IJICTR
International Journal of Information & Communication Technology Research

Volume 16- Number 3 – 2024 (25 -34)

Claim Detection Dataset for Persian Tweets

Mohammad Hadi Bokaei*
ICT Research Institute
Tehran, Iran
mh.bokaei@itrc.ac.ir

Minoo Nassajian
Alumni of Computational Linguistics,
Sharif University of Technology
minoonassajian@alum.sharif.edu

Mojgan Farhoodi
ICT Research Institute
Tehran, Iran
farhoodi@itrc.ac.ir

Mona Davoudi Shamsi
ICT Research Institute
Tehran, Iran
davoudi@itrc.ac.ir

Received: 14 October 2023 – Revised: 2 March 2024 - Accepted: 18 April 2024

Abstract—**The proliferation of false information on social media has profound negative impacts across various aspects of people's lives. To mitigate these effects, numerous studies have focused on developing automated fact-checking systems aimed at enhancing the accuracy and reliability of news and information. Claim detection, recognized as the initial stage in constructing such systems, has been explored in several languages. In our paper, we introduce a corpus of Persian tweets annotated with 11 labels derived from linguistic analysis, representing different types of claims. Additionally, we establish a baseline claim detection model to assess the dataset. This study frames claim detection as a classification task and employs a transformer-based approach to train a multi-label classifier capable of identifying various types of claims in Persian texts.**

Keywords: Automatic Claim Detection, Persian Language Processing, Low-resource language, Multi Label Classification

Article type: Research Article

Publisher: ICT Research Institute

I. INTRODUCTION

Social media has become the most crucial platform for quickly and easily disseminating information today. It enables people to share content in various formats, such as texts, pictures, videos, and audios, in the shortest possible time. Users can access different news sources with a cellphone and effortlessly share them with others.

Despite the numerous benefits of technological advancements, they have also introduced significant disadvantages. The internet hosts a vast amount of information, some of which is incorrect or biased, presenting only one side of contentious issues. This problem is recognized as a growing concern in public [3], [4], as incorrect information can manipulate people's perceptions of reality, influence conscious and unconscious attitudes, and alter behaviors [5]. This misinformation leads to mistrust in various domains [6], especially during crises [7], affecting areas such as health behavior [8], [9], [10], [11], political attitudes and voting behaviors [12], [13], financial markets [14],

* Corresponding Author

[15], [16], and cognitive psychology [17], [18], [19]. Journalists and fact-checkers work diligently to verify and correct misinformation. Given the potential of AI-based tools to reduce the burden and time required for these activities, various studies have focused on developing fact-checking tools [20], [21], [22]. Fact-checking provides accurate and unbiased analysis of claims to enhance individuals' understanding of important issues [23], [24]. A claim is a statement or assertion typically made without providing evidence, forming the central part of an argument [25], [26]. Claim detection is the first step in the automated fact-checking process [27]. Although extensive research has been conducted on fact-checking under different names like disinformation detection, fake news detection, or rumor detection, claim detection is a relatively new area in natural language processing (NLP) that has gained researchers' attention in recent years.

As Twitter is a prominent platform for the dissemination of fake news and no prior research has focused on Persian tweets, our primary emphasis was on exploring the content on this platform. The main objective of this study is to devise an automated system for detecting claims on Twitter. By identifying these claims, we can further analyze their stance and evaluate their veracity or falsity in subsequent stages. To accomplish this, we curated a dataset of Persian tweets and employed an annotation schema that effectively encompasses various types of claims and non-claims, relying on linguistic analysis. The dataset consists of 4,910 tweets, annotated by two individuals. This corpus played a pivotal role in developing an automated claim detection system based on transformers.

The rest of the paper is organized as follows: Section 2 reviews previous work related to this study. Section 3 discusses the different types of claims used to classify Persian tweets. Section 4 introduces our methodology and corpus information. Section 5 presents the experimental reports, including evaluation metrics, error analysis, and results. Finally section 6 concludes the study and suggests future directions in this area.

II. Literature Review

Claim detection is considered a sub-task within the broader fields of argumentation mining (AM) [30] and fact-checking [32]. In these tasks, analyzing claims is crucial for providing structured data for computational models of arguments and reasoning engines [31] (in AM) and for assessing the truthfulness of claims (in fact-checking) [32]. The automated fact-checking process, particularly in the context of computational journalism, has been extensively discussed by [33], [34], [35]. This study has also gained significant attention in the field of NLP.

In AI-based research, the automated fact-checking process typically consists of three stages, the first of which is claim detection [36]. While some methods assume that a claim is provided as input, fully automated fact-checking must identify claims within articles or social media comment sections. Therefore, the initial step is to determine what constitutes a claim [37].

Recent studies on claim detection have been conducted in a limited number of languages, including English [48], [49], [50], [51], Arabic [52], Turkish [53], [55], and Dutch [56]. Most of these studies have framed claim detection as a classification task. For instance, [48] employed a multi-class SVM classifier, [57] utilized a deep model with a Feed-Forward Neural Network (FNN) as the final layer to rank labels, and [58] and [59] used transformer-based models. However, some works have applied sequence labeling techniques [60], [61] and an unsupervised learning approach [62].

These studies have generally applied two approaches to annotated datasets. The first approach relies on the concepts of check-worthiness and claim importance, defining three labels to classify sentences: 1) claim, 2) non-claim, and 3) unimportant claim. The second approach avoids subjective concepts of check-worthiness or importance, leaving this determination to fact-checkers. Researchers linguistically analyzed sentences and defined multiple labels to identify different types of claims [26].

For Persian, which is considered a low-resource language, there has been no prior research on automatic claim detection. Related studies have focused on stance detection [64], fake news detection [65], and rumor verification [69]. This paper aims to address this gap by focusing on Persian claim detection and developing a dataset for this purpose for the first time.

