Enhanced Conditional Random Field Models for Cause and Effect Detection in Financial Documents

Satyam Agrawal, Ashitosh Phadatare
Artificial Intelligence
Department of Information Technology
National Institute of Technology Karnataka, Surathkal
satyamagrawal.211ee251@nitk.edu.in, ashitoshphadatare.211ai007@nitk.edu.in
Anand Kumar Madasamy
Department of Information Technology
National Institute of Technology Karnataka, Surathkal
m_anandkumar@nitk.edu.in

***Abstract*—This paper addresses the Financial Document Causality Detection Task (FinCausal-2023), aiming to uncover the intricate causal relationships embedded in financial texts. The methodology proposed incorporates diverse natural language processing techniques, including Word2Vec embeddings, BERT embeddings, contextual encoding with BERT, token classification using SVM, and Conditional Random Fields (CRF). The study proposes a novel method to enhance the performance of CRFs using features created from a contextual based classification model and compares the same with the SOTA methods. Evaluation metrics, including precision, recall, F1-score, and an exact match percentage, assess the effectiveness of the proposed methodologies.**

The literature review section provides insights into previous work in financial causality detection, coveringshared tasks, and models such as sequence labeling, etc. The paper concludes with a presentation of results and a discussion of their implications, contributing to the ongoing discourse on causality detection in financial narratives.

I. Introduction

In the field of Finance, the amount of textual data is huge which poses syntactic and semantic problems while processing the data for any task. Amidst this pool of large data is another feature that is found very commonly in data from the field of finance, this feature is the element of Causality present. Any decision making task or process has to take into account the complex structure of dependency present in the data in order to provide results that are meaningful and make sense. In order to do that, it is necessary to extract the parts of the sentences that represent Cause and the related Effect, which is the prime goal of this paper.

This paper presents a set of solutions for the Financial Document Causality Detection Task (FinCausal-2023), the main goal of which was again to identify the causal relationships present in the sentences that were taken from financial documents in order to determine what are the changes that take place in the market (effect) along with what causes those changes (cause).

The paper presents multiple approaches that can be applied to the problem, ranging from modification in the embeddings used, to the way that they are being used to represent the sentences. Advanced methods such as the usage of DistilBERT is being compared to the more traditional but feature enhanced methods such as Conditional Random Fields. Baseline methods using SVM along with multiple form of embeddings have also been established. Metrics such as F1 score along with exact match percentage has been used to compare these approaches.

The goal of the research is to provide the reader with an overview of methods that are suitable for the task of Causality detection along with the methods that are not, while also showcasing the utility of the tradition CRF when annotated with features relevant to the data. The subsequent paper is structured as follows, Section II lists the literature survey followed by Methodology employed in Section III. Section IV of the paper deals with the results and analysis of the approaches, ending with a note concluding the paper as well as highlighting the future areas of research in Section V.

II. Literature Review

NTUNLPL Pei-Wei Kao et al. uses Stanford Core NLP stanza for tokenizing and POS tagging which was used to solve two sub tasks ie. causal meanings detection and detection of cause and effect from the given text. BIO scheme was also used to label the causal events in the sentences. The paper compares two Baseline models for the detection task, a conditional Random Fields(CRF) and a BERT model. CRF uses neighbouring labels during calculation and is popular for sequential labelling. POS information as a one hot vector was added as an output to the BERT(base) model and concatenated it with the final layer. Paper uses Viterbi decoder, an algorithm that finds the most probable sequence of labels, which is used on top of the fine tuned BERT classifier. Output of the linear classifier used emission matrix and pre-defined transition matrix based on BIO scheme. This improved the performance for exact match ratio.

