BERT-Based Emotion Recognition:
A Comprehensive Study in Dari Texts

Group Number: 02

RA: Ehsanur Rahman Rhythm

ST: Mehnaz Ara Fazal

241057 Malika Muradi

21141064 Basit Hussain

Introduction:

- Low resource language with over 110 million speaker around the world
- Importance of Dari language in Businesses
- Good progress

Literature review:

- EmoPars and ArmanEmo two human labeled dataset,
    - F1-score 0.76 and o.81
- Compare Pars BERT and Multilingual BERT
    - Performance evaluated in NLP tasks
    - Pars BERT better performance

Dataset

A. Data collection:

- ArmanEmo
- Comprehensive manually labeled emotion dataset
- 7,000+ Dari Sentences
- Categorized into 7 Emotion Classes (Anger, Fear, Happiness, Hatred, Sadness, Wonder, Other)
- Source of comments from (Twitter, Instagram, Digikala (Iranian e-commerce platform))

B. Data Augmentation:

- Random Oversampling Technique

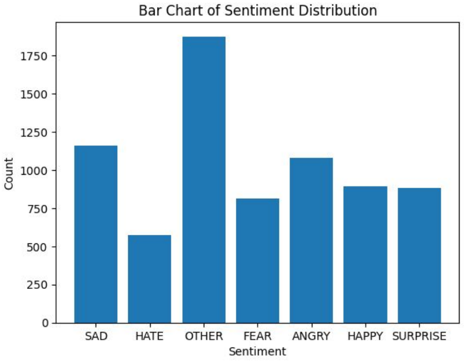


Fig 1: Sentiment Distribution before Augmentation

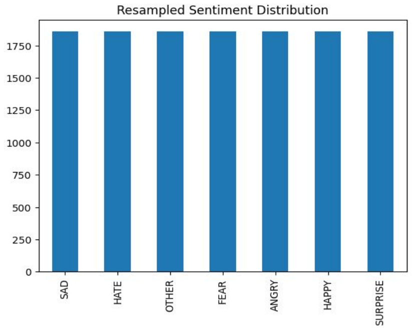


Fig 1: Sentiment Distribution after Augmentation

Dataset (Con…)

C. Data Preprocessing:

- Data Cleaning
- Tokenization Techniques (ParsBERT Tokenizer and Multilingual BERT Tokenizer)
- Mapping Tokens to Word IDs
- Representative Split for Training and Testing 85% for Training Data 15% for Testing Data

Methodology

A. ParsBERT Model:

. **Overview**

- Utilization: Pre-trained Pars-BERT model
- Domain: Emotion detection within Dari text
- Foundation: Based on Google BERT architecture
- **Pre-training Data:**
  - Extensive Persian corpora
  - Diverse writing styles (scientific, literary, journalistic)
- **Statistics**:
  - 3.9 million documents
  - 73 million sentences
  - Lexical repository of 1.3 billion words

Methodology (Con...)

Model Architecture

- Embedding Layer:
    - Maps input tokens to vectors
- Transformer Blocks:
    - 12 bidirectional blocks
    - Multi-head self-attention mechanism
- Training Details:
    - Batch size: 16
    - Learning rate: 2e-5
    - Training epochs: 4

Evaluation and Interpretability

- Evaluation Process
- LIME XAI Method

Methodology (Con...)

B. Multilingual BERT Model:

- Comparative analysis:
    - Utilization: BERT-base-multilingual-cased variant
    - Task: Emotion detection in Dari text
- Training Details:
    - Batch size: 16
    - Training epochs: 4
- Fine-tuning:
    - Dari text dataset with input IDs, token type IDs, attention masks, and emotion labels