III. Dataset Description

To provide annotation guidance for labeling our corpus, we used 4,910 Persian tweets and conducted a linguistic analysis. Unlike some studies that rely on the concept of the check-worthiness of claims, resulting in subjective interpretation, we decided to leave the judgment of importance to fact-checkers and journalists. Instead, we categorized tweets using 10 labels for different types of claims and one label for non-claims.

Drawing inspiration from prior research, such as [27], which identified 19 sub-categories for claim types, we have simplified the categorization by defining 10 labels for the most frequent claim types observed in Persian tweets. Additionally, we classify less frequent claim types under the "other claim" label. This approach ensures a more manageable and practical classification scheme while capturing the essential variations in claim types encountered in the dataset.

Since a tweet can contain multiple sentences or different types of information, it can be annotated with multiple labels. In the next section, we introduce the sub-categorization of claims.

The sub-categorization of claims

To categorize claims, we utilize linguistic features and define 10 classes based on syntactic, semantic, and pragmatic analysis. These classes are:

1. Action: Statements describing actions or events.
2. Prediction: Statements predicting future events or outcomes.
3. Support/Oppose: Statements expressing support or opposition.
4. Causation/Correlation: Statements indicating a cause-and-effect or correlation relationship.

5. Quantity: Statements involving numerical data or quantities.
6. Comparison: Statements making comparisons between entities or events.
7. Quote: Statements quoting another source or individual.
8. Trait: Statements describing characteristics or traits.
9. Law/Rule: Statements about laws, rules, or regulations.
10. Other Claim: Statements that don't fit into the above categories but still constitute a claim.

A non-claim label is also defined for statements that do not constitute a claim. Below, we will describe each category in detail.

1) Action claims

The action claim type pertains to statements describing actions that have been carried out in the past, are presently occurring, or are anticipated to take place in the near future. This category encompasses a broad range of expressions involving events and activities. The below examples respectively show the past (as shown in example 1 and 4), present (as in example 2), and future tenses (as seen in example 3) in Persian that are annotated by the action claim label.

(1) *500 pezešk sāle gozašte māliyāt pardāxt nakardand.*
"500 doctors did not pay tax last year"

(2) *In kešvarhā dar hāle tote'ečini barāye hamle be suriye hastand.*
"These countries are planning to attack Syria."

(3) *Fardā tahrimhā e`māl xāhad šod.*
"Sanctions will be imposed tomorrow."

(4) *Agar budje be šerkat taxsis miyāft, sāzmān varšekast nemišod.*
"If the budget had been allocated to the company, it would not have gone bankrupt."

In addition to the above structures, claims can be made in the forms of presuppositional structures. According to 6 different types of presupposition [72], active presupposition and lexical presupposition consist of words (such as *'know'*, *'regret'*, *'realize'*, *'start'*, *'stop'*, *'again'* etc.) showing an assumption about taking an action in the past. Examples 5 and 6 presuppose actions taken only in the past.

(5) *In kešvar mojaddad e`māle tahrimhā rā elayhe Irān āgāz kard.*
"This country again imposing sanctions against Iran starts"

(6) *Dolat digar be panāhandegān ejāzeye vorud nemidahad.*
"The government no longer allows refugees to enter."

2) Prediction claims

Prediction claims include syntactic and semantic structures predicting events in the future. These kinds of simple and complex sentences (exemplified in 7 to 9) can contain adverbs for future (such as *by the end of this week/month/year, soon, in 2 days, etc.*). Moreover, some expressions or words showing a prediction or an expectation (such as *it is predicted/ expected that..., it is possible/likely...., etc.*) given in 10. Future tense (as seen in 7), present tense (illustrated in 8 and 9), and subjunctive forms (as shown in 10 and 11) are used for Persian verbs to show taking actions in the future.

(7) *Mo`āmelāte bāzāre sahām tā pāyāne sāle jāri tahte ta`sire tavarrom qarār xāhad gereft.*
"Stock market transactions will be affected by inflation by the end of this year."

(8) *Bā taxsise budje ta 2 māhe āyande proje rā be etmām miresānim.*
"By allocating the budget, we will complete the project in the next 2 months."

(9) *Agar budje taxsis yābad, tā 2 māhe āyande proje rā be etmām miresānim.*
"If the budget is allocated, we will complete the project in the next two months."

(10) *Pi`šbini mišavad pišrafte in proje emsāl be biš az 60% beresad.*
"The progress of this project is expected to reach more than 60% this year."

(11) *Ehtemāl dārad āmrikā dar āyandeye nazdik bā jange dāxeli movājeh šavad.*
"It is possible America face a civil war in the near future."

3) Support/Oppose claims

This type of claim encompasses statements that either support, oppose, or remain neutral regarding a particular issue or an individual's opinion. Such claims play a crucial role in expressing different perspectives on a subject matter and can have varying degrees of impact on the overall argument. By presenting viewpoints that align with or contradict a given position, support/oppose claims contribute significantly to the discourse surrounding the topic under consideration (in 12 to 14).

(12) *Rand Paul moxālefe tavāfoqe haste`i bā Iran ast.*
"Rand Paul opposes the nuclear deal with Iran."

(13) *Demokrāthā be lāyeheye zirsāxtha ra`ye moxālef dādand.*
"Democrats voted against the infrastructure bill."

(14) *Donald Trump farmāne ejrā`i e`māle tahrimhāye jadid alayhe Irān rā emzā kard.*
"Donald Trump signed an executive order imposing new sanctions on Iran."

4) Causation/Correlation claims

This type of claim aims to capture sentences asserting at least 2 events occurring. In causation claims, one event causes occurring another one (as shown in 15 to 17). In correlation claims, there is a correlation between two events (as given in 18). To make this claim, if-then structures, prepositional phrases to express one of the events, and causative verbs can be used in Persian.