NER using CRF proposed by Nita Patil et al. introduces a new CRF based NER method for Urdu language. They used a novel dataset that had NE types manually annotated using IOBES encoding scheme. It used feature templates that used language dependent and language independent POS and context window of words. They reported the proposed methodology outperforms the baseline model. The paper explains the theoretical background of CRF and while implementation they used Mallet toolkit. The paper uses future directions for improving the performance of NER tags for URDU language such as using gazzeteers, stemming, lemmatization and neural networks.

Xianchao Wu et al. proposed a system for span based extraction of causality from finanacial news document. Examines pretrained models like ALBERT-xxlarge, BERT-large, and RoBERTa-large models fine-tuned under SQuAD2.0, a question answering dataset. For combining n-best predictions from five checkpoints, the paper uses a grid based ensemble learning algorithm. This gave considerable amount of improvement in F1 score and exact match score. Relates with other paper which used fine tuning based system architecture using projection network and layer of CRF on top of it. The paper also introduces some methods for ensemble learning like majority voting, weighted voting, and stacking.

The paper "Economic Causal-Chain Search and Economic Indicator Prediction using Textual Data" by Kiyoshi Izumi et al. uses causal information extraction for predicting economical indicators. The model automatically extracts causal information using ML and NLP techniques based on economic causality dataset. The paper word similarity between result and cause expression to generate a chain of causal words. These chains are used to predict how economical indicators will change in the future to spillover effects mentioned in the paper. The paper proposes an algorithms that generates causal sequence that were derived from phrases representing specific events and economic indicators related to spillover effects.

III. Methodolgy

The dataset is taken from FinCausal-2023 Shared Task: "Financial Document Causality Detection". The data consisted of approx. 2700 sentences, each with 3 columns, namely the text, cause and effect. Cause and Effect were contiguous pieces of sentences from the text data. The problem was initially framed as that of a classification problem, where the target values were the word locations where the Cause and Effect parts were starting from and ending at. An elementary example will illustrate the point made in a better manner:

Text: The storm caused the wall to break.
Cause: The storm (start: 0, end: 2)
Effect: The wall to break (start: 4, end: 6)

A. Word2Vec

Pretrained embeddings from "word2vec-google-news-300" were used to create features for each individual word after the initial preprocessing that converted sentence into word tokens, removing the special characters and converting all into lowercase. To convert individual word tokens into sentence representations, multiple approaches were employed which are mentioned in the part that follows.

1) **Concatenation**
In order to represent the sentences from the 'text' column, the first approach was that of concatenating the word embeddings to represent the sentences. The number of tokens used were fixed at 80 after looking at a distribution graph in order to retain maximum number of sentences. Padding using NaN was carried out to create a uniform length sentence vector for each of the examples.

2) **Concatenation with Causal Conjunctions**
To represent the sentences from the 'text' column, the second approach was that of concatenating the word embeddings to represent the sentences along with a special focus on the causal conjunctions, the most common of which were [”because”, ”since”, ”as”, ”for”, ”given”, ”seeing”, ”owing”, ”due”, ”thanks”, ”therefore”, ”thus”, ”consequently”, ”hence”, ”so”, ”accordingly”, "reason”].
The rest of the method was similar to that mentioned in 1A. To model the special focus on the causal conjunctions, the embeddings for these conjunctions were assigned a higher weightage for the model to be able to focus more on these particular words.

3) **TF-IDF**
In this method for representing sentences, TF-IDF was used to assign weights to individual words and then take a weighted mean of the individual tokens to produce a sentence representation. Term Frequency - Inverse Document Frequency was used to calculate these weights that were used to combine the words to get the final embedding, with an aim to better capture the important tokens that might be crucial for the performance of the models trained.

4) **TF-IDF with Causal Conjunctions**
In this method for representing sentences, along with TF-IDF to assign weights to individual words, causal conjunction mentioned earlier were also taken into account. Term Frequency - Inverse Document Frequency was used to calculate these weights which were then modified on the basis of whether or not the word was a causal conjunction, giving more weight in case it was. This was done to combine the efficacy of both the weight assignment methods and provide better embeddings.