Results and Discussion

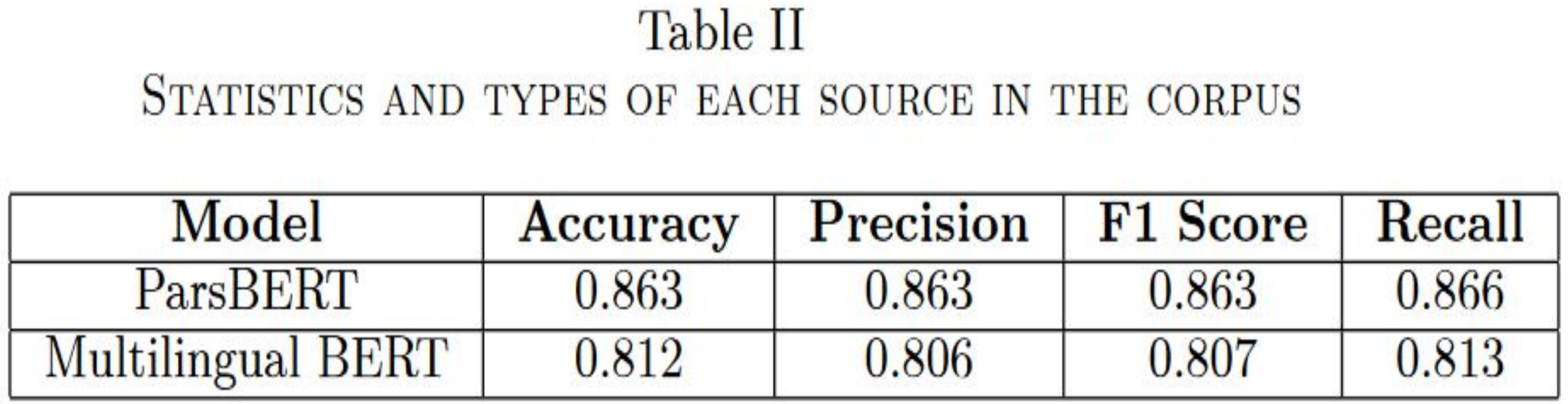
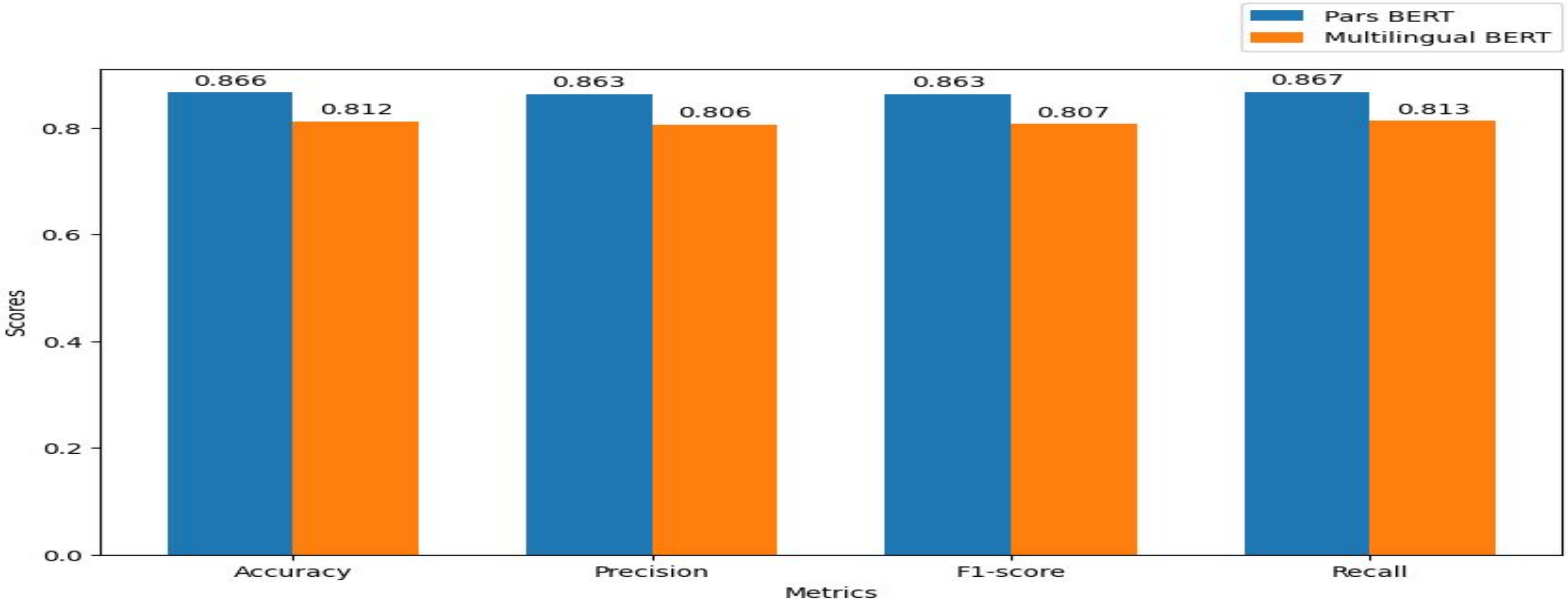


