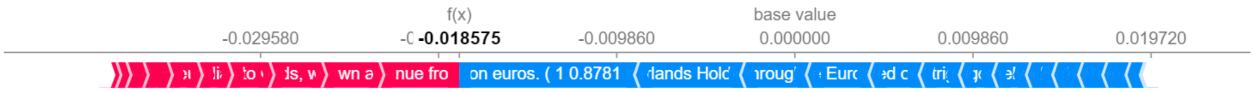
Note: Performance of equal-weighted and value-weighted long/short portfolios formed by top/bottom 50 out-of-sample predicted returns. Each panel reports annualized returns, annualized standard deviations, and annualized Sharpe ratios for long, short, and long-short portfolios. We also report monthly turnover of each strategy.

Example 1

- News-Pretrained BERT



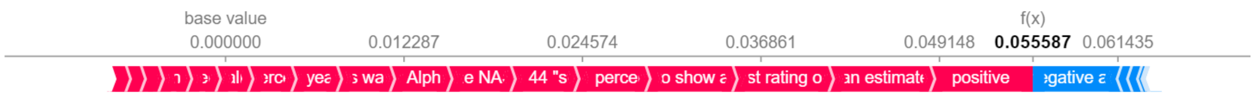
- Raw BERT



AMSTERDAM, Jan 3 (Reuters) Google moved 19.9 billion euros (22.7 billion) through a Dutch to Bermuda in 2017, as part of an arrangement that allows it to reduce its foreign tax bill, according to documents filed at the Dutch Chamber of Commerce. The amount channelled through Google Netherlands Holdings BV was around 4 billion euros more than in 2016, the documents, filed on Dec. 21, showed. Google did not immediately respond to an email and a phone call The subsidiary in the Netherlands is used to shift revenue from royalties earned outside the United States to Google Ireland Holdings, an affiliate based in Bermuda, pay no tax. The tax strategy, known as the "Double Irish, Dutch Sandwich", is legal and allows Alphabetowned Google GOOGL.O to avoid triggering U.S. taxes or European withholding taxes on the funds, which represent the bulk of its overseas profits. However, under pressure from the European Union and the United States, Ireland in 2014 decided to phase out the arrangement, ending Google's tax advantages in 2020. Google Netherlands Holdings BV paid 3.4 million euros in taxes in the Netherlands in 2017, the documents showed, on a gross profit of 13.6 million euros. (1 0.8781 euros) (Reporting

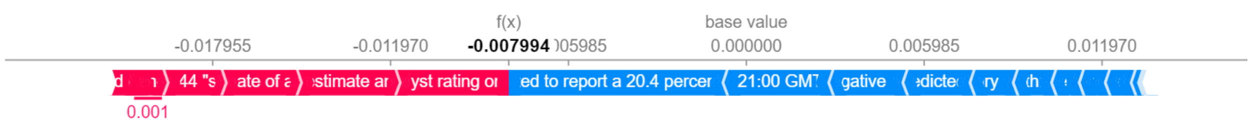
Example 2

- News-Pretrained BERT



Alphabet Inc GOOGL.OQ GOOGL.O is expected to show a rise in quarterly revenue when it reports results. The Mountain View, is expected to report a 20.4 percent increase in revenue to 38.93 billion from 32.32 billion a year ago, according to the mean estimate of 30 analysts, based on Refinitiv data. The analyst mean estimate for Alphabet Inc is for earnings of 10.82 per share. For the same quarter last year, reported earnings of 9.70 per share. The current average analyst rating on the shares is "buy" and the breakdown of is 44 "strong buy" or "buy," 1 "hold" and no "sell" or "strong sell." The StarMine predicted earnings surprise, which is the difference between Wall Street's mean estimate and StarMine's estimate of its highest rated analysts, is positive for Alphabet at 0.75 percent. Predicted revenue surprise is negative at 0.06 percent. Alphabet Inc shares had risen by 7.7 percent this quarter. The mean earnings estimate of analysts was unchanged in the last three months. Alphabet Inc belongs to the NASDAQ Composite Index. This summary was machine generated February 2 at 21:00 GMT.

- Raw BERT

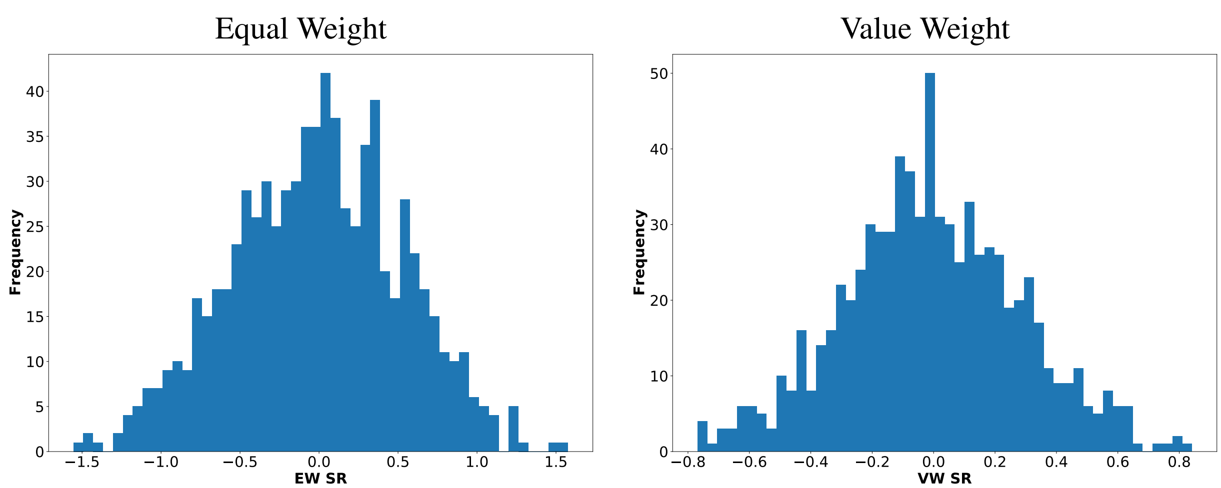


Alphabet Inc GOOGL.OQ GOOGL.O is expected to show a rise in quarterly revenue when it reports results. The Mountain View, is expected to report a 20.4 percent increase in revenue to 38.93 billion from 32.32 billion a year ago, according to the mean estimate of 30 analysts, based on Refinitiv data. The analyst mean estimate for Alphabet Inc is for earnings of 10.82 per share. For the same quarter last year, reported earnings of 9.70 per share. The current average analyst rating on the shares is "buy" and the breakdown of is 44 "strong buy" or "buy," 1 "hold" and no "sell" or "strong sell." The StarMine predicted earnings surprise, which is the difference between Wall Street's mean estimate and StarMine's estimate of its highest rated analysts, is positive for Alphabet at 0.75 percent. Predicted revenue surprise is negative at 0.06 percent. Alphabet Inc shares had risen by 7.7 percent this quarter. The mean earnings estimate of analysts was unchanged in the last three months. Alphabet Inc belongs to the NASDAQ Composite Index. This summary was machine generated February 2 at 21:00 GMT.

We apply SHAP values to interpret the BERT(LOGIT) and RAW BERT(LOGIT) models on two examples. In the first news example that Google exercised a strategy to reduce foreign tax bill. BERT(LOGIT) manages to recognize some important positive sentiments in the news, including “as part of an arrangement”, “shift revenue”, “is legal and allows”, and “a gross profit”. The model is able to detect the tone of legal certainty and assigns positive sentiment to those highlighted phrases. On the other hand, the RAW BERT model, without knowledge in business, is only able to pick up negative sentiment from "Google did not immediately respond to an email", and "under pressure from the European Union and the United States", and so on. The second news example reports a promising prospect of Google stock price before the quarterly revenue report. Although RAW BERT is able to detect some positive sentences such as "The current average analyst rating on the shares is 'buy'", the overall sentiment prediction of this news is negative. In contrast, the news-pretrained BERT almost assigns all positive sentiments to the whole article, with the only exception of a negative weight on the negative predicted revenue surprise. It is also worth mentioning that the news-pretrained BERT specifically assigns a positive sentiment to the word "positve" regarding the positive predicted earnings surprise.

5.1.2 *Portfolio Performance of 768 Signals from the BERT Embedding*

Figure 5.1: Histogram of SR for the Raw BERT Embedding



Note: This figure shows the histogram of annualized SR for all 768 signals.

As the BERT embedding is a vector of length 768, we show the long-short portfolio performance of each of the 768 signal from 1996 to 2019. Fig. 5.1 shows the histogram of SR. It turns out that the Sharpe ratios of portfolios from individual embedding values forms a normal distribution in terms of both equal-weighted and value-weighted, with the majority showing insignificant return predictability.

5.2 Advantage of BERT on News with Negation Words

In this section, we show that BERT has higher out-of-sample prediction accuracy than BOW and W2V models on news with negation words. Each negation word has a corresponding head word. For example, in the sentence "The company does not increase the production", the head word corresponding to the negation word "not" is "increase". We select news containing negation words with head words included in the 400 sentiment words of the BOW200 model since bag-of-words models are only able to make predictions on news with pre-selected sentiment words.

Table 5.2: Model Performance on News with Negation Words

	Count	BERT(LOGIT)		BERT(OLS)		BOW100		BOW200		W2V(LOGIT)		W2V(OLS)	
		Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr
not	24478	52.60	9.84	52.81	10.32	51.58	6.20	51.36	6.77	50.76	4.30	51.31	5.61
no	21513	52.42	9.56	52.57	9.80	51.64	6.52	51.37	7.04	50.40	4.39	51.14	5.53
nor	4227	53.58	15.06	53.39	15.37	53.25	13.85	52.61	14.40	49.52	2.02	51.01	6.42
neither	2280	51.62	6.03	51.10	5.49	51.89	6.62	51.45	7.85	49.47	3.10	50.18	4.18
n't	1984	50.81	3.44	51.11	3.90	51.51	7.43	51.81	7.61	52.07	8.26	52.32	7.26
never	1607	50.16	4.17	50.78	5.74	52.71	5.01	52.08	3.66	51.52	4.12	51.03	5.73
little	831	48.26	1.88	49.34	2.23	51.38	4.34	51.50	5.80	48.38	3.24	49.70	4.33
nt	813	49.20	0.80	49.82	1.47	50.31	0.44	49.57	2.24	48.22	0.84	49.69	-0.27
few	762	53.94	9.71	53.41	10.41	52.36	8.14	52.49	8.95	49.61	4.76	50.00	4.12
none	702	56.13	5.59	55.41	6.29	53.99	6.12	51.71	3.93	48.01	-2.93	49.29	1.46
nothing	514	51.17	7.61	51.75	7.73	52.14	7.50	52.14	8.24	51.75	7.29	51.95	7.18
Overall	29863	52.28	9.43	52.41	9.84	51.45	5.96	51.16	6.31	50.74	4.47	51.29	5.49

Note: This table shows the news count, out-of-sample prediction accuracy and rank correlation for each model grouped by each negation word. The bottom row reports the average prediction accuracy for all news including negation words.

Table. 5.2 shows the news count, out-of-sample prediction accuracy and rank correlation for each model. The count column reports the number of articles with at least one occurrence of the corresponding negation word. The top negation words that appear the most frequently are "not", "no", and "nor". Then for each group of articles containing the negation word, we calculate the out-of-sample prediction accuracy and rank correlation for each model. The overall accuracy is 52.3% for BERT(LOGIT) and 52.4% for BERT(OLS), beating the best classical model (BOW100 of 51.5%) by almost 1%. The average cross-sectional rank correlations between the predicted sentiment and the realized returns are over 9% for both BERT models, higher than around one half of the best classical models (BOW of 6.3). One thing that's worth mentioning is that the existence of some negation words can severely negatively affect the models' prediction performance. For example, of 831 articles with word "nt", 5 out of 6 models perform worse than random guess, with accuracy less than 50% and correlation close to zero. W2V models generally underperform compared to BOW. The reason could be that BOW models heavily relies on the words assigned with sentiments (100 words in BOW100 and 200 words in BOW200), which might not be the head word of the negation word, thus escaping the negative impact. On the other hand, the vocabulary of W2V models nearly covers all words in the article, including the head word of the negation words, thus assigning wrong weighted to the negated word.

Table 5.3: Model Performance on News with Negation Words Grouped By Positive Head Words

	Count	BERT(LOGIT)		BERT(OLS)		BOW100		BOW200		W2V(LOGIT)		W2V(OLS)	
		Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr
assurance	5960	53.91	10.18	53.91	10.68	53.09	7.65	52.58	8.53	51.09	5.88	51.43	7.46
exceed	1865	49.33	0.19	49.28	0.62	53.89	5.11	52.55	4.93	51.47	2.93	52.12	3.77
materialize	990	55.76	14.59	55.96	14.44	51.82	5.36	52.63	6.91	52.32	4.62	53.54	5.77
representation	722	55.26	13.03	54.85	12.53	49.86	-0.38	51.94	1.14	45.71	-1.80	48.89	1.65
strictly	529	56.71	15.99	57.28	16.13	57.47	11.30	55.39	10.69	53.31	6.02	54.06	5.79
substitute	458	56.33	17.78	56.11	17.34	51.53	5.95	51.97	5.12	53.71	8.23	54.37	9.25
approve	435	51.72	13.69	52.87	14.41	53.79	8.26	52.18	8.45	56.32	11.60	54.48	12.14
strong	280	51.79	16.48	52.14	16.90	53.93	8.37	50.00	7.89	50.36	4.10	48.57	1.77
tender	251	52.59	2.29	50.60	3.48	54.18	4.10	52.19	-2.51	47.41	5.43	51.79	8.62
buy	235	54.47	17.19	52.77	19.05	49.36	12.52	51.06	13.85	50.21	0.97	50.64	4.54
repurchase	232	59.48	11.27	60.34	11.78	49.57	-2.95	53.45	2.32	49.57	-0.26	51.72	-2.39
dividend	226	45.58	11.00	48.23	11.22	51.77	16.69	50.44	12.64	50.88	16.94	50.88	15.77
raise	217	53.92	13.48	52.53	12.69	56.68	6.03	51.15	2.84	49.77	-3.73	47.47	-4.04
consideration	217	47.93	2.01	49.31	1.21	46.08	-6.87	53.00	3.89	58.06	14.49	58.99	13.67
record	199	61.81	21.70	61.81	21.17	47.24	6.90	50.25	13.26	55.28	5.11	53.27	3.22
demonstrate	172	52.33	-0.50	52.33	1.29	48.26	-0.46	48.26	-3.18	46.51	-5.82	45.35	-3.92
contingent	153	61.44	27.57	60.78	30.29	64.05	38.69	61.44	36.77	51.63	1.28	56.21	14.14
authorize	121	51.24	14.19	52.07	14.34	55.37	27.35	57.85	21.66	57.85	25.99	52.89	23.26
fulfill	101	57.43	9.86	55.45	10.68	47.52	-3.40	47.52	4.80	52.48	12.15	49.50	1.52
selling	97	58.76	8.54	60.82	10.89	55.67	12.33	53.61	13.07	55.67	11.33	55.67	21.80
Overall	15961	53.12	10.11	53.30	10.59	52.42	6.59	52.28	7.26	51.29	4.43	51.61	5.62

Note: This table shows the news count, out-of-sample prediction accuracy and rank correlation for each model grouped by head words with positive sentiments. The bottom row reports the average prediction accuracy for all news including negation words.

In order to see which words are most affected by the negation words. We group articles with the same the head word in terms of words with positive (Table. 5.3) and negative sentiments (Table. 5.4), respectively. The top three positive words are "assurance", "exceed", and "materialize"

Table 5.4: Model Performance on News with Negation Words Grouped By Negative Head Words

	Count	BERT(LOGIT)		BERT(OLS)		BOW100		BOW200		W2V(LOGIT)		W2V(OLS)	
		Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr	Accy	Corr
longer	6893	51.68	9.69	51.52	9.56	50.12	5.36	50.05	6.00	50.81	5.86	51.46	6.61
miss	439	49.20	-1.91	49.20	-0.17	50.34	11.73	47.84	7.30	48.29	0.27	49.20	-1.15
shall	425	49.18	10.14	48.24	11.57	49.41	17.54	48.24	17.11	45.41	-9.79	47.06	3.54
solicitation	409	53.55	13.13	53.55	12.29	55.26	18.24	55.50	20.34	51.34	8.76	55.75	13.81
prospectus	388	48.20	9.25	49.48	10.50	51.55	-6.17	51.80	-4.57	49.23	-8.05	49.74	-6.18
vouch	286	47.55	8.24	47.55	9.69	47.20	1.03	46.50	5.65	48.95	7.29	50.70	5.86
leave	244	51.23	-3.72	50.41	-2.33	51.64	-1.67	51.64	-9.41	53.28	4.29	52.05	3.10
fall	236	46.19	-10.81	46.61	-9.22	39.41	-6.89	41.53	-12.95	37.71	-13.75	40.68	-14.58
damage	223	52.02	5.17	55.61	7.77	52.91	8.39	47.98	8.16	44.39	-0.46	45.29	-2.03
repeat	183	59.56	20.73	59.02	21.60	51.37	6.13	54.10	6.90	48.09	4.39	49.73	1.46
cut	166	50.60	17.27	53.01	18.23	56.02	7.09	56.02	13.55	53.61	8.16	53.61	9.89
lose	159	50.31	12.08	54.09	12.39	54.72	10.14	57.23	5.92	49.69	-2.77	50.94	4.20
offering	129	49.61	13.26	50.39	12.94	55.04	11.52	54.26	9.36	52.71	5.28	54.26	5.40
reveal	121	52.89	1.17	53.72	1.98	47.11	-5.83	47.93	-9.43	49.59	-4.45	47.93	-6.15
involved	114	43.86	5.77	45.61	3.92	48.25	0.47	45.61	3.03	53.51	-2.72	53.51	2.85
worry	111	53.15	10.35	52.25	10.13	49.55	2.89	45.05	0.78	58.56	18.35	58.56	19.89
hurt	107	57.94	16.07	58.88	13.04	42.06	18.06	44.86	14.10	51.40	18.98	47.66	12.25
doubt	94	46.81	1.34	46.81	0.01	46.81	-6.33	46.81	-5.76	45.74	-3.59	47.87	-10.71
mandatory	90	51.11	23.45	52.22	22.41	53.33	2.63	51.11	9.62	51.11	20.00	53.33	32.10
closed	90	65.56	36.43	67.78	35.29	55.56	20.56	62.22	25.51	56.67	18.05	60.00	15.58
Overall	13902	51.30	8.60	51.38	8.93	50.35	5.23	49.86	5.37	50.09	4.45	50.93	5.27

Note: This table shows the news count, out-of-sample prediction accuracy and rank correlation for each model grouped by head words with negative sentiments. The bottom row reports the average prediction accuracy for all news including negation words.

while the top three negative words are “longer”, “miss”, and “shall”. All models generally perform worse on negative head words in front of negation words.

5.3 Long News or Short News?

In this section, we share some of our findings on whether the BERT models makes better prediction based on long news or short news. Note that the fact that the portfolio performance based on news alert is higher than that on news articles doesn’t suffice to show that BERT models are more accurate on short news. A main reason is the more timely delivery of new stock information by alerts than by news.

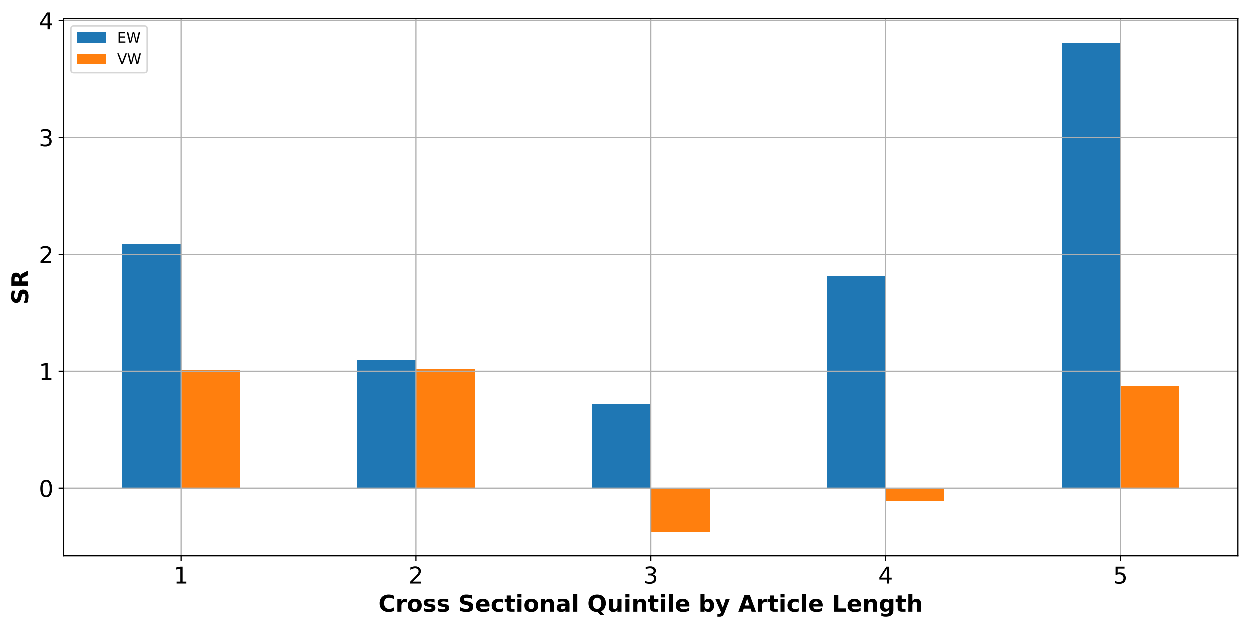
We group news cross-sectionally each day into quitiles of article lengths. The summary statistics of quintiles of article lengths is reported in Table. 5.5. There are around 220k news articles in each quintile. The median article lengths range from 285 to 9,438. The longest news article contains 964,060, and is in fact the whole earning report.

