Togedemaru at SemEval-2023 Task 8: Causal Medical Claim Identification and Extraction from Social Media Posts

Andra-Maria Oică
Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi, Romania
andra.oica@gmail.com

Daniela Gîfu
Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi, Romania
Institute of Computer Science, Romanian Academy - Iasi Branch
daniela.gifu@iit.academiaromana-is.ro

Diana Trandăbăț
Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi, Romania
dtrandabat@info.uaic.ro

Abstract

The "Causal Medical Claim Identification and Extraction from Social Media Posts" task at SemEval 2023 competition focuses on identifying and validating medical claims in English, by posing two subtasks on causal claim identification and PIO (Population, Intervention, Outcome) frame extraction. In the context of SemEval, we present a method for sentence classification in four categories (*claim*, *experience*, *experience_based_claim* or *a question*) based on BioBERT model with a MLP layer. The website from which the dataset was gathered, Reddit, is a social news and content discussion site. The evaluation results show the effectiveness of the solution of this study (83.68%).

1 Introduction

One of the most brainstorming tasks facing natural language processing (NLP) is the information extraction (Khurana, D. et al., 2023; Landosi, M.Y. et al., 2023) with important applicability in the medical field (Zhang, T. et al., 2021). In fact, an essential step for various medical decision-making processes is the identification and automatic verification of specific claims from unstructured user-generated text data (Doppalaudi, S. et al., 2022). It refers to the idea of cause and effect. The causal relation extraction has been mostly treated as a subtask of relation extraction, being helpful in the context of personalized healthcare (Khetan, V. et al., 2022). Extracted causal information from clinical notes, as a crucial step within medical decision-making, can and have to be combined (Yang, J. et al., 2022). The aim of this paper is focused on identification of causal claims[1], experience, etc., in a provided multi (or single) sentence text snippet. The rest of the paper is organized as follows: section 2 briefly presents studies related to claim identification, section 3 provides information about the system designed to classify text sentences, section 4 describes the experimental setups. Section 5 resumes the results of the conducted experiments, with their interpretations, followed by section 6 with the conclusions.

2 Background

This topic has attracted significant attention in recent years, evidenced by increasing number of workshops (e.g., Workshop on Curative Power of MEdical Data - MEDA 2017; 2018; 2020, Workshop on Events and Stories in the News 2018) (Cohen, K.B et al., 2020., Gifu, D. et al., 2019, Tommaso, C. et al., 2018).

Competitions such as SemEval-2023 Task 8: Causal Medical Claim Identification and Extraction from Social Media Posts are attractive, especially since the problem of labeled data is somewhat solved, since the automatic identification and automatic verification of medical claims depends on them.

[1] SemEval-2023 Task 8 overview is available at: https://causalclaims.github.io/

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 913–921
July 13-14, 2023 ©2023 Association for Computational Linguistics

In order to automatically identify and validate the four main classes (claim, personal experience, question and claim by personal experience), in text snippets extracted from the Reddit platform (this representing the first task within the competition), we researched similar studies. In fact, there are many approaches based on NLP techniques trying to solve this (Sumner, P. et al., 2014, Mausam, 2016), while there is still a considerable number of challenges along the way.

Our solution implies the usage of BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text Mining). In fact, it is a BERT-based model pre-trained on large-scale biomedical corpora which outperforms BERT on biomedical named entity recognition (0.62% F1 score improvement), biomedical relation extraction (2.80% F1 score improvement) and biomedical question answering (12.24% MRR improvement). Starting from an existing model-generating code base[2] (Yu, B. et al., 2019), changes[3] have been brought in the label naming and implementations have been added in the direction of preprocessing and postprocessing of the datasets of interest.

SemEval-2023 Task 8 implies that for a provided user-generated English-language text, it is requested to identify the span of text that is either a claim, personal experience, personal experience-based claim, or a question (Figure 1.1, Figure 1.2).