(15) *Agar budje be šerkat taxsis miyāft, sāzmān varsekast nemisod.*

"If the budget had been allocated to the company last year, it would not have gone bankrupt."

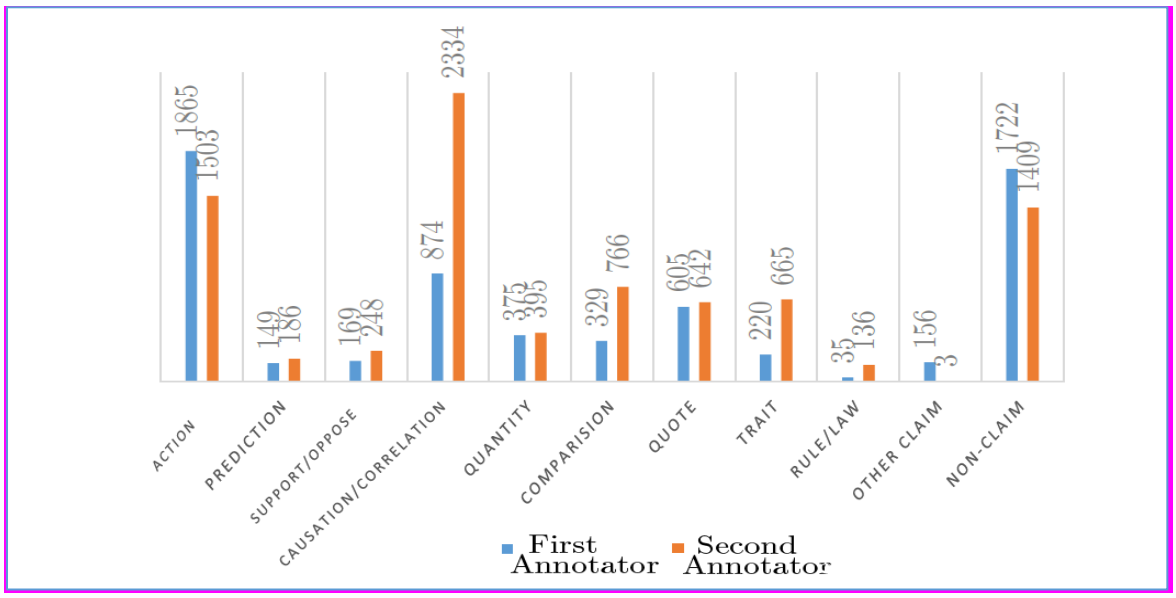

Figure 1. The distribution of 11 classes labeled by annotator 1 and annotator

(16) *40% az šerkathā bā e`māle tahrimhā varšekast sodand.*
"40% of companies have gone bankrupt due to sanctions."

(17) *Vāksane koronā bā`ese nābārvari mišavad.*
"Covid-19 vaccine causes infertility."

(18) *Har zamān haddeaqal hoquq rā afzāyeš dādim, rošdi dar mašāqel mošāhede šod.*
"Every time we've increased the minimum wage, we've seen a growth in jobs."

5) Quantity claims

This subcategory encompasses ratio and percentage, ranking, date, and numerical and statistical analyses. The below examples (19 to 22) show this kind of claim.

(19) *Nerxe bikāri dar Irān sale gozašte 9.5 darsad bud.*
"Last year, unemployment rate was 9.5 percent in Iran."

(20) *Mā dovvomin sāderkonandeye naft dar jahān hastim.*
"We are the second oil exporter in the world."

(21) *Ānhā 10 sāl māliyāt pardāxt nakardeand.*
"They have not paid taxes for 10 years."

(22) *Tahrimhā 40 miliyārd dolār manābe`e arzi rā masdud kard.*
"Sanctions froze $ 40 billion worth of foreign currencies."

(28) *Ra`is jomhure āmrikā e`lām kard tahrimhāye jadidi alayhe bānke melli Irān emruz e`māl mišavad.*
"The president of America said new sanctions against Iran's national bank will be imposed today."

8) Trait claim

6) Comparison claims

This subcategory includes all comparative structures such as the comparison between 2 or more things (as seen in 23), relationship between qualities in/over time (in 24), uniqueness (as given in 25), similarity and difference (exemplified in 26).

(23) *Espāniyā bištarin tedāde javānāne bikār rā dārad.*
"Spain has the most unemployed young people."

(24) *Nerxe bikāri dar dolate fe`li kamtar az dolate qabli ast.*
"The unemployment rate in the current government is lower than that in the previous government."

(25) *Āmrikā tanhā kesvari ast ke nerxe bikāri dar ān sefr ast.*
"America is the only country where the unemployment rate is zero."

(26) *Bar xalāfe dolate qabli Nerxe bikari dar dolate fe`li xeili pāyin ast.*
"Unlike the previous government, the unemployment rate in the current government is very low."

7) Quote claims

This class encompass claims that repeat or paraphrase what an entity said. The following examples show a quote claim and its paraphrase.

(27) *Trump goft: tahrimhāye jadid bānke melli Irān rā emruz e`māl mikonim."*
"Trump said: we impose new sanctions against Iran's national bank today."

This type of claim covers different properties of an entity such as strength, weakness, capability, qualification etc. using especial verbs such as "able, be capable, can (as seen in 29) and modifiers such as adjective phrases (as in 30).

(29) *In kešvar qāder ast dar āyandeh`i nazdik be selāhe haste`i dast yabad.*

"This country is capable of acquiring nuclear weapons in the near future."

(30) *Išān za'iftarin ra'is jomhure tārixe in kesvar hastand.*
"He is the weakest president in the history of this country."