Table 5.5: Summary Statistics of Article Length of Cross-Sectional Quintiles

Quintile	count	mean	std	min	25%	50%	75%	max
Shortest	223096	299	115	100	212	285	367	1865
2	221775	775	321	241	527	685	975	2965
3	221842	1972	570	450	1568	1912	2344	5114
4	221775	4169	1706	1190	3233	3820	4517	19145
Longest	222696	14942	18622	2959	6128	9438	19778	964060

Note: This table shows the summary statistics of the out-of-sample news article length based on cross-sectional quintiles.

Figure 5.2: SR of Quintiles by Article Length



Note: This figure shows the SR of long-short portfolio of quintiles by article length.

To find out the associate between news length and return predictability, we form long-short portfolios for each quintiles. Figure. 5.2 shows the bar chart of equal-weighted and value-weighted Sharpe ratio for each quintile. Surprisingly, the Sharpe ratios of quintiles show a “U” shape in terms of both equal-weighted and value-weighted portfolios.

5.3.1 Summarization using XSUM Model

When human readers read an article, most likely they will manually generate a summary of the story in their minds and intuitively form a sentiment to the article just read. Except for a small amount of news with numerical analysis, most human readers tend to focus more on the conclusion while forgetting about the exact details of the news, for example, they only care about the Tesla's price is up instead of the exact price it closes, or a great deal of a merge instead of the exact cost of the merge. With the help of Narayan et al. (2018), which proposed an approach to summarize a single document into one sentence, we are able to measure the impact of news details on the portfolio performance.

We apply the pre-trained XSUM model proposed by Narayan et al. (2018) to summarize all news articles into a one-sentence news summary answering the question "What is the article about?". Then instead of using the original news, we use the summarized news to train separate BERT models.

Example News

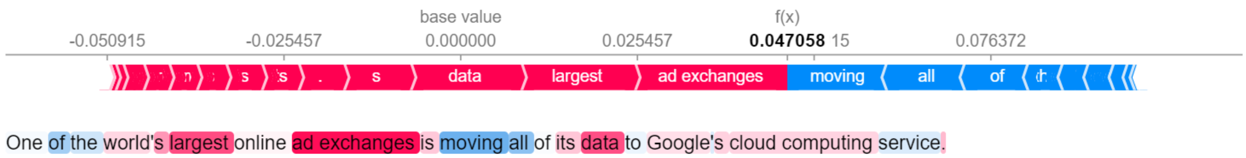
- **Headline:** Ad exchange to $110 mln Google Cloud deal
- **Body:** Privately held ad software firm OpenX announced on Thursday a fiveyear agreement totaling more than $110 million with Alphabet Inc's GOOGL.O Google to use its services in what described as a first for the online ads industry. OpenX said it has begun transferring data to Google servers from its own and that it would the first major online ad exchange to move fully to the cloud. said it signed the deal in the fourth quarter after evaluating cloud providers including Inc AMZN.O and Microsoft Corp MSFT.O, both of which are far ahead of Google in market share. Its to Google represents a minimum and the tab will grow as the firm uses more artificial intelligence services, OpenX Chief Technology Officer Paul Ryan said in an interview. Among the few that disclose cloud contracts, chat app developer Snap Inc SNAP.N has said it must spend at least $400 million annually on Google Cloud for five

years. Ad exchanges use automated software to match millions of ad requests each second from online publishers with realtime bids from advertisers.

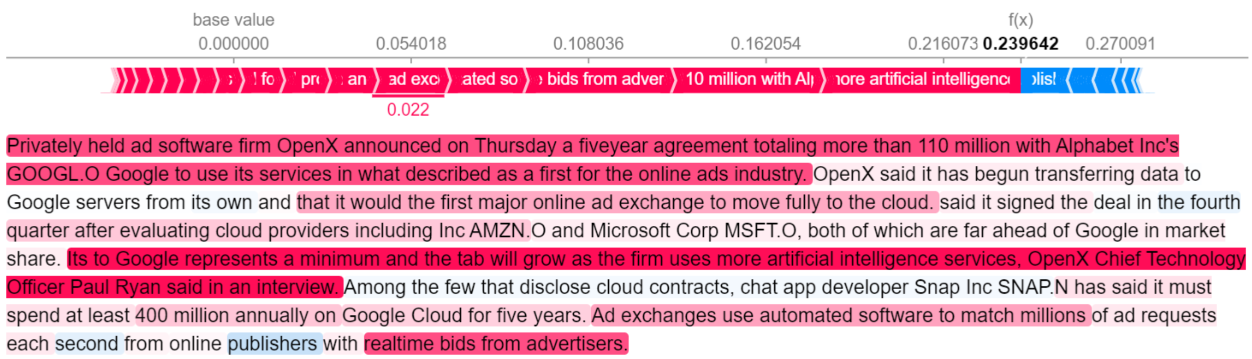
- **Summary: One of the world's largest online ad exchanges is moving all of its data to Google's cloud computing service.**

Interpretation of News

- Summarized News



- Raw News



In the example shown above, the news article reports that OpenX announced a five-year agreement with Google Cloud. The output of XSUM correctly summarizes the news (bold face). When interpreting the BERT models on the original and the summarized models, we note the major difference here. While BERT(LOGIT) on the original article highlights the agreement and the growth of the underlying company as providing positive sentiments, BERTXSUM(LOGIT) on the summarized news is only able to pick up some key words, e.g. "largest", "ad exchanges", and "data". The final predicted sentiment score of BERTXSUM(LOGIT) is 0.05, much lower than that of BERT(LOGIT).

Table 5.6: Summary Statistics of Article Length before and after Summarization

	count	mean	std	min	25%	50%	75%	max
Raw Article	1111184	4435	9981	100	533	1912	4451	964060
Summary	1111184	107	92	0	29	95	155	818

Note: This table shows the summary statistics of the out-of-sample news article length before and after the summarization.

Table 5.7: Portfolio Performance

	BERT(LOGIT)							BERT(OLS)					
	EW			VW				EW			VW		
	Long	Short	L-S	Long	Short	L-S		Long	Short	L-S	Long	Short	L-S
Ret	0.41	-0.18	0.59	0.27	0.08	0.19	Ret	0.41	-0.20	0.61	0.26	0.08	0.18
Std	0.18	0.18	0.13	0.16	0.18	0.12	Std	0.18	0.18	0.13	0.16	0.17	0.12
SR	2.29	-1.00	4.60	1.66	0.47	1.60	SR	2.24	-1.08	4.58	1.60	0.47	1.52
Turnover			22.51			25.45	Turnover			22.49			25.50
	BERTXSUM(LOGIT)							BERTXSUM(OLS)					
	EW			VW				EW			VW		
	Long	Short	L-S	Long	Short	L-S		Long	Short	L-S	Long	Short	L-S
Ret	0.29	0.04	0.26	0.25	0.19	0.06	Ret	0.33	0.03	0.29	0.24	0.17	0.07
Std	0.18	0.17	0.12	0.16	0.17	0.11	Std	0.18	0.18	0.13	0.16	0.17	0.11
SR	1.68	0.21	2.24	1.53	1.08	0.56	SR	1.76	0.20	2.21	1.48	1.01	0.61
Turnover			22.60			25.38	Turnover			22.57			25.35

Note: Performance of equal-weighted and value-weighted long/short portfolios formed by top/bottom 50 out-of-sample predicted returns. Each panel reports annualized returns, annualized standard deviations, and annualized Sharpe ratios for long, short, and long-short portfolios. We also report monthly turnover of each strategy.

To have an overall view of the impact of missing information of news articles, we form portfolios based on the new model prediction on summarized news. Figure. 5.2 reports the portfolio performance for the two BERT models on summarized news, i.e. BERTXSUM(LOGIT) and BERTXSUM(OLS), and the baseline models BERT(LOGIT) and BERT(OLS). Compared with the baseline models on the original news, the Sharpe ratios of the BERTXSUM models on summarized news are less than one half in terms of both equal-weighted (2.2) and value-weighted portfolios (0.6). Indeed, the result shows that the detailed information within the news articles actually provides additional value for return predictability.

5.4 BERT By Emotion Scores

In this section, we apply the pre-trained BERT model from Demszky et al. (2020) to obtain emotion scores on US news articles and form long-short portfolios based on the rankings of scores. The model is trained on a dataset called GoEmotions, included 58k manually annotated English Reddit comment, labeled for 27 emotion categories or Neutral. The 27 emotions consist of admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, and surprise. The taxonomy is derived from Ekmans 6 emotion categories, including anger, surprise, disgust, enjoyment, fear, and sadness, and extended through multi-round process with feedbacks from annotators. Similar or scarce selected emotions were removed.

Table 5.8: Correlation and SR of Go Emotions

	Corr		SR			Corr		SR	
	Mean	Tstat	EW	VW		Mean	Tstat	EW	VW
approval	0.43	4.31	1.10	0.16	caring	-0.01	-0.07	0.25	-0.08
excitement	0.32	3.27	0.91	0.31	anger	-0.04	-0.39	0.14	0.02
gratitude	0.29	2.95	0.73	0.07	confusion	-0.06	-0.57	0.24	-0.02
admiration	0.26	2.64	0.73	0.16	neutral	-0.06	-0.62	-0.46	0.04
joy	0.22	2.25	0.51	0.32	disgust	-0.10	-1.01	0.20	0.17
amusement	0.17	1.75	0.58	0.09	realization	-0.10	-1.07	0.18	-0.02
optimism	0.15	1.55	0.66	0.08	disapproval	-0.14	-1.48	-0.06	-0.12
pride	0.15	1.52	0.77	0.03	remorse	-0.15	-1.50	0.15	-0.04
desire	0.13	1.32	0.73	-0.08	grief	-0.19	-1.91	0.00	-0.01
love	0.12	1.22	0.73	0.19	nervousness	-0.19	-1.94	0.12	-0.19
curiosity	0.11	1.16	0.43	0.02	annoyance	-0.19	-1.96	-0.20	0.07
surprise	0.07	0.74	0.53	0.03	embarrassment	-0.22	-2.24	-0.19	-0.10
relief	0.04	0.44	0.51	0.03	sadness	-0.35	-3.65	-0.28	-0.13
fear	-0.00	-0.02	0.40	-0.39	disappointment	-0.46	-4.69	-0.65	-0.17

Note: This table shows the cross-sectional rank correlation between the model predictions and 3-day returns and SR of long-short portfolios based on the 28 emotion scores predicted by the pre-trained BERT model from Demszky et al. (2020). The news are from 1996 to 2019. The table is sorted in ascending order of correlation t-stats.

Table. 5.8 reports the cross-sectional rank correlation between the model predictions and 3-day returns, the corresponding t-stats, and the long-short SR for each emotion. We see from the table that negative emotions (disappointment, sadness, embarrassment, disapproval, remorse, grief, nervousness, annoyance) tend to have negative correlations and SRs and positive emotions (approval, excitement, admiration, joy, gratitude, pride, amusement, optimism, love) tend to have positive correlations and SRs. Five positive sentiments (approval, excitement, gratitude, admiration, and joy) show significant positive t-stats, while four negative sentiments (disappointment, sadness, embarrassment, and annoyance) show significant negative t-stats (under 5% significance level). The top performer "approval" has equal-weighted Sharpe ratio of 1.10 and "disppointment" has equal-weighted Sharpe ratio of -0.65. Please refer to the section "Emotion News Example" in the Appendix for example news articles for each emotion.

CHAPTER 6

CONCLUSION

In this paper, we propose and analyze an approach to extract sentiment from news based on BERT embedding. In contrast to the classical methods, our BERT-based method is able to learn contextual features, thus yielding better prediction performance. In addition, the advantage in sentiment extraction is well transformed into higher profits of long-short portfolio based on the strategy. Furthermore, we expand the BERT models into international markets and show the methodology is robust across different market and languages.

To better interpret the BERT models, we analyze the relative performance on news with negation words. Results show that BERT models, due to its contextual understanding, performs better on the articles with negations words than classical models. Also, we show that article emotions directly leads to price movements in the corresponding directions. Simply trading based on the emotion scores like approval can produce significant Sharpe ratios.

References

Araci, Dogu, 2019, Finbert: Financial sentiment analysis with pre-trained language models, *arXiv preprint arXiv:1908.10063* .

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio, 2014, Neural machine translation by jointly learning to align and translate, *arXiv preprint arXiv:1409.0473* .

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov, 2017, Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics* 5, 135–146.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, 2020, The structure of economic news, Technical report, National Bureau of Economic Research.

Chew, Ming, Sahil Puri, Arsh Sood, and Adam Wearne, 2017, Using natural language processing techniques for stock return predictions, *Available at SSRN 2940564* .

Cong, Lin William, Tengyuan Liang, and Xiao Zhang, 2019, Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information, *Interpretable, and Data-driven Approach to Analyzing Unstructured Information (September 1, 2019)* .

Demszky, Dorottya, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi, 2020, Goemotions: A dataset of fine-grained emotions, *arXiv preprint arXiv:2005.00547* .

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* .

Finkel, Jenny Rose, Trond Grenager, and Christopher D Manning, 2005, Incorporating non-local information into information extraction systems by gibbs sampling, in *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)*, 363–370.

Grave, Edouard, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov, 2018, Learning word vectors for 157 languages, in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Harris, Zellig S, 1954, Distributional structure, *Word* 10, 146–162.

Jegadeesh, Narasimhan, and Di Wu, 2013, Word power: A new approach for content analysis, *Journal of financial economics* 110, 712–729.

Jha, Manish, Hongyi Liu, and Asaf Manela, 2020, Does finance benefit society? a language embedding approach, *A Language Embedding Approach (July 18, 2020)* .

Ke, Zheng Tracy, Bryan T Kelly, and Dacheng Xiu, 2019, Predicting returns with text data, Technical report, National Bureau of Economic Research.

Kölbel, Julian F, Markus Leippold, Jordy Rillaerts, and Qian Wang, 2020, Ask bert: How regulatory disclosure of transition and physical climate risks affects the cds term structure, *Swiss Finance Institute Research Paper* .

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of finance* 66, 35–65.

Lundberg, Scott M, and Su-In Lee, 2017, A unified approach to interpreting model predictions, in *Proceedings of the 31st international conference on neural information processing systems*, 4768–4777.

Luong, Minh-Thang, Hieu Pham, and Christopher D Manning, 2015, Effective approaches to attention-based neural machine translation, *arXiv preprint arXiv:1508.04025* .

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, 2013, Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* .

Mikolov, Tomas, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin, 2018, Advances in pre-training distributed word representations, in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Narayan, Shashi, Shay B. Cohen, and Mirella Lapata, 2018, Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization, *ArXiv* abs/1808.08745.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning, 2014, Glove: Global vectors for word representation, in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Tetlock, Paul C, 2007, Giving content to investor sentiment: The role of media in the stock market, *The Journal of finance* 62, 1139–1168.

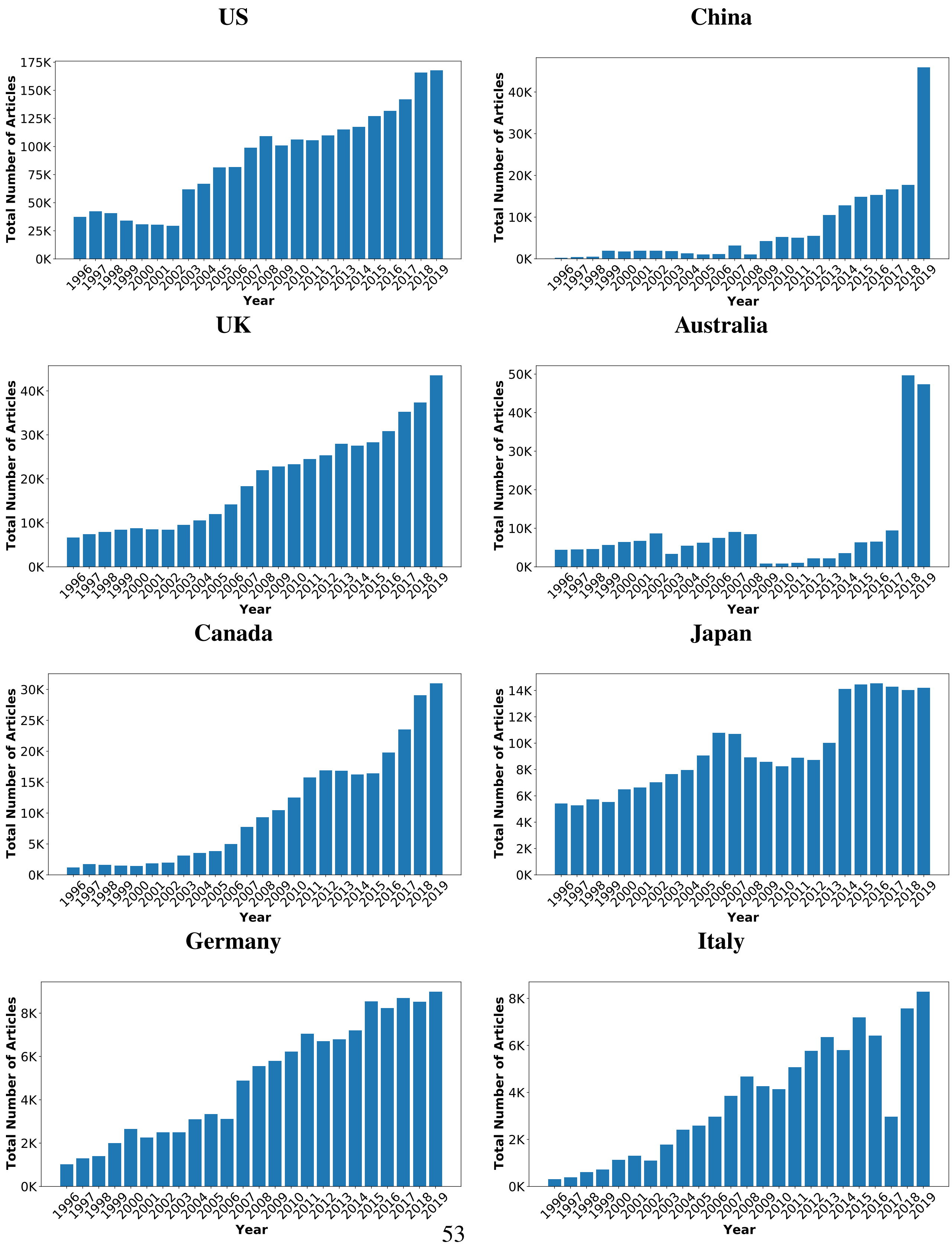
Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, 2017, Attention is all you need, in *Advances in neural information processing systems*, 5998–6008.