The dataset that this competition provided us with consists of 7121 Reddit posts. As input, the post IDs and specific annotations are given, to be further used in constructing the final datasets.

Social media Post	Claim(s)	Personal Experience(s)
Cytoxan and prednisone Rheumatologist says cellcept now I have developed lupus n messed up my hips so badly th replacedI dont want to get ba its to bring the inflammation	rheumatologist says its inflammation down in m the side effects sound id lupus flare.	Ive never been on Cytoxan
Can I just not eat food anymo I'm getting a lot of mixed info eat. One source tells me bean the other says they're terrible. One article says cherry juice h study says it does nothing.	beans and plant protei and the other says they article says cherry juice uric acid, another stud nothing.	

Figure 1.1 – Causal claim identification - The given input (social media post as a text snippet) and the requested classes to be identified for the first subtask: claim(s), personal experience(s) (…)

Question(s)	Claim(s) based on Personal Experience(s)
How am I supposed to be positive this?	Prednisone messed up my hips that they both need to be repla
Can I just not eat food anymore? ; to eat?	getting a lot of mixed informati what I can and can't eat

Figure 2.2 – Causal claim identification - question(s), claim(s) based on personal experience(s) (cont.)

An extracting script[4] shared by the task's organizers was used to obtain a Reddit collection of around 5696 posts (for training) and 1425 posts (for testing) which, after preprocessing and splitting by annotations, implied around 21652 sentences (for training) and 13567 sentences (for testing). The initial labels have been changed into integers in the interest of an easier manipulation. (0 – "claim", 1 – "per_exp", 2 – "question", 3 – "claim_per_exp"). As output, the predicted labels for a test dataset, consisting of sentences, were shared between the words of the text snippet of interest (Figure 2.3 - each word in the sentence occupying a different row, but sharing the same label as the whole sentence, as required by the task organizers).

0	**post_id, subreddit_id, stage1_labels** s1jpia,t5_2s23e,"[{""crowd-entity-annotation"":{""entities"":[{""endOffset"":858,""label"":""per_exp"",""startOffset"":661},{""endOffset"":2213,""label"":""per_exp"",""startOffset"":1861}, …}}
1	**post_id, subreddit_id, stage1_labels, text** (Same as '0', but followed by the post text itself)
2	**sentence, label** It seems like so long ago now,1 but one morning I woke up and my left side wasnt responding as fast as my right side,1 It felt heavy,1

[2] BioBERT model-generating code base is available at: https://github.com/junwang4/causal-language-use-in-science

[3] The solution code repository is available at: https://github.com/dinosaph/SPLN_Togedemaru_Semeval_Task_8

[4] The script used to extract the Reddit posts: https://drive.google.com/file/d/10D5VKvdKcIJvtC47vE7IcQQl_2f9qvG4/view

3	**post_id, subreddit_id, words, labels** pwns5j,t5_2r876,Speeding,per_exp pwns5j,t5_2r876,fine,per_exp pwns5j,t5_2r876,I,per_exp pwns5j,t5_2r876,know,per_exp

Figure 2.3 – Dataset manipulation: 0 – Initial state, 1 – After extraction, 2 – After preprocessing, 3 – Results

Causal research can help improve existing medical decision-making processes. There are many studies focused on the causal relation extraction, leaving a nice starting point for further training, and testing of CSC tasks (Causal Sentence Classification; Tan, F.A. et al., 2021, Yu, B. Y et al., 2019) that provides a good base for BioBERT embeddings fed through one layer of MLP, which serves as a classifier. The solution described in this paper also describes the results obtained by choosing BioBERT as this model proved to outperform BERT in several medical related tasks which fits our task's purpose. Yu B and colleagues (Yu, B. Y et al., 2019) provided an open-source code that was used as a starting point in this development. It was required to also set up a sklearn[5] related repository dealing with imbalanced classes for BERT. (Pedregosa et al., 2011)

3 System overview

The strategy picked for this task represents a causal approach to BERT (BioBERT + MLP (Figure 3) as found in the architecture (Tan, F. A et al., 2021). The starting point in the development of this solution was the model-generating code made public by Yu, B., Li, Y. and Wang, J. (2019), on Github. The only significant changes were done to the labels, from 0 – "none", 1 – "causal", 2 – "cond", 3 – "corr" to 0 – "claim", 1 – "per_exp", 2 – "question", 3 – "claim_per_exp".