Table II

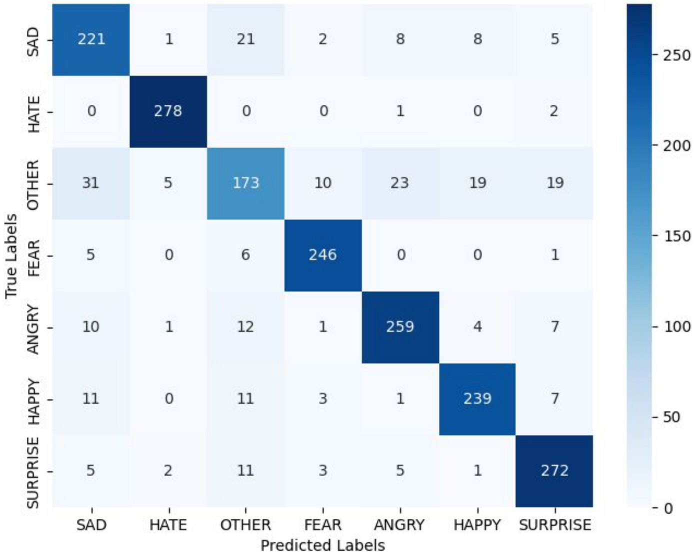
STATISTICS AND TYPES OF EACH SOURCE IN THE CORPUS

Model	**Accuracy**	**Precision**	**F1 Score**	**Recall**
ParsBERT	0.863	0.863	0.863	0.866
Multilingual BERT	0.812	0.806	0.807	0.813

Results and Discussion (Con...)

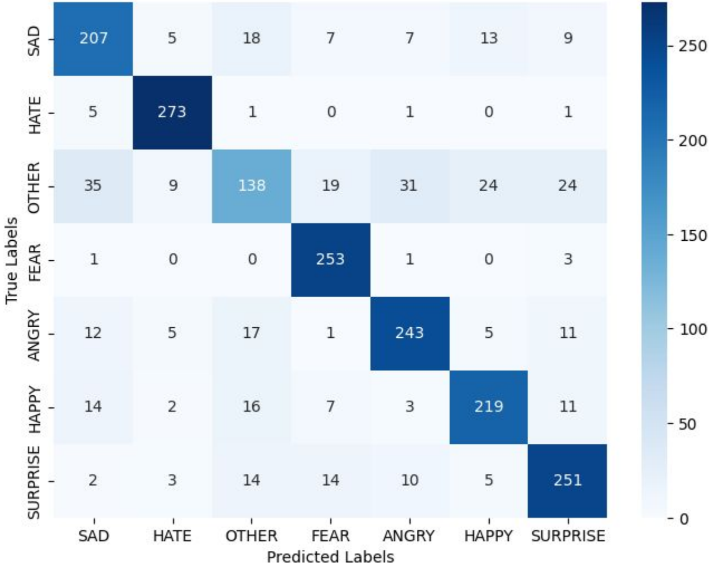


Results and Discussion (Con...)