APPENDIX A

APPENDIX

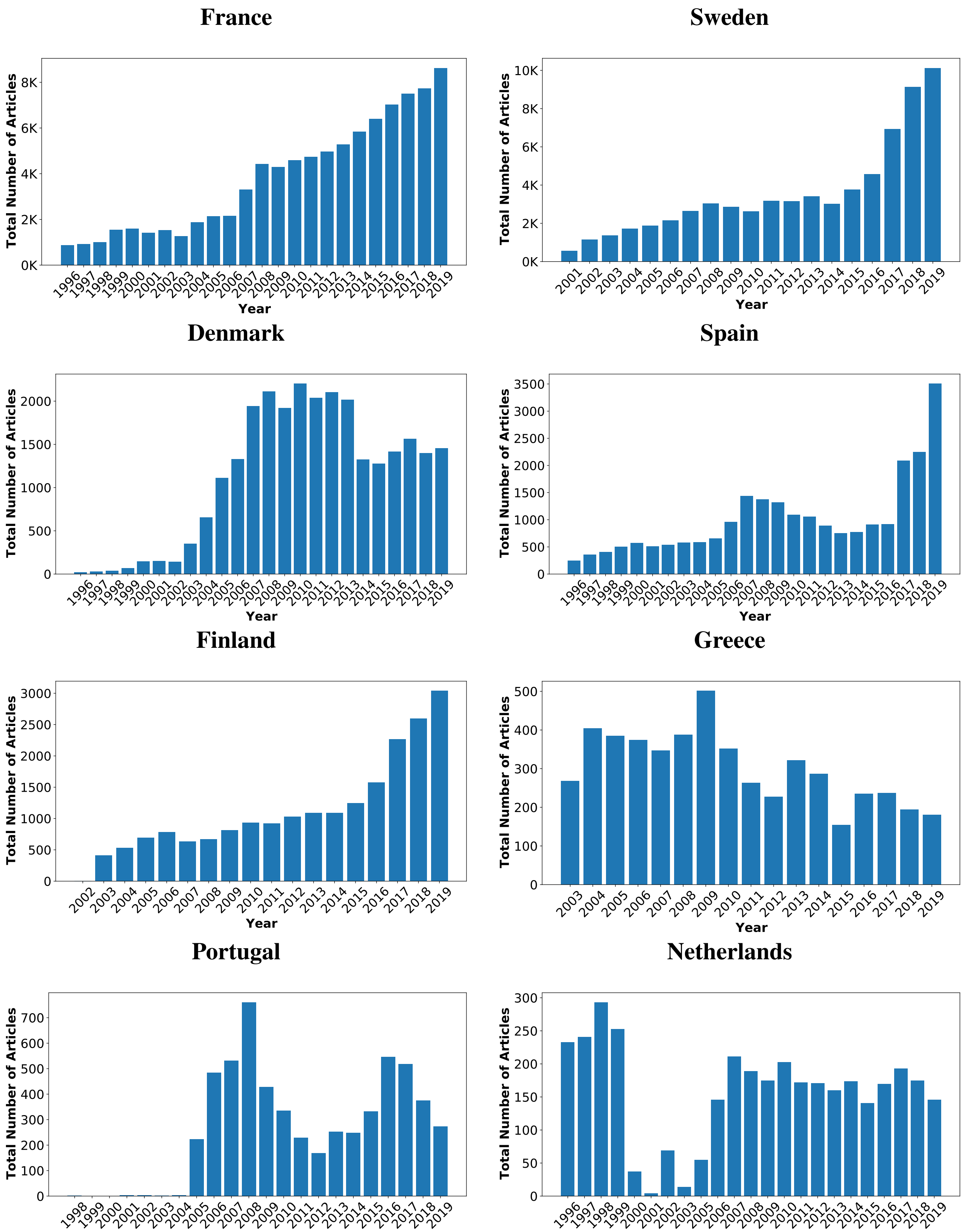
A.1 News Article Counts

Figure A.1: Annual Time Series of the Total Number of Articles (Part 1/2)



Note: Each figure plots the annual time series of the total number of articles per year. We only provide an estimate for 2019 (highlighted in red), by annualizing the total number of articles of the few months we observe, since we do not have a whole year's data for this year.

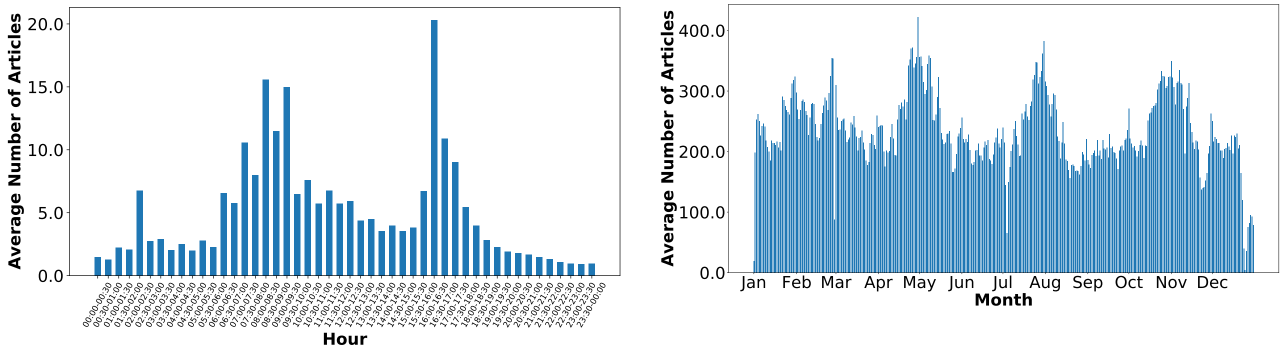
Figure A.2: Annual Time Series of the Total Number of Articles (Part 2/2)



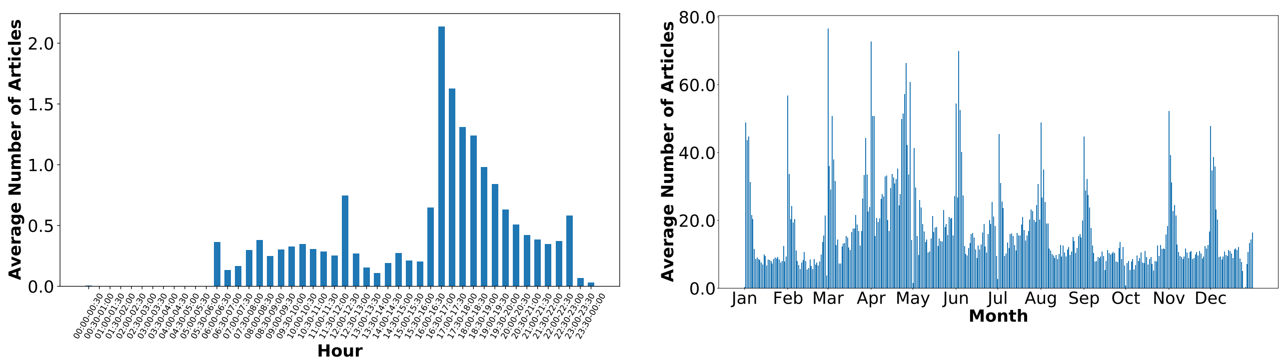
Note: Each figure plots the annual time series of the total number of articles per year. We only provide an estimate for 2019 (highlighted in red), by annualizing the total number of articles of the few months we observe, since we do not have a whole year's data for this year.

Figure A.3: Average Article Counts (Part 1/4)

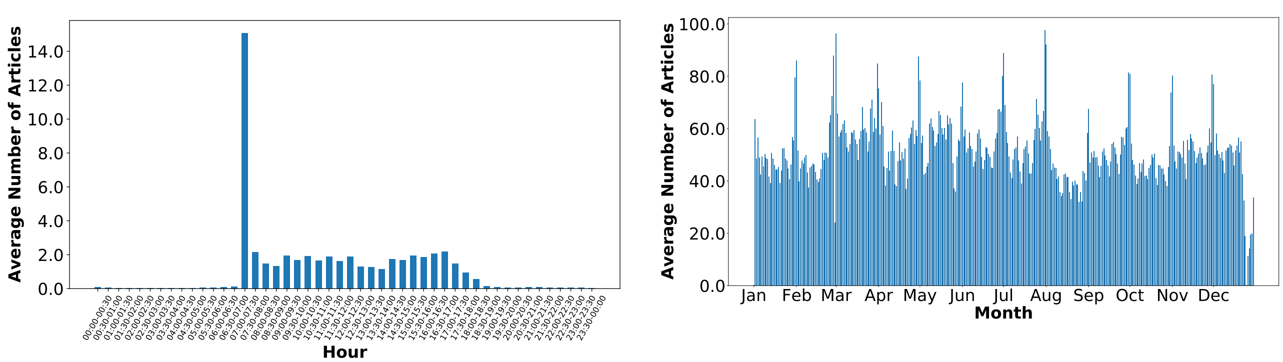
US



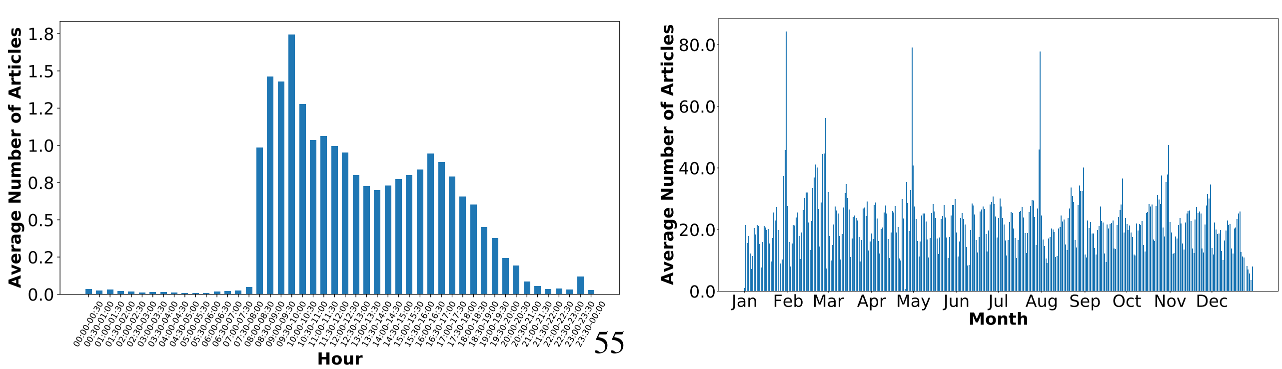
China



UK



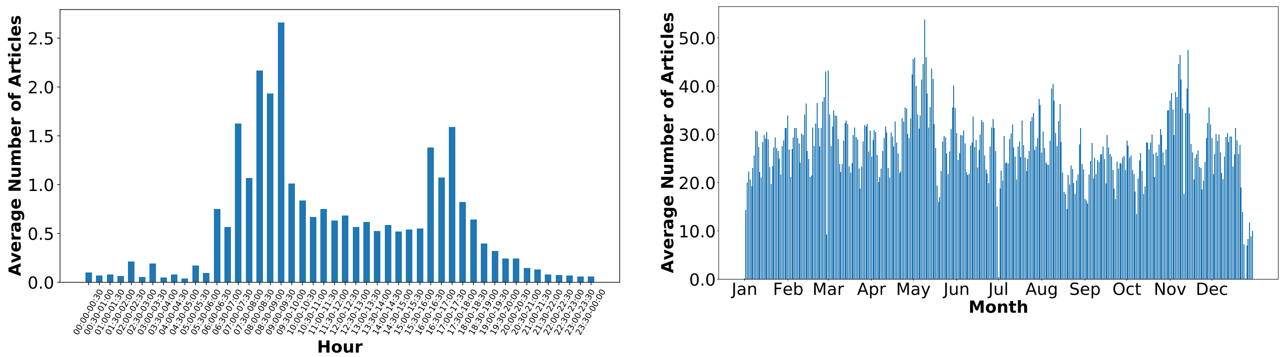
Australia



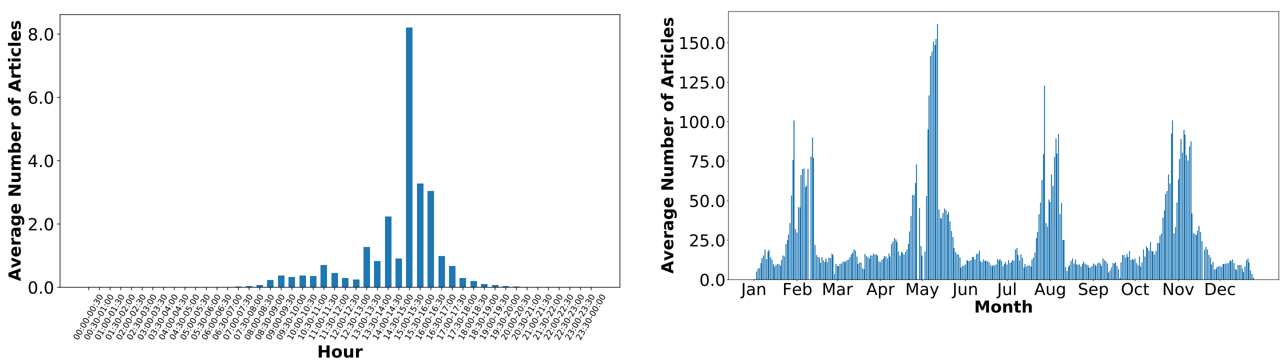
Note: The left column plots the average numbers of articles per half an hour (24 hour local time). The right column plots the average numbers of articles per calendar day.

Figure A.4: Average Article Counts (Part 2/4)

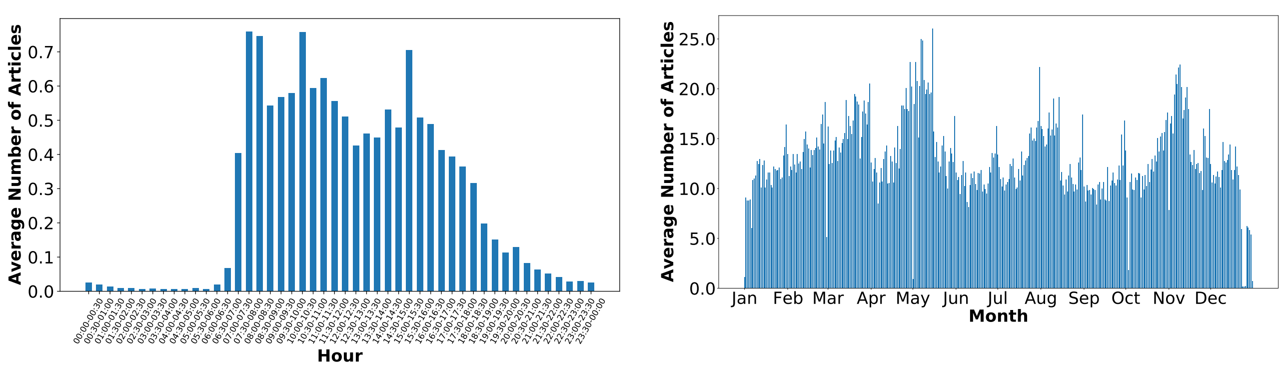
Canada



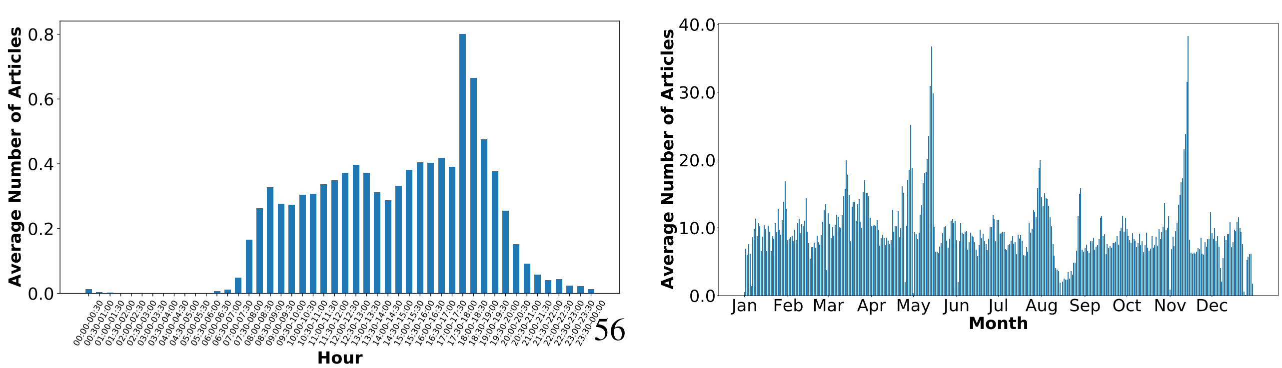
Japan



Germany



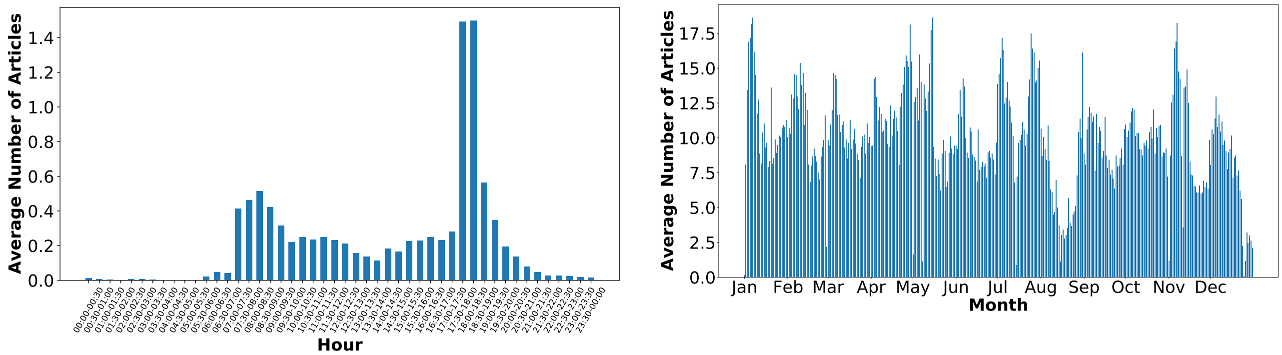
Italy



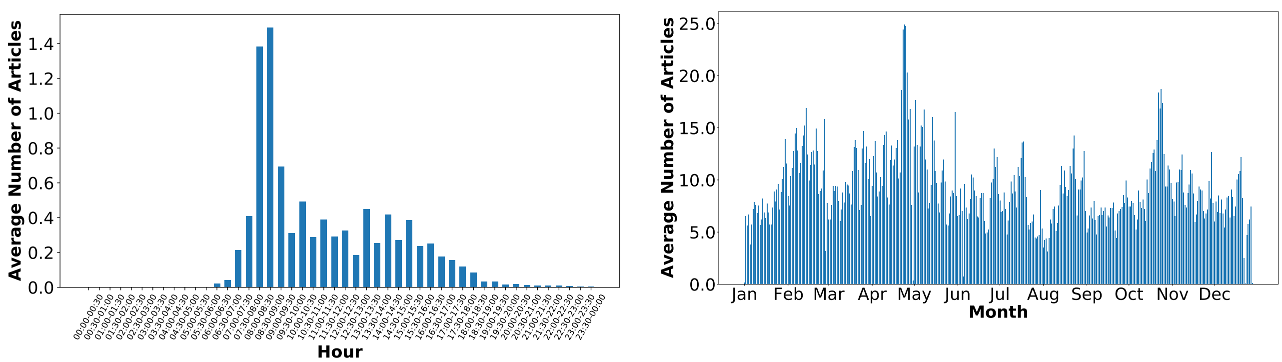
Note: The left column plots the average numbers of articles per half an hour (24 hour local time). The right column plots the average numbers of articles per calendar day.