B. Bert Embeddings

The textual dataset is preprocessed using the methodology, which includes lower-casing, removing non-alphanumeric characters, and tokenizing text using the BERT tokenizer. The task is modelled in the same way as it was mentioned previously, that is in the form of a classification task based on start and end positions of Cause and Effect.

1) **Mean**
On obtaining the Bert representation for each token from the ‘bert-base-uncased’ model after pruning the token limit to 80 due to reasons mentioned earlier, the obtained embeddings were used to represent the sentence. The individual embeddings were summed up and a mean was taken in order to normalise the values.

2) **Mean with Causal Conjunction**
causation_conjunctions = [”because”, ”since”, ”as”, ”for”, ”given”, ”seeing”, ”owing”, ”due”, ”thanks”, ”therefore”, ”thus”, ”consequently”, ”hence”, ”so”, ”accordingly”, “reason”] were used along with BERT embeddings (the ”bert-base-uncased” model) to tokenize the text. Tokens connected to causal conjunctions in the given list are given higher weights. The computation of a mean embedding for every sentence is done using this weighted representation, which emphasizes the significance of words related to causal relationships.

3) **TF-IDF**
Using BERT embeddings (the ”bert-base-uncased” model), this technique tokenizes the text and seeks to locate the positions of the ”Cause” and ”Effect” phrases within the text data. The weighted-mean embedding for every sentence is calculated using this representation. It uses TF-IDF vectorization to provide the weights to calculate the weighted mean in order to better capture the importance of words present in the sentence.

4) **TF-IDF with Causal Conjunctions**
After tokenizing the text into tokens using BERT embeddings (also known as the ”bert-base-uncased” model), this method connects the TF-IDF approach to causal conjunctions by giving causal conjunctions of the list previously mentioned, higher weights. The computation of a mean embedding for every sentence is done using this weighted representation, which emphasizes the significance of words related to causal relationships. .

For the subsequent sections, the problem was remodelled into a token classification rather than that of Start and End. In order to do so, the BIO tagging scheme was utilised which marks the Beginning, Interior and the Ending of the Cause and Effect part of the text. Each word was given one of these 7 labels, which are beginning of cause (BC), interior cause (IC), end of cause (EC), beginning of effect (BE), interior effect (IE), end of effect (EE), or marked as ”O” if it does not fall into either cause or effect category. In the subsequent sections, this remodelled problem is then approached in two ways, one based on that of Token Classification using Transformers, the other based on the probabilistic Conditional Random Fields.

C. *DistilBERT based Token Classification*

In the data preprocessing phase, we took several steps to ensure the quality and coherence of our dataset. Duplicates and null values were systematically removed, creating a cleaner and more reliable foundation for subsequent model training.

The Text, Cause, and Effect columns underwent thorough tokenization and cleaning processes, enhancing the model’s ability to grasp the underlying patterns within the textual data.

To facilitate the identification of cause and effect relationships, a BIO tagging scheme was implemented. This scheme classified words in the text as the beginning (BC), interior (IC), or end (EC) of causes and effects, resulting in structured BIO sequences for each sentence in the dataset. This approach aimed to provide the model with explicit labels for understanding the contextual nuances of causal relationships. DistilBERT, a state-of-the-art transformer-based model, was employed for tokenization. This process converted tokenized words into input sequences suitable for training the model. The core of our methodology involved the utilization of a DistilBERT-based Token Classification model for training. Fine-tuning was executed with a train-validation split, optimizing the model for token classification accuracy. The binary cross-entropy loss function was employed to effectively train the model for multi-label token classification, considering the intricate relationships inherent in cause and effect tagging.