9) Law/Rule claim

This type of claim contains statements that express laws and regulations and consider actions permissible and impermissible. Examples 31 and 32 show this kind of claim.

(31) *Dolat be šerkathāye xāreji ejāze midahad dar bāzar mošārekat konand.*
"The government allows foreign companies to participate in its market."

(32) *Sāxtosāz dar in rustā qeire qānuni ast.*
"The constructions of this village are illegal."

10) Other claims

This type of claim contains statements that do not fit into any of the previous categories (as given in 33 and

(33) *Āmāre bikārān čālešbarangiz va negarānkonande ast.*
"Unemployment statistics are challenging and worrying."

(34) *Hadafe mā jang nist.*
"Our goal is not war."

11) Non-claim

There are some tweets such as people's personal opinion, personal experiences, advises, poems etc. that are not considered as claim. We define a non-claim label to annotate these sentences (as exemplified in 35 and 36).

(35) *Sāle no mobārak.*
"Happy new year."

(36) *Inbār behtarin tajrobeye safaram ra dāštam.*
"I had the best travel experience this time."

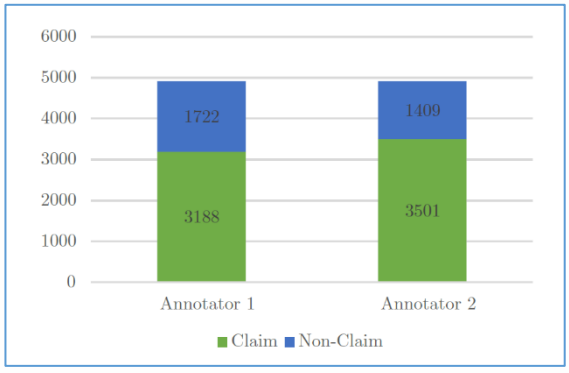

Figure 2. Distribution of binary class (claim/non-claim)

IV. METHODOLOGY

To ensure diversity in our dataset, we initially sampled tweets from 120 accounts. We filtered these tweets to include only those in Persian, excluding any containing non-Persian texts. Additionally, we applied a length threshold, removing sentences with fewer than 8 words. We selected tweets with significant engagement, specifically those with over 30 retweets and 30 likes. Finally, we included tweets from the 60 most followed public figures on Twitter in Iran, encompassing politicians, news and television presenters, actors, and popular social media influencers. This dataset comprises tweets posted between May 2016 and January 2020.

The shortest and longest tweets in the corpus consist of 8 and 70 tokens, respectively, with an average tweet length of approximately 40 tokens. In this study, we employed classification techniques for the claim detection task, exploring both binary and multi-label classification approaches. Subsequent sections will provide detailed information on corpus statistics and the model used.

A. Data and Annotation

To annotate tweets, we utilized Doccano, an open-source data labeling tool tailored for machine learning practitioners. This platform supports collaborative annotations and accommodates various languages. Initially, we developed a preliminary annotation guideline based on linguistic analysis of 200 tweets. We refined this guideline through three iterations to establish the final annotation criteria. Two annotators were provided with the guideline, detailed explanations, and examples for 11 categories of claims.

The agreement percentage among labels was 33.90%, while the agreement percentage for at least one label was 64.69%. For the binary claim/non-claim annotation task, Cohen's Kappa statistic [73] was calculated to be 0.58, and Krippendorff's alpha [74] was also computed, resulting in a value of 0.57. These metrics demonstrate a significant improvement over [27]. Unlike their approach of mapping some claim types into the non-claim category to achieve higher agreement metrics, we categorized all types of claims as claims in our binary classification.

Figure 1 illustrates the distribution of the 11 annotated classes by annotators 1 and 2, showcasing high agreement in distinguishing claim classes from non-claim classes. Figure 2 displays the distribution of tweets in the binary claim class. The figures highlight that while there is substantial agreement between annotators, most disagreements occurred in categorizing the causation/correlation class.

V. EXPERIMENTS

We begin by training a model to differentiate between claims and non-claims within the dataset, considering both binary and multi-label classifications. This section starts with a discussion on the model and experimental setup. Following this, we present the outcomes of each classifier under both binary and multi-label conditions. Lastly, we delve into the model's errors, paving the way for future studies on the topic.

Model

We utilized a transformer-based model, ParsBert [76], for conducting experiments on both binary and multi-label classification tasks. ParsBert, a monolingual language model tailored for the Persian language,

shares its architecture with Bert [77]. It was initially pre-trained on various texts including news, novels, and scientific documents. We fine-tuned ParsBert specifically using the claim tweet corpus, augmenting it with a fully-connected network to align ParsBert's outputs with the tag space.

TABLE I. THE RESULTS FOR CLAIM DETECTION EXPERIMENTS, SEPARETD INTO BINARY MULTILABEL EVALUATIONS. THE BEST RESULT IS BOLDED

Evaluation	Task	Precision	Recall	F1
Binary classification	Union Labels	83.63%	74.87%	79.00%
	Intersection Labels	90.78%	87.95%	**89.34%**
Multi-label classification	Union Labels	77.70%	67.14%	69.34%
	Intersection Labels	77.27%	63.64%	**69.79%**
	Annotator 1 Labels	67.49%	51.17%	58.20%
	Annotator 2 Labels	69.60%	58.81%	62.56%

Experimental Setting

The corpus consists of 197,480 tokens and 4,910 tweets. We divided the corpus into training (60%), validation (20%), and test (20%) sets. The model was optimized using Adam [78] with a learning rate of 5e-05, a batch size of 32, and trained for 10 epochs. Implementation was done using the TensorFlow library [79]. Hyperparameters were tuned based on performance metrics on the validation set, aiming for the highest F1-score. A dropout rate of 0.1 was applied, and the output layer utilized 11 sigmoid functions to predict tweet labels. Binary cross-entropy served as the loss function.