Figure A.5: Average Article Counts (Part 3/4)

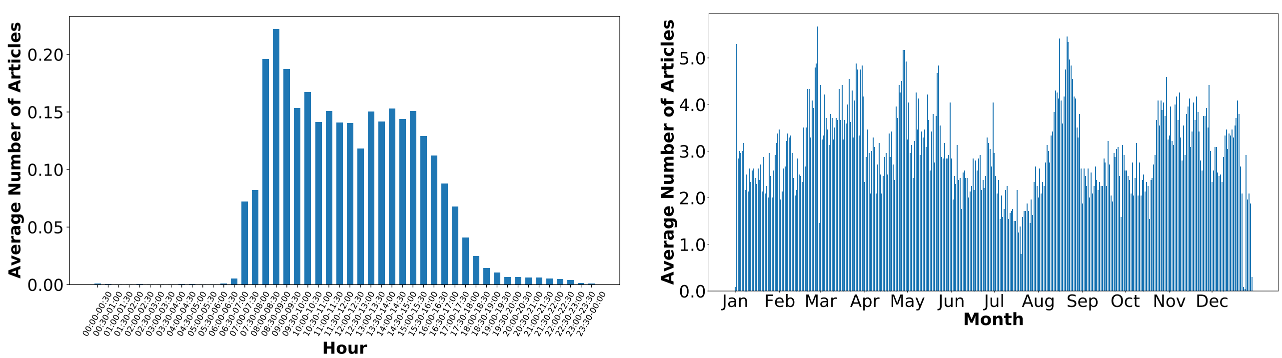
France



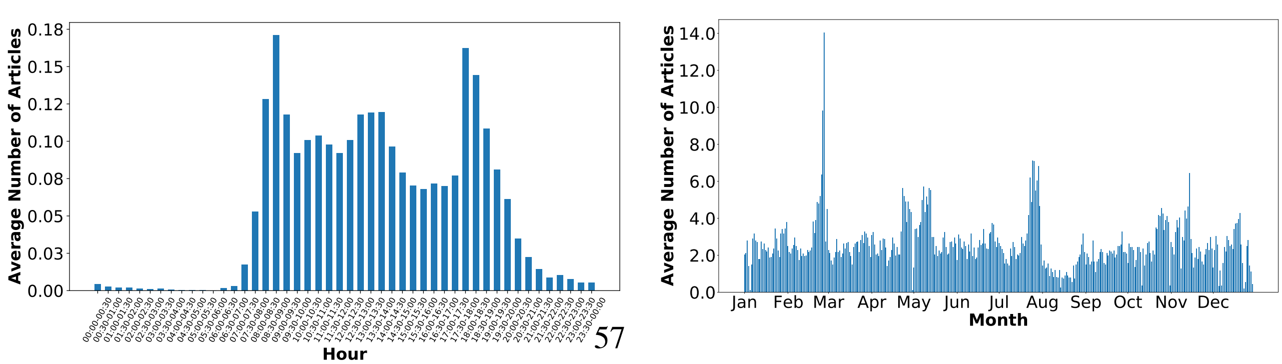
Sweden



Denmark



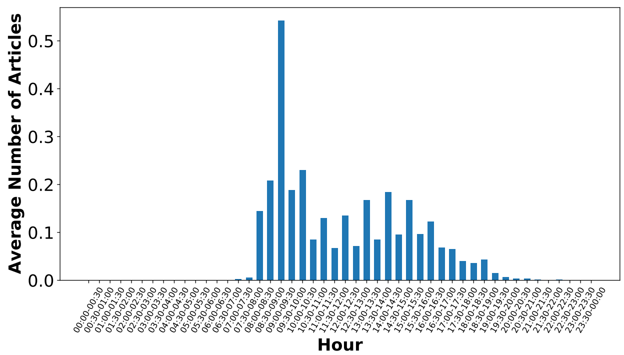
Spain

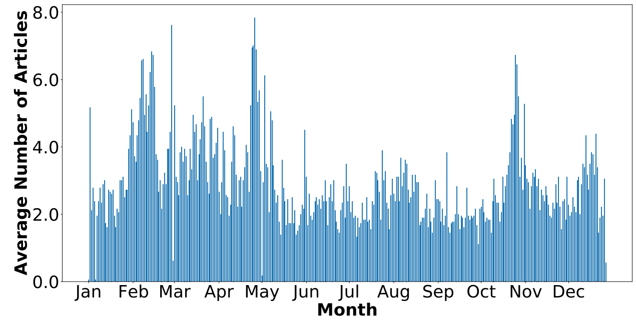


Note: The left column plots the average numbers of articles per half an hour (24 hour local time). The right column plots the average numbers of articles per calendar day.

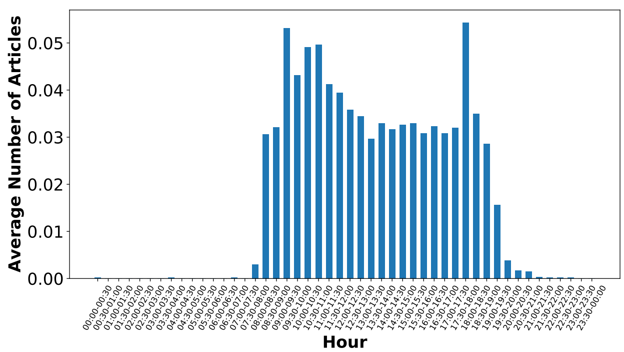
Figure A.6: Average Article Counts (Part 4/4)

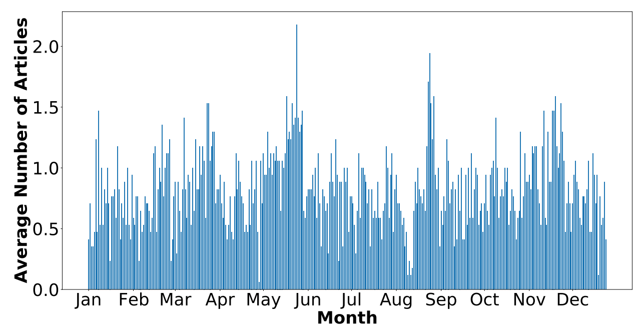
Finland



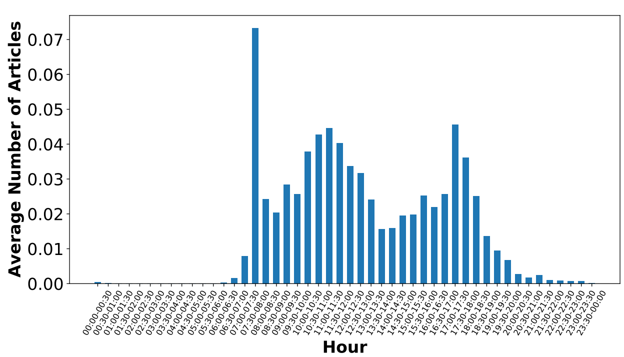


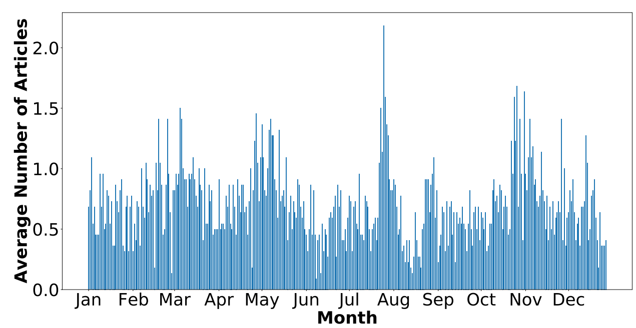
Greece



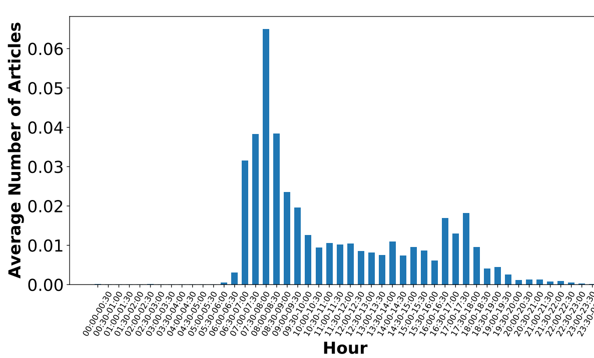


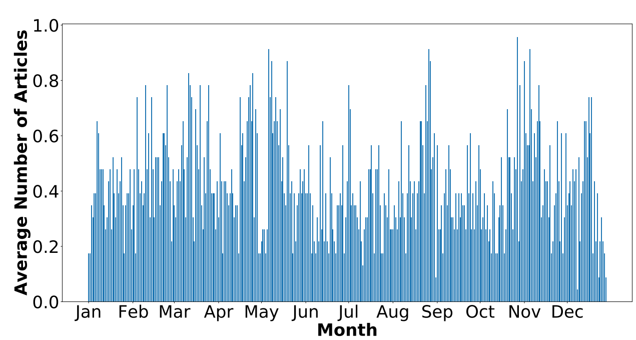
Portugal





Netherlands





Note: The left column plots the average numbers of articles per half an hour (24 hour local time). The right column plots the average numbers of articles per calendar day.

A.2 Additional Summary Statistics

Table A.1: Percentiles of News Length Before Filtering Short News

	count	mean	min	1%	5%	10%	25%	50%	75%	90%	95%	99%	max
US	4755247	2952	1	31	52	53	357	1119	3348	6202	10377	28392	2778015
China	2086045	74	1	2	4	6	8	16	27	32	49	1080	415138
UK	906705	4871	1	173	333	420	834	2044	4130	8319	19709	49237	878939
Australia	388585	5856	1	43	215	263	263	645	3293	13093	22753	80836	4198858
Canada	481891	5145	1	151	291	433	1122	3166	5880	10642	17789	40081	822346
Japan	405341	333	5	86	143	166	196	298	418	494	562	1246	13999
Germany	238577	2451	2	68	189	424	869	1450	2881	4869	7029	16078	164861
Italy	173250	3483	1	46	164	388	753	1507	3814	7323	11988	26681	705566
France	174917	3090	1	186	328	473	823	1571	3426	6189	9547	28006	745840
Sweden	126211	2403	8	188	291	357	693	1738	2857	4725	6896	15412	57366
Denmark	53056	1665	1	88	260	377	590	1125	2012	3296	4386	9498	68574
Spain	47541	2373	1	43	123	182	258	728	1391	2623	3861	33340	874748
Finland	38163	4298	35	407	572	720	1213	1938	3541	9131	17255	41123	295803
Portugal	11284	1188	10	125	430	475	542	843	1542	2415	3161	4663	27183
Greece	10093	1378	42	156	230	379	592	1014	1864	2793	3379	4974	25476
Netherlands	7137	2574	39	202	384	497	806	1568	2913	5102	8755	17040	56818

Note: This table reports the news count and percentiles of news length for all countries.

Table A.2: Percentiles of News Length

	count	mean	std	min	1%	5%	10%	25%	50%	75%	90%	95%	99%	max
US	2061640	3605	7932	100	148	227	290	521	1593	3825	7446	14729	31604	964060
China	152668	581	4841	30	30	30	30	31	40	140	921	2538	3698	415138
UK	450920	4395	10764	100	182	284	371	694	1593	3525	7941	20739	46748	878939
Australia	191057	6323	23628	100	150	207	263	263	869	3834	13947	25840	99726	964060
Canada	239103	5014	7569	100	176	270	392	1107	3101	5839	10346	16738	39480	822346
Japan	221181	332	170	100	118	156	174	208	325	429	486	532	624	8899
Germany	114633	2727	4809	100	175	325	492	848	1428	3003	5476	7873	23327	164861
Italy	84205	3525	7746	100	152	222	399	707	1444	4026	7509	12879	25274	685494
France	87392	3107	6825	100	173	306	451	781	1409	3478	6539	9898	27720	745840
Sweden	63084	2698	3435	101	209	307	381	870	1904	3105	5325	7880	17842	57366
Denmark	26191	1624	2421	100	133	260	344	529	966	1843	3364	4769	10592	68574
Spain	22801	1964	14738	100	113	169	203	338	683	1133	1911	2782	24057	874748
Finland	19062	5777	9449	103	453	699	904	1421	2300	5643	13975	26403	45067	295803
Portugal	5616	1402	897	100	323	485	587	844	1220	1717	2411	3027	4414	27183
Greece	5041	1207	1338	100	150	208	306	513	816	1462	2473	3223	5574	25476
Netherlands	3564	2856	4069	112	194	356	464	735	1624	3190	6330	10327	19412	56818

Note: This table reports final news count and percentiles of news length of the dataset we use after filtering out short news and news with greater than the median cosine similarity.

Table A.3: Percentiles of Cosine Similarity

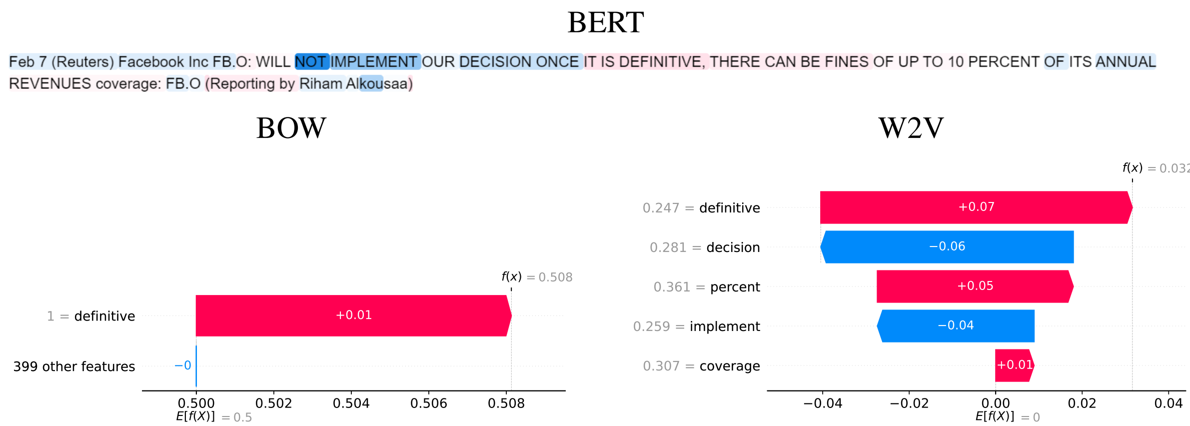
	count	mean	std	min	25%	Median	75%	max
US	4123279	0.44	0.38	0.0	0.01	0.36	0.84	1.0
China	305335	0.48	0.46	0.0	0.00	0.43	1.00	1.0
UK	901838	0.50	0.41	0.0	0.08	0.44	1.00	1.0
Australia	382114	0.46	0.43	0.0	0.00	0.34	1.00	1.0
Canada	478205	0.41	0.40	0.0	0.00	0.31	0.86	1.0
Japan	399185	0.32	0.40	0.0	0.00	0.00	0.75	1.0
Germany	229265	0.40	0.37	0.0	0.00	0.31	0.73	1.0
Italy	168410	0.42	0.36	0.0	0.03	0.37	0.75	1.0
France	174784	0.33	0.33	0.0	0.00	0.26	0.57	1.0
Sweden	126168	0.31	0.32	0.0	0.00	0.27	0.56	1.0
Denmark	52381	0.39	0.34	0.0	0.00	0.35	0.66	1.0
Spain	45597	0.40	0.38	0.0	0.00	0.32	0.80	1.0
Finland	38123	0.42	0.39	0.0	0.00	0.41	0.80	1.0
Portugal	11231	0.61	0.40	0.0	0.22	0.72	1.00	1.0
Greece	10082	0.43	0.39	0.0	0.00	0.48	0.78	1.0
Netherlands	7128	0.34	0.38	0.0	0.00	0.12	0.74	1.0

Note: Following Ke et al. (2019), we calculate the cosine similarity using vocabulary of all words appearing more than 1k times after removing news with length less than 100 (30 for China).

A.3 Sample News that BERT Outperform

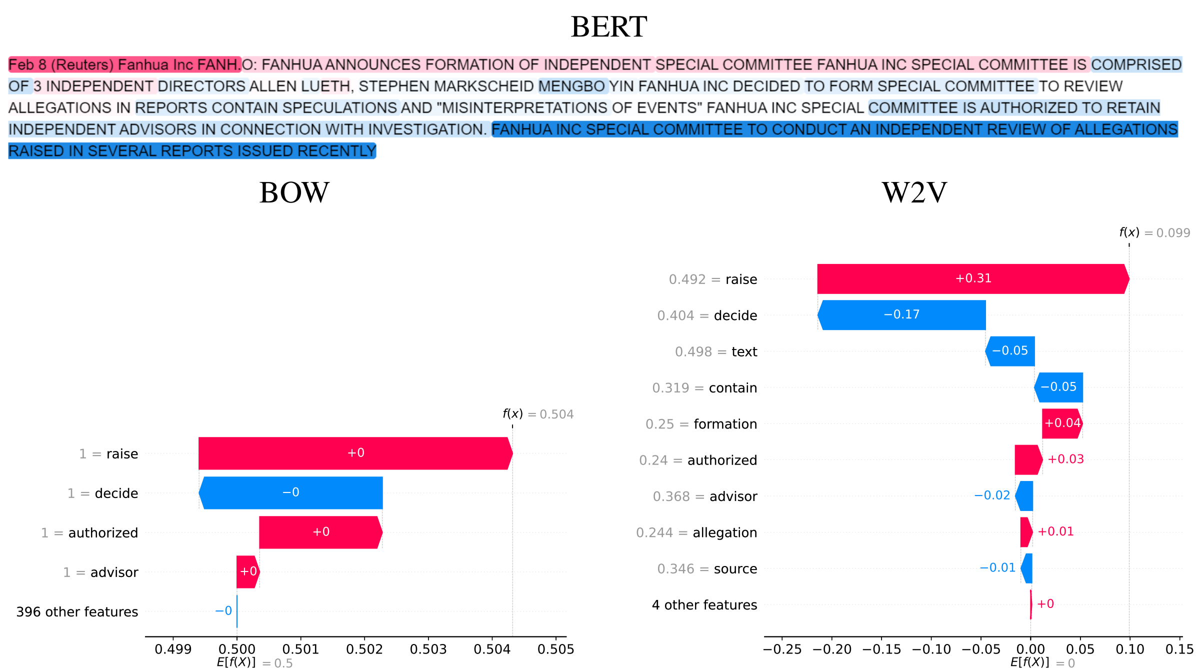
We use a method called SHAP (SHapley Additive exPlanations) by Lundberg and Lee (2017) to interpret individual predictions. SHAP is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. The SHAP values can be viewed as expected change in predictions conditional on specific features. When interpreting BERT, segments with positive SHAP values are highlighted in red and segments with negative SHAP values are highlighted in blue. The darker the color is, the larger magnitude the SHAP value is. When interpreting Bag of Words (BOW) and Words to Vectors (W2V), we use a waterfall plot which shows all features contributing to the prediction. The blue features push the prediction to the negative side while the red features push the prediction to the positive side.

Figure A.7: Interpretation of Negative News 1



News Text: Facebook Inc FB.O: will not implement our decision once it is definitive, there can be fines of up to 10 percent of its annual revenues.

Figure A.8: Interpretation of Negative News 2



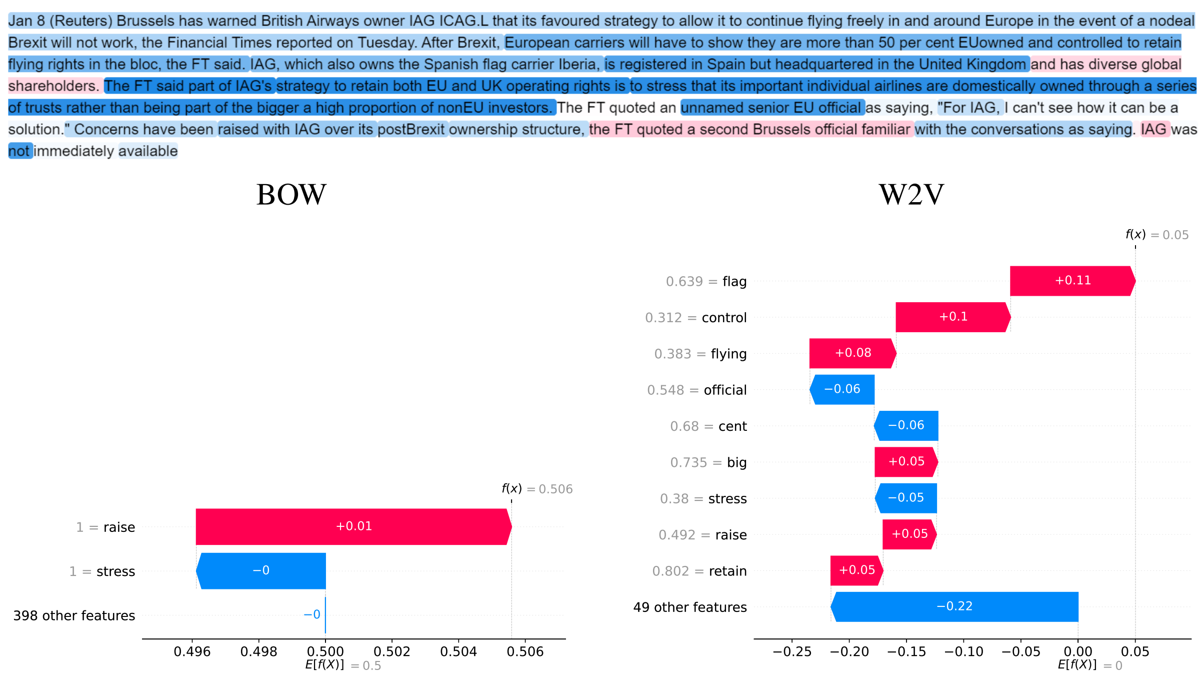
News Text: Fanhua Inc FANH.O: Fanhua announces formation of independent special committee Fanhua Inc special committee is comprised of 3 independent directors Allen Lueth, Stephen Markscheid Mengbo Yin Fanhua Inc decided to form special committee to review allegations in reports contain speculations and "misinterpretations of events" Fanhua Inc special committee is authorized to retain independent advisors in connection with investigation. Fanhua Inc special committee to conduct an independent review of allegations raised in several reports issued recently