D. *Conditional Random Fields - CRF*

First, a sourced textual dataset is used with Conditional Random Fields (CRFs) in our methodology to address the task of cause and effect identification. Because they are probabilistic models skilled at structured prediction, CRFs are appropriate for sequence labeling tasks. Finding cause and effect connections in a given text is the aim of this project. We used the BIO tagging scheme, which labels each word as either a beginning of cause (BC), interior cause (IC), end of cause (EC), beginning of effect (BE), interior effect (IE), end of effect (EE), or marked as ”O” if it does not fall into either cause or effect category, to turn this task into a token classification problem.

1) **CRF without Causal Conjunctions**
We used feature engineering to make the training of a Conditional Random Fields (CRF) model easier. We used a set of features to represent every word in a sentence, such as:
• word[-3:]: The word’s final three characters.
• word[-2:]: The word’s final two characters.
• word: The actual word.
• word.isdigit(): A binary characteristic that indicates whether or not a word has digits.
• postag: The word’s part-of-speech tag.
Furthermore, we included contextual features:
• -1:postag: The preceding word’s part-of-speech tag.
• +1:postag: The next word’s part-of-speech tag.
The features were created to capture contextual information from nearby words as well as the inherent qualities of individual words. After this a conditional random field was trained for the data and results noted.

2) **CRF with Causal Conjunction** :
We included causal conjunctions (CC) as a critical component in our methodology to improve the model’s

comprehension of cause and effect relationships.

causal conjunctions = ['since', 'as', 'for','so', 'therefore', 'thus', 'consequently', 'hence', 'accordingly', 'henceforth','sooner', 'due', 'henceforward', 'whence', 'then', 'wherefore', 'ergo','subsequently','resulting','resultantly']

C0: A feature that is binary and indicates if a word is a causal conjunction or not.
C+1: A binary characteristic that indicates whether a causal conjunction comes next.
C-1: A binary characteristic that indicates if the word before it was a causal conjunction or not.were created to record contextual information from nearby words as well as the inherent qualities of individual words.

CC on Current: This variation entails determining whether the word at hand is a causal conjunction, giving the model instant access to information about possible indicators of cause and effect.
CC on Current and Next: This variation captures circumstances where a causal conjunction in the current word is followed by additional context that could emphasize its significance by expanding the context to include the next word.
CC on Prev and Current: The model can identify patterns where a conjunction might continue a causal relationship by looking at the conjunction's presence in both the previous and current words.

3) **CRF with Feature Engineering**
The novelty of the paper lies in this particular approach that was employed. An SVM trained on contextual nature of the data was leveraged to generate features for the CRF model, greatly enhancing its performance as we will see in the results section.
The SVM was trained on the first 500 instances of that data in a way so that it could capture the contextuality of the dataset with respect to the BIO tag that it gets. BERT embeddings for the current word, along with previous and next words were concatenated to form the feature set. The previous and next embedding were added to provide a context for the SVM model to exploit. The target value was set as the BIO tag for the current word and the model was trained. The final model was able to predict the BIO tag for a given word, represented as a concatenation of previous, current and next word embeddings. The said model could not be used independently for token classification due to a lack of continuous sequence which is required in the problem mentioned. Notwithstanding this drawback, our method makes use of BERT embeddings to capture complex contextual dependencies, offering a useful viewpoint on cause and effect relationships in the provided text. And this is what we use to generate the features for our CRF model.
We incorporated the output from the BERT-based SVM model into our Conditional Random Fields (CRF) model in order to solve the tag discontinuity problem. Predictions covering the probabilities for every BIO tag were produced by the trained SVM model and used as extra features for the CRF model. At each token position in the sequence, the corresponding features, such as "O," "BC," "IC," "EC," "BE," "IE," and "EE," were given probabilities. These 7 probabilities were now added as features for the CRF which added contextual information to the CRF model, improving its ability to capture the subtleties of cause-and-effect relationships. In addition to conventional features like words, word suffixes, and part-of-speech tags, the augmented feature set now contains the probabilities determined by the SVM model. One should note that the 500 sentences used to train the SVM were not used while training the CRF model in order to escape any bias that may creep in due to the same.