Results

Table 1 presents the results for both binary and multi-label classifications on the test set. For the binary classification task, we assessed two sets of gold tags: one comprising the union and the other the intersection of annotators' labels. In the multi-label classification task, we compared predicted labels against each annotator's tag set.

The results indicate that the intersection label approach achieved a more balanced precision and recall, yielding the highest F-score in the binary setting at 89.34%. Similarly, in the multi-label classification, the intersection label approach outperformed the union label approach by 0.45%.

In the intersection dataset, the number of reference tags per instance ranged from 0 to 4, averaging 1.12 tags per instance. In contrast, the union dataset exhibited a wider range of gold tags, ranging from 1 to 6 tags per instance, averaging 2.26 tags per instance.

Notably, our primary dataset is the union dataset, which includes all tags identified by each annotator. This approach ensures a more comprehensive and diverse representation of claim types present in the data. By aggregating tags from multiple annotators, we capture a richer spectrum of claim categories and their occurrences within tweets.

Analysis

Based on the model's optimal performance in both binary and multi-label contexts, we infer that the model effectively identified sentences containing advisory content and correctly classified them as non-claims, as shown in (a).

(a) Din ādam rā mostaqel bār miāvarad va u rā rošd midahad.

“Religion makes a person independent and develops him.”

Transformer-based models benefit from contextualized embeddings that effectively capture syntactic relationships [80, 81]. Moreover, these models demonstrate proficiency in recognizing various syntactic structures. They accurately identify causative structures and future tenses, correctly classifying them as causation and prediction claims.

Causative structures typically denote actions caused to occur, prompting the model to assign the action claim label to such sentences. For instance, the model correctly labeled sentence (b) as involving causative/correlation and action claims.

(*b)Vaz'iyate bohrāni dar in šahr natijeye adame tavajoh be hošdārhā ast!*

“The critical situation in this city is the result of not paying attention to the warnings!”

Furthermore, the model effectively utilizes morphological features to discern comparative and superlative adjectives, accurately categorizing sentences as comparison claims (c).

(c)Tasmimāte ra'is jomhure fe'li bohrāne eqtesādi rā badtar mikonad.

“The current president's decisions are worsening economic crisis.”

Additionally, the model adeptly identifies words and punctuation indicating quotations or paraphrases, enabling it to accurately recognize quote claims, as demonstrated in sentences (d) and (e).

(d) Irnā: 30% as bimārāne viruse koronā dar in šahr mosāfer hastand.

“Irna: 30% of COVID-19 patients in this city are travelers.”

(e) Keyhān nevešt ke xabarnegāre panāhande eslāhāt talab ast.

“Kayhan wrote that the refugee journalist is a reformist.”

The model also demonstrates strong capability in detecting Law/Rule claims, effectively recognizing words that express rules within tweets (f).

(f) Tarhe dolat barāye sāderāte xodro tasvib sod.

“The government's plan to export cars was approved.”

The model adeptly captures words that express support and opposition, as evidenced in sentence (g).

(g) Namāyandegān az tarhe jadide dolat hemāyat kardand.

"Members of parliaments supported the government's new plan."

The model effectively captures tense as a syntactic feature, significantly aiding in the detection of prediction claims, exemplified in (h).

(h) In kešvar az haqe mardome xod darbāreye mozākerāte haste'i kutāh naxāhad āmad.

"This country will never step back from the rights of their people in nuclear negotiations."

In spite of the aforementioned points, there are some frequent and important errors in the performance of the model. Although the attention heads in transformer-based models could well capture the syntactic structures of sentences [80, 81], the model made several errors:

1. Considering morphological features, the model labeled some tweets containing ordinal numbers as quantity claim. For instance, (i) is not a claim but the model's tag is quantity claim.

(i) Emruz noxostin ruze qarne pānzdahom ast.

"Today is the first day of the 15th century."

2. As we mentioned in section 3.1.4, causative structures can be formed by if-then structures in Persian. However, all if-then structures are not causative. The model incorrectly labeled (j) as causation/correlation claim.

(j) Agar tavāne moqābele ba vaz'iyate bohraniye fe'li ra nadārid, este'fā dahid.

"If you cannot cope with the current critical situation, resign."

3. There are some tweets incorrectly labeled by annotators. However, the model could correctly label them. For instance, the model correctly tagged (k) as trait claim.

(k) Modire bānk ideye eqtesādi nadārad.

"The bank manager has no economic idea."

VI. Conclusion and Future Work

We've devised the initial annotation schema for Persian claim detection, grounded in linguistic analysis. Our annotated dataset comprises sentences extracted from Persian tweets by Iranian public figures. Using the transformer-based ParsBert model, we conducted experiments on binary and multi-label claim detection tasks. The results demonstrate the model's adeptness in capturing syntactic features to identify claim types. However, the model exhibits weaknesses in semantically analyzing sentences to discern claim types with identical syntactic structures.

Our overarching goal is to develop a fact-checking tool tailored for Persian tweets. The claim detection model serves as a foundational element for subsequent tasks, including stance detection and fake news identification. By harnessing this model, we aim to facilitate thorough and precise analyses in verifying tweets and assessing information credibility.

References

[1] Herman, Edward S., and Noam Chomsky. Manufacturing consent: The political economy of the mass media. Random House, 2010.

[2] Bradshaw, Samantha, and Philip N. Howard. "The global disinformation order: 2019 global inventory of organised social media manipulation." (2019).

[3] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19, no. 1 (2017): 22-36.

[4] Zubiaga, Arkaitz, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. "Detection and resolution of rumours in social media: A survey." ACM Computing Surveys (CSUR) 51, no. 2 (2018): 1-36.

