Қазақстан Республикасы Ғылым және жоғары білім министрлігі

«Л.Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ

Шекербек Ержан Қуандықұлы

**«****Мәтіннің тоналдылығын талдау үшін машиналық оқыту әдістерін қолдану (Sentiment Analysis)»**

*тақырыбында*

ДИПЛОМДЫҚ ЖҰМЫС

6В06103 - «Ақпараттық жүйелер» білім беру бағдарламасы

Астана, 2025

Қазақстан Республикасы Ғылым және жоғары білім министрлігі

«Л.Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ

«Қорғауға жіберілді»  
«\_\_\_\_»\_\_\_\_\_\_\_ 2025 ж.

«Ақпараттық жүйелер»

кафедрасының меңгерушісі

\_\_\_\_\_\_\_\_\_ Муханова А. А.

**ДИПЛОМДЫҚ ЖҰМЫС**

**Тақырыбы: «Мәтіннің тоналдылығын талдау үшін машиналық оқыту әдістерін қолдану (Sentiment Analysis)»**

6В06103 – «Ақпараттық жүйелер» білім беру бағдарламасы бойынша

Орындаған: Шекербек Е. Қ.

Ғылыми жетекшісі:   
PhD, аға оқытушы Серикбаева С. К.

Астана, 2025

Аңдатпа

Бұл жұмыста мәтіннің тоналдылығын талдау үшін машиналық оқыту әдістерін қолдану қарастырылады. Эмоционалдық талдаудың теориялық негіздері мен заманауи модельдері (BERT, RoBERTa, GPT) салыстырылып, IMDb, Twitter, Yelp деректері негізінде тәжірибелік зерттеу жүргізілді. Python арқылы бағдарламалық құрал әзірленіп, трансформерлік модельдердің жоғары тиімділігі анықталды. Зерттеу нәтижесінде қазақ тіліне бейімделген модельдер әзірлеу қажеттілігі ұсынылды.

Аннотация

В работе рассматривается применение методов машинного обучения для анализа тональности текста. Описаны теоретические основы и сравнены современные модели (BERT, RoBERTa, GPT). На основе данных IMDb, Twitter и Yelp проведено практическое исследование. Разработано программное средство на Python, подтверждена высокая эффективность трансформерных моделей. По результатам обоснована необходимость разработки моделей для казахского языка.

Annotation

This thesis examines the use of machine learning methods for sentiment analysis of text. Theoretical foundations and modern models (BERT, RoBERTa, GPT) were compared. Practical experiments were conducted using IMDb, Twitter, and Yelp datasets. A Python-based software tool was developed, confirming the high efficiency of transformer models. The study highlights the need to develop models adapted for the Kazakh language.

МАЗМҰНЫ

[Кіріспе 7](#_Toc195095015)

[1 Мәтіндердің эмоционалдық талдау ұғымы мен міндеттері 9](#_Toc195095017)

[1.1 Эмоционалдық талдауға әдістемелік тәсілдер](#_Toc195095018) 9

[1.2 Отандық және шетелдік зерттеулердің салыстырмалы талдауы 13](#_Toc195095019)

[1.3 Зерттеу әдіснамасы мен тәсілдері 15](#_Toc195095020)

[2. Эмоцияналдық талдаудың заманауи модельдері мен зерттеу әдістері 18](#_Toc195095021)

[2.1 Бағдарламалық қамтамасыз етуді іске асыру 18](#_Toc195095022)

[2.2 Эмоционалдық талдаудың заманауи әдістері мен модельдері 22](#_Toc195095023)

[2.3](#_Toc195095024) Модельдерді салыстырмалы талдау 24

[2.4 Заманауи модельдердің тиімділігін бағалау 25](#_Toc195095025)

[3 Бағдарламалық жасақтама әзірлеу және нәтижелерді бағалау 2](#_Toc195095025)8

[3.1 Бағдарламалық жасақтаманы жетілдіру жолдары 2](#_Toc195095026)8

[3.2 Зерттеудің негізгі қорытындылары 30](#_Toc195095027)

[3.3 Эмоционалдық талдау модельдерінің шектеулері мен шешілмеген мәселе 3](#_Toc195095028)2

[3.4 Зерттеу нәтижелерінің қорытындылары мен болашақтағы зерттеулер 3](#_Toc195095029)4

[3.5 Мәтіндерді эмоционалдық талдау сапасын арттыру әдістері 3](#_Toc195095030)6

[3.6 Бағдарламалық жасақтаманың негізгі функциялары 3](#_Toc195095031)9

[3.7 Мәселелер мен шешімдердің таксономиясы 4](#_Toc195095032)1

[3.8 Бағдарлама кодының сипаттамасы 41](#_Toc195095033)

[Қорытынды 4](#_Toc195095034)7

[Пайдаланылған әдебиеттер тізімі 4](#_Toc195095035)8

[Қосымша 5](#_Toc195095036)1

Қазақстан Республикасы Ғылым және жоғары білім министрлігі

«Л.Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ

Ақпараттық технологиялар факультеті

Ақпараттық жүйелер кафедрасы

6В06103 - «Ақпараттық жүйелер» білім беру бағдарламасы

**«Бекітемін»**

«Ақпараттық жүйелер»

кафедрасының меңгерушісі

\_\_\_\_\_\_\_\_\_\_ Муханова А. А.