E. Evaluation Metrics

1) Cause Identification: To evaluate the effectiveness of the model in determining causes, we utilized precision, recall, and F1-score metrics. These metrics were computed through a comparison of the actual and predicted causes. In particular, we took into account the following:

- **Precision:** A measure of the accuracy of identified causes, expressed as the ratio of true positive causes to all predicted causes.
- **Recall:** A measure of the model's capacity to account for all pertinent causes, expressed as the ratio of true positive causes to all actual causes.
- **F1 Score:** A balanced indicator of the model's overall performance in cause identification, calculated as the harmonic mean of precision and recall.
- **Exact Match Percentage:** This indicates the proportion of cases in which the actual and predicted causes matched exactly, providing information about how accurate the model is at identifying causes.

2) Effect Identification: In the same way, we used the same set of evaluation metrics for effect identification. The assessment procedure contrasted expected and real effects, taking into account:

- **Precision:** Measuring the ratio of actual positive effects to all predicted effects to reflect the accuracy of identified effects.
- **Recall:** Determining the ratio of true positive effects to the total number of actual effects to evaluate the model's capacity to capture all pertinent effects.
- **F1 Score:** Using the harmonic mean of precision and recall, this score offers a fair assessment of the model's overall effectiveness in effect identification.

- **Exact Match Percentage:** This measure illustrates the proportion of cases in which the actual and predicted effects matched exactly, providing insight into the accuracy of effect identification.

IV. Results

The first and foremost result from this analytical study was the inefficiency of Concatenation or Mean based embeddings in capturing the positional information for the sentences while performing a classification task. The Bert embeddings used by taking the mean perform extremely well when it comes to the Precision metric while giving a poor performance in the rest of the metrics. Notably, the percentage of exact match obtained was equal to 0 for methods using Bert Embeddings, be it the TF-IDF or the Causal Conjunctions version. The high precision and low recall observed in the BERT embeddings model can be attributed to its inclination towards minimizing false positives at the expense of higher false negatives. This behavior suggests that the model is conservative in predicting positive instances, being highly accurate when it asserts a cause or effect relationship. However, the downside is that it tends to miss several instances of actual causes or effects, resulting in a substantial number of false negatives. This approach, while reducing the risk of making incorrect positive predictions, leads to an incomplete identification of the true positive instances, ultimately causing the observed low recall.

Model	Cause Precision	Cause Recall	Cause F1	Cause Exact-Match
BERT - Mean	0.9205	0.4751	0.6267	0.2247
BERT - Mean - CC	0.9266	0.3964	0.5552	0.2247
BERT - TFIDF - CC	0.9266	0.3964	0.5552	0.2247
BERT - TFIDF	0.9138	0.5388	0.6779	0.0

TABLE I
RESULTS FOR BERT - CAUSE

Model	Effect Precision	Effect Recall	Effect F1	Effect Exact-Match
BERT - Mean	0.9161	0.4062	0.5629	0.0
BERT - Mean - CC	0.9161	0.4062	0.5629	0.0
BERT - TFIDF - CC	0.9161	0.4062	0.5629	0.0
BERT - TFIDF	0.9161	0.4062	0.5629	0.0

TABLE II
RESULTS FOR BERT - EFFECT

The models that use Word2Vec embeddings did not perform well due to the lack of positional information present in the representations used. TF-IDF and Mean do not lead to a representation with sufficient positional information of the words. Even the inclusion of Positional Encodings did not help improve the result forcing us to look for a better alternative. The performance score after these initial attempts forced us to reconsider the problem and reframe it as a token classification problem based on BIO tagging.

The DistilBERT model applied to token classification for BIO tags yielded promising results in cause-effect identification. The metrics indicate a relatively high precision, implying that when the model predicts a cause or effect, it is accurate the majority of the time. However, the recall and exact match percentages are comparatively lower. The model might struggle with identifying the entire set of actual effects, leading to a lower recall. These results highlight the efficacy of the DistilBERT model in accurately classifying and distinguishing between different components of the cause-effect relationships within the given text.