[5] Bastick, Zach. "Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation." Computers in human behavior 116 (2021): 106633.

[6] Ecker, Ullrich KH, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga, and Michelle A. Amazeen. "The psychological drivers of misinformation belief and its resistance to correction." Nature Reviews Psychology 1, no. 1 (2022): 13-29.

[7] Plotnick, Linda, Starr Hiltz, Sukeshini Grandhi, and Julie Dugdale. "Real or fake? User behavior and attitudes related to determining the veracity of social media posts." arXiv preprint arXiv:1904.03989 (2019).

[8] Featherstone, Jieyu Ding, C. A. Davis, and J. Zhang. "Correcting Vaccine Misinformation on Social Media Using Fact-checking Labels."

In APHA's 2019 annual meeting and expo. 2019.

[9] Roozenbeek, Jon, Claudia R. Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. "Susceptibility to misinformation about COVID-19 around the world." Royal Society open science 7, no. 10 (2020): 201199.

[10] Greene, Ciara M., and Gillian Murphy. "Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation." Journal of Experimental Psychology: Applied (2021).

[11] Zhou, Cheng, Haoxin Xiu, Yuqiu Wang, and Xinyao Yu. "Characterizing the dissemination of misinformation on social media in health emergencies: An empirical study based on COVID-19." Information Processing and Management 58, no. 4 (2021): 102554.

[12] Van Der Linden, Sander, Costas Panagopoulos, and Jon Roozenbeek. "You are fake news: political bias in perceptions of fake news." Media, Culture and Society 42, no. 3 (2020): 460-470.

[13] Muqsith, Munadhil Abdul, R. Ridho Pratomo, Anna Gustina Zaina, and Ana Kuswanti. "Fake News as a Tool to Manipulate the Public with False Information." In 2nd International Indonesia Conference on Interdisciplinary Studies (IICIS 2021), pp. 118-127. Atlantis Press, 2021.

[14] Polak, Mateusz. "The misinformation effect in financial markets: An emerging issue in behavioural fianance." e-Finanse: Financial Internet Quarterly 8, no. 3 (2012): 55-61.

[15] Kogan, Shimon, Tobias J. Moskowitz, and Marina Niessner. "Fake news: Evidence from financial markets." Available at SSRN 3237763 (2019).

[16] Rich, Patrick R., and Maria S. Zaragoza. "The continued inuence of implied and explicitly stated misinformation in news

reports." Journal of experimental psychology: learning, memory, and cognition 42, no. 1 (2016): 62.

[17] Thorson, Emily. "Belief echoes: The persistent effects of corrected misinformation." Political Communication 33, no. 3 (2016): 460-480.

[18] Desai, Saoirse Connor, and Stian Reimers. "Some misinformation is more easily countered: An experiment on the continued inuence effect." In CogSci. 2018.

[19] Wang, Xuezhi, Cong Yu, Simon Baumgartner, and Flip Korn. "Relevant document discovery for fact-checking articles." In Companion Proceedings of the The Web Conference 2018, pp. 525-533. 2018.

[20] Jo, Saehan, Immanuel Trummer, Weicheng Yu, Xuezhi Wang, Cong Yu, Daniel Liu, and Niyati Mehta. "Verifying text summaries of relational data sets." In Proceedings of the 2019 International Conference on Management of Data, pp. 299-316. 2019.

[21] Trokhymovych, Mykola, and Diego Saez-Trumper. "Wikicheck: An end-to-end open source automatic fact-checking api based on wikipedia." In Proceedings of the 30th ACM International Conference on Information and Knowledge Management, pp. 4155-4164. 2021.

[22] Graves, Lucas, Brendan Nyhan, and Jason Reier. "Understanding innovations in journalistic practice: A field experiment examining motivations for fact-checking." Journal of Communication 66, no. 1 (2016): 102-138.

[23] Graves, D. "Understanding the promise and limits of automated factchecking." (2018).

[24] Lippi, Marco, and Paolo Torroni. "Context-independent claim detection for argument mining." In Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.

[25] Stab, Christian, and Iryna Gurevych. "Parsing argumentation structures in persuasive essays." Computational Linguistics 43, no. 3 (2017): 619-659.

[26] Konstantinovskiy, Lev, Oliver Price, Mevan Babakar, and Arkaitz Zubiaga. "Toward automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection." Digital Threats: Research and Practice 2, no. 2 (2021): 1-16.

[27] Shin, Jieun, Lian Jian, Kevin Driscoll, and Francois Bar. "The diffusion of misinformation on social media: Temporal pattern, message, and source." Computers in Human Behavior 83 (2018): 278-287.

[28] Wu, Liang, Fred Morstatter, Kathleen M. Carley, and Huan Liu. "Misinformation in social media: definition, manipulation, and detection." ACM SIGKDD Explorations Newsletter 21, no. 2 (2019): 80-90.

[29] Levy, Ran, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. "Context dependent claim detection." In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 1489-1500. 2014.

[30] Lippi, Marco, and Paolo Torroni. "Argument mining: A machine learning perspective." In International Workshop on Theory and Applications of Formal Argumentation, pp. 163-176. Springer, Cham, 2015.

[31] Vlachos, Andreas, and Sebastian Riedel. "Fact checking: Task definition and dataset construction." In Proceedings of the ACL 2014 workshop on language technologies and computational social science, pp. 18-22. 2014.

[32] Daniel, Anna, Terry Flew, and Christina Spurgeon. "The promise of computational journalism." In Proceedings of the Australian and New Zealand Communication Association (ANZCA) Conference 2010: Media, Democracy and Change, pp. 1-19. Australia and New Zealand Communication Association, 2010.

[33] Cohen, Sarah, Chengkai Li, and Jun Yang. "C. Yu. Computational journalism: A call to arms to database researchers." CIDR, 2011.