True Labels \ Predicted Labels	SAD	HATE	OTHER	FEAR	ANGRY	HAPPY	SURPRISE
SAD	221	1	21	2	8	8	5
HATE	0	278	0	0	1	0	2
OTHER	31	5	173	10	23	19	19
FEAR	5	0	6	246	0	0	1
ANGRY	10	1	12	1	259	4	7
HAPPY	11	0	11	3	1	239	7
SURPRISE	5	2	11	3	5	1	272

Fig 1: Confusion Metric for ParsBERT Model

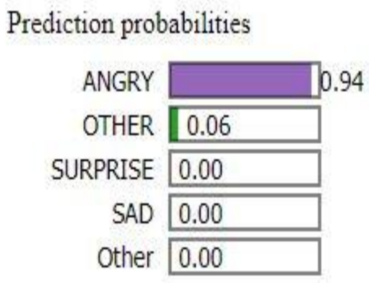


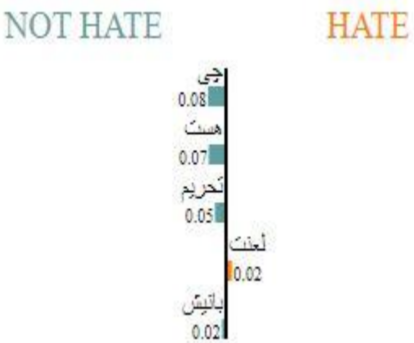
True Labels \ Predicted Labels	SAD	HATE	OTHER	FEAR	ANGRY	HAPPY	SURPRISE
SAD	207	5	18	7	7	13	9
HATE	5	273	1	0	1	0	1
OTHER	35	9	138	19	31	24	24
FEAR	1	0	0	253	1	0	3
ANGRY	12	5	17	1	243	5	11
HAPPY	14	2	16	7	3	219	11
SURPRISE	2	3	14	14	10	5	251

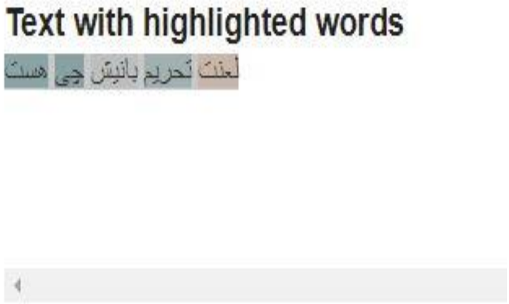
Fig 2: Confusion Metric for Multilingual BERT Model