Figure A.9: Interpretation of Negative News 3

BERT

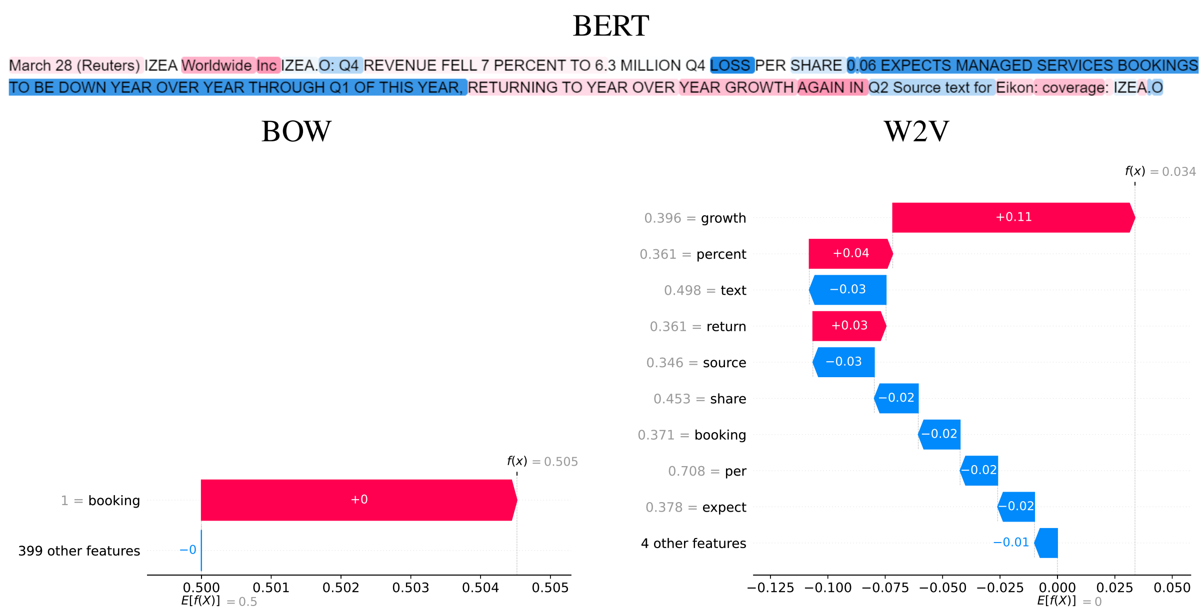
BOW

W2V



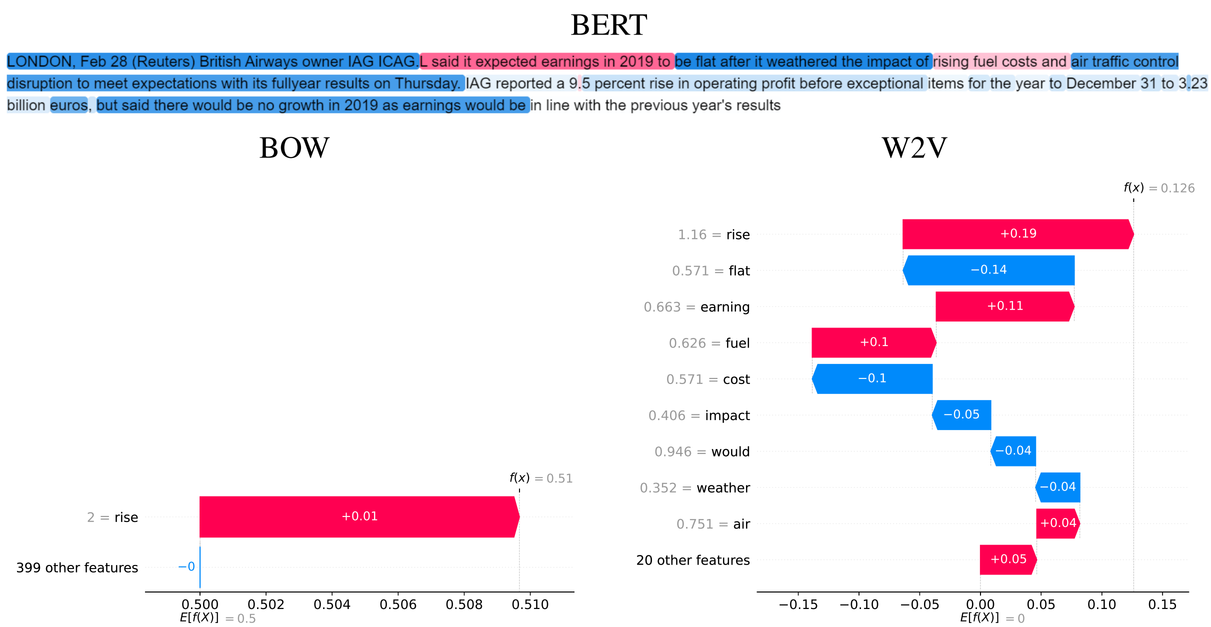
News Text: Brussels has warned British Airways owner IAG ICAG.L that its favoured strategy to allow it to continue flying freely in and around Europe in the event of a nodeal Brexit will not work, the Financial Times reported on Tuesday. After Brexit, European carriers will have to show they are more than 50 per cent EUowned and controlled to retain flying rights in the bloc, the FT said. IAG, which also owns the Spanish flag carrier Iberia, is registered in Spain but headquartered in the United Kingdom and has diverse global shareholders. The FT said part of IAG's strategy to retain both EU and UK operating rights is to stress that its important individual airlines are domestically owned through a series of trusts rather than being part of the bigger a high proportion of nonEU investors. The FT quoted an unnamed senior EU official as saying, ”For IAG, I can't see how it can be a solution.” Concerns have been raised with IAG over its postBrexit ownership structure, the FT quoted a second Brussels official familiar with the conversations as saying. IAG was not immediately available

Figure A.10: Interpretation of Negative News 4



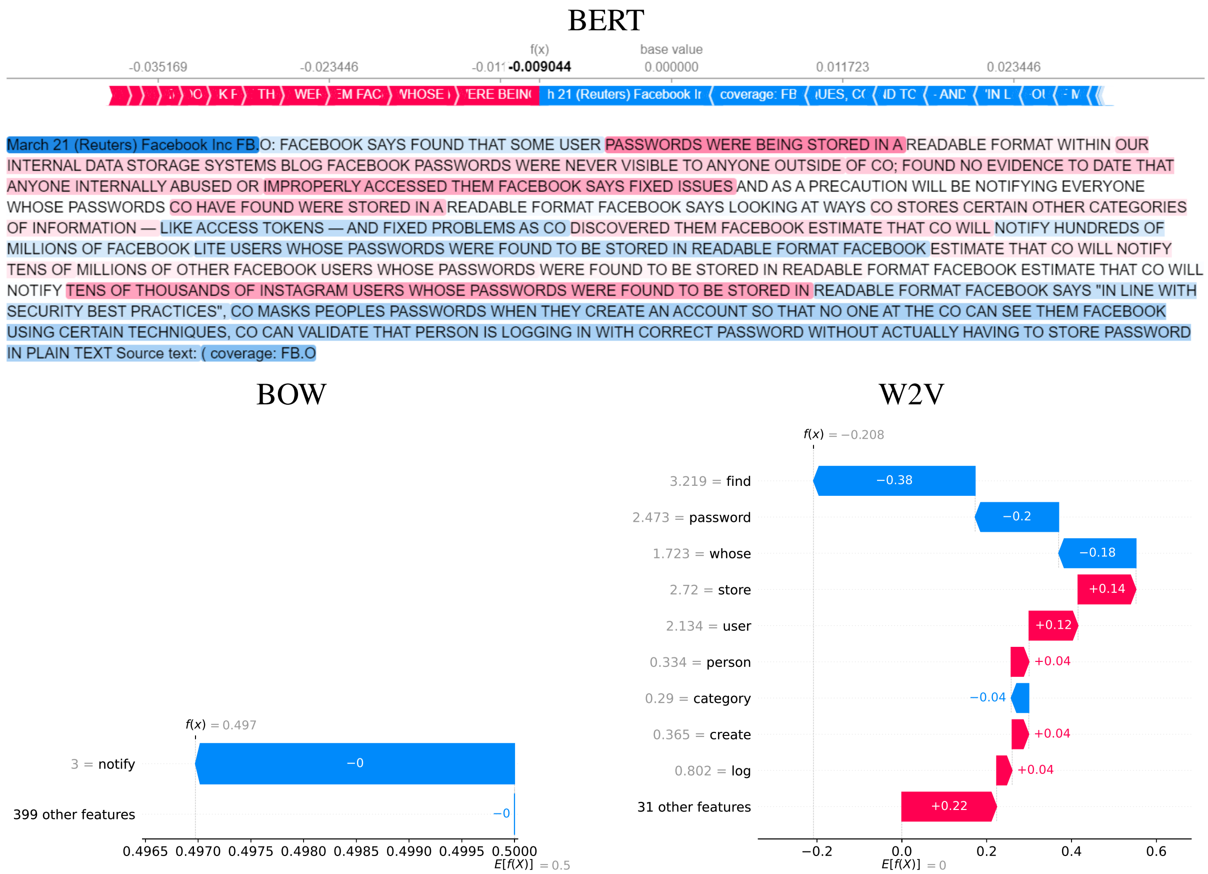
News Text: IZEA Worldwide Inc IZEA.O: Q4 REVENUE FELL 7 PERCENT TO 6.3 MILLION Q4 LOSS PER SHARE 0.06 EXPECTS MANAGED SERVICES BOOKINGS TO BE DOWN YEAR OVER YEAR THROUGH Q1 OF THIS YEAR, RETURNING TO YEAR OVER YEAR GROWTH AGAIN IN Q2

Figure A.11: Interpretation of Negative News 5



News Text: British Airways owner IAG ICAG.L said it expected earnings in 2019 to be flat after it weathered the impact of rising fuel costs and air traffic control disruption to meet expectations with its fullyear results on Thursday. IAG reported a 9.5 percent rise in operating profit before exceptional items for the year to December 31 to 3.23 billion euros, but said there would be no growth in 2019 as earnings would be in line with the previous year's results.

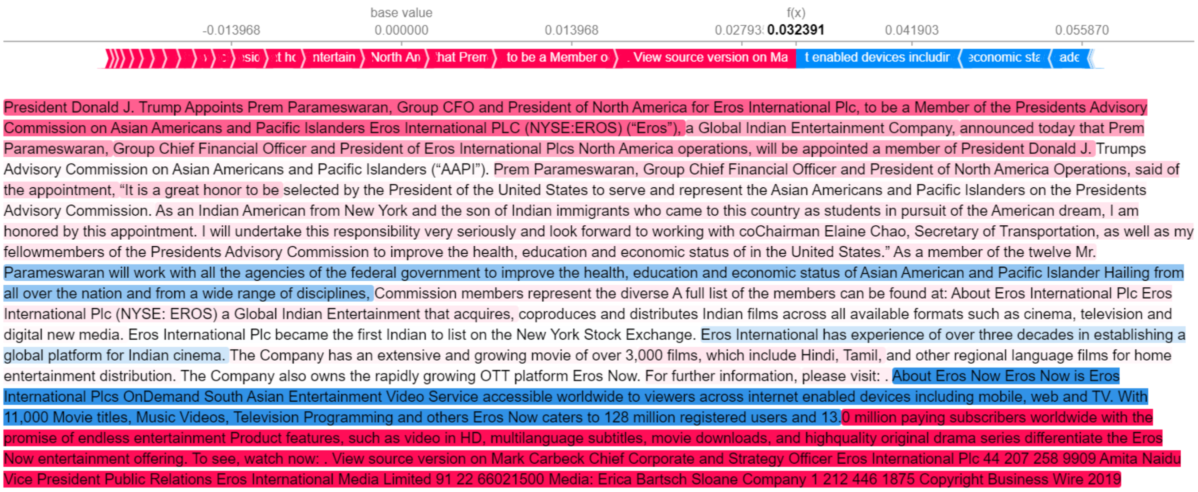
Figure A.12: Interpretation of Positive News 1



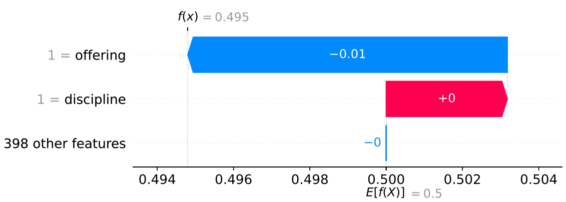
News Text: Facebook inc FB.O: Facebook says found that some user passwords were being stored in a readable format within our internal data storage systems blog facebook passwords were never visible to anyone outside of co; found no evidence to date that anyone internally abused or improperly accessed them facebook says fixed issues and as a precaution will be notifying everyone whose passwords co have found were stored in a readable format facebook says looking at ways co stores certain other categories of information — like access tokens — and fixed problems as co discovered them facebook estimate that co will notify hundreds of millions of facebook lite users whose passwords were found to be stored in readable format facebook estimate that co will notify tens of millions of other facebook users whose passwords were found to be stored in readable format facebook estimate that co will notify tens of thousands of instagram users whose passwords were found to be stored in readable format facebook says ”in line with security best practices”, co masks peoples passwords when they create an account so that no one at the co can see them facebook using certain techniques, co can validate that person is logging in with correct password without actually having to store password in plain text

Figure A.13: Interpretation of Positive News 2

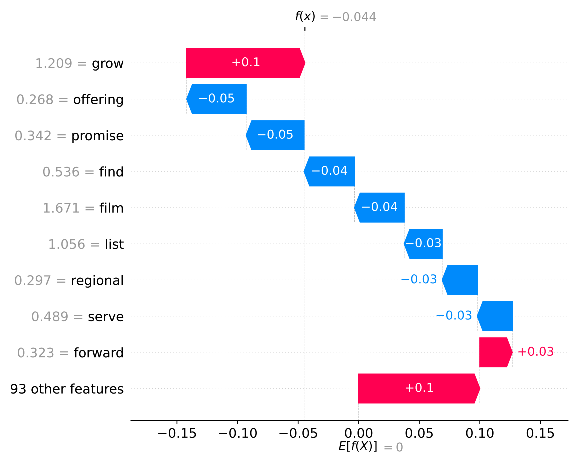
BERT



BOW

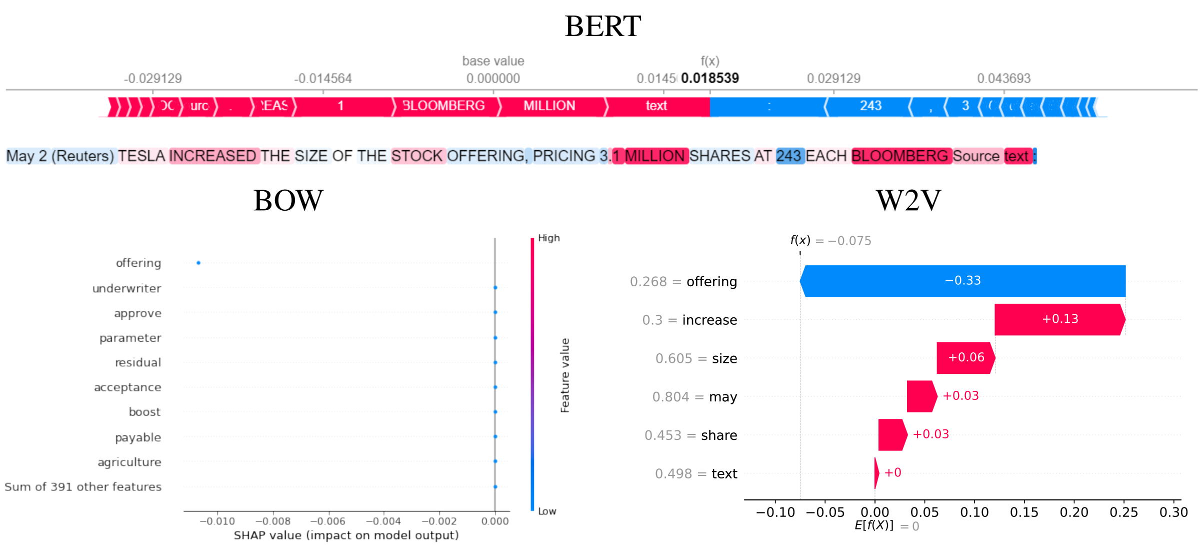


W2V



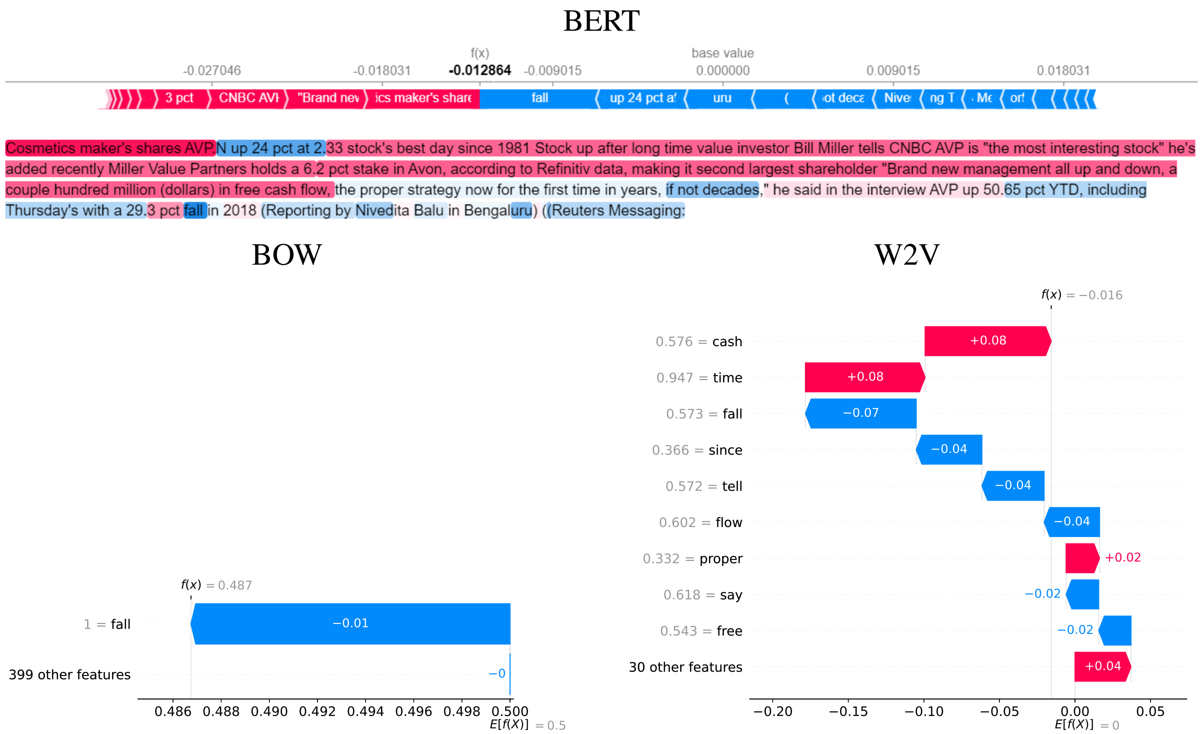
News Text: President Donald J. Trump Appoints Prem Parameswaran, Group CFO and President of North America for Eros International Plc, to be a Member of the Presidents Advisory Commission on Asian Americans and Pacific Islanders Eros International PLC (NYSE:EROS) (“Eros”), a Global Indian Entertainment Company, announced today that Prem Parameswaran, Group Chief Financial Officer and President of Eros International Plcs North America operations, will be appointed a member of President Donald J. Trumps Advisory Commission on Asian Americans and Pacific Islanders (“AAPI”). Prem Parameswaran, Group Chief Financial Officer and President of North America Operations, said of the appointment, “It is a great honor to be selected by the President of the United States to serve and represent the Asian Americans and Pacific Islanders on the Presidents Advisory Commission. As an Indian American from New York and the son of Indian immigrants who came to this country as students in pursuit of the American dream, I am honored by this appointment. I will undertake this responsibility very seriously and look forward to working with coChairman Elaine Chao, Secretary of Transportation, as well as my fellowmembers of the Presidents Advisory Commission to improve the health, education and economic status of in the United States.” As a member of the twelve Mr. Parameswaran will work with all the agencies of the federal government to improve the health, education and economic status of Asian American and Pacific Islander Hailing from all over the nation and from a wide range of disciplines, Commission members represent the diverse A full list of the members can be found at: About Eros International Plc Eros International Plc (NYSE: EROS) a Global Indian Entertainment that acquires, coproduces and distributes Indian films across all available formats such as cinema, television and digital new media. Eros International Plc became the first Indian to list on the New York Stock Exchange. Eros International has experience of over three decades in establishing a global platform for Indian cinema. The Company has an extensive and growing movie of over 3,000 films, which include Hindi, Tamil, and other regional language films for home enter-

Figure A.14: Interpretation of Positive News 3



News Text: Tesla increased the size of the stock offering, pricing 3.1 million shares at $243 each

Figure A.15: Interpretation of Positive News 4



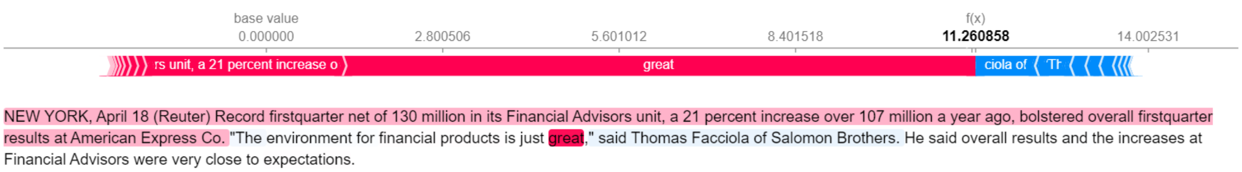
News Text: Cosmetics maker's shares AVP.N up 24 pct at 2.33 stock's best day since 1981 Stock up after long time value investor Bill Miller tells CNBC AVP is ”the most interesting stock” he's added recently Miller Value Partners holds a 6.2 pct stake in Avon, according to Refinitiv data, making it second largest shareholder ”Brand new management all up and down, a couple hundred million (dollars) in free cash flow, the proper strategy now for the first time in years, if not decades,” he said in the interview AVP up 50.65 pct YTD, including Thursday's with a 29.3 pct fall in 2018

A.4 Emotion News Example

For each news, the interpretation of the model on the corresponding emotion is also plotted with the red color highlighting the positive impact on the prediction.