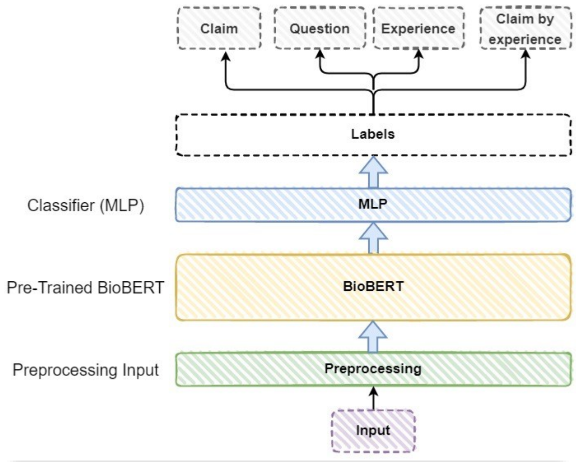

Figure 3 – Proposed architecture: Pre-trained BioBERT + MLP layer; Preprocessing class handling input development

The datasets were constructed from the given collection of post ids and annotations by automatic extraction from the Reddit platform, preprocessed and further used for training the BioBERT + MLP model. The training process took around 2.5 hours and it implied a combination of 5 folds and 5 epochs.

The subtask provided an extraction script, mentioned in the previous sections, to be used for extracting the required training social posts/text snippets. The given input files for training and testing were only files describing the path from which we had to extract the text by post id with preset Reddit credentials, process that took around 3 hours for the posts required by the task of interest (Figure 4). The purpose was to extract the text and match the substrings by the labels specified by ranges of "<startOffset>" and "<endOffset>" (Figure 5).

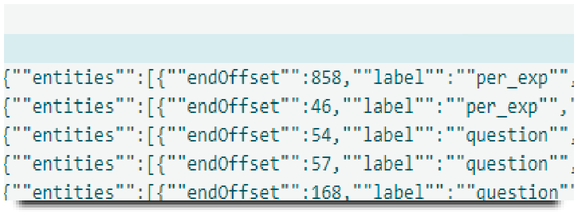

Figure 4 – Initial input with label range, <endOffset> - <startOffset>

[5] Sklearn library details are available at: https://scikit-learn.org/stable/

```
datasets > st1_train_inc_text.csv
post_id,subreddit_id,stage1_labels,text
s1jpia,t5_2s23e,"[{""crowd-entity-annotation"":{""entities"":[{""en
I wrote this a few years ago and just found it again.  I thought I'

&#x200B;

When I was 17, it was a very good year. Like the opening line of th

Growing up I was an only child with older parents, which is probabl

After several hours at the hospital, several days of blood tests, a

De-Nial is NOT just a river in Egypt.

I went home feeling a bit shocked, a lot overwhelmed, and A LOT of

Looking back I realize that, what the doctor said is probably true

I went on with my life as usual. Relapsing Remitting MS is a horrib

De-Nial is NOT just a river in Egypt.
```

Figure 5 – Input dataset after posts extraction

The resulted texts still needed to pass through a preprocessing class that would prepare them for the model training (Figure 6). Further development of the solution is shown in the implementation of such class that modifies the input dataset in the preferred way. The text has been split into sentences sharing the initial text snippet label, the unnecessary symbols and the empty spans were removed and finally, the labels were converted into integers, leaving a clean input dataset (Figure 7).

```
if __name__ == '__main__':

    file_train = PreprocessExpert.get_prep_reddit_train_df(r
    file_test = PreprocessExpert.get_prep_reddit_test_df(r'd
```

