Framing in the Presence of Supporting Data: A Case Study in U.S. Economic News

Anonymous ACL submission

Abstract

The mainstream media has a lot of leeway in what it chooses to cover and how it covers it. These choices have real-world consequences on what people know, and subsequent key personal and societal behaviors. However, the lack of objective measures for editorial choices makes research in this area particularly difficult. In this paper, we argue that there are newsworthy topics where objective measures exist in the form of supporting data, and propose a computational framework to analyze framing in this setup. We focus on the economy because the reporting of economic indicators, an omnipresent supporting data when discussing the topic, presents us with a relatively easy way to determine both the selection and framing of various publications. Their values provide a ground truth of how the economy is doing relative to how the publications cover it. To do this, we define frame prediction as a set of interdependent tasks. At the article level, we learn to identify the reported stance towards the general state of the economy. Then, for every numerical quantity reported in the article, we learn to identify whether it corresponds to an economic indicator and whether it's being reported in a positive or negative way. To perform our analysis, we track six American publishers and every article that appeared on the top 10 slots on their landing page between 2015 and 2023.

1 Introduction

The mainstream media has a lot of leeway in what it chooses to cover and how it covers it. Should the top article at any given point in time be about a domestic political fight, an international incident, or the state of the economy? Conditional on choosing to cover the economy, should the article focus on the indicator that is up, the one that is down, or eschew quantitative measure altogether? Should it be optimistic or pessimistic towards the future? These choices have real-world consequences affecting outcomes such as voting behavior (Druckman and Parkin, 2005; Gerber et al., 2009), gun purchases (Krupenkin et al., 2023), and attitudes towards immigrants (Krupenkin et al., 2020; Guo et al., 2023). Most selection and framing decisions lack objective measures for a researcher to weigh if the coverage reflects the underlying reality, which makes research on this topic difficult.

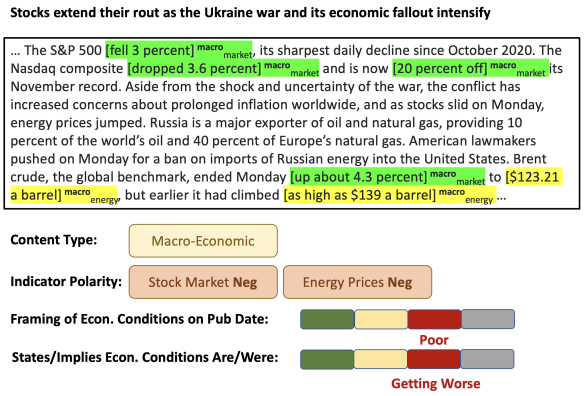

Figure 1: The Frame Prediction Framework

In this paper, we make the observation that there are a select number of newsworthy topics where objective measures *do* exist. For example, there is a finite set of economic indicators that experts use to define the state of the economy, such as the change in non-farm payroll and the gross domestic product. In order to research the selection and framing of publications it is not enough to just have supporting objective data points, but also a relatively well defined set to choose from. Similarly, crime has well established data points maintained by the FBI and local jurisdictions. Building a system capable of tracking a large number of publications, and identifying *which* indicators are reported and *how* they are being reported has enormous research potential. Such as system would enable us to define the full picture of how a publication is covering a topic, how their coverage differs from other publications, how it evolves over time, and how it compares to the accepted ground truth of experts in the field.

Following this rationale, we propose a computational framework to predict frames *in the presence of supporting data*. For a given article, our goal is to automatically identify *how* the general topic is being portrayed, *which* indicators are being reported to support this view, and *how* each of these indicators are being presented. We focus specifically on news articles about the U.S. economy, however, we argue that this framework could be easily adapted to other domains where numerical indicators are used, such as reports on crime, climate change, and public opinion.

To model framing in news about the U.S. economy, we decompose frame prediction into a set of interdependent tasks (See Fig. 1). At the article level, we learn to identify the relevance of the article to the economy, the type of economic information it covers (e.g., macro economic, industry-specific, firm-specific), and two measures of reported economic sentiment: the general state of the economy (e.g., good, fair, poor), and the direction in which the economy is heading (e.g., better, same, worse). Then, for every numerical quantity reported in the article, we learn to predict whether it corresponds to an economic indicator (e.g., market numbers), and its reported polarity (e.g., "The S&P 500 fell 3 percent" has a negative polarity).

Most previous computational framing analysis approaches have conceptualized and interpreted frames as high-level relevant topics and themes (Ali and Hassan, 2022). For example, Boydstun et al. (2014) and (Card et al., 2015) proposed 15 broad dimensions that aim to capture ways in which policy issues are discussed in news articles. The dimensions include themes like "economic" and "public opinion". In their position paper, Ali and Hassan (2022) convincingly argue that these broad dimensions often fail to capture a frame's nuances. For example, the "economic frame" focuses on anything related to the economy, which is insufficient to answer the question of *how* different aspects of the economy are being presented. In contrast, our framework captures three levels: if the economy is being discussed, which economic indicators are highlighted, and the framing of those indicators. We can then measure not just topic selection, but how the publications differ from the objective reality at that time, and how this compares within publication across time, and between publications.

We make the following contributions: (1) We propose a computational framework to model framing in the presence of supporting data and apply it to news about the U.S. economy. (2) We collect a novel dataset of landing page news articles published by a set of major U.S. news outlets between 2015 and 2023, and provide high-quality annotations of our proposed frame components for a small subset of examples. (3) We evaluate two automated methods to detect each of the proposed frame components under low-supervision settings. (4) We demonstrate that our framework can be used to track *when*, *how often* and *how* different news outlets report on specific aspects of the economy, and compare these reports to the ground truth for an example indicator: the job numbers. **All of our code and data will be released to the community under an open source license.**

2 Background and Related Work

In this section, we cover the relevant background on selection and framing, explain what economic indicators are, and discuss previous computational approaches to analyze framing.