Admiration

- *1996-04-18 17:33:10.820000, American Express Company (AXP)* NEW YORK, April 18 (Reuter) Record firstquarter net of \$130 million in its Financial Advisors unit, a 21 percent increase over \$107 million a year ago, bolstered overall firstquarter results at American Express Co. ”The environment for financial products is just great,” said Thomas Facciola of Salomon Brothers. He said overall results and the increases at Financial Advisors were very close to expectations.

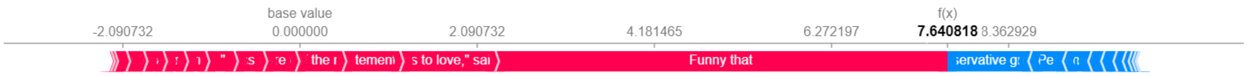


- *2005-07-26 09:08:12.221000, Darden Restaurants (DRI)* On Monday, Red Lobster received a “Choice in Chains” Award from “Restaurants Institutions” magazine for being voted “best seafood restaurant” in a survey of more than 3,200 consumers. Respondents rated 200 of the nation’s largest on food quality, menu variety, cleanliness, service, value, atmosphere and reputation. Red Lobster is the only restaurant to have been named best in its category every year since the seafood category was created in 1989. “We are honored to receive this award because it is based on a national survey of the restaurant critics who matter most our guests,” said Red Lobster President Kim Lopdrup. “We’re focusing on delighting every guest with a ‘simply great’ seafood dining experience. Our guest satisfaction scores are at best ever levels.” In its 37 year history, Red Lobster has introduced America to many seafood dishes including live Maine lobster, snow crab legs, jumbo shrimp and tilapia. Recently, received national recognition for its LightHouse Menu, which features greattasting dishes that are low in fat, carbs and calories, as well as its kid's menu that provides healthy tastes of the sea, including steamed crab legs, grilled fish and fresh vegetables. Red Lobster operates over 670 restaurants in the United States and Canada, serving more than 141 million guests a year.

On Monday, Red Lobster received a "Choice in Chains" Award from "Restaurants Institutions" magazine for being voted "best seafood restaurant" in a survey of more than 3,200 consumers. (Logo:) Respondents rated 200 of the nation's largest on food quality, menu variety, cleanliness, service, value, atmosphere and reputation. Red Lobster is the only restaurant to have been named best in its category every year since the seafood category was created in 1989. "We are honored to receive this award because it is based on a national survey of the restaurant critics who matter most our guests," said Red Lobster President Kim Lopdrup. "We're focusing on delighting every guest with a 'simply great' seafood dining experience. Our guest satisfaction scores are at bestever levels." In its 37year history, Red Lobster has introduced America to many seafood dishes including live Maine lobster, snow crab legs, jumbo shrimp and tilapia. Recently, received national recognition for its LightHouse Menu, which features greattasting dishes that are low in fat, carbs and calories, as well as its kid's menu that provides healthy tastes of the sea, including steamed crab legs, grilled fish and fresh vegetables. Red Lobster operates over 670 restaurants in the United States and Canada, serving more than 141 million guests a year.

Amusement

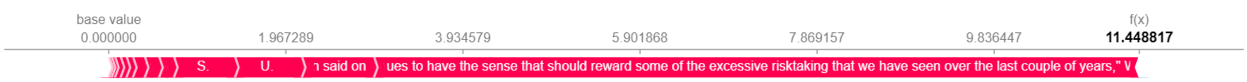
- *2012-02-03 16:42:51.057000, J.C. Penney (JCP)* J.C. Penney Co Inc JCP.N said on Friday said it fully backs its partnership with chat show host Ellen DeGeneres after a conservative group urged the retailer to reconsider hiring DeGeneres as a spokeswoman because she is a lesbian. A spokeswoman for the retailer declined on the issue but did say in an email to Reuters, "jcpenney stands behind its partnership with Ellen DeGeneres" and added that its announcement of the agreement last week sums up view of the popular TV personality. In that statement on Jan. president Michael Francis called DeGeneres "one of the most fun and vibrant people in entertainment today, with great warmth and a downtoearth attitude." Penney's decision to hire DeGeneres spurred conservative group One Million Moms, a division of the American Family Association, to slam for not being "neutral in the culture war." "Funny that JC Penney thinks hiring an open homosexual spokesperson will help their business when most of their customers are traditional families. More sales will be lost than gained unless they replace their spokesperson quickly," the organization posted on their website, urging supporters to call their local J.C. Penney store manager to against hiring DeGeneres. Gay and lesbian rights group GLAAD launched their own countercampaign, Stand Up For Ellen, in response to One Million Moms, attracting more than 25,000 signatures in their petition to J.C. Penney to keep DeGeneres as their spokesperson, and applauded the retail store for sticking to their decision. ”This week Americans spoke out in overwhelming support of LGBT people and J.C. Penney’s decision not to fire Ellen simply for who she happens to love,” said Herndon Graddick, senior director of programs at GLAAD.



J.C. Penney Co Inc JCP.N said on Friday said it fully backs its partnership with chat show host Ellen DeGeneres after a conservative group urged the retailer to reconsider hiring DeGeneres as a spokeswoman because she is a lesbian. A spokeswoman for the retailer declined on the issue but did say in an email to Reuters, "jcpenney stands behind its partnership with Ellen DeGeneres" and added that its announcement of the agreement last week sums up view of the popular TV personality. In that statement on Jan. president Michael Francis called DeGeneres "one of the most fun and vibrant people in entertainment today, with great warmth and a downtoearth attitude." Penney's decision to hire DeGeneres spurred conservative group One Million Moms, a division of the American Family Association, to slam for not being "neutral in the culture war." "Funny that JC Penney thinks hiring an open homosexual spokesperson will help their business when most of their customers are traditional families. More sales will be lost than gained unless they replace their spokesperson quickly," the organization posted on their website, urging supporters to call their local J.C. Penney store manager to against hiring DeGeneres. Gay and lesbian rights group GLAAD launched their own countercampaign, Stand Up For Ellen, in response to One Million Moms, attracting more than 25,000 signatures in their petition to J.C. Penney to keep DeGeneres as their spokesperson, and applauded the retail store for sticking to their decision. "This week Americans spoke out in overwhelming support of LGBT people and J.C. Penney's decision not to fire Ellen simply for who she happens to love," said Herndon Graddick, senior director of programs at GLAAD.

Anger

- *2010-02-03 13:37:59.686000, American International Group (AIG)* U.S. President Barack Obama is ”frustrated and angry” over on Wall Street, his spokesman said on Wednesday. ”The president is frustrated and angry that Wall Street continues to have the sense that should reward some of the excessive risktaking that we have seen over the last couple of years,” White House spokesman Bill Burton told a news conference.



U.S. President Barack Obama is "frustrated and angry" over on Wall Street, his spokesman said on Wednesday. "The president is frustrated and angry that Wall Street continues to have the sense that should reward some of the excessive risktaking that we have seen over the last couple of years," White House spokesman Bill Burton told a news conference.

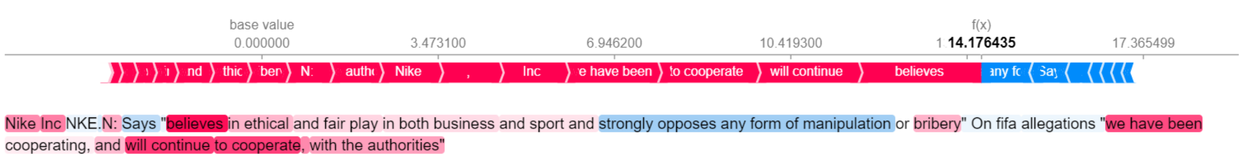
- *2014-05-28 09:31:26.683000, BlackRock (BLK)* Blackrock Inc BLK.N Chief Executive Larry Fink, who runs the world’s largest asset manager, said on Wednesday he has fielded angry phone calls over a letter he sent in March to SP 500 executives that warned them about the perils of shortterm thinking. ”I’ve had some really angry phone calls,” Fink said at a New York investment conference hosted by Sanford Bernstein. He did not name any of the angry callers. In the March 21 letter, Fink warned against relying too much on dividends and buybacks to produce quick returns at the expense of longterm investment. Blackrock oversees more than $4 trillion in client assets.

Annoyance

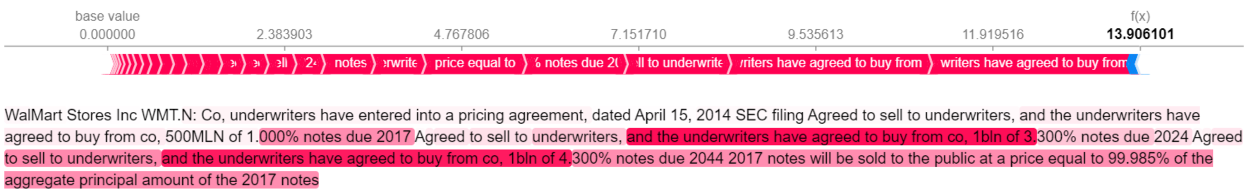
- *1997-12-09 12:32:04.668000, Boeing (BA)* National Transportation Safety Board (NTSB) Chairman Jim Hall said on Tuesday he was frustrated by the slow pace of Boeing Co.'s BA.N voluntary inspection program for center fuel tanks on its 747 planes. At a hearing into the explosion of TWA Flight 800, which killed 230 people nearly 17 months ago, Hall also expressed irritation over the Federal Aviation Administration's slow pace in making the Boeing program mandatory for 747s and other airliners.

Approval

- *2015-05-27 16:25:57.943000, Nike Inc (NKE)* Nike Inc NKE.N: Says "believes in ethical and fair play in both business and sport and strongly opposes any form of manipulation or bribery" On fifa allegations "we have been cooperating, and will continue to cooperate, with the authorities" This is a good example to show BERT's understanding of news. "strongly opposes any form of manipulation" has negative effects on emotion Approval but the overall prediction is based on the whole context.

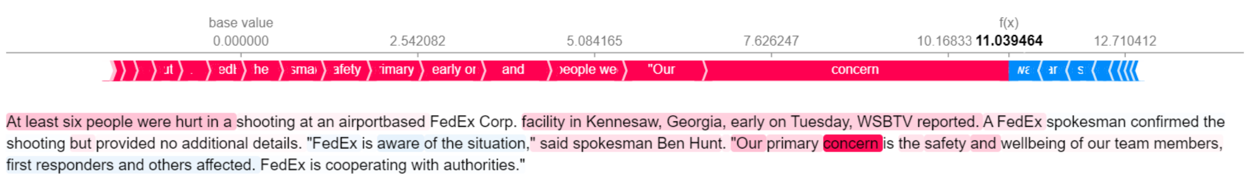


- *2014-04-21 17:24:48.480000, WalMart (WMT)* WalMart Stores Inc WMT.N: Co, underwriters have entered into a pricing agreement, dated April 15, 2014 SEC filing Agreed to sell to underwriters, and the underwriters have agreed to buy from co, \$500MLN of 1.000% notes due 2017 Agreed to sell to underwriters, and the underwriters have agreed to buy from co,\$1bln of 3.300% notes due 2024 Agreed to sell to underwriters, and the underwriters have agreed to buy from co,\$1bln of 4.300% notes due 2044 2017 notes will be sold to the public at a price equal to 99.985% of the aggregate principal amount of the 2017 notes



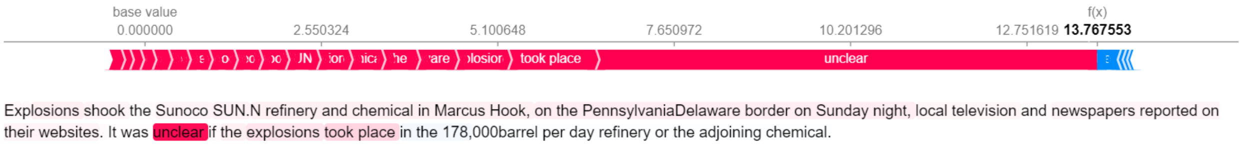
Caring

- *2014-04-29 08:02:33.184000, FedEx Corp. (FDX)* At least six people were hurt in a shooting at an airportbased FedEx Corp. facility in Kennesaw, Georgia, early on Tuesday, WSBTV reported. A FedEx spokesman confirmed the shooting but provided no additional details. "FedEx is aware of the situation," said spokesman Ben Hunt. "Our primary concern is the safety and wellbeing of our team members, first responders and others affected. FedEx is cooperating with authorities."

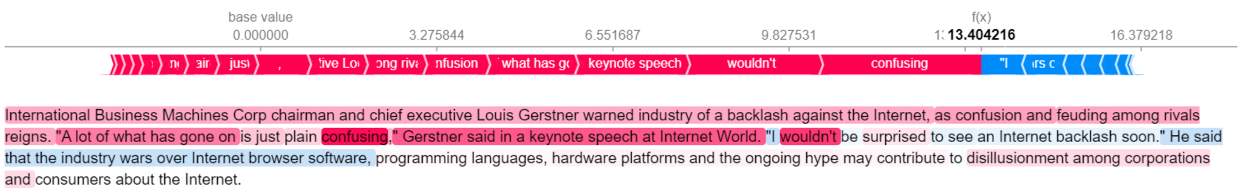


Confusion

- *2009-05-18 00:23:06.551000, Sunoco (SUN)* Explosions shook the Sunoco SUN.N refinery and chemical in Marcus Hook, on the PennsylvaniaDelaware border on Sunday night, local television and newspapers reported on their websites. It was unclear if the explosions took place in the 178,000 barrel per day refinery or the adjoining chemical

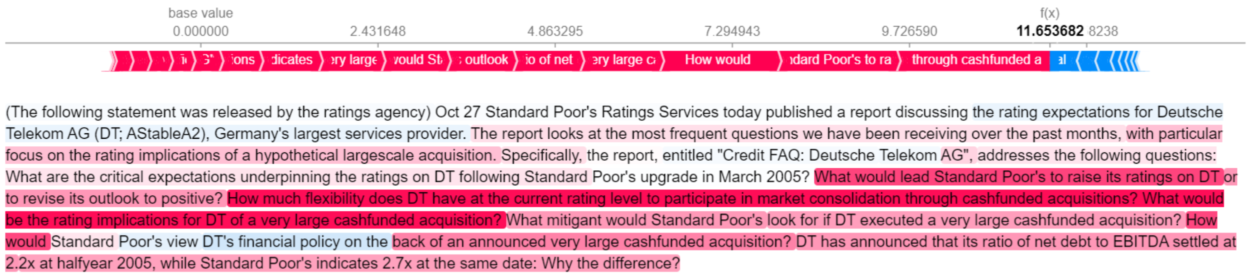


- *1996-12-11 14:32:16.860000, IBM Corp. (IBM)* International Business Machines Corp chairman and chief executive Louis Gerstner warned industry of a backlash against the Internet, as confusion and feuding among rivals reigns. ”A lot of what has gone on is just plain confusing,” Gerstner said in a keynote speech at Internet World. ”I wouldn’t be surprised to see an Internet backlash soon.” He said that the industry wars over Internet browser software, programming languages, hardware platforms and the ongoing hype may contribute to disillusionment among corporations and consumers about the Internet.



Curiosity

- *2005-10-27 09:30:40.793000, Deutsche Telekom AG (DTE)* (The following statement was released by the ratings agency) Oct 27 Standard Poor’s Ratings Services today published a report discussing the rating expectations for Deutsche Telekom AG (DT; AStableA2), Germany’s largest services provider. The report looks at the most frequent questions we have been receiving over the past months, with particular focus on the rating implications of a hypothetical largescale acquisition. Specifically, the report, entitled ”Credit FAQ: Deutsche Telekom AG”, addresses the following questions: What are the critical expectations underpinning the ratings on DT following Standard Poor’s upgrade in March 2005? What would lead Standard Poor’s to raise its ratings on DT or to revise its outlook to positive? How much flexibility does DT have at the current rating level to participate in market consolidation through cashfunded acquisitions? What would be the rating implications for DT of a very large cashfunded acquisition? What mitigant would Standard Poor's look for if DT executed a very large cashfunded acquisition? How would Standard Poor's view DT's financial policy on the back of an announced very large cashfunded acquisition? DT has announced that its ratio of net debt to EBITDA settled at 2.2x at halfyear 2005, while Standard Poor's indicates 2.7x at the same date: Why the difference?



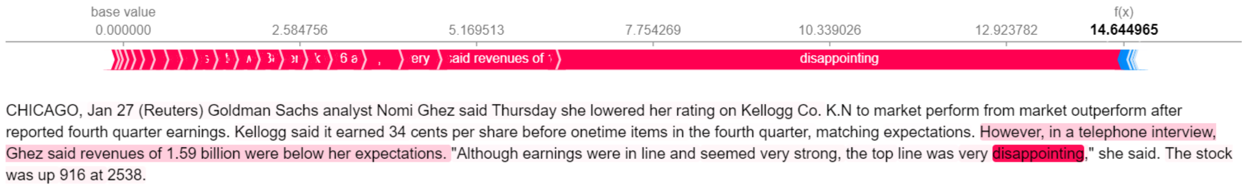
Desire

- *1996-07-16 03:57:14.550000, MEDIA ASIA PACIFIC LIMITED (MAS)* At a Board Meeting today, the resignation of Mr Interlandi was accepted. Mr Interlandi wishes to pursue other interests and has recently accepted additional responsibility in his chosen profession. The Board wishes to thank Mr Interlandi for his efforts and will be seeking a replacement Director immediately.

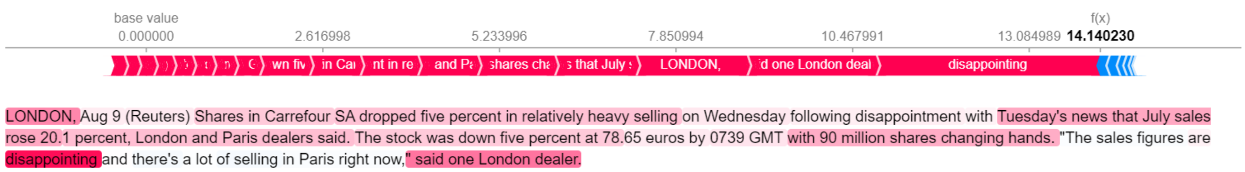
Disappointment

- *2000-01-27 11:10:05.609000, Kellogg Company (K)* CHICAGO, Jan 27 (Reuters) Goldman Sachs analyst Nomi Ghez said Thursday she lowered her rating on Kellogg Co. K.N to market perform from market outperform after reported fourth quarter earnings. Kellogg said it earned 34 cents per share before onetime items in the fourth quarter, matching expectations.

However, in a telephone interview, Ghez said revenues of $1.59 billion were below her expectations. ”Although earnings were in line and seemed very strong, the top line was very disappointing,” she said. The stock was up 916 at 2538.