Implementation Of Explainable AI

Explainable AI for Pars BERT model

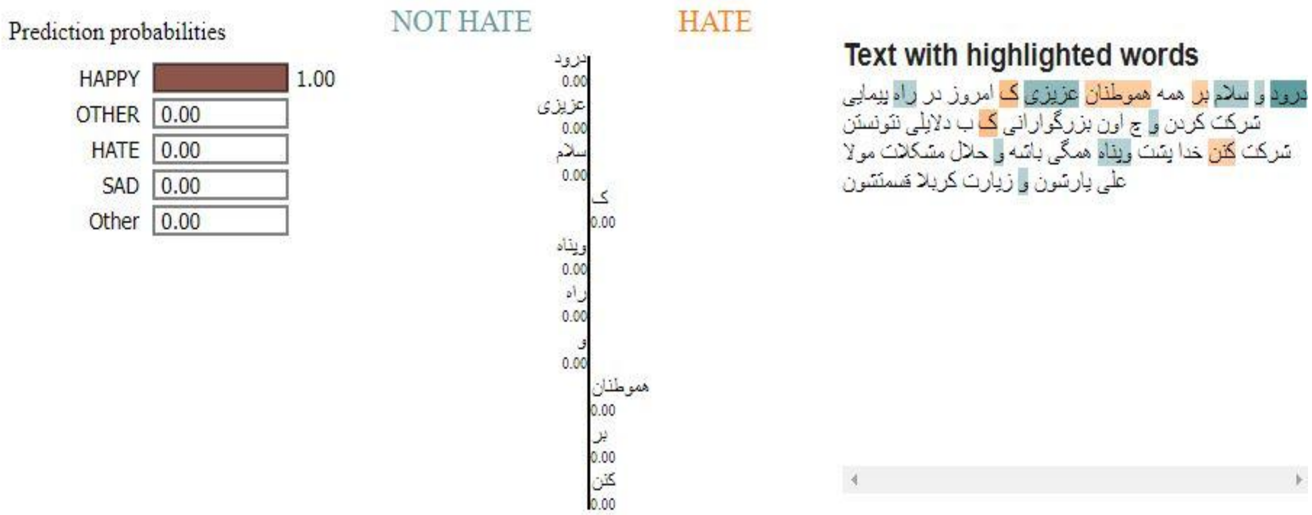





Text with highlighted words

لعنت تحریم بانیش چی هست

- Greetings and salutations to all my dear compatriots who participated in the march today and to those honorable people who could not participate for some reason, may God protect them all and solve the problems of Mullah Ali Yaar and their pilgrimage to Karbala.



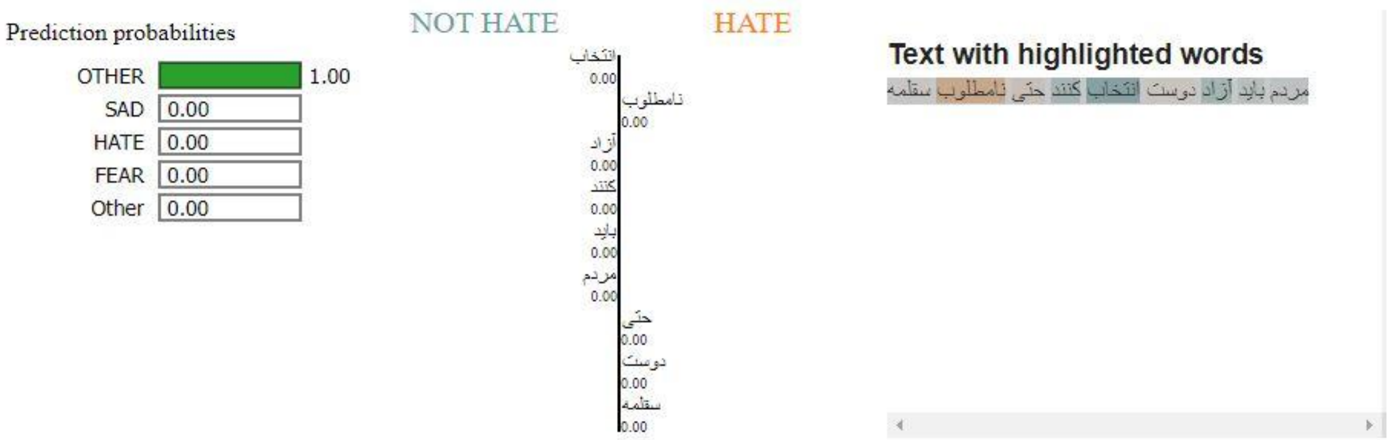
Prediction probabilities

HAPPY	1.00
OTHER	0.00
HATE	0.00
SAD	0.00
Other	0.00

Text with highlighted words

درود و سلام بر همه هموطنان عزیزی ک امروز در راه پیمایی
شرکت کردن و ج اون بزرگوارانی ک ب دلایلی نتونستن
شرکت کنن خدا پشت وپناه همگی باشه و حلال مشکلات مولا
علی یارشون و زیارت کربلا قسمتشون

- people should be free to choose friends, even when they are undesirable



Prediction probabilities

OTHER	1.00
SAD	0.00
HATE	0.00
FEAR	0.00
Other	0.00

Text with highlighted words

مردم باید آزاد دوست انتخاب کنند حتی نامطلوب سقلمه

```
explanation.save_to_file('lime-parsBERT.html')
```

THANK YOU