2.1 Theory of Selection and Framing

National news publications have a massive choice-set of topics they could cover any given day: it may feel like the publications are reacting to a defined set of events that recently happened, but in reality, outside of a few major events, they choose what the news is on any given day. In other words, news do not present an objective view of current affairs, and are influenced by the *selection* made by media professionals (McQuail, 1992). Selection is driven by what researchers have called news values (Galtung and Ruge, 1965; Harcup T and O'Neill D, 2001, 2017), which determine whether a story is newsworthy. While there is variation in the taxonomies that have been proposed to determine newsworthiness, the most recent ones include factors such as: the power of elite (stories concerning powerful entities), bad and good news (stories with particular emotional overtones), and magnitude (stories that are perceived as significant for a wide audience) (Harcup T and O'Neill D, 2017). More interestingly, news values vary from publication to publication. A study of the run-up to the 2022 election showed the venerable *New York Times* and *Washington Post* averaged just two overlapping topics each morning in their printed editions, out of about six articles (Rothschild et al., 2023).

Conditional on selecting a topic, there is addi-

tional leeway on how to cover it, as there is a wide-range of facts and opinions that could spin the take-away of the readers. This phenomenon, referred to as *framing*, has been widely studied in communication studies. While there is not an agreed upon definition of framing, recent surveys (Vallejo et al., 2023) have identified the following prevalent definitions: equivalence framing (presenting the same exact information in different ways) (Cacciatore et al., 2015), emphasis framing (highlighting specific aspects of an event to promote a particular interpretation) (Entman, 2007), and story framing (leveraging established narratives to convey information) (Hallahan, 1999). When it comes to the economy, a publication could present a rosy or terrible version of the economy, without lying, just by picking the right economic indicator with the right perspective. For example, there are months where non-farm payroll is up more than the average but it is below the market expectation: either perspective is technically true. But, with a relatively finite set of indicators, and an established way for economists to judge them, there is still a clear ground truth to compare if any given indicator with any given perspective represents the consensus understanding of the economy at that time.

2.2 Economic Indicators

Understanding the health and wealth of the U.S. economy is critical for investment and growth: while some data goes back earlier, most of the modern economic indicators have been tracked for decades, codified in the by the mid-20th century. Example include: jobs numbers (e.g., nonfarm payroll), prices (e.g., consumer price index), and macro economic (e.g., gross domestic product). With the exception of market prices (e.g., S&P500), which are a reflection of the underlying assets, the core indicators are tracked by a few government agencies (e.g, the Bureau of Economic Analysis for GDP, spending, income, trade, and the Bureau of Labor Statistics for unemployment, prices, productivity), which conduct regular, large scale polling.

These economic indicators are not just key for investors, but also inform the public on the health of the economy. The jobs numbers are front page news when they are released at 8:30 AM ET on the first Friday of each month, the state of the markets is constantly in the news, and indicators such as the consumer price index pop up whenever the media finds them most important, interesting, and engaging. Most people have limited views of the economy outside of their close social circle, so these values are the key to their understanding of the health of the economy as a whole. And, traditionally, the health of the economy has tracked *very closely* with the approval of incumbent political leaders (Hummel and Rothschild, 2014).

2.3 Computational Framing Analysis

With the proliferation of NLP techniques, scholars have increasingly adopted computational approaches to study framing in diverse forms of media. By automating parts of the process, studies have been able to scale to very large media repositories. In most cases, researchers adopt unsupervised techniques such as topic modeling to identify latent themes (DiMaggio et al., 2013; Nguyen et al., 2015; Gilardi et al., 2021). However, topic models are by design limited in to capture framing nuances. By defining topics as simple distributions over words, they lack the semantic and discursive conceptualization needed to answer *how* issues and events are being presented (Ali and Hassan, 2022).

On the other hand, there is a body of work leveraging supervised learning (Johnson et al., 2017; Khanehzar et al., 2019; Kwak et al., 2020; Huguet Cabot et al., 2020; Mendelsohn et al., 2021) and lexicon expansion techniques (Field et al., 2018; Roy and Goldwasser, 2020) to analyze framing. For this to be feasible, authors need to define a concrete taxonomy of what frames are relevant and how they are represented in data. Most of these approaches rely on the taxonomy of framing dimensions proposed by Boydstun et al. (2014); Card et al. (2015). These dimensions correspond to broad themes such as "economic", and "public opinion". While these themes have shown to be useful to model a select set of contentious political issues like abortion and immigration, they are too broad to capture *how* the economy and other data-driven topics are being presented in the media.

3 Data

In this section, we describe our data collection process, as well as our annotation schema, annotation guidelines, and quality assurance process. Finally, we present statistics for our resulting dataset.

3.1 Data Collection

All of the data for this study came from the Internet Archive's Wayback Machine (Archive, 1996-2023).

We consider six American publishers: The Wall Street Journal, Fox News, The New York Times, The Washington Post, HuffPost and Breitbart. For each of our publishers' landing pages, we retrieved a list of every Wayback Machine entry from January 1st 2015 to January 1st 2023. We segmented the list of entries by the hour, and downloaded the earliest available document in that hour. Each of these documents is considered "front page".

For each front page, we parsed the HTML to identify individual articles. We only recorded articles in the top 20 positions. We extracted plain text versions of the article document In some cases, the main body of the article could not be extracted either because the link no longer exists, or because the content is behind a paywall. We removed all such instances. We discarded all duplicates, keeping only the first appearance of each article.

In some cases, there were errors when extracting headlines. To deal with these cases, we used T5 (Raffel et al., 2020), a generative language model, to generate headlines using the body of the article. To do this, we used a model that was trained on 500,000 articles with their headlines [*].