Model	Cause Precision	Cause Recall	Cause F1	Cause Exact Match
Word2Vec - TFIDF	0.5306	0.6728	0.5933	2.9213
Word2Vec - TFIDF - CC	0.5447	0.5679	0.5561	2.6966
Word2Vec - Concat	0.5560	0.4737	0.5116	2.2471
Word2Vec - Concat - CC	0.5569	0.4371	0.4898	1.7977

TABLE III
RESULTS FOR WORD2VEC - CAUSE

Model	Effect Precision	Effect Recall	Effect F1	Effect Exact Match
Word2Vec - TFIDF	0.4874	0.4687	0.4779	2.4719
Word2Vec - TFIDF - CC	0.4884	0.5172	0.5024	2.0224
Word2Vec - Concat	0.4884	0.5172	0.5024	2.0224
Word2Vec - Concat - CC	0.4982	0.7417	0.5961	0.4494

TABLE IV
RESULTS FOR WORD2VEC - EFFECT

The CRF models exhibited a notable improvement in both Exact Match and overall F1 score compared to other methods in the token classification task, particularly in the context of BIO tagging. This enhancement suggests that the Conditional Random Fields (CRF) approach effectively captures the sequential dependencies in the data, enabling a more accurate prediction of the token labels. By considering the relationships between neighboring tokens in the BIO tagging schema, the CRF models demonstrate a superior ability to achieve both precision and recall, resulting in a substantial boost in Exact Match and F1 score metrics. This underscores the efficacy of CRF models for token-level classification tasks, such as identifying causes and effects in text. Further attempt to improve the CRF model was done by using Causal Conjunctions by adding it to the feature. This information was binary in nature and as seen in the table, CRF with CC in any of the 4 formats in general outperformed the CRF without CC. However, a much more detailed experimentation is required to analyse the true effect.

The novelty of this paper lies in the last and best performing model of all when it came to the Exact Match metric, which could arguably be defined as the most important metric of the four that are used. The CRF model used with feature engineering using probabilities obtained from SVM was able to attain the highest number of exact matches even though it was trained on the least amount of data. The probability features obtained using the pretrained SVM model on the first 500 data points seemed to have added to the efficiency of the CRF model by providing it with a token level classification information along with context.

Model	Cause Precision	Cause Recall	Cause F1	Cause Exact Match
CRF - CC Feature Engineering	0.6813	0.6933	0.6872	33.6986
DistilBERT	0.8102	0.8058	0.808	32.4712
CRF - CC on current and next word	0.5918	0.6744	0.6304	22.3655
CRF - CC on current word	0.5967	0.6573	0.6256	22.1505
CRF - CC on current and prev word	0.5896	0.6562	0.6211	21.7204
CRF - CC on current , prev and next	0.5935	0.6646	0.6271	21.5053
CRF without CC	0.5911	0.6328	0.6113	20.8602

TABLE V
RESULTS FOR CRF AND DISTILBERT - CAUSE

Model	Cause Precision	Cause Recall	Cause F1	Cause Exact Match
CRF - CC Feature Engineering	0.7286	0.6837	0.7055	30.1369
DistilBERT	0.8236	0.8609	0.8418	29.885
CRF - CC on current and next word	0.6821	0.5957	0.636	23.2258
CRF - CC on current word	0.6824	0.6131	0.6459	23.0107
CRF - CC on current and prev word	0.6846	0.5815	0.6289	22.5806
CRF - CC on current , prev and next	0.6676	0.6137	0.6395	22.5806
CRF without CC	0.6767	0.5941	0.6327	21.9354