[34] Graves, D. "Understanding the promise and limits of automated factchecking." (2018).

[35] Guo, Zhijiang, Michael Schlichtkrull, and Andreas Vlachos. "A survey on automated fact-checking." Transactions of the Association for Computational Linguistics 10 (2022): 178-206.

[36] Lazarski, Eric, Mahmood Al-Khassaweneh, and Cynthia Howard. "Using NLP for Fact Checking: A Survey." Designs 5, no. 3 (2021)

[37] Walton, Douglas. "Argumentation theory: A very short introduction." In Argumentation in artificial intelligence, pp. 1-22. Springer, Boston, MA, 2009.

[38] Moens, Marie-Francine, Erik Boiy, Raquel Mochales Palau, and Chris Reed. "Automatic detection of arguments in legal texts." In Proceedings of the 11th international conference on Artificial intelligence and law, pp. 225-230. 2007.

[39] Palau, Raquel Mochales, and Marie-Francine Moens. "Argumentation mining: the detection, classification and structure of arguments in text." In Proceedings of the 12th international conference on arti-ficial intelligence and law, pp. 98-107. 2009.

[40] Saint-Dizier, Patrick. "Processing natural language arguments with the< TextCoop> platform." Argument and Computation 3, no. 1 (2012): 49-82.

[41] Cabrio, Elena, and Serena Villata. "Natural language arguments: A combined approach." In ECAI 2012, pp. 205-210. IOS Press, 2012.

[42] Stylianou, Nikolaos, and Ioannis Vlahavas. "Transformed: End-to-End transformers for evidence-based medicine and argument mining in medical literature." Journal of Biomedical Informatics 117 (2021):103767.

[43] Al Khatib, Khalid, Tirthankar Ghosal, Yufang Hou, Anita deWaard, and Dayne Freitag. "Argument mining for scholarly document processing: Taking stock and looking ahead." In Proceedings of the Second Workshop on Scholarly Document Processing, pp. 56-65. 2021.

[44] Mebane, Waleed. "Detection of Claims and Supporting Evidence in Wikipedia Articles on Controversial Topics." PhD diss., 2017.

[45] Lawrence, John, and Chris Reed. "Combining argument mining techniques." In Proceedings of the 2nd Workshop on Argumentation Mining, pp. 127-136. 2015.17

[46] Budzynska, Katarzyna, Mathilde Janier, Juyeon Kang, Chris Reed, Patrick Saint-Dizier, Manfred Stede, and Olena Yaskorska. "Towards argument mining from dialogue." In Computational Models of Argument, pp. 185-196. IOS Press, 2014.

[47] Patwari, Ayush, Dan Goldwasser, and Saurabh Bagchi. "Tathya: A multi-classifier system for detecting check-worthy statements in political debates." In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 2259-2262. 2017.

[48] Hassan, Naeemul, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan et al. "Claimbuster: The first-ever end-to-end fact-checking system." Proceedings of the VLDB Endowment 10, no. 12 (2017): 1945-1948.

[49] Gencheva, Pepa, Preslav Nakov, Liuis Marquez, Alberto Barron Cede, and Ivan Koychev. "A context-aware approach for de-

tecting worth-checking claims in political debates." In Proceedings of the International Conference Recent Advances in Natural Language Processing,

RANLP 2017, pp. 267-276. 2017.

[50] Hanto, Vigdis, and Mats Tostrup. "Towards Automated Fake News Classification-On Building Collections for Claim Analysis Research." Master's thesis, NTNU, 2018.

[51] Atanasova, Pepa, Preslav Nakov, Georgi Karadzhov, Mitra Mohtarami, and Giovanni Da San Martino. "Overview of the CLEF-2019 CheckThat! Lab: Automatic Identification and Verification of Claims. Task 1: Check-Worthiness." CLEF (Working Notes) 2380 (2019).

[52] Shaar, Shaden, Alex Nikolov, Nikolay Babulkov, Firoj Alam, Alberto Barron-Cedeno, Tamer Elsayed, Maram Hasanain et al. "Overview of CheckThat! 2020 English: Automatic identification and verification of claims in social media." In CLEF (Working Notes). 2020.

[53] Atanasova, Pepa, Alberto Barron-Cedeno, Tamer Elsayed, Reem Suwaileh, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, and Preslav Nakov. "Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims. Task 1: Check-worthiness." arXiv preprint arXiv:1808.05542 (2018).

[54] Barron-codeno, Alberto, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari et al. "Overview of CheckThat! 2020: Automatic identification and verification of claims in social media." In International Conference of the Cross-Language Evaluation Forum for European Languages, pp. 215-236. Springer, Cham, 2020.

[55] Kartal, Yavuz Selim, and Mucahid Kutlu. "TrClaim-19: The first collection for Turkish check-worthy claim detection with annotator rationales." In Proceedings of the 24th Conference on Computational Natural Language Learning, pp. 386-395. 2020.

[56] Berendt, Bettina, Peter Burger, Rafael Hautekiet, Jan Jagers, Alexander Pleijter, and Peter Van Aelst. "FactRank: Developing automated claim detection for Dutch-language fact-checkers." Online Social Networks and Media 22 (2021): 100113.

[57] Atanasova, Pepa, Preslav Nakov, Luis Marquez, Alberto Barron-Cede~no, Georgi Karadzhov, Tsvetomila Mihaylova, Mitra Mohtarami, and James Glass. "Automatic fact-checking using context and discourse information." Journal of Data and Information Quality (JDIQ) 11, no. 3 (2019): 1-27.18

[58] Williams, Evan, Paul Rodrigues, and Valerie Novak. "Accenture at CheckThat! 2020: If you say so: post-hoc fact-checking of claims using transformer-based models." arXiv preprint arXiv:2009.02431 (2020).