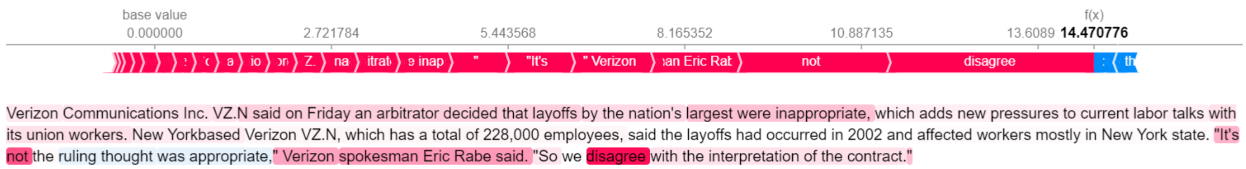


- *2000-08-09 03:44:15.276000, Carrefour (CA)* LONDON, Aug 9 (Reuters) Shares in Carrefour SA dropped five percent in relatively heavy selling on Wednesday following disappointment with Tuesday’s news that July sales rose 20.1 percent, London and Paris dealers said. The stock was down five percent at 78.65 euros by 0739 GMT with 90 million shares changing hands. ”The sales figures are disappointing and there’s a lot of selling in Paris right now,” said one London dealer.

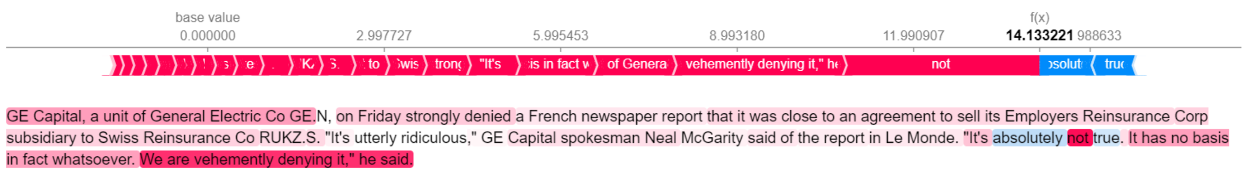


Disapproval

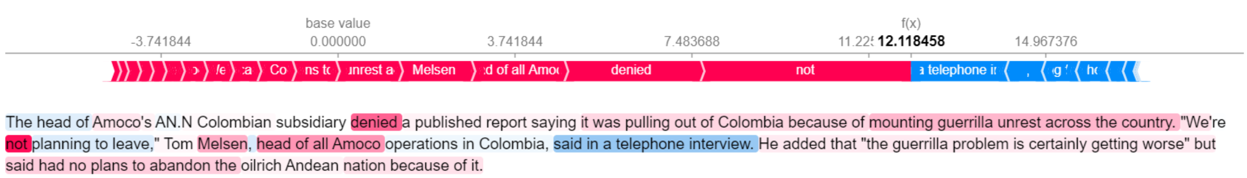
- *2003-07-11 12:28:40.697000, Verizon Communications Inc. (VZ)* Verizon Communications Inc. VZ.N said on Friday an arbitrator decided that layoffs by the nationś largest were inappropriate, which adds new pressures to current labor talks with its union workers. New Yorkbased Verizon VZ.N, which has a total of 228,000 employees, said the layoffs had occurred in 2002 and affected workers mostly in New York state. ”Itś not the ruling thought was appropriate,” Verizon spokesman Eric Rabe said. ”So we disagree with the interpretation of the contract.”



- *1997-10-17 14:43:36.898000, General Electric Co. (GE)* GE Capital, a unit of General Electric Co GE.N, on Friday strongly denied a French newspaper report that it was close to an agreement to sell its Employers Reinsurance Corp subsidiary to Swiss Reinsurance Co RUKZ.S. ”Itś utterly ridiculous,” GE Capital spokesman Neal McGarity said of the report in Le Monde. ”Itś absolutely not true. It has no basis in fact whatsoever. We are vehemently denying it,” he said.



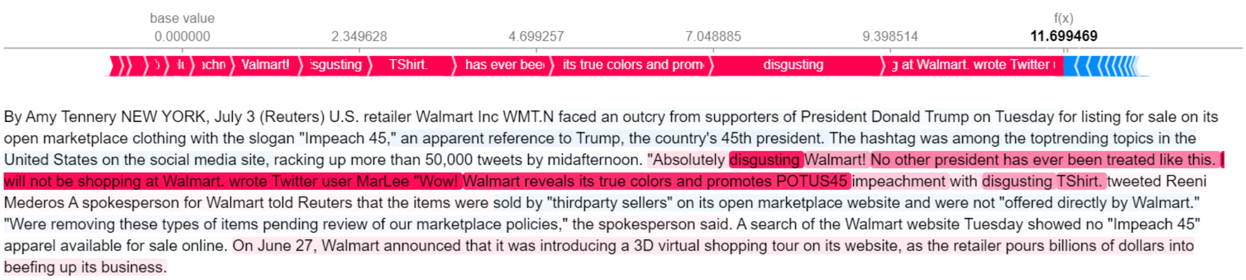
- *1996-10-04 11:44:24.210000, Amoco (AN)* The head of Amoco’s AN.N Colombian subsidiary denied a published report saying it was pulling out of Colombia because of mounting guerrilla unrest across the country. ”Wee not planning to leave,” Tom Melsen, head of all Amoco operations in Colombia, said in a telephone interview. He added that ”the guerrilla problem is certainly getting worse” but said had no plans to abandon the oilrich Andean nation because of it.



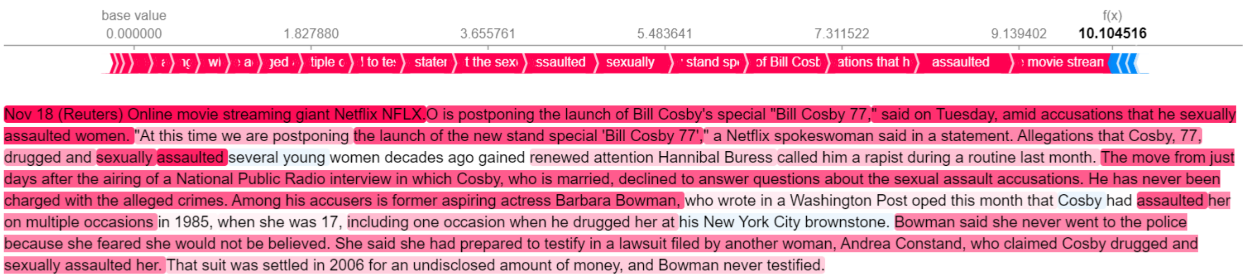
Disgust

- *2018-07-03 16:06:33.553000, Walmart (WMT)* By Amy Tennery NEW YORK, July 3 (Reuters) U.S. retailer Walmart Inc WMT.N faced an outcry from supporters of President Donald

Trump on Tuesday for listing for sale on its open marketplace clothing with the slogan ”Impeach 45,” an apparent reference to Trump, the country’s 45th president. The hashtag was among the toptrending topics in the United States on the social media site, racking up more than 50,000 tweets by midafternoon. ”Absolutely disgusting Walmart! No other president has ever been treated like this. I will not be shopping at Walmart. wrote Twitter user MarLee ”Wow! Walmart reveals its true colors and promotes POTUS45 impeachment with disgusting TShirt. tweeted Reeni Mederos A spokesperson for Walmart told Reuters that the items were sold by ”thirdparty sellers” on its open marketplace website and were not ”offered directly by Walmart.” ”Were removing these types of items pending review of our marketplace policies,” the spokesperson said. A search of the Walmart website Tuesday showed no ”Impeach 45” apparel available for sale online. On June 27, Walmart announced that it was introducing a 3D virtual shopping tour on its website, as the retailer pours billions of dollars into beefing up its business.



- *2014-11-19 00:54:00.357000, Netflix (NFLX)* Nov 18 (Reuters) Online movie streaming giant Netflix NFLX.O is postponing the launch of Bill Cosbyś special ”Bill Cosby 77,” said on Tuesday, amid accusations that he sexually assaulted women. ”At this time we are postponing the launch of the new stand special B́ill Cosby 77,́” a Netflix spokeswoman said in a statement. Allegations that Cosby, 77, drugged and sexually assaulted several young women decades ago gained renewed attention Hannibal Buress called him a rapist during a routine last month. The move from just days after the airing of a National Public Radio interview in which Cosby, who is married, declined to answer questions about the sexual assault accusations. He has never been charged with the alleged crimes. Among his accusers is former aspiring actress Barbara Bowman, who wrote in a Washington Post oped this month that Cosby had assaulted her on multiple occasions in 1985, when she was 17, including one occasion when he drugged her at his New York City brownstone. Bowman said she never went to the police because she feared she would not be believed. She said she had prepared to testify in a lawsuit filed by another woman, Andrea Constand, who claimed Cosby drugged and sexually assaulted her. That suit was settled in 2006 for an undisclosed amount of money, and Bowman never testified.

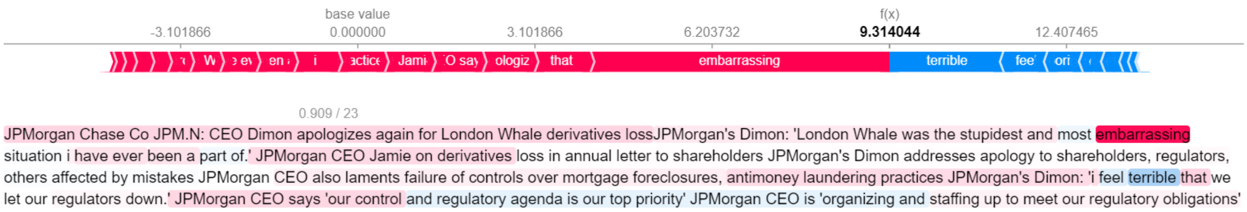


Embarrassment

- *1998-03-31 15:19:16.888000, Union Pacific Corp. (UNP)* OMAHA, Neb., March 31 (Reuters) Union Pacific Corp.'s top executive Tuesday said the railroad industry faces a severe capacity problem and apologized for the difficulty and delay has faced in correcting a severe service disruption crisis. "I am acutely embarrassed, and is embarrassed, at the time it has taken to recover from our congestion crisis," said Richard Davidson, chairman and chief executive officer of Union Pacific.

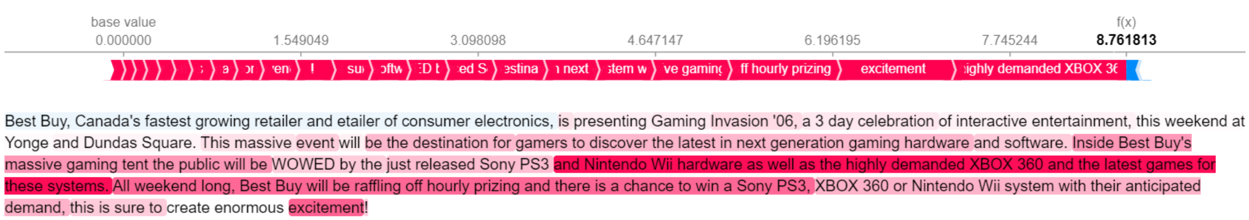
- *2013-04-10 14:24:57.426000, JPMorgan Chase Co (JPM)* JPMorgan Chase Co JPM.N: CEO Dimon apologizes again for London Whale derivatives loss JPMorgan's Dimon: 'London Whale was the stupidest and most embarrassing situation i have ever been a part of.'

JPMorgan CEO Jamie on derivatives loss in annual letter to shareholders JPMorgan's Dimon addresses apology to shareholders, regulators, others affected by mistakes JPMorgan CEO also laments failure of controls over mortgage foreclosures, antimoney laundering practices JPMorgan's Dimon: 'i feel terrible that we let our regulators down.' JPMorgan CEO says 'our control and regulatory agenda is our top priority' JPMorgan CEO is 'organizing and staffing up to meet our regulatory obligations'



Excitement

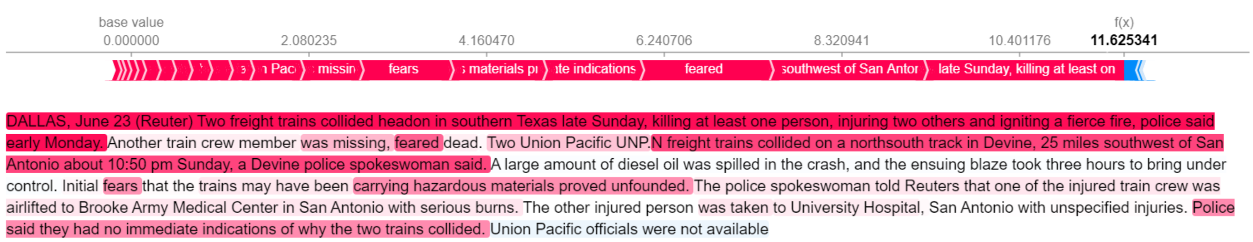
- *2006-11-15 06:00:09.436000, Best Buy (BBY)* Best Buy, Canada's fastest growing retailer and etailer of consumer electronics, is presenting Gaming Invasion '06, a 3 day celebration of interactive entertainment, this weekend at Yonge and Dundas Square. This massive event will be the destination for gamers to discover the latest in next generation gaming hardware and software. Inside Best Buy's massive gaming tent the public will be WOWED by the just released Sony PS3 and Nintendo Wii hardware as well as the highly demanded XBOX 360 and the latest games for these systems. All weekend long, Best Buy will be raffling off hourly prizing and there is a chance to win a Sony PS3, XBOX 360 or Nintendo Wii system with their anticipated demand, this is sure to create enormous excitement!



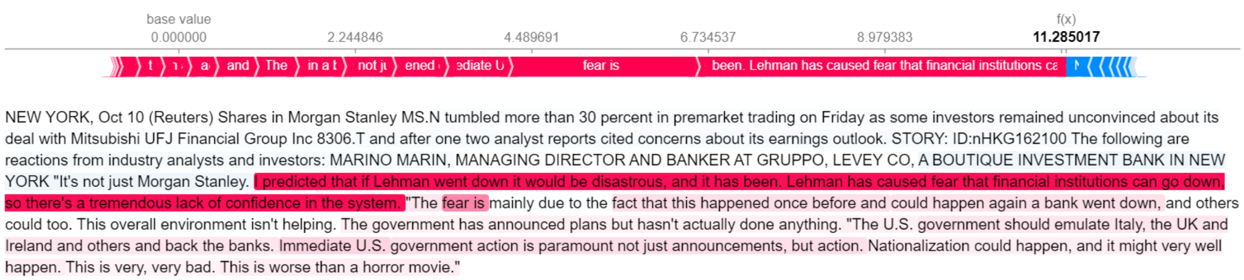
- *2016-06-16 12:02:17.314000, Walmart (WMT)* WalMart WMT.N is excited about the opportunities in China, is tough, chief executive Doug McMillon said on Thursday. "China is a tremendous opportunity and I am very excited about China, bullish on China, but itś very involved, so itś going to be tough,” he told a conference.

Fear

- *1997-06-23 03:57:45.390000, Union Pacific (UNP)* DALLAS, June 23 (Reuter) Two freight trains collided headon in southern Texas late Sunday, killing at least one person, injuring two others and igniting a fierce fire, police said early Monday. Another train crew member was missing, feared dead. Two Union Pacific UNP.N freight trains collided on a northsouth track in Devine, 25 miles southwest of San Antonio about 10:50 pm Sunday, a Devine police spokeswoman said. A large amount of diesel oil was spilled in the crash, and the ensuing blaze took three hours to bring under control. Initial fears that the trains may have been carrying hazardous materials proved unfounded. The police spokeswoman told Reuters that one of the injured train crew was airlifted to Brooke Army Medical Center in San Antonio with serious burns. The other injured person was taken to University Hospital, San Antonio with unspecified injuries. Police said they had no immediate indications of why the two trains collided. Union Pacific officials were not available.



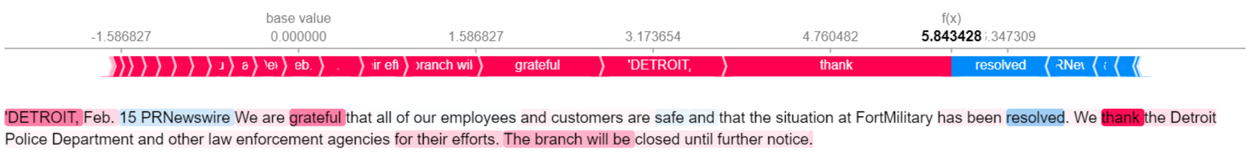
- *2008-10-10 09:00:33.789000, Morgan Stanley (MS)* Shares in Morgan Stanley MS.N tumbled more than 30 percent in premarket trading on Friday as some investors remained unconvinced about its deal with Mitsubishi UFJ Financial Group Inc 8306.T and after one two analyst reports cited concerns about its earnings outlook. STORY: ID:nHKG162100 The following are reactions from industry analysts and investors: MARINO MARIN, MANAGING DIRECTOR AND BANKER AT GRUPPO, LEVEY CO, A BOUTIQUE INVESTMENT BANK IN NEW YORK ”Itś not just Morgan Stanley. I predicted that if Lehman went down it would be disastrous, and it has been. Lehman has caused fear that financial institutions can go down, so thereś a tremendous lack of confidence in the system. ”The fear is mainly due to the fact that this happened once before and could happen again a bank went down, and others could too. This overall environment isnt́ helping. The government has announced plans but hasnt́ actually done anything. ”The U.S. government should emulate Italy, the UK and Ireland and others and back the banks. Immediate U.S. government action is paramount not just announcements, but action. Nationalization could happen, and it might very well happen. This is very, very bad. This is worse than a horror movie.”



- *2001-10-07 21:09:40.262000, Straits Times (STI)* SINGAPORE, Oct 8 (Reuters) Singapore shares opened sharply lower on Monday with investors anxious after the United States and Britain launched a wave of retaliation to the attacks on New York and Washington by bombing Afghanistan. The bellwether Straits Times Index .STI opened 1.67 percent lower and then lost ground to be down 2.19 percent or 30.35 points down at 1,355.10 in early trade. Losers overwhelmed gainers 84 to eight, with volume brisk at 11 million shares. The losses were led by heavyweights across the board. The United States and Britain bombed bases, airports and guerrilla training camps across Afghanistan, plunging Kabul into darkness and panic but leaving Washington’s prime suspect, Osama bin Laden, and the ruling Taliban’s leader unscathed. Fearing possible reprisal attacks by Islamic militants, countries across the world tightened security.

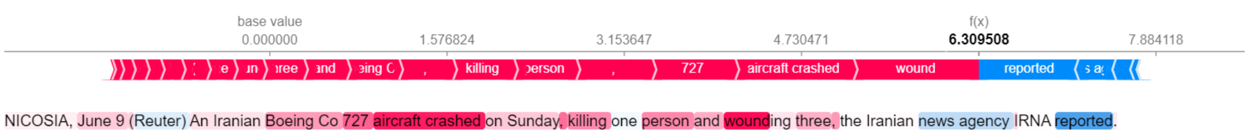
Gratitude

- *2006-02-15 18:08:10.444000, Comerica Bank (CMA)* DETROIT, Feb. 15 PRNewswire We are grateful that all of our employees and customers are safe and that the situation at FortMilitary has been resolved. We thank the Detroit Police Department and other law enforcement agencies for their efforts. The branch will be closed until further notice.

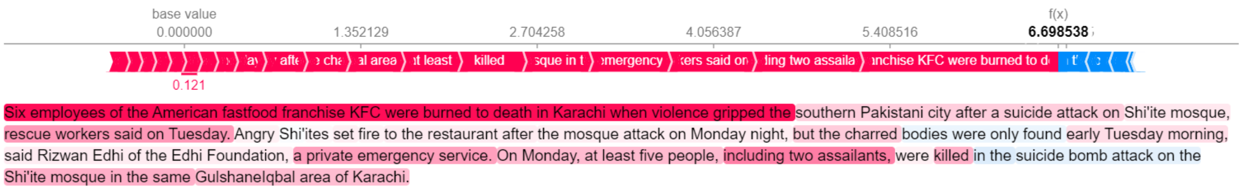


Grief

- *1996-06-09 15:06:56.610000, Boeing (BA)* NICOSIA, June 9 (Reuter) An Iranian Boeing Co 727 aircraft crashed on Sunday, killing one person and wounding three, the Iranian news agency IRNA reported.