To identify the subset of articles that discuss the economy, we curated lexicon of economic terms. We sourced these terms from the Bureau of Labor Statistics and the FRED database operated by the Federal Reserve bank of St. Louis.[†‡] These sources provide a fairly comprehensive list of common metrics of economic activity. We consider an article to be relevant to the economy if it contains at least three sentences mentioning any economic term from our lexicon. Figure 2 shows the number of valid, unique articles extracted per source, as well as the resulting number of economic articles. The lexicon of economic terms used to identify articles is included in Appendix A.1.

3.2 Data Annotation

We employed a group of six annotators to label our dataset following the guidelines outlined in App. A.2. Our annotators comprised 1 senior, 1 post-doctoral and 2 pre-doctoral researchers, as well as 2 undergraduate research assistants. The fields of study of the annotators include: economics, computational social science, and communications. They all received training from the authors of the paper, and several rounds of calibration were performed

[*]huggingface.co/Michau/t5-base-en-generate-headline
[†]https://www.bls.gov/data/
[‡]https://fred.stlouisfed.org/

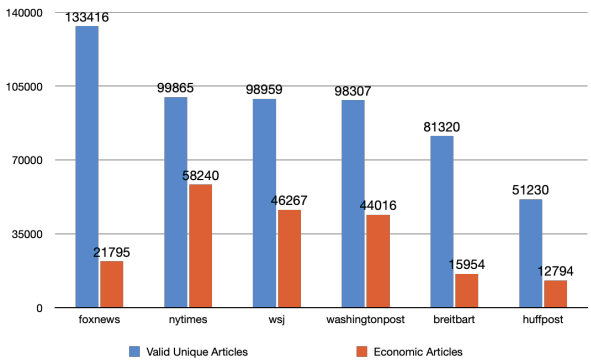

Figure 2: Economic Articles per Publisher

before annotation began. Each week, annotators were assigned a new set of articles to cover. These articles were chosen such that there would be some overlap between coders (in order to check quality), as well as a decent breadth of coverage. To promote topical breadth, we followed the sampling procedure introduced in Pacheco et al. (2022b, 2023).

For each article in their batch, annotators had to select: (1) what *type* of economic information was the most prominent (macro, government, industry-specific, business-specific or personal), (2) the framing of the general economic *conditions*, if expressed in the article (good, fair, bad), and (3) the framing of the *direction* that the economy is heading, if expressed in the article (better, same, worse). Then, for every numerical quantity reported in the article, annotators had to select: (1) what type of economic data it reports, if any (macro, government, industry-specific, business-specific or personal), (2) if the quantity is macro, then which indicator was reported (e.g., jobs, prices. See App. A.2 for the full list of options), and (3) the reported polarity of each quantity (positive, negative, neutral). We used regular expressions for identifying candidate quantities, and used a set of manually defined rules to get rid of dates and other irrelevant quantities. We report resulting statistics in Tab. 1, and inter-annotator agreement in Tab. 2. Lastly, we show the distribution for all annotated frame components in Fig. 3.

We find that inter-annotator agreement is generally good for quantity-level information (Krippenforff's α 0.56-0.83). Meanwhile, article-level judgements are considerably harder for annotators, and exhibit moderate agreement (Krippenforff's α 0.42-0.48). This is expected and in line with other framing tasks (Card et al., 2015; Roy et al., 2021; Mendelsohn et al., 2021).

Publisher	Econ Arts.	Quants.	Human Annotations	
			Art-level	Quant-level
New York Times	58,240	370,723	516	1,117
Wall Street Journal	46,267	551,726	206	493
Washington Post	44,016	274,197	231	421
Fox News	21,795	76,074	45	42
Breitbart	15,954	66,572	91	149
HuffPost	12,794	75,180	82	210
Total	199,066	1,414,472	1,171	2,414
Cross-Annotated			270	689

Table 1: Resulting Dataset Statistics

Annotation	K's α	% of Agreement	
		Full	Partial
Article Type	0.48	57.81	90.62
Econ. Conditions	0.43	55.13	85.90
Econ. Direction	0.42	47.44	82.05
Quantity Type	0.75	80.84	88.97
Quantity Polarity	0.56	68.04	76.48
Macro Ind.	0.83	83.42	92.78

Table 2: Inter-annotator Agreement. Here, Krippendorff's $\alpha = 1$ indicates perfect reliability, $\alpha = 0$ indicates the complete absence of reliability, i.e., judgments are random, and $\alpha < 0$ indicates that disagreements are systematic and exceed what can be expected by chance.

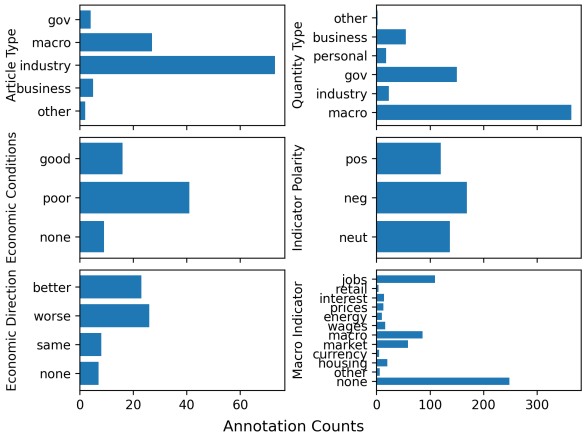

Figure 3: Distribution of annotation labels

4 Frame Prediction

We evaluate two methods to predict frames using the schema introduced in Sec. 3: (1) Individual classifiers for each frame component, using fine-tuned LMs, and (2) A relational model where we explicitly express the dependencies between the different components. In this section, we motivate and describe these two methods, as well as the experimental settings used for evaluation. Then, we present and discuss our final results.

General Experimental Settings We consider all articles with human annotations, and test only on articles that were cross-annotated and take the majority vote. We perform 5-fold cross-validation in all scenarios, and assume a multi-class setup for each frame component. We report the average weighted and macro F1 scores for all folds. All experiments were run on a server node with a single NVIDIA Tesla A100 GPU with 40 GB of memory, an AMD 64C 2.8GHz CPU with 192 GiB RAM.