Figure 6 – Preprocessing actions

```
datasets > st1_train_inc_text_prep_out.csv
sentence,label
It seems like so long ago now,1
but one morning I woke up and my left side wasnt responding
It felt heavy,1
it was hard to wash my hair,1
it lasted a few hours,1
I went on with my life as usual,1
Relapsing Remitting MS is a horrible monster,1
it lulls you into a false sense of security,1
believing nothing is really wrong,1
My original symptoms had resolved and I went on as I always
```

Figure 7 – Preprocessed input dataset

```
pred > EMNLP_biobert_train > K5_epochs5.csv
c0,c1,c2,c3,confidence,winner,sentence,label
0.001,0.772,0.009,0.218,0.772,c1,It felt heavy,1
0.002,0.550,0.012,0.436,0.550,c1,it was hard to was
0.001,0.317,0.007,0.674,0.674,c3,it lasted a few ho
0.005,0.708,0.125,0.162,0.708,c1,believing nothing
0.004,0.455,0.029,0.511,0.511,c3,but overall,3
0.001,0.513,0.011,0.476,0.513,c1,bad vertigo,3
```

Figure 8 – Output (predictions) sample: 0 – “claim”, 1 – “per_exp”, 2 – “question”, 3 – “claim_per_exp”.

```
results > st1_pred.csv
post_id,subreddit_id,words,labels
pwns5j,t5_2r876,Speeding,per_exp
pwns5j,t5_2r876,fine,per_exp
pwns5j,t5_2r876,I,per_exp
pwns5j,t5_2r876,know,per_exp
```

Figure 9 – Postprocessed results, as required for the task's submission – words share the same label as the sentence.

A main class called “CausalExtractor” has been implemented which handles all model related actions and loads the pretrained model by initialization. This class provides a baseline algorithm description, but also uses our chosen BioBERT algorithm to perform the predictions on the required test data (Figure 8). One of its functions, “generate_submission_st1” further controls the predicted output and generates the required results format for the subtask (Figure 9). It splits the sentences into words, each word sharing the same label, while the labels are converted back into the initial string format.

4 Experimental setups

The datasets used in the development of this solution have gone through some changes, before the training of the model and after the generation of the predictions. The following tables describe the training dataset going through different changes, from its initial state to its preprocessed form:

INITIAL DATASET:

post_id, subreddit_id, stage1_labels
s1jpia,t5_2s23e,"[{""crowd-entity-annotation"":{""entities"":[{""endOffset"":858,""label"":""per_exp"",""startOffset"":661},{""endOffset"":2213,""label"":""per_exp"",""startOffset"":1861},{""endOffset"":2407,""label"":""per_exp"",""startOffset"":2255},{""endOffset"":3254,""label"":""claim_per_exp"",""startOffset"":2697},{""endOffset"":3620,""label"":""claim_per_exp"",""startOffset"":3294},{""endOffset"":3751,""label"":""claim_per_exp"",""startOffset"":3621},{""endOffset"":4480,""label"":""per_exp"",""startOffset"":3752},{""endOffset"":4759,""label"":""q

uestion"","startOffset"":4482}]}}]
"
(…)

INPUT DATASET AFTER EXTRACTION:

post_id, subreddit_id, stage1_labels, text
s1jpia,t5_2s23e,"[{""crowd-entity-annotation"":{""entities"":[{""endOffset"":858,""label"":""per_exp"",""startOffset"":661},{""endOffset"":2213,""label"":""per_exp"",""startOffset"":1861},{""endOffset"":2407,""label"":""per_exp"",""startOffset"":2255},{""endOffset"":3254,""label"":""claim_per_exp"",""startOffset"":2697},{""endOffset"":3620,""label"":""claim_per_exp"",""startOffset"":3294},{""endOffset"":3751,""label"":""claim_per_exp"",""startOffset"":3621},{""endOffset"":4480,""label"":""per_exp"",""startOffset"":3752},{""endOffset"":4759,""label"":""question"",""startOffset"":4482}]}}]","De-Nial "I wrote this a few years ago and just found it again. I thought I'd share... When I was 17, it was a very good year. Like the opening line of the old Frank Sanatra song, when I was 17, it truly was a very good year. I was getting ready to graduate high school, I had deeply bonded friends, I loved where I lived, and I had amazing parents. (…)"