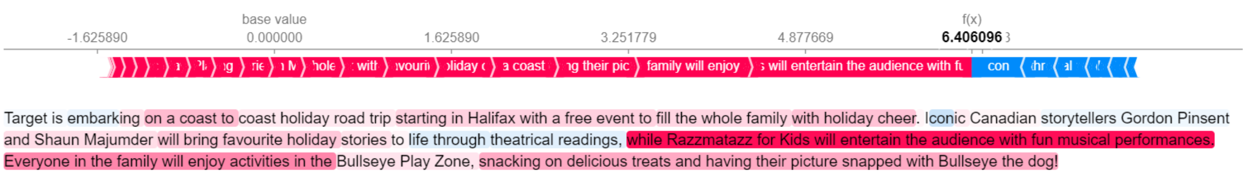


- *2005-05-31 01:25:57.827000, Yum! Brands, Inc. (YUM)* Six employees of the American fastfood franchise KFC were burned to death in Karachi when violence gripped the southern Pakistani city after a suicide attack on Shi'ite mosque, rescue workers said on Tuesday. Angry Shi'ites set fire to the restaurant after the mosque attack on Monday night, but the charred bodies were only found early Tuesday morning, said Rizwan Edhi of the Edhi Foundation, a private emergency service. On Monday, at least five people, including two assailants, were killed in the suicide bomb attack on the Shi'ite mosque in the same GulshaneIqbal area of Karachi.



Joy

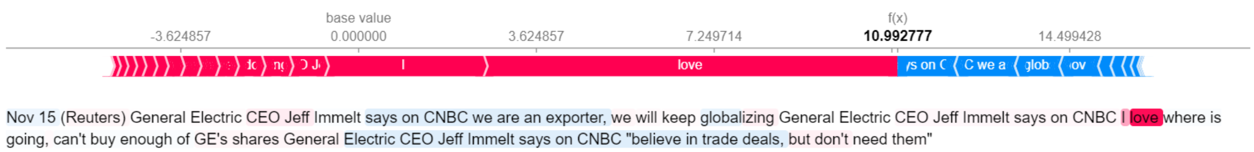
- *2012-11-29 08:00:18.484000, Target (TGT)* Target is embarking on a coast to coast holiday road trip starting in Halifax with a free event to fill the whole family with holiday cheer. Iconic Canadian storytellers Gordon Pinsent and Shaun Majumder will bring favourite holiday stories to life through theatrical readings, while Razzmatazz for Kids will entertain the audience with fun musical performances. Everyone in the family will enjoy activities in the Bullseye Play Zone, snacking on delicious treats and having their picture snapped with Bullseye the dog!



- *1997-03-04 12:36:03.330000, Microsof Corp. (MSFT)* SEATTLE, March 4 (Reuter) Microsoft Corp Treasurer Greg Maffei Tuesday said the has sold five million Office 97 licenses since the product was launched in January. Speaking to a Piper Jaffray investorś conference here, he said one million of those licenses had been presold. He executives were "exceedingly happy" with the results so far. Office 97 is the upgrade to marketleading bundle of productivity applications, including Word and Excel.

Love

- *2016-11-15 11:41:15.414000, General Electric (GE)* Nov 15 (Reuters) General Electric CEO Jeff Immelt says on CNBC we are an exporter, we will keep globalizing General Electric CEO Jeff Immelt says on CNBC I love where is going, can't buy enough of GE's shares General Electric CEO Jeff Immelt says on CNBC ”believe in trade deals, but don't need them”



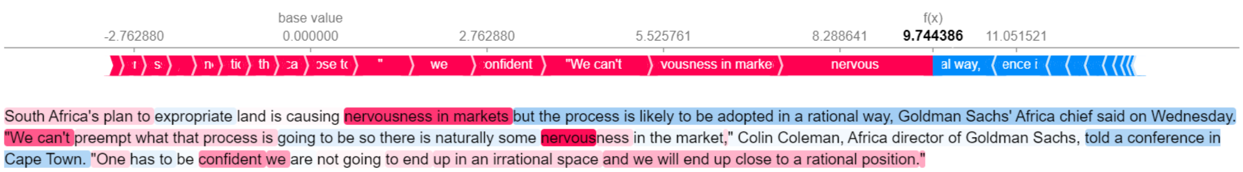
- *2011-03-16 15:39:26.046000, Intercontinental Exchange (ICE)* Highfrequency traders account for about 10 percent of the volume in markets, the president of ICE Futures US ICE.N said on Wednesday, arguing they help markets and are unfairly blamed for volatility. ”We love them,” said Tom Farley, who spoke at a Futures Industry Association conference. ”We think that they dampen volatility.”

Nervousness

- *2002-10-07 07:36:04.252000, Mattel Inc (MAT)* Even though toy giant Mattel Inc MAT.N would traditionally start to slow production now after its holiday season rampup, the maker of the Barbie and Harry Potter lines said on Monday a prolonged work stoppage at West Coast U.S. ports could hurt the Christmas period. ”We worry about the (lockout) on the

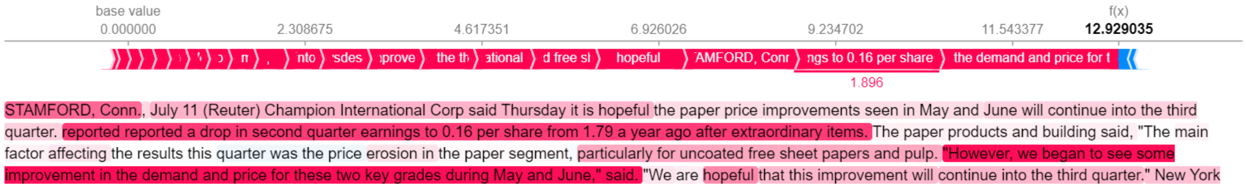
West Coast because we do have products that are there on ships, and we do have a few more weeks of shipments to make, so we are worried about it," said David Lewis, senior vice president at Mattel Asia Pacific Sourcing Ltd in Hong Kong. "Christmas will a worry if the lockout doesn't end soon," he said in a telephone interview.

- *2018-03-07 04:52:29.283000, Goldman Sachs (GS)* South Africa's plan to expropriate land is causing nervousness in markets but the process is likely to be adopted in a rational way, Goldman SachsÁfrica chief said on Wednesday. "We can't preempt what that process is going to be so there is naturally some nervousness in the market," Colin Coleman, Africa director of Goldman Sachs, told a conference in Cape Town. "One has to be confident we are not going to end up in an irrational space and we will end up close to a rational position."



Optimism

- *1996-07-11 12:47:49.074000, Champion International Corp (CHA)* STAMFORD, Conn., July 11 (Reuter) Champion International Corp said Thursday it is hopeful the paper price improvements seen in May and June will continue into the third quarter. reported reported a drop in second quarter earnings to $0.16 per share from 1.79$ a year ago after extraordinary items. The paper products and building said, "The main factor affecting the results this quarter was the price erosion in the paper segment, particularly for uncoated free sheet papers and pulp. "However, we began to see some improvement in the demand and price for these two key grades during May and June,” said. ”We are hopeful that this improvement will continue into the third quarter.” New York

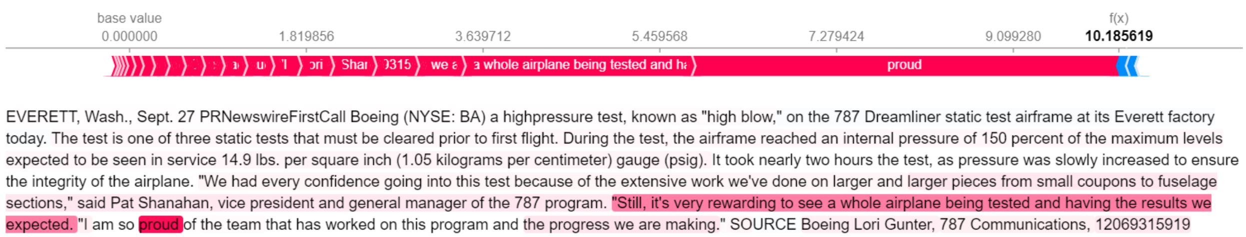


- *1997-10-08 16:50:02.993000, Boeing (BA)* EVERETT, Wash., Oct. 8 (Reuter) Boeing Chairman and Chief Executive Officer Phil Condit said Wednesday he was hopeful that China would place a longawaited aircraft order sometime within the next two months. But Condit told reporters, after a speech to a business group here, that he had no indication the order would be placed within the next two days. ”We are hopeful that somewhere over the next couple of months there might be something,” he said.

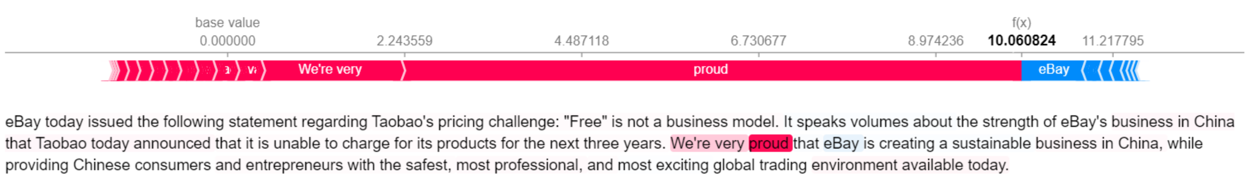
- *2018-09-18 08:49:38.028000, Apple Inc (AAPL)* Apple Inc AAPL.O Chief Executive Tim Cook, whose products were spared from new U.S. tariffs on Chinese goods imposed on Monday, said he is optimistic that the United States and China will eventually work through their trade differences. ”Iḿ optimistic because trade is one of those things where itś not a zerosum game,” Cook told ABC News”́Good Morning America” program on Tuesday. ”Iḿ optimistic that the two countries will sort this out and life will go on.”

Pride

- *2008-09-27 18:33:29.720000, Boeing (BA)* Boeing (NYSE: BA) a highpressure test, known as ”high blow,” on the 787 Dreamliner static test airframe at its Everett factory today. The test is one of three static tests that must be cleared prior to first flight. During the test, the airframe reached an internal pressure of 150 percent of the maximum levels expected to be seen in service 14.9 lbs. per square inch (1.05 kilograms per centimeter) gauge (psig). It took nearly two hours the test, as pressure was slowly increased to ensure the integrity of the airplane. ”We had every confidence going into this test because of the extensive work wéve done on larger and larger pieces from small coupons to fuselage sections,” said Pat Shanahan, vice president and general manager of the 787 program. ”Still, itś very rewarding to see a whole airplane being tested and having the results we expected. ”I am so proud of the team that has worked on this program and the progress we are making.”

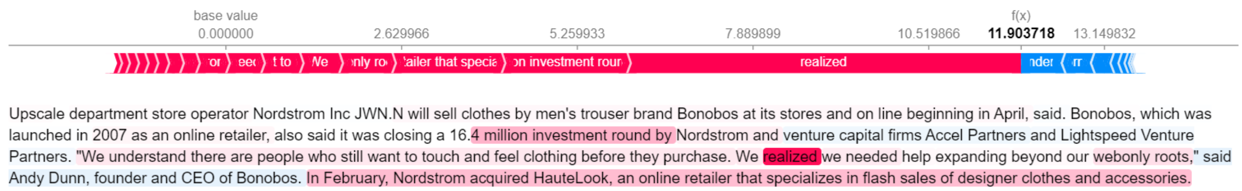


- *2005-10-19 19:07:02.407000, eBay (EBAY)* eBay today issued the following statement regarding Taobaoś pricing challenge: ”Free” is not a business model. It speaks volumes about the strength of eBayś business in China that Taobao today announced that it is unable to charge for its products for the next three years. Weŕe very proud that eBay is creating a sustainable business in China, while providing Chinese consumers and entrepreneurs with the safest, most professional, and most exciting global trading environment available today.



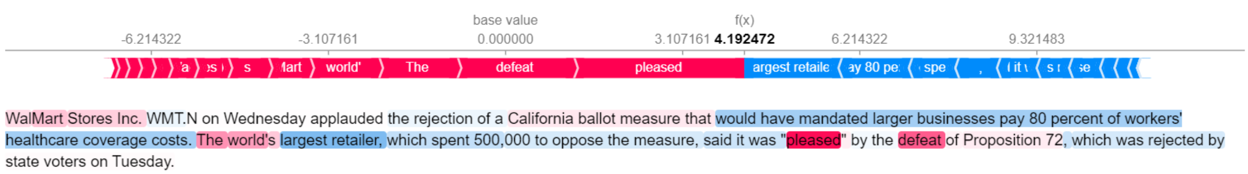
Realization

- *2012-04-12 11:52:17.925000, Nordstrom (JWN)* Upscale department store operator Nordstrom Inc JWN.N will sell clothes by menś trouser brand Bonobos at its stores and on line beginning in April, said. Bonobos, which was launched in 2007 as an online retailer, also said it was closing a $16.4 million investment round by Nordstrom and venture capital firms Accel Partners and Lightspeed Venture Partners. ”We understand there are people who still want to touch and feel clothing before they purchase. We realized we needed help expanding beyond our webonly roots,” said Andy Dunn, founder and CEO of Bonobos. In February, Nordstrom acquired HauteLook, an online retailer that specializes in flash sales of designer clothes and accessories.



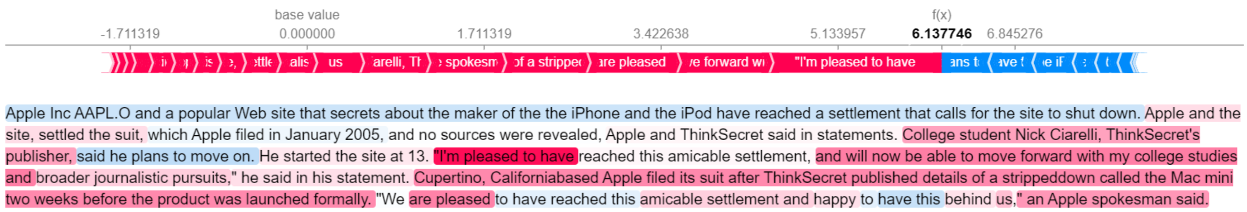
Relief

- *2004-11-03 13:01:16.888000, WalMart (WMT)* WalMart Stores Inc. WMT.N on Wednesday applauded the rejection of a California ballot measure that would have mandated larger businesses pay 80 percent of workers´healthcare coverage costs. The worldś largest retailer, which spent $500,000 to oppose the measure, said it was ”pleased” by the defeat of Proposition 72, which was rejected by state voters on Tuesday.



- *2007-12-20 13:26:53.361000, Apple Inc (AAPL)* Apple Inc AAPL.O and a popular Web site that secrets about the maker of the the iPhone and the iPod have reached a settlement that calls for the site to shut down. Apple and the site, settled the suit, which Apple filed in January 2005, and no sources were revealed, Apple and ThinkSecret said in statements.

College student Nick Ciarelli, ThinkSecretś publisher, said he plans to move on. He started the site at 13. ”Iḿ pleased to have reached this amicable settlement, and will now be able to move forward with my college studies and broader journalistic pursuits,” he said in his statement. Cupertino, Californiabased Apple filed its suit after ThinkSecret published details of a strippeddown called the Mac mini two weeks before the product was launched formally. ”We are pleased to have reached this amicable settlement and happy to have this behind us,” an Apple spokesman said.



Remorse

- *2018-04-26 07:10:08.016000, Facebook (FB)* Facebook did not intend its behaviour towards the media to be interpreted as trying to stop the truth about a data scandal out, chief technology officer said on Thursday. Asked by a lawmaker on a British whether Facebook would apologise for its ”bullying” behaviour towards the press, Mike Schroepfer said: ”I am sorry that journalists feel that we are trying to prevent them from getting the truth out. ”That is not the intent, so I’m sorry,” he added.

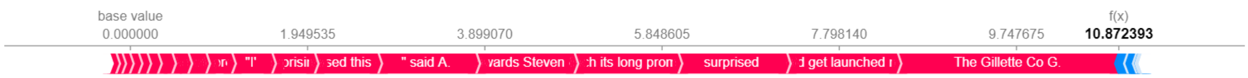
- *2009-11-17 15:41:46.563000, Goldman Sachs Group Inc (GS)* NEW YORK, Nov 17 (Reuters) Goldman Sachs Group Inc GS.N chief executive Lloyd Blankfein said his firm ”participated in things that were clearly wrong” in the leadup to the financial crisis, Bloomberg News reported on Tuesday. Blankfein apologized during a conference in New York hosted by Directorship magazine, Bloomberg said. ”We participated in things that were clearly wrong and have reason to regret," Blankfein said. "We apologize." As a result of the crisis, Goldman Sachs received billions of dollars in bailouts from taxpayers. Goldman Sachs has repaid the $10billionitborrowedfromtheU.S.governmentandhasreportedmorethan3$ billion in profits during each of the past two quarters. Its quick turnaround and potential for outsized bonuses so soon after the crisis have brought a public relations problem for Goldman.

Sadness

- *2011-04-14 15:33:17.465000, Dow Chemical (DOW)* NEW YORK, April 14 (Reuters) A Dow Chemical employee DOW.N died on Thursday after falling inside Midland, Michigan Dow said in a statement. Dow, the largest chemical maker in the United States, declined to release the manś name until his family could be notified. The last fatality at Dowś Michigan was in 1998. "We are greatly saddened by the passing of one of our own," said Earl Shipp, vice president of Dowś Michigan Operations. "This is an extremely difficult time for us, and our deepest sympathies, prayers and support go out to the family and friends." Dow said the worker fell from an "elevated location", but the Midland Daily News reported that he fell 30 feet from a catwalk, citing information from police scanners. Another local newspaper, the Bay City Times, reported that Michigań Occupational Safety and Health Administration had begun an investigation into the death.

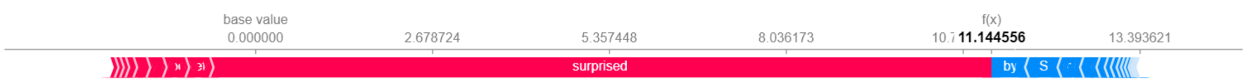
Surprise

- *1998-01-12 16:06:34.314000, Gillette (G)* The Gillette Co G.N said Monday it would launch its long promised mens̈ razor on July 1, surprising analysts who expected the new product to be released this winter. ”Im̈ a bit surprised,” said A.G. Edwards Steven East. ”I thought it would get launched right about now.” The Boston based consumer products group told a conference in Miami that it would launch its new product in North America and other selected markets on July 1. A spokesman for said he could not recall Gillette announcing a specific release date for a new product before.



The Gillette Co G.N said Monday it would launch its long promised men's razor on July 1, surprising analysts who expected the new product to be released this winter. "I'm a bit surprised," said A.G. Edwards Steven East. "I thought it would get launched right about now." The Bostonbased consumer products group told a conference in Miami that it would launch its new product in North America and other selected markets on July 1. A spokesman for said he could not recall Gillette announcing a specific release date for a new product before.

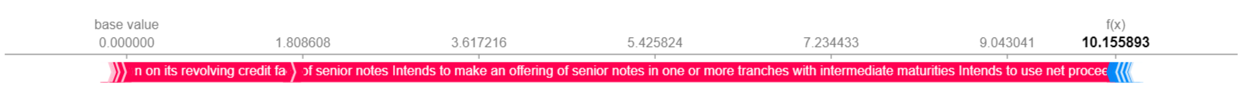
- *1998-06-19 01:59:31.161000, Boeing (BA)* TOKYO, June 19 (Reuters) Ron Woodard, president of Boeing Co’s airplane group, said on Friday he was very pleased and pleasantly surprised by joint U.S.Japanese intervention in the foreign exchange markets to bolster the yen.



TOKYO, June 19 (Reuters) Ron Woodard, president of Boeing Co's airplane group, said on Friday he was very pleased and pleasantly surprised by joint U.S.Japanese intervention in the foreign exchange markets to bolster the yen.

Neutral

- *2016-08-11 09:34:40.124000, Fidelity National Information Services Inc (FIS)* Aug 11 (Reuters) Fidelity National Information Services Inc FIS.N FIS announces proposed offering of senior notes Intends to make an offering of senior notes in one or more tranches with intermediate maturities Intends to use net proceeds to repay all or portion of about \$2.2 billion on its revolving credit facility.



Aug 11 (Reuters) Fidelity National Information Services Inc FIS.N FIS announces proposed offering of senior notes Intends to make an offering of senior notes in one or more tranches with intermediate maturities Intends to use net proceeds to repay all or portion of about 2.2 billion on its revolving credit facility.

- *2013-09-16 16:43:50.428000, JPMorgan Chase Co (JPM)* Sept 16 (Reuters) JPMorgan Chase Co JPM.N: Announces doddfrank midyear stress test results Results for severely adverse scenario do not incorporate feedback received from federal reserve bank in April 2013 Under severely adverse scenario,minimum stressed ratio of 8.5%, tier 1 capital ratio of 9.6%,tier 1 leverage ratio of 5.6% Says caculated 9quarter cumulative preprovision net revenue of $57 billion; provisions of $36.9 billion Says calculated 9quarter umulative loan losses of $32.1 billion 2013 yearend ccar will reflect enhanced forecasting methodologies and processes in response to frb feedback received in Q2 Source text coverage JPM.N