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 2025 ж.

**Дипломдық жұмысты (жобаны) орындауға берілген**

**ТАПСЫРМА**

Студент Шекербек Ержан Қуандықұлы, 4 курс, АЖ-45 тобы, «Ақпараттық жүйелер», күндізгі оқу бөлімі.

1. Дипломдық жұмыс (жобаның) тақырыбы «Мәтіннің тоналдылығын талдау үшін машиналық оқыту әдістерін қолдану (Sentiment Analysis)», № 71-п ректордың бұйрығымен бекітілген «16» қаңтар 2025 ж.
2. Білім алушының аяқталған жұмысты тапсыру мерзімі «22».04.2025 ж.
3. Жұмысқа қажетті бастапқы дерктер (заңдар, әдебиет көздері, зертханалық және өндірістік мәліметтер):

* Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool.
* Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.
* Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,8(4).

1. Дипломдық жұмыста (жобада) жасалатын тақырыптар тізімі

* Мәтіннің тоналдығын талдау үшін машиналық оқыту әдістері негізінде құрылған модельдерге талдау және салыстырма жүргізу;

1. Графикалық материалдар тізбесі (сызбалар, кестелер, диаграммалар және т.б.):
2. Ұсынылатын негізгі әдебиеттер тізімі:

* Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, 28(2), 15–21.
* Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O’Reilly Media.
* Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

1. Жұмыс бойынша кеңестер (оларға қатысты жұмыс бөлімдерін көрсете отырып)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Бөлімнің (тараудың) нөмірі, атауы | Ғылыми жетекші, консультант | Тапсырма алу мерзімі | Тапсырма берілді  (қолы) | Тапсырма қабылданды  (қолы) |
| Кіріспе | Серикбаева С. К. |  |  |  |
| Машиналық оқыту теориялық негіздері | Серикбаева С. К. |  |  |  |
| Машиналық оқыту негізінде мәтіннің тоналдығын анықтау модельдерін зерттеу | Серикбаева С. К. |  |  |  |
| Мәтіннің тоналдығын тоналдығын анықтауға арналған модельдерді қолдану аясы | Серикбаева С. К. |  |  |  |
| Мәтіннің тоналдығын анықтауға бағытталған бағдарламалық жасақтама жасау | Серикбаева С. К. |  |  |  |
| Қорытынды | Серикбаева С. К. |  |  |  |

1. Дипломдық жұмысты (жобаны) орындау кестесі

| № | Жұмыс сатысы | Жұмыс кезеңдерін орындау мерзімдері | Ескерту |
| --- | --- | --- | --- |
| 1 | Дипломдық жұмыстың (жобаның) тақырыбын бекіту |  |  |
| 2 | Дипломдық жұмысты (жобаны) дайындау үшін материалдар жинау |  |  |
| 3 | Дипломдық жұмыстың (жобаның) теориялық бөлігін дайындау (1-бөлім) |  | Практикаға кеткенге дейін |
| 4 | Дипломдық жұмыстың (жобаның) талдамалық бөлігін (2-3 бөлімдер) дайындау |  | Практика кезінде |
| 5 | Дипломдық жұмыстың (жобаның) толық мәтінінің жоба нұсқасын аяқтау |  | Практика аяқталғаннан кейінгі бірінші аптада |
| 6 | Алдын ала қорғауға дипломдық жұмысты (жобаны) ұсыну |  | Шолу дәрістері (консультациялар) кезінде |
| 7 | Рецензияға дипломдық жұмысты жіберу |  |  |
| 8 | Дипломдық жұмысты қорғау (жобаның) |  |  |
| 9 | Дипломдық жұмысты қорғау |  | МАК кестесіне сәйкес |

Тапсырманың берілген күні: «16» қаңтар 2025 ж.

Ғылыми жетекші: PhD, аға оқытушы: \_\_\_\_\_\_\_ Серикбаева С. К.

Тапсырманы қабылдады:студент \_\_\_\_\_\_\_ Шекербек Е. Қ.

# Кіріспе

Цифрлық дәуірде мәтіндік деректердің көлемі керемет қарқынмен өсіп келе жатқанда, мәтіннің тоналдылығын талдау қоғамдық пікірді, пайдаланушылардың қалауын және ақпараттың эмоционалды реңкін түсінудің негізгі құралдарының біріне айналды. Сентименталды талдау — бұл мәтіннің эмоционалды бояуын автоматты түрде анықтау процесі, яғни оң, теріс немесе бейтарап екенін анықтау. Бұл тапсырма маркетинг, беделді басқару, тұтынушылардың пікірлерін талдау, әлеуметтік зерттеулер және тіпті саяси талдау сияқты салаларда кеңінен қолданылады.

Жасанды интеллект пен машиналық оқыту технологияларының дамуымