

КЛАССИФИКАЦИЯ ТЕКСТОВ ПО ЭМОЦИОНАЛЬНОЙ ОКРАСКЕ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ TRANSFORMERS

Ботиралиев Бахтиёр Баходир угли

(Ульяновский государственный

технический университет ,

группа: ИИПАмд-11 , направление:

Программная инженерия)

Аннотация: В статье рассматривается задача классификации эмоциональной окраски текстов с помощью преобученной модели BERT (с использованием тахрирчи-модели для узбекского языка). Реализована система обучения и оценки модели с логированием использования ресурсов памяти. Результаты показывают высокую точность классификации эмоций в многоклассовой среде.

Ключевые слова: классификация текста, BERT, узбекский язык, эмоции, глубокое обучение, Transformers, анализ текста.

1. Введение

Анализ эмоциональной составляющей текста является одной из важнейших задач современной компьютерной лингвистики и обработки естественного языка (NLP). Определение эмоциональной окраски текстов позволяет глубже понять контекст сообщений, улучшить взаимодействие с пользователями в диалоговых системах, оптимизировать рекомендательные системы и осуществлять мониторинг общественного мнения. Особую актуальность

представляет разработка систем анализа эмоций для языков с ограниченными ресурсами, к которым относится узбекский язык. Несмотря на значительный прогресс в области NLP для крупных мировых языков, таких как английский, китайский или русский, малоресурсные языки остаются недостаточно изученными, что создает значительный пробел в цифровом языковом равенстве [1].

В данной работе исследуется возможность применения современных архитектур глубокого обучения, а именно – моделей на основе трансформеров BERT, для задачи классификации эмоциональной окраски текстов на узбекском языке. Ключевым аспектом исследования является использование предварительно обученной мультязычной модели BERT с последующей тонкой настройкой на специализированном корпусе узбекских текстов.

Основные задачи исследования включают:

1. Обзор современных подходов к анализу эмоций в текстах и их применимость к узбекскому языку
2. Формирование и разметка корпуса узбекских текстов для обучения и тестирования модели
3. Разработка и реализация архитектуры классификатора на основе BERT
4. Оценка производительности модели и анализ результатов
5. Исследование эффективности использования вычислительных ресурсов при обучении и применении модели

2. Обзор литературы

2.1 Анализ эмоций в текстах

Задача анализа эмоций в текстах (emotion analysis или emotion detection) представляет собой подзадачу анализа тональности (sentiment analysis), но с

более детальной классификацией эмоциональных состояний [2]. Если анализ тональности обычно ограничивается определением положительной, отрицательной или нейтральной окраски текста, то анализ эмоций направлен на выявление конкретных эмоциональных состояний, таких как радость, грусть, гнев, страх, удивление и других [3].

Ранние подходы к анализу эмоций основывались на словарных методах, где каждому слову или фразе присваивалась определенная эмоциональная метка [4]. Позднее получили развитие статистические методы, включая SVM, Naive Bayes и логистическую регрессию [5]. С развитием глубокого обучения приоритет получили нейросетевые архитектуры, такие как рекуррентные нейронные сети (RNN), сети с долгой краткосрочной памятью (LSTM) и сверточные нейронные сети (CNN) [6]. В последние годы доминирующее положение в задачах обработки текста заняли модели на основе архитектуры Transformer, такие как BERT, GPT, RoBERTa и XLNet [7]. Эти модели демонстрируют значительное превосходство над предыдущими подходами благодаря механизму самовнимания (self-attention) и предварительному обучению на масштабных корпусах текстов.

2.2 Особенности обработки узбекского языка

Узбекский язык относится к тюркской языковой семье и имеет ряд специфических особенностей, которые необходимо учитывать при разработке систем NLP. К этим особенностям относятся:

1. **Агглютинативность** – образование грамматических форм путем последовательного присоединения аффиксов к основе слова
2. **Гармония гласных** – фонетическое явление, при котором гласные в слове гармонируются по определенным признакам
3. **Отсутствие грамматического рода**
4. **Порядок слов SOV** (подлежащее-дополнение-сказуемое)

5. **Богатая морфология** с большим количеством падежей и временных форм

Исследования в области NLP для узбекского языка ограничены по сравнению с более ресурсными языками. Тем не менее, в последние годы наблюдается рост интереса к разработке инструментов и ресурсов для обработки узбекского языка [8, 9].

Одним из прорывов стало появление мультязычных моделей трансформеров, таких как mBERT (multilingual BERT) [10], XLM-RoBERTa [11] и других, которые были предварительно обучены на корпусах текстов, включающих материалы на узбекском языке. Кроме того, появились специализированные модели для узбекского языка, такие как тахрирчи-BERT, которые демонстрируют более высокую производительность на узбекоязычных текстах [12].

2.3 Модели классификации на основе BERT

BERT (Bidirectional Encoder Representations from Transformers) представляет собой архитектуру нейронной сети, разработанную компанией Google в 2018 году [13]. Ключевыми особенностями BERT являются:

1. **Двунаправленность** – модель учитывает контекст слова как слева, так и справа от него
2. **Предварительное обучение** на масштабных корпусах текстов с использованием задач маскированного языкового моделирования (Masked Language Model, MLM) и предсказания следующего предложения (Next Sentence Prediction, NSP)
3. **Тонкая настройка** (fine-tuning) предварительно обученной модели на конкретных задачах, включая классификацию текстов

Для задачи классификации текстов по эмоциональной окраске модель BERT может быть дополнена классификационным слоем, который принимает на вход

векторное представление токена [CLS] (сгенерированное BERT) и выдает вероятностное распределение по классам эмоций [14].

Среди преимуществ использования BERT для классификации эмоций можно выделить:

- Способность к глубокому пониманию контекста и семантики текста
- Эффективная работа с длинными последовательностями
- Возможность переноса знаний (transfer learning) с одного языка на другой
- Устойчивость к омонимии и полисемии

3. Методология

3.1 Формирование датасета

Для обучения и оценки модели классификации эмоций был создан специализированный корпус текстов на узбекском языке. Корпус включает материалы из следующих источников:

1. Сообщения из социальных сетей (Twitter, Telegram, Facebook)
2. Комментарии к новостным статьям
3. Отзывы о продуктах и услугах
4. Фрагменты литературных произведений
5. Диалоги из кинофильмов и телесериалов

Общий объем собранного корпуса составил 15 000 текстовых фрагментов различной длины (от 1 до 50 предложений). Для разметки текстов по эмоциональным категориям была использована модель Плутчика [15], которая выделяет восемь базовых эмоций:

1. Радость (quvonch)

2. Доверие (ishonch)
3. Страх (qo'rquv)
4. Удивление (hayrat)
5. Печаль (g'amginlik)
6. Отвращение (nafrat)
7. Гнев (g'azab)
8. Ожидание (kutish)

Разметка осуществлялась с привлечением пяти независимых экспертов-лингвистов, носителей узбекского языка. Для обеспечения высокого качества разметки применялась методика межэкспертного согласования с использованием коэффициента каппа Коэна [16]. В итоговый набор данных были включены только те тексты, по которым коэффициент согласия экспертов составил не менее 0.75.

Распределение классов в итоговом датасете представлено в таблице 1.

Таблица 1. Распределение классов эмоций в корпусе текстов

Эмоция	Количество текстов	Доля в корпусе (%)
Радость	2800	18.7
Доверие	1950	13.0
Страх	1750	11.7
Удивление	1600	10.7
Печаль	2100	14.0
Отвращение	1500	10.0
Гнев	1900	12.6
Ожидание	1400	9.3
Всего	15000	100.0

Датасет был разделен на обучающую (70%), валидационную (15%) и тестовую (15%) выборки с сохранением пропорционального распределения классов.

3.2 Архитектура модели

В качестве базовой архитектуры была выбрана модель тахрирчи-BERT – адаптация модели BERT для узбекского языка [12]. Тахрирчи-BERT имеет ту же архитектуру, что и оригинальная модель BERT-base:

- 12 слоев трансформеров
- 12 механизмов внимания
- Скрытый размер 768
- Общее количество параметров: 110 миллионов

Для решения задачи классификации эмоций базовая модель была дополнена следующими компонентами:

1. **Слой пулинга** – для получения единого векторного представления всего текста на основе выходного представления токена [CLS]
2. **Слой dropout** с вероятностью 0.1 – для предотвращения переобучения
3. **Полносвязный слой** размерностью 768×8 , преобразующий выходное представление в вероятностное распределение по 8 классам эмоций
4. **Функция активации softmax** – для нормализации выходных значений и получения вероятностей принадлежности к каждому классу

Общая архитектура модели представлена на рисунке 1.

3.3 Процесс обучения

Обучение модели проводилось с использованием следующих гиперпараметров:

- Оптимизатор: AdamW с параметрами $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e-8$
- Функция потерь: категориальная кросс-энтропия
- Начальная скорость обучения: $2e-5$
- Планировщик скорости обучения: линейное уменьшение
- Размер батча: 16
- Количество эпох: 5
- Максимальная длина последовательности: 512 токенов

Для предотвращения переобучения использовалась техника ранней остановки (early stopping) с мониторингом точности на валидационной выборке и терпением в 2 эпохи. Для улучшения работы модели с несбалансированным распределением классов применялось взвешивание классов при расчете функции потерь, где вес класса определялся как обратно пропорциональный его частоте в обучающем наборе данных. Обучение проводилось на графическом процессоре NVIDIA Tesla V100 с 32 ГБ памяти. Для мониторинга использования вычислительных ресурсов применялась библиотека `torch.cuda.memory_stats()`, которая позволяла отслеживать пиковое потребление памяти и эффективность ее использования.

3.4 Метрики оценки

Для оценки производительности модели использовались следующие метрики:

1. **Accuracy** (точность) – доля правильно классифицированных текстов среди всех текстов
2. **Precision** (точность) – доля правильно классифицированных текстов среди всех текстов, отнесенных к данному классу

3. **Recall** (полнота) – доля правильно классифицированных текстов среди всех текстов данного класса
4. **F1-score** – гармоническое среднее между precision и recall
5. **Confusion matrix** (матрица ошибок) – для визуализации результатов классификации
6. **Cohen's Kappa** – для оценки согласованности между предсказаниями модели и экспертной разметкой

Эти метрики рассчитывались как для отдельных классов эмоций, так и в среднем по всем классам (с использованием macro-averaging для учета несбалансированности классов).

4. Результаты и обсуждение

4.1 Общие результаты классификации

Результаты оценки модели на тестовой выборке представлены в таблице 2.

Таблица 2. Результаты классификации эмоций на тестовой выборке

Метрика	Значение
sAccuracy	0.834
Macro Precision	0.827
Macro Recall	0.811
Macro F1-score	0.818
Cohen's Kappa	0.798

Как видно из результатов, модель демонстрирует высокую точность классификации эмоций в текстах на узбекском языке. Значение коэффициента каппа Коэна близко к 0.8, что соответствует высокому уровню согласованности между предсказаниями модели и экспертной разметкой.

Детальные результаты по каждому классу эмоций представлены в таблице

3.

Таблица 3. Метрики по отдельным классам эмоций

Эмоция	Precision	Recall	F1-score
Радость	0.891	0.875	0.883
Доверие	0.816	0.804	0.810
Страх	0.847	0.823	0.835
Удивление	0.822	0.835	0.828
Печаль	0.858	0.842	0.850
Отвращение	0.776	0.761	0.768
Гнев	0.826	0.797	0.811
Ожидание	0.781	0.755	0.768

Анализ результатов показывает, что модель наиболее успешно распознает тексты с эмоциями радости и печали, что может быть связано с более четкими лексическими маркерами этих эмоций в узбекском языке. Наименее точно определяются эмоции отвращения и ожидания, что может быть обусловлено их более тонкой и контекстуальной природой.

4.2 Анализ матрицы ошибок

Матрица ошибок (confusion matrix) представлена на рисунке 2. Анализ матрицы показывает, что наиболее частыми ошибками классификации являются:

1. Смешение эмоций радости и доверия
2. Смешение эмоций страха и печали
3. Смешение эмоций гнева и отвращения
4. Смешение эмоций удивления и ожидания

Это соответствует психологическим исследованиям, указывающим на близость этих пар эмоций в эмоциональном спектре человека [17].

Интересно отметить, что модель редко путает эмоции из противоположных секторов круга Плутчика (например, радость и печаль, страх и гнев), что свидетельствует о ее способности к различению базовых эмоциональных противоположностей.

4.3 Анализ использования вычислительных ресурсов

В ходе обучения и применения модели проводился мониторинг использования вычислительных ресурсов. Результаты представлены в таблице 4.

Таблица 4. Использование вычислительных ресурсов

Этап	Пиковое использование GPU памяти (ГБ)	Среднее время обработки одного текста (мс)
Обучение	14.2	-
Валидация	8.7	32
Тестирование	8.7	31
Производственное применение	6.5	28

Наблюдается значительное потребление памяти GPU во время обучения, что обусловлено необходимостью хранения градиентов и промежуточных состояний модели. При этом потребление памяти во время тестирования и производственного применения значительно ниже благодаря отключению расчета градиентов.

Среднее время обработки одного текста на производственном сервере составляет 28 мс, что позволяет обрабатывать около 35 текстов в секунду на одном GPU. Это делает возможным применение модели в системах реального времени, например, для

анализа потока комментариев или сообщений в социальных сетях.

Для оптимизации использования ресурсов были применены следующие техники:

1. Динамическое выделение памяти (PyTorch lazy initialization)
2. Смешанная точность вычислений (mixed precision training)
3. Оптимизация размера батча
4. Освобождение кэша CUDA после каждой итерации

4.4 Сравнение с базовыми моделями

Для оценки эффективности предложенного подхода было проведено сравнение с рядом базовых моделей:

1. TF-IDF + SVM
2. FastText + логистическая регрессия
3. BiLSTM с GloVe-эмбедингами
4. Мультиязычная модель mBERT без дополнительной настройки

Результаты сравнения представлены в таблице 5.

Таблица 5. Сравнение различных моделей (F1-score)

Модель	Macro F1-score
TF-IDF + SVM	0.612
FastText + логистическая регрессия	0.684
BiLSTM с GloVe-эмбедингами	0.723
mBERT без дополнительной настройки	0.775
Тахирчи-BERT (наша модель)	0.818

Предложенная модель на основе тахрирчи-BERT демонстрирует значительное превосходство над базовыми подходами. Особенно заметно превосходство над традиционными методами (TF-IDF + SVM и FastText), что подтверждает эффективность архитектуры трансформеров для задач анализа текстов на узбекском языке. Интересно отметить, что даже мультязычная модель mBERT без дополнительной настройки показывает достаточно высокие результаты, что свидетельствует о значительном потенциале трансферного обучения для малоресурсных языков.

5. Анализ конкретных примеров

Для более глубокого понимания работы модели были проанализированы конкретные примеры классификации из тестового набора данных.

Пример 1. Текст с эмоцией радости:

"Bugun imtihondan a'lo baho bilan o'tdim! Oilam bilan bayram qilamiz, hammasi juda ajoyib!" (Сегодня я сдал экзамен на отлично! Мы отпразднуем с семьей, всё просто замечательно!)

Предсказание модели: Радость (вероятность 0.92)

В данном примере модель уверенно определяет эмоцию радости, опираясь на позитивную лексику ("a'lo", "bayram", "ajoyib").

Пример 2. Текст с эмоцией страха:

"Qorong'i ko'chadan yurganda kimdir orqamdan kelayotganini sezdim. Yuragim qinidan chiqib ketay dedi." (Когда я шёл по тёмной улице, я почувствовал, что кто-то идёт за мной. Моё сердце чуть не выскочило из груди.)

Предсказание модели: Страх (вероятность 0.87)

Модель правильно определяет эмоцию страха, учитывая контекст ситуации и физиологические проявления страха, описанные в тексте.

Пример 3. Сложный пример со смешанными эмоциями:

"Yangi ishga kirganim uchun ham quvonchli, ham xavotirli. Jamoaga moslasha olamanmi yo'qmi, bilmauman." (Я и рад, и волнуюсь из-за новой работы. Не знаю, смогу ли адаптироваться к команде.)

Предсказание модели: Ожидание (вероятность 0.45), Радость (вероятность 0.32), Страх (вероятность 0.18)

В этом случае модель определяет ожидание как основную эмоцию, что соответствует содержанию текста, где автор выражает неопределённость относительно будущего. При этом модель также определяет значительные компоненты радости и страха, что отражает смешанный характер эмоций в тексте.

Пример 4. Ошибка классификации:

"Bu film meni umuman ta'sirlamadi, vaqtni bekorga sarfladim." (Этот фильм совершенно не впечатлил меня, я зря потратил время.)

Предсказание модели: Печаль (вероятность 0.61), Отвращение (вероятность 0.27) Экспертная разметка: Отвращение

В данном примере модель ошибочно определяет эмоцию печали как основную, хотя эксперты отнесли этот текст к категории отвращения. Такая ошибка может быть связана с тем, что выражение отрицательных эмоций в узбекском языке часто имеет схожие лексические маркеры, и разграничение подобных эмоций требует более тонкого анализа контекста.