PREPROCESSED DATASET:

sentence, label
It seems like so long ago now,1 but one morning I woke up and my left side wasnt responding as fast as my right side,1 It felt heavy,1 it was hard to wash my hair,1 it lasted a few hours,1 I went on with my life as usual,1 Relapsing Remitting MS is a horrible monster,1 it lulls you into a false sense of security,1 (…)

The provided code repository contains a class called "PreprocessExpert" that deals with the preprocessing of both training and testing datasets. Each function in this class serves the purpose of cleaning the texts, organizing the data in such a way that aids the training or the testing. The training data is split by label ranges ("<endOffset> - <startOffset>", such as in Figure 10), then split by sentence and finally filtered (keeping only sentences with more than 3 tokens/words) and cleaned (removing messy symbols, empty spaces). Finally, this dataset's columns consist of only "sentence" and "label".

Similar cleaning and splitting processes take place for the testing dataset, but this time the columns of interest remain "post_id", "subreddit_id" and "text" (Figure 11).

The main run of the solution is executed through Python by calling the main script, "main.py", which prepares the datasets (train and test), creates a "CausalExtractor" instance (previously described in Section 3) and generates scores and predictions as requested (Figure 12).

5 Results

In order to explore the efficacy of the chosen solution, tests have been run on the input data within a sklearn dummy classifier[6] through multiple modes which returned the accuracies shown in Figure 13.

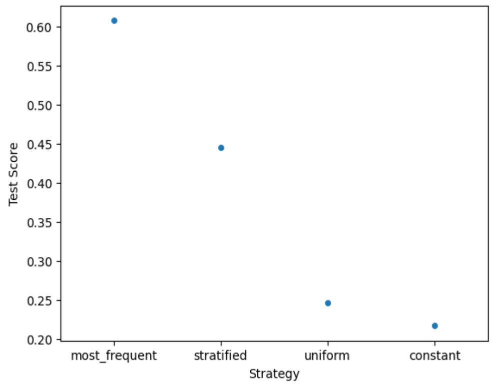

Figure 13 – Accuracies returned by the sklearn dummy classifier for different modes: ~0.61 ("most_frequent" mode), ~0.45 ("stratified" mode), ~0.25 ("uniform" mode), ~0.22 ("constant" mode)

[6] Sklearn dummy classifier details available at: https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html

By comparing the scores of our BioBERT model with the dummy classifier, the chosen model is a good classifier for the given data (Figure 14).

```
Testing: 100%|

Loss: 0.4778, Accuracy: 83.68%
83.68226600985221
```

Figure 14 – BioBERT model score overview – Accuracy: 83.68%

Due to choosing a model outperforming Bert in the medical tasks, the solution's model performs better than the classic baseline methods for the testing data, though improvements can be brought in the preprocessing stage and in studying the impact of specific tokens in the final classification. We managed to create a light solution that fulfills the need of automatically identifying and validating medical claims in social media posts. By studying the results, we can see that the best performing combination of k-fold on 5 epochs was reached at the 3rd epoch with an F1 accuracy score of around 0.899 (Figure 15).

BERT follows an interesting technique and due to its complexity, we might need to further review the way in which we have preprocessed the text and the edits made on the base code for BioBERT, as the results are not 100% correct and we are aware of the need of their improvement. The fulfillment of the next subtask is what interests us in the future. We will be posting updates on the task's repository on which we will reorganize our thoughts and ideas into a better development of the results for both subtasks.

6 Conclusion

For performing automatic span identification from textual documents (a Reddit collection) that is either a claim, experience, experience-based claim, or a question in an unstructured user-generated English text, we developed a BioBERT model with a MLP layer. While the solution for the subtask of interest received the 7th place out of 7 entries, the results show that it generated better scores than the classic baseline methods, so it provides a good start in this study, while being light and flexible for improvement. In future work, we would like to develop new approaches based on current rule-based systems with rules tailored to the linguistic features associated with medical claim.

References

Khurana, D., Koli, A., Khatter, K, Singh, S. 2023. Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl* 82, pages 3713–3744. https://doi.org/10.1007/s11042-022-13428-4.

Landolsi, M.Y., Hlaoua, L. & Ben Romdhane, L. 2023. Information extraction from electronic medical documents: state of the art and future research directions. *Knowl Inf Syst* 65, pages 463–516. https://doi.org/10.1007/s10115-022-01779-1.

Zhang, T., Lin, H., Tadesse, M.M. et al. Chinese medical relation extraction based on multi-hop self-attention mechanism. Int. J. Mach. Learn. & Cyber. 12, 355–363 (2021). https://doi.org/10.1007/s13042-020-01176-6.

Doppalapudi, S., Wang, T. and Qiu, R. 2022. Transforming unstructured digital clinical notes for improved health literacy, *Digital Transformation and Society*, Vol. 1 No. 1, pages 9-28. https://doi.org/10.1108/DTS-05-2022-0013.

Khetan, V., Rizvi, M.I., Huber, J., Bartusiak, P., Sacaleanu, B., and Fano, A. 2022. MIMICause: Representation and automatic extraction of causal relation types from clinical notes. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 764–773, Dublin, Ireland. Association for Computational Linguistics.

Yang, J., Wan, Y., Ni, Q., Zuo, J., Wang, J., Zhang, X., and Zhou, L. 2022. Quantifying causal effects from observed data using quasi-intervention. *BMC Med Inform Decis Mak* 22, 337 (2022). https://doi.org/10.1186/s12911-022-02086-z.

Cohen, K.B., Gîfu, D., Li, Y., Ripple, A. and Xia, J., 2020, August. MEDA 2020: The curative power of medical data. In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (pp. 575-576).

Gîfu, D., Trandabăț, D., Cohen, K. and Xia, J., 2019. Special Issue on the Curative Power of Medical Data. *Data*, *4*(2), p.85.

Tommaso, C., Miller, B., van Erp, M., Vossen, P., Palmer, M., Hovy, E., Mitamura, T., Caswell, D., Brown, S. W., Bonial, C. 2018. Proceedings of the Workshop Events and Stories in the News 2018, Association for Computational Linguistics, New Mexico, US.

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F., Chambers, C. D. 2014. The association between exaggeration in health-related science news and academic press releas-es: retrospective observational study. *BMJ*, 349, g7015.

Tan, F. A., Hazarika, D., Ng, S.-K., Poria, S. and Zimmermann, R. 2021. Causal Augmentation for Causal Sentence Classification. *Proceedings of CI+NLP: First Workshop on Causal Inference and NLP*, ACL, pages 1–20. https://doi.org/10.18653/v1/2021.cinlp-1.1.

Yu, B., Li, Y., and Wang, J. 2019. Detecting Causal Language Use in Science Findings*, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, and the 9th International Joint Conference on Natural Language Processing*, ACL, pages 4664–4674. https://doi.org/10.18653/v1/D19-1473.

Pedregosa et al. 2011. Scikit-learn: Machine Learning in Python, JMLR 12, *pp. 2825-2830*

A. Appendices

```python
for i in range(len(df)):
    labeled_sections = eval(df.iloc[i]['stage1_labels'])[0]['crowd-entity-annotation']['entities']
    for j in range(len(labeled_sections)):
        s = df.iloc[i]['text'][:labeled_sections[j]['startOffset']].rfind(' ')
        e = df.iloc[i]['text'][:labeled_sections[j]['endOffset']].rfind(' ')
        sentences = re.split('[,.?!]', df.iloc[i]['text'][s:e])
        sentences = [st for st in sentences if st != '']
        if len(sentences) > 0:
            for sentence in sentences:
                if len(sentence.split(' ')) > 2:
                    sentence_n_label.append({
                        'sentence': sentence.strip(),
                        'label': PreprocessExpert.LABELS_TO_INT[labeled_sections[j]['label']]
                    })
```

Figure 10 – “PreprocessExpert” class – “get_prep_reddit_train_df” function code snippet.

```python
for i in range(len(df)):
    current_post = df.iloc[i]
    post_sentences = re.split('[,.?!]', current_post['text'])
    post_sentences = [st for st in post_sentences if st != '']
    if len(post_sentences) > 0:
        for ps in post_sentences:
            if len(ps.strip()) > 0 and len(ps.split(' ')) > 3:
                data_n_sentences.append({
                    'post_id': current_post['post_id'],
                    'subreddit_id': current_post['subreddit_id'],
                    'text': ps.strip(),
                })

df_new = pd.DataFrame.from_records(data_n_sentences)
df_new['text'] = df_new['text'].progress_apply(PreprocessExpert.clean_text)
```

Figure 11 - “PreprocessExpert” class – “get_prep_reddit_test_df” function code snippet.

```python
# ==========================================================================================
# RUN / TEST
# ==========================================================================================

if __name__ == '__main__':

    file_train = PreprocessExpert.get_prep_reddit_train_df(r'datasets/st1_train_inc_text.csv')
    file_test = PreprocessExpert.get_prep_reddit_test_df(r'datasets/st1_test_inc_text.csv')

    ce = CausalExtractor()
    ce.get_baseline(file_train)
    ce.get_predictions_st1(file_test)
```

Figure 12 – Main solution run code snippet.

	Acc	weight	size	P	R	F1	P_0	R_0	F1_0	P_1	R_1	F1_1	P_2	R_2	F1_2	P_3	R_3	F1_3
0	0.725	0.2	4331	0.626	0.640	0.624	0.439	0.385	0.410	0.852	0.722	0.782	0.870	0.887	0.878	0.344	0.566	0.428
1	0.740	0.2	4330	0.650	0.627	0.629	0.512	0.338	0.407	0.842	0.761	0.799	0.897	0.881	0.889	0.350	0.527	0.421
2	0.741	0.2	4330	0.663	0.658	0.654	0.566	0.462	0.508	0.846	0.749	0.795	0.893	0.905	0.899	0.345	0.518	0.415
3	0.714	0.2	4330	0.604	0.605	0.598	0.385	0.323	0.351	0.831	0.731	0.778	0.876	0.854	0.865	0.325	0.513	0.398
4	0.730	0.2	4330	0.622	0.622	0.615	0.422	0.331	0.371	0.842	0.740	0.788	0.879	0.889	0.884	0.343	0.529	0.416
avg	0.730	0.2	4330	0.633	0.631	0.624	0.465	0.368	0.410	0.843	0.741	0.788	0.883	0.883	0.883	0.341	0.531	0.415

Figure 15 – Chosen solution accuracies.

```
sxy06a,t5_2saq9,until,per_exp
sxy06a,t5_2saq9,I,per_exp
sxy06a,t5_2saq9,tried,per_exp
sxy06a,t5_2saq9,with,per_exp
sxy06a,t5_2saq9,my,per_exp
sxy06a,t5_2saq9,pulse,per_exp
sxy06a,t5_2saq9,ox,per_exp
sxy06a,t5_2saq9,and,per_exp
sxy06a,t5_2saq9,got,per_exp
sxy06a,t5_2saq9,the,per_exp
sxy06a,t5_2saq9,same,per_exp
sxy06a,t5_2saq9,thing,per_exp
sxy06a,t5_2saq9,they,per_exp
sxy06a,t5_2saq9,always,per_exp
sxy06a,t5_2saq9,used,per_exp
```

Figure 16 – Task submission results sample