TABLE VI
RESULTS FOR CRF AND DISTILBERT - EFFECT

V. CONCLUSION AND FUTURE WORK

In conclusion, the comprehensive analysis of cause and effect identification methods reveals diverse strategies for transforming textual data into meaningful numerical representations. While BERT embeddings exhibit high precision, they suffer from low recall, emphasizing the need for a balanced approach. The utilization of CRF models, especially when enhanced with feature engineering and contextual information from SVM models, proves effective in capturing sequential dependencies and achieving superior performance in token-level classification. The integration of DistilBERT for token classification offers promising results, demonstrating its potential in discerning cause and effect relationships. There is a lot of scope for further enhancement of the methods undertook in this study along with varied approaches. The ratio used to trained SVM for features, which was 500 is arbitrary and more experiment is required to find the optimised ratio for best performance. Further improvement can be made by experimenting with more advanced classification models to generate probabilities for the CRF model to work with and improve.

REFERENCES

[1] Antonio Moreno-Sandoval (UAM, Spain), Blanca Carbajo Coronado (UAM, Spain), Doaa Samy (UCM, Spain), Jordi Porta (UAM, Spain), Dominique Mariko (Yseop, France) FinCausal-2023 Shared Task: Financial Document Causality Detection, FNP 2023, Organized by lli-uam

[2] Dominique Mariko, Hanna Abi-Akl, Estelle Labidurie, Stephane Durfort, Hugues de Mazancourt, Mahmoud El-Haj2 - Financial Document Causality Detection Shared Task (FinCausal 2020)

[3] Hiroki Sakaji, Kiyoshi Izum - Financial Causality Extraction Based on Universal Dependencies and Clue Expression .

[4] Tapas Naik, Soumya Sharma, Yash Butala, Koustuv DasGupta, Pawan Goyal, Niloy Ganguly. - A Generative Approach for Financial Causality Extraction

[5] KSohom Ghosh, Sudip Kumar Naskar, Fidelity Investments, Jadavpur University- LIPI at FinCausal 2022: Mining Causes and Effects from Financial Text

[6] Guillaume Becquin guillaume.becquin@gmail.co - GBe at FinCausal 2020, Task 2: Span-based Causality Extraction for Financial Document

[7] Charles Sutton and Andrew McCallum (2012), "An Introduction to Conditional Random Fields", Foundations and Trends® in Machine Learning: Vol. 4: No. 4, pp 267-373.

[8] Nita Patil, Ajay S. Patil, and B. V. Pawar (2016) "Issues and Challenges in Marathi Named Entity Recognition" International Journal of Natural Language Computing (IJNLC), 5(6):15-31.

[9] Maithilee L. Patawar, M. A. Potey (2016) "Named Entity Recognition from Indian tweets using Conditional Random Fields based Approach", International Journal of Advanced Research in Computer Engineering Technology (IJARCET) 5(5): 1541-1545.

[10] Kao et al., NTUNLPL at FinCausal 2020, Task 2:Improving Causality Detection Using Viterbi Decoder https://aclanthology.org/2020.fnp-1.11 , FNP 2020

[11] Izumi et al., Economic Causal-Chain Search and Economic Indicator Prediction using Textual Data (https://aclanthology.org/2021.fnp-1.3)

[12] Victor Sanh and Lysandre Debut and Julien Chaumond and Thomas Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,2020

[13] Jacob Devlin and Ming-Wei Chang and Kenton Lee and Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019

[14] Tomas Mikolov and Kai Chen and Greg Corrado and Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space, 2013

[15] Xianchao Wu,NVJPFSI at FinCausal 2021 Span-based Causality Extraction Task

[16] Gunjan Haldar et al. DSC-IITISM at FinCausal 2021: Combining POS tagging with Attention-based Contextual Representations for Identifying Causal Relationships in Financial Documents.

[17] Ziwei Xu et al. iLab at FinCausal 2022: Enhancing Causality Detection with an External Cause-Effect

[18] Satyam Kumar et al. Causal Inference for Banking, Finance, and Insurance