6. Возможные улучшения и направления дальнейших исследований

На основе полученных результатов можно предложить следующие направления для улучшения и развития исследования:

1. Расширение датасета

- Увеличение объема аннотированных данных, особенно для недостаточно представленных классов эмоций
- Включение текстов различных жанров и стилей
- Создание специализированных корпусов для отдельных доменов (литература, новости, социальные медиа)

2. Улучшение архитектуры модели

- Применение более современных архитектур трансформеров (RoBERTa, XLM-R, DeBERTa)
- Интеграция механизмов внимания, специфичных для эмоциональной классификации
- Включение морфологического анализа, учитывающего агглютинативный характер узбекского языка

3. Многозадачное обучение

- Совместное обучение на задачах классификации эмоций, анализа тональности и определения интенсивности эмоций
- Использование вспомогательных задач, таких как определение стиля текста или авторства

4. Мультимодальный анализ

- Интеграция текстового анализа с аудио- и визуальными модальностями
- Применение в системах распознавания речи для анализа эмоций в устной речи

5. Оптимизация производительности

- Дистилляция знаний для создания более легковесных моделей
- Квантизация модели для ускорения инференса

- Оптимизация для развертывания на мобильных устройствах

6. **Кросс-языковой перенос**

- Исследование возможностей переноса знаний между родственными тюркскими языками
- Создание единой модели для группы близкородственных языков

7. Заключение

В данной работе представлен подход к классификации эмоциональной окраски текстов на узбекском языке с использованием модели тахрирчи-BERT. Результаты экспериментов показывают, что предложенная модель достигает высокой точности в многоклассовой классификации эмоций, значительно превосходя традиционные методы и базовые нейросетевые архитектуры.

Ключевыми факторами успеха предложенного подхода являются:

1. Использование предварительно обученной модели, специфичной для узбекского языка
2. Тщательная подготовка и аннотация корпуса текстов с высоким межэкспертным согласием
3. Оптимизация гиперпараметров обучения и архитектуры классификационной надстройки
4. Применение техник эффективного использования вычислительных ресурсов

Разработанная система открывает новые возможности для анализа текстов на узбекском языке в различных прикладных сферах, включая мониторинг социальных медиа, анализ отзывов пользователей, разработку диалоговых систем и персонализированных рекомендательных сервисов.

Дальнейшие исследования будут направлены на расширение возможностей системы для обработки более сложных эмоциональных состояний, улучшение производительности модели и адаптацию предложенного подхода к другим малоресурсным языкам.

Список литературы:

1. Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 6282-6293.
2. Mohammad, S. M. (2022). Practical and ethical considerations in the effective use of emotion and sentiment lexicons. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, 1675-1684.
3. Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys*, 50(2), 1-33.
4. Strapparava, C., & Valitutti, A. (2004). WordNet-Affect: an affective extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluation, 1083-1086.
5. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
6. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.
7. Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, 8, 842-866.
8. Kuriyozov, E., Matlatipov, S., & Alonso, P. (2023). Natural language processing for low-resource Turkic languages: A comprehensive survey. *ACM Computing Surveys*, 55(12), 1-38.

9. Matlatipov, G., & Vetulani, Z. (2022). Building language resources for Uzbek natural language processing. In Proceedings of the International Conference on NLP Applications for Turkic Languages, 47-55.
10. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171-4186.
11. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8440-8451.
12. Madrahimov, B., Absalomov, L., & Khodjaev, S. (2023). Tahririchi: Pre-trained transformer models for Uzbek language. In Proceedings of the 3rd Workshop on NLP for Turkic Languages, 112-121.
13. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171-4186.
14. Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification? In China National Conference on Chinese Computational Linguistics, 194-206.
15. Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience*, 1, 3-33.
16. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
17. Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science*, 11(1), 31-34.

18. Abdurahmonova, N., & Tulakov, U. (2022). Corpus linguistics in Uzbek language: Current state and perspectives. *Journal of Turkic Language Processing*, 7(3), 145-159.
19. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
20. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems*, 5753-5763.
21. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
22. Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., & Kumar, R. (2019). Predicting the type and target of offensive posts in social media. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1415-1420.
23. Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2), 310-328.
24. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2227-2237.
25. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.
26. Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2018). Advances in pre-training distributed word representations. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, 52-55.

27. Alimova, I., & Tuychiev, G. (2022). Emotion detection in Uzbek social media texts using transformer-based models. In Proceedings of the 4th International Conference on NLP Applications for Turkic Languages, 78-86.
28. Akhmedova, D., & Khodjaev, J. (2023). Transfer learning for sentiment analysis in low-resource Uzbek language. *Journal of Natural Language Processing*, 15(2), 87-103.
29. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
30. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... & Rush, A. (2020). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38-45.