[59] Hasanain, Maram, and Tamer Elsayed. "bigIR at CheckThat! 2020: Multilingual BERT for Ranking Arabic Tweets by Check-worthiness." In CLEF (Working Notes). (2020).

[60] Goudas, Theodosis, Christos Louizos, Georgios Petasis, and Vangelis Karkaletsis. "Argument extraction from news, blogs, and social media." In Hellenic Conference on Artificail Intelligence, pp. 287-299. Springer, Cham, (2014).

[61] Sardianos, Christos, Ioannis Manousos Katakis, Georgios Petasis, and Vangelis Karkaletsis. "Argument extraction from news." In Proceedings of the 2nd Workshop on Argumentation Mining, pp. 56-66. (2015).

[62] Levy, Ran, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim. "Unsupervised corpus {wide claim detection." In Proceedings of the 4th Workshop on Argument Mining,

pp. 79-84. (2017).

[63] Josue, Caraballo. \PolitiTax A Taxonomy of Political Claims.". (2018). https://https://perma.cc/4RQF-FCPV.

[64] Zarharan, Majid, Samane Ahangar, Fateme Sadat Rezvaninejad, Mahdi Lotfi Bidhendi, Mohammad Taher Pilevar, Behrouz Minaei, and Sauleh Eetemadi. "Persian Stance Classification Data Set." In TTO. 2019.

[65] Samadi, Mohammadreza, Maryam Mousavian, and Saeedeh Momtazi. "Persian fake news detection: Neural representation and classification at word and text levels." Transactions on Asian and Low-Resource Language Information Processing 21, no. 1 (2021): 1-11.

[66] Mottaghi, Vahid, Mahdi Esmaeili, Ghasem Ali Bazaee, and Mohammadali Afshar Kazemi. "A decision-making system for detecting fake persian news by improving deep learning algorithms{case study of Covid-19 news." Journal of applied research on industrial engineering 8, no. Special Issue (2021): 1-17.

[67] Sadr, Mohammad Mohsen, Afshin Mousavi Chelak, Soraya Ziaei, and Jafar Tanha. "A predictive model based on machine learning methods to recognize fake persian news on twitter." International Journal of Nonlinear Analysis and Applications 11 (2020): 119-128.

[68] Sadr, Mohammad Mohsen. "The Use of LSTM Neural Network to Detect Fake News on Persian Twitter." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 11 (2021): 6658-6668.

[69] Jahanbakhsh-Nagadeh, Zoleikha, Mohammad-Reza Feizi-Derakhshi, and Arash Sharifi. "A semi-supervised model for Persian rumor verification based on content information." Multimedia Tools and Applications 80, no. 28 (2021): 35267-35295.

[70] Jahanbakhsh-Nagadeh, Zoleikha, Mohammad-Reza Feizi-Derakhshi, and Arash Sharifi. "A Deep Content-Based Model for Persian Rumor Verification." Transactions on Asian and Low-Resource Language Information Processing 21, no. 1 (2021): 1-29.19

[71] Thoyyibah, Luthfiyatun. "Presupposition Triggers-a Comparative Analysis Between Oral News and Written Online News Discourse." JALL (Journal of Applied Linguistics and Literacy) 1, no. 2 (2017): 10-23.

[72] Cohen, Jacob. "A coefficient of agreement for nominal scales." Educational and psychological measurement 20, no. 1 (1960): 37-46.

[73] Krippendorff , Klaus. "Validity in content analysis." (1980): 69.

[74] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[75] Farahani, Mehrdad, Mohammad Gharachorloo, Marzieh Farahani, and Mohammad Manthouri. "Parsbert: Transformer-based model for persian language understanding." Neural Processing Letters 53, no. 6 (2021): 3831-3847.

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

[77] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[78] Abadi, Martin, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado et al. "Tensorow: Largescale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016).

[79] Tenney, Ian, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R. Thomas McCoy, Najoung Kim et al. "What do you

learn from context? probing for sentence structure in contextualized word representations." arXiv preprint arXiv:1905.06316 (2019).

[80] Hewitt, John, and Christopher D. Manning. "A structural probe for finding syntax in word representations". In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4129-4138. (2019).

[81] Ehsan Doostmohammadi, Minoo Nassajian, and Adel Rahimi. "Persian Ezafe Recognition Using Transformers and Its Role in Part-Of-Speech Tagging". In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 961{971, Online. Association for Computational Linguistics. (2020).

Mohammad Hadi Bokaei received the B.Sc. and M.Sc. degrees from Iranian University of Science and Technology (2008) and Sharif University of Technology (2011), respectively, and the Ph.D. degree in Artificial Intelligence from Sharif University of Technology in 2015. He is currently an Associate Professor at the Iran Telecommunication Research Center, Tehran, Iran. His research interests include the area of Deep Learning, Machine Learning, Spoken Language Processing and Natural Language Processing.

Minoo Nassajian received the B.Sc. and M.Sc. degrees from Azad University (2012) and Sharif University of Technology (2019) respectively. Se is currently a PhD student in Computational Linguistics in Charles University, Prague, Czech Republic. Her research interests include Linguistics, Corpus Linguistics and Computational Linguistics.

Mojgan Farhoodi received her B.Sc. degree in Software Engineering and her M.Sc. and Ph.D. degree in IT Engineering and IT management Respectively. She has been working as a researcher at the ICT Research Institute since 2010. Currently, she is head of AI lab and faculty member of the ICT research institute. Her areas of expertise are Information Retrieval, Data Mining, Natural Language Processing and Artificial Intelligence.

Mona Davoudi Shamsi received her B.Sc. degree from Shahid Beheshti University in the field of software engineering. Since 1380, she has been engaged in research activities at the ICT Research Institute.