Individual Classifiers One of the main challenges that we face for predicting frames is that we have a very small amount of supervision. However, we have the advantage that we have access to a very large repository of in-domain unlabeled examples (all 199,066 economic articles collected). Further pre-training with domain-specific unlabeled data has been shown to improve the performance of pre-trained LMs such as RoBERTa (Gururangan et al., 2020). This method is referred to as domain-adaptive pre-trainingn (DAPT) and is conducted by randomly masking words in a domain-specific training set and training the the LM to predict a random subset of masked words.

We use three individual classifiers to predict article-level components in our framework: the article type, the economic conditions and the economic direction. In all cases, we fine-tune RoBERTa-base with and without DAPT. For training, we use the AdamW optimizer, the cross-entropy loss, and a learning rate of $2e-5$. For early stopping, we use the macro F1 on the validation set, consisting of ten percent of the training articles for each fold.

To predict quantity-level components we also use three individual classifiers. In this case, we also fine-tune RoBERTa-base with and without DAPT. To represent the context surrounding the quantity, we use the sentence containing the excerpt and the two surrounding sentences, to the left and right. We concatenate the CLS embedding of this contextual information with the average embedding (final-layer) of all tokens within the excerpt containing the quantity before passing it to the classifier. We use the same training configuration as the article-level classifiers.

Relational Model Statistical Relational Learning (SRL) methods attempt to model a joint distribution over relational, inter-dependent data (Richardson and Domingos, 2006; Bach et al., 2017; Pacheco and Goldwasser, 2021). These methods have proven particularly effective in tasks where contextualizing information and interdependent decisions

can compensate for a low number of annotated examples (Deng and Wiebe, 2015; Roy et al., 2021; Pacheco et al., 2022a).

We take advantage of the relations between the values of the frame components for a given article, and implement a relational model using Probablistic Soft Logic (PSL) (Bach et al., 2017). PSL is a declarative framework for specifying and learning relational models. Dependencies in PSL are expressed using weighted logical rules of the form: $w_r : P_1 ... P_{n-1} \rightarrow P_n$. Where w_r indicates the importance of the rule in the model, and can be learned from data, while predicates P_i correspond to decisions and observations. In PSL, rules are compiled into a Hinge-Loss Markov random field and weights are learned using maximum likelihood estimation. For an in-depth explanation of PSL, we refer the reader to the original paper. To model our problem, we consider the following rules:

(r_1) **Dependency between the polarity of quantities and economic conditions.** Our intuition is that if many negative quantities are reported in the article, then the article is likely to frame economic conditions as negative, and vice-versa for positive quantities: $\texttt{Reports}(\texttt{a}, \texttt{q}) \wedge \texttt{Polarity}(\texttt{q}, \texttt{p}) \Rightarrow \texttt{EconConditions}(\texttt{a}, \texttt{p})$

(r_2) **Dependency between the polarity of quantities and economic direction.** Our intuition is that if many negative quantities are reported in the article, then the article is likely to frame the direction of the economy as negative, and vice-versa for positive quantities: $\texttt{Reports}(\texttt{a}, \texttt{q}) \wedge \texttt{Polarity}(\texttt{q}, \texttt{p}) \Rightarrow \texttt{EconDirection}(\texttt{a}, \texttt{p})$

(r_3) **Dependency between the quantity types and the article type.** Our intuition is that if many quantities of specific type are reported in an article (e.g. macro, government, industry), then that article is likely have that same dominant type: $\texttt{Reports}(\texttt{a}, \texttt{q}) \wedge \texttt{Type}(\texttt{q}, \texttt{t}) \Rightarrow \texttt{Type}(\texttt{a}, \texttt{t})$

(r_4) **Consistency between the quantity type and the macro-indicator.** Finally, we enforce that if the quantity of a type is macro, then a macro-indicator must be predicted, and vice-versa: $\texttt{Reports}(\texttt{a}, \texttt{q}) \wedge \texttt{Type}(\texttt{q}, \texttt{macro}) \Leftrightarrow \neg\texttt{MacroIndicator}(\texttt{q}, \texttt{none})$

Rules r_{1-3} are modeled as *soft* constraints, and weights are learned through PSL. This way, we allow for quantities that do not conform to the templates to be active. Rule r_4 is modeled as a *hard* constraint, and is always enforced. To initialize PSL, we include the predictions of the initial classifiers as priors. This strategy has been repeatedly used to combine PSL with strong classifiers, with consistent success (Sridhar et al., 2015; Pacheco and Goldwasser, 2021; Roy et al., 2021).

Model	Article-level			Quantity-level		
	Type	Cond	Dir	Type	Ind	+/−
RoBERTa	0.695	**0.733**	0.396	0.924	0.857	0.677
RoBERTa+DAPT	0.675	0.728	0.395	**0.935**	0.882	**0.783**
Relational (Best)	**0.722**	**0.733**	**0.460**	0.935	**0.907**	**0.783**

Table 3: General Results: Average Weighted F1 Scores after 5-Fold Cross Validation

We present our general results in Tab. 3. First, we find that domain-adaptive pre-training is considerably helpful for quantity-level predictions, but it does not help for article-level predictions. We hypothesize that it is hard for RoBERTa to model the long context of full articles to take advantage of the masked language modeling objective. Next, we find that the relational model generally gives us a boost in performance, with the exception of economic conditions and quantity polarity. This improvement supports our hypothesis that decomposing frame prediction into inter-dependent subtasks is beneficial, particularly in the face of low supervision. While Tab. 3 reports weighted F1, we find the same trend for macro F1. These results are included in App. A.3.

In Tab. 4 we include an ablation study for the PSL rules. All rules, with the exception of r_1 (polarity and economic conditions) contribute to an increase in performance for their corresponding decisions predicates (see bold scores). However, when combined together into a single model, they fail to complement each other in any meaningful way. We hypothesize that this is due to the fact that some of these decisions are in competition with each other. This is expected given that PSL learns a general weight for each rule template, and does not provide a simple way to incorporate instance-specific information. Going forward, we would like to experiment with other relational learning frameworks like DRaiL (Pacheco and Goldwasser, 2021), that allow for a more seamless integration with expressive, distributed representations.

Lastly, we include fine-grained results for predicting the different macro-indicators using our best model in Tab. 5. We find that we have generally good performance for most of them, with the exception of retail sales, energy prices, and wages. We attribute this to the lack of sufficient supervi-

Rule	Article-level			Quantity-level		
	Type	Cond	Dir	Type	Ind	$+/-$
No rules	0.695	**0.733**	0.396	**0.935**	0.882	**0.783**
r_1	0.695	0.733	0.396	0.935	0.882	0.781
r_2	0.695	0.733	**0.459**	0.935	0.882	0.779
r_3	**0.722**	0.733	0.396	0.908	0.882	0.783
r_4	0.695	0.733	0.396	0.935	**0.907**	0.783
$r_1 + r_4$	0.695	0.733	0.396	0.935	0.905	0.781
$r_2 + r_4$	0.695	0.733	**0.460**	0.934	0.905	0.778
$r_3 + r_4$	**0.722**	0.733	0.396	0.908	0.906	0.783
$r1 + r_2 + r_3 + r_4$	0.714	0.654	0.449	0.912	0.906	0.781

Table 4: Ablation Results for the Relational Model

Macro-Indicator	Prec.	Recall	F1
Job Numbers (jobs, unemployment)	0.920	0.954	0.937
Retail Sales	0.333	0.25	0.286
Interest Rates (Fed, mortgage)	1.0	0.857	0.923
Prices (CPI, PPI)	0.619	1.0	0.765
Energy Prices (gas, oil, etc.)	1.0	0.4	0.571
Wages	0.8	0.5	0.615
Macro Economy (GDP, etc)	0.938	0.884	0.910
Market Numbers (any financial market)	0.873	0.932	0.901
Currency Values	0.667	0.8	0.727
Housing (start, sales, pricing)	0.909	1.0	0.952
None	0.956	0.960	0.957
Accuracy	0.910	0.910	0.910
Macro Average	0.780	0.739	0.739
Weighted Average	0.914	0.910	0.907

Table 5: Macro-Indicator Results for our Best Model

sion for these classes (See Fig. 3). Going forward, we need to expand our annotations for these classes to be able to support the large-scale analysis of these indicators in the news.

5 Framing Analysis

In this section we present a brief analysis of how the methods outlined in this paper can be used to study editorial choices. To do this, we use our best model to predict all frame components for the full dataset of 199,066 articles. We then use these frames to show how economic messaging shifts in response to exogenous shocks and changes in the balance of political power. In particular, we show how our framework can help us understand how the different choices these major publications make, in just how they handle one aspect of the economy, can produce a very different view of economic conditions.

Article-level framing with macro-indicators We generate article-level frames for specific economic indicators using our quantity predictions. First, we select the subset of articles that include at least two quantities in the category of interest. Then, we assign a positive indicator frame to articles with at least twice as many positive quantities as negative quantities and vice versa for negative frames. Articles that meet neither threshold are assigned a "neutral" frame.

5.1 Shifts in Framing

Our preliminary evidence suggests that the framing of articles on any given topic by any given publication is both sensitive to exogenous factors as well as decisions made by the publication itself.

In Figure 4, we can see how the *New York Times*, *Washington Post*, and the the *Wall Street Journal* framed their articles with jobs data from 2015 to 2023: it is striking to note just how much the Times and Post differ in their framing to that of the Journal. First, focus on the purple dashed lines to see that the Times and the Post both have a much greater percentage of their economic coverage devoted to jobs numbers (the Journal has a lot more coverage of markets, which makes sense as they cover the daily fluctuations). And while all three publications chose to promptly respond to the massive job loss caused by the pandemic, dramatically increasing their percent of coverage devoted to jobs, the Journal dipped back down quickly, the Post slowly dipped to their pre-pandemic norm, and the Times slowly fell below pre-pandemic norms (not shown, the Times had a relatively higher surge in their coverage of prices in this post-pandemic era). Further, it is not just quantity, but the spin that makes the frame: the Journal has very malleable framing, with heavily positive coverage of jobs numbers as they soured in the early Biden administration, while the Times coverage continued to be heavily negative despite the historic run of jobs recovery.

Comparing between publications allows us to show that the editors have considerable leeway when choosing what indicator to cover at any given time, and how to cover it. But, the black dotted lines also let us explore how these choices differ from the ground truth. It makes sense that as the jobs numbers literally go off the chart, there is a lot of coverage of them. But it is striking that as the recovery proved to be historically quick and consistent, the Journal moved to positive framing, while the Post dipped down to their historical level of negative framing, but the Times remained more negative than historical levels.

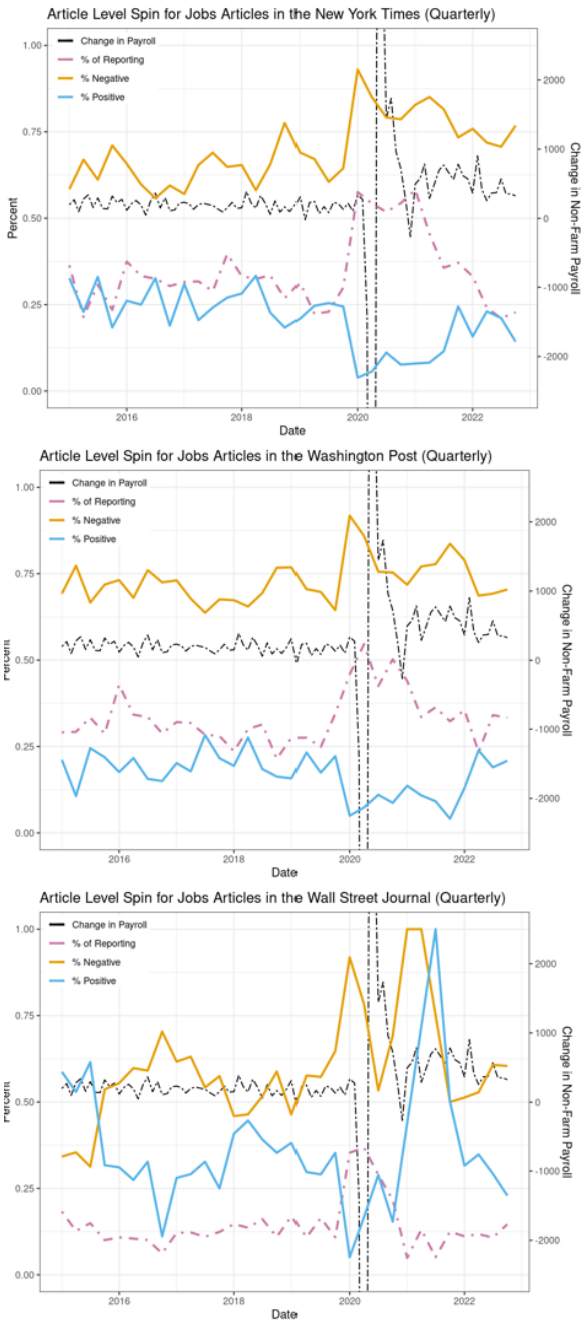

Figure 4: Framing of articles referencing job numbers from 2015 through 2023 in the New York Times, Washington Post, and Wall Street Journal

6 Conclusion and Future Work

In this paper, we presented a novel framework for modeling framing in news articles in the presence of supporting data. We focused our analysis on news about the U.S. economy, and proposed an annotation schema to identify framing information at three levels of abstraction: the general framing of the economic conditions, the economic indicators that are highlighted, and the framing of each of those indicators. We showed that we can predict the components of our schema with relatively good performance, even in very low supervision settings. Finally, we demonstrated how our framework can be used to perform a large scale exploration of the frame choices for a given economic indicator. We are able to show that different publications cover economics using different indicators, for example the *New York Times* covers jobs numbers less than normal during a strong jobs recovery, instead switching to more coverage of prices which have risen. And, unlike the generally negativity bias of the Times and the *Washington Post*, the *Wall Street Journal* shifts into very positive coverage of jobs when the numbers rise and negative when they fall, better reflecting the underlying objective truth.

Going forward, our efforts are three-fold. First, we are working on expanding our schema to capture a larger set of economic aspects. While we focused our analysis on macro-economic indicators, there are other indicators that we could track to get a more holistic view of the economy. For example, for governmental indicators we could keep track of different types of expenditures (e.g. social security), revenue (e.g. taxes), as well as debt and deficit. Similarly, we could track industry-specific data points (e.g. mining, manufacturing). Second, we want to improve our predictive models to improve the confidence of our analysis. This will require scaling up our annotations to have more supervision, as well as exploring alternative techniques for low-resource learning, such as few-shot in-context learning and data augmentation. Finally, we want to perform a large-scale, holistic analysis of economic news framing. In this paper, we presented a brief demonstration of how our framework can be used to analyze shifts in framing across time for different publications. Next, we want to do this for a larger set of economic indicators and publications. Additionally, we want to adapt and validate our framework for other domains of interest, such as crime, climate change and public opinion.

7 Limitations

The work presented in this paper has two main limitations. (1) Obtaining high-quality annotations for our frame structure is very expensive and time consuming. For this reason, we had to work with a low amount of supervision. While we showed that we could obtain relatively good performance in this constrained scenario, to be able to use our framework for a large scale, holistic analysis, we need to address this issue. This is particularly important given that our annotations are considerably skewed. For example, we have way more supervision for job numbers and market numbers (which occur more often) than we do for energy prices (which occur less often). This considerably affects our performance for more long-tail indicators, and hence limits the types of analysis that we can perform. Our current efforts are dedicated to increase the amount of annotations for all components of our frame structure, as well as incorporating alternative semi-supervised strategies. We are hoping to release an extended version of the annotated dataset to the community in the next few months. (2) The fact that we are using automated techniques for frame prediction necessarily carries some uncertainty. Even if we improve our models considerably, our large scale analysis has a margin of error. It is important to acknowledge this when presenting our findings.

8 Ethical Considerations

To the best of our knowledge, no code of ethics was violated during the development of this project. We used publicly available tools to collect our dataset, and discarded any instances that were unreachable. The annotators were paid $15 per hour, and no personally identifiable information was collected or recorded during annotation. We did a thorough evaluation of our dataset, which is presented in the paper. We reported all pre-processing steps, learning configuration, hyperparameters, and additional technical details. In the interest of reproducibility, the data and code will be released to the community. Due to space constraints, some information was relegated to the Appendix. The results reported in this paper support our claims and we believe that they are reproducible. The analysis reported in Section 5 is done using the outputs of a machine learning model and do not represent the authors personal views. The uncertainty of these predictions was adequately acknowledged in the Limitations Section, and the estimated accuracy was reported in Section 4.

References

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Internet Archive. 1996-2023. The wayback machine. Accessed Continuously from January 2022.

Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2017. Hinge-loss markov random fields and probabilistic soft logic. *J. Mach. Learn. Res.*, 18(1):3846–3912.

Amber E. Boydstun, Dallas Card, Justin Gross, Paul Resnick, and Noah A. Smith. 2014. Tracking the Development of Media Frames within and across Policy Issues.

Michael Cacciatore, Dietram Scheufele, and Shanto Iyengar. 2015. The end of framing as we know it ... and the future of media effects. *Mass Communication Society*, 19.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Lingjia Deng and Janyce Wiebe. 2015. Joint prediction for entity/event-level sentiment analysis using probabilistic soft logic models. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 179–189, Lisbon, Portugal. Association for Computational Linguistics.

Paul DiMaggio, Manish Nag, and David Blei. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of u.s. government arts funding. *Poetics*, 41:570–606.

James N. Druckman and Michael Parkin. 2005. The impact of media bias: How editorial slant affects voters. *Journal of Politics*, 67(4):1030–1049.

Robert Entman. 2007. Framing bias: Media in the distribution of power. *Journal of Communication*, 57:163 – 173.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computa-

tional analysis of intricate political strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

Johan Galtung and Mari Holmboe Ruge. 1965. The structure of foreign news: The presentation of the congo, cuba and cyprus crises in four norwegian newspapers. *Journal of Peace Research*, 2(1):64–90.

Alan S. Gerber, Dean Karlan, and Daniel Bergan. 2009. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52.

Fabrizio Gilardi, Charles R. Shipan, and Bruno Wüest. 2021. Policy diffusion: The issue-definition stage. *American Journal of Political Science*, 65(1):21–35.

Lei Guo, Chris Chao Su, and Hsuan-Ting Chen. 2023. Do news frames really have some influence in the real world? a computational analysis of cumulative framing effects on emotions and opinions about immigration. *The International Journal of Press/Politics*, 0(0):19401612231204535.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.

Kirk Hallahan. 1999. Seven models of framing: Implications for public relations. *Journal of Public Relations Research*, 11(3):205–242.

Harcup T and O'Neill D. 2001. What Is News? Galtung and Ruge revisited. *Journalism Studies*, 2:261–280. Accessed on 2023/12/15.

Harcup T and O'Neill D. 2017. What is news? News values revisited (again). *Journalism Studies*, 18:1470–1488. Accessed on 2023/12/15.

Pere-Lluís Huguet Cabot, Verna Dankers, David Abadi, Agneta Fischer, and Ekaterina Shutova. 2020. The Pragmatics behind Politics: Modelling Metaphor, Framing and Emotion in Political Discourse. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4479–4488, Online. Association for Computational Linguistics.

Patrick Hummel and David Rothschild. 2014. Fundamental models for forecasting elections at the state level. *Electoral Studies*, 35:123–139.

Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging behavioral and social information for weakly supervised collective classification of political discourse on Twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 741–752, Vancouver, Canada. Association for Computational Linguistics.

Shima Khanehzar, Andrew Turpin, and Gosia Mikolajczak. 2019. Modeling political framing across policy issues and contexts. In *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association*, pages 61–66, Sydney, Australia. Australasian Language Technology Association.

Masha Krupenkin, Shawndra Hill, and David Rothschild. 2020. Beyond agenda setting: Does media coverage of immigration lead to anti-immigrant behavior.

Masha Krupenkin, Elad Yom-Tov, and David Rothschild. 2023. Fear and loathing in st. louis: Gun purchase behavior as backlash to black lives matter protests.

Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. 2020. A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. In *12th ACM Conference on Web Science*, WebSci '20, page 305–314, New York, NY, USA. Association for Computing Machinery.

D. McQuail. 1992. *Media Performance: Mass Communication and the Public Interest*. Sage Publications (CA).

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1438–1448, Beijing, China. Association for Computational Linguistics.

Maria Leonor Pacheco and Dan Goldwasser. 2021. Modeling content and context with deep relational learning. *Transactions of the Association for Computational Linguistics*, 9:100–119.

Maria Leonor Pacheco, Tunazzina Islam, Monal Mahajan, Andrey Shor, Ming Yin, Lyle Ungar, and Dan Goldwasser. 2022a. A holistic framework for analyzing the COVID-19 vaccine debate. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5821–5839, Seattle, United States. Association for Computational Linguistics.

Maria Leonor Pacheco, Tunazzina Islam, Lyle Ungar, Ming Yin, and Dan Goldwasser. 2022b. Interactively uncovering latent arguments in social media platforms: A case study on the covid-19 vaccine debate. In *Proceedings of the Fourth Workshop on Data Science with Human-in-the-Loop (Language Advances)*, pages 94–111, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Maria Leonor Pacheco, Tunazzina Islam, Lyle Ungar, Ming Yin, and Dan Goldwasser. 2023. Interactive concept learning for uncovering latent themes in large text collections. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5059–5080, Toronto, Canada. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1).

Matthew Richardson and Pedro Domingos. 2006. Markov logic networks. *Machine Learning*, 62(1-2):107–136.

David M. Rothschild, Elliot Pickens, Gideon Heltzer, Jenny Wang, and Duncan J. Watts. 2023. Warped front pages.

Shamik Roy and Dan Goldwasser. 2020. Weakly supervised learning of nuanced frames for analyzing polarization in news media. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7698–7716, Online. Association for Computational Linguistics.

Shamik Roy, Maria Leonor Pacheco, and Dan Goldwasser. 2021. Identifying morality frames in political tweets using relational learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9939–9958, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 116–125, Beijing, China. Association for Computational Linguistics.

Gisela Vallejo, Timothy Baldwin, and Lea Frermann. 2023. Connecting the dots in news analysis: A cross-disciplinary survey of media bias and framing.

A Appendix

A.1 Additional Data Collection Details

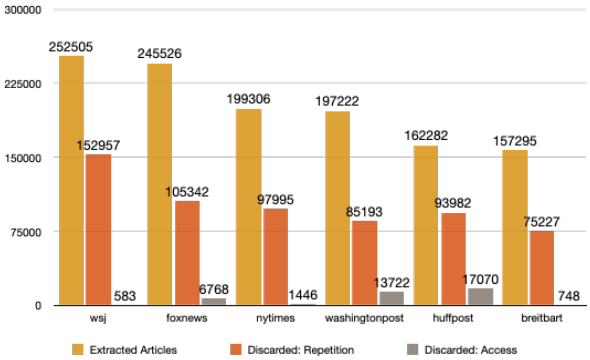

Figure 5: Discarded Articles per Source

Fig. 5 shows the number of articles that were discarded due to either repetition or access issues. The number of discarded articles is quite large, but given that our dataset is based on hourly snapshots for the archive, it is to be expected. We aimed to over collect during the early collection steps to maximize the completeness of our final dataset.

bond*, budget, business, consumer, consumer price index, cost, cpi, currency debt, deficit, demand, dow jones, earning*, econo*, employ*, expenditure*, export*, fed, financ*, fiscal, gdp, gross domestic product, housing, import*, income, ipi, ipp, industrial production index, inflation, interest, invest*, international price program, job*, labor, market, monet*, mortgage, nasdaq, poverty, ppi, price*, productivity, producer price index, retail, revenue, s&p 500, sales, securities, small cap 2000, stimulus, stock, supply, tax, trade, trading, treasur*, unemploy*, wage*, wti, west texas intermediate

Table 6: Economic Lexicon

Tab. 6 holds the lexicon of economic terms used to determine whether or not an article contained economic content.

A.2 Annotation Guidelines

In the following sections we replicate the codebook provided to the coders. This codebook was used during training, and as a reference for the coders to use as needed.

A.2.1 Instructions

To annotate specific ARTICLES we will:

1. Click “Edit Article” for the first 5 articles assigned to each topic

2. Annotate key quantities (first 5 relevant quantities in an article + 2 scattered throughout the article, if present):

   - Each quantity identified will be highlighted in yellow and clickable
   - If the quantity is a date or otherwise not relevant, skip it
   - If the quantity is relevant, click it and follow the instructions (code-book will be outlined below)
   - Data points may not fit into one of the groups. These should be labeled "Other" and a short comment should be left to explain it

A.2.2 Codebook

1. **Macro-economic**

   - Macro-economic indicators aggregate data according to a set of demographics.
   - Data Guide
     - For the purpose of this study, these will mostly be values that are aggregated up to a national level (e.g. national GDP). There may be, however, other demographic groups, like states, that fall into this category

2. Types of macro-indicators:

   - Jobs Numbers (Jobs, Unemployment)
   - Market Numbers (any financial market)
   - Housing (Start, Sales, Pricing)
   - Macro Economy (GDP, etc.)
   - Wages
   - Prices (CPI, PPI)
   - Confidence
   - Retail Sales
   - Interest Rates (Fed., Mortgage)
   - Currency Values
   - Energy Prices (Gas, Oil, etc.)
   - Other (Specify)

3. **Firm-Specific**

   - A firm-specific data point is a data point associated with a particular firm or company.
   - Data Guide:
   - Examples include: stock prices, debt offerings, and capital investments

4. **Industry-Specific**

- Industry level articles/quantities describe an entire industry rather than individual businesses
- Data Guide:
  - Examples include: Chip Manufacturers Post Strong Growth Numbers
- Types of Industries:
  - Agriculture, forestry and hunting
  - Mining
  - Utilities
  - Construction
  - Manufacturing
  - Wholesale trade - *selling products in bulk to other businesses*
  - Retail trade - *selling products directly to the end consumer*
  - Transportation and warehousing
  - Information
  - Finance, insurance, real estate, rental and leasing
  - Professional and business services
  - Educational services, health care, and social assistance
  - Arts, entertainment, recreation, accommodation and food services
  - Other (Except government)

5. Government Revenue and Expenditures
   - Any value that describes how a government earned or spent its income falls into this category.
   - Data Guide:
     - A few examples are: taxes, budgets, and treasury issuances
   - Government Level:
     - Federal
     - State and local
   - Types of Expenditures
     - Social Security and Public Welfare
     - Health and Hospitals
     - National Defense
     - Police
     - Transportation
     - Research
     - Education
     - Employment
     - Housing
     - Corrections
     - Courts
     - Net Interest
     - Other
   - Types of Revenue
     - Taxes and other compulsory transfers imposed by government units Property income derived from the ownership of assets
     - Sales of goods and services
     - Voluntary transfers received from other units

6. Personal
   - If an article/quantity focuses on the economic condition of a single person, or a group of individuals that is not large enough to represent an entire demographic then we consider it personal. E.g. an individual's struggle to find work, or the grocery/gas budget of a single family
   - Data Guide:
     - E.g. household expenditures, personal debts
   - A short comment should be added to explain the frame/quantity

A.2.3 Additional Notes

Data Points:

- Each candidate data point is highlighted and clickable
- Each data point has a general type:
  - Macro-economic:
  - Firm-specific
  - Industry-specific
  - Government revenue and expenditures
  - Personal

General Frame:

- While an article can reference different data points, the goal is to identify the dominant frame
- An article can have only one dominant frame
  - Macro-economic
  - Firm-specific
  - Industry-specific
  - Government revenue and expenditures
  - Personal

- Other

- An article may frame the economy in a certain light – each article will be tied to the following two questions:
  - How **does this article** rate economic conditions in the US – Excellent; Good; Only Fair; Poor; No Opinion; Not relevant to the US economy
  - **Does this article** state/imply that economic conditions in the US as a whole are…? {Getting better; Getting worse; Same; No Opinion; Not relevant to the US economy}

- A comment should be added to explain rationale for choosing the frame class / economic outlook

A.3 Additional Results for Frame Prediction

In Tab. 7, we include the macro F1 results for our models.

Model	Article-level			Quantity-level		
	Type	Cond	Dir	Type	Ind	+/−
RoBERTa	0.357	0.671	0.361	0.679	0.671	0.670
RoBERTa+DAPT	0.315	0.641	0.374	0.702	0.712	0.778
Relational (Best)	0.355	0.671	0.361	0.698	0.740	0.778

Table 7: General Results: Average Macro F1 Scores after 5-Fold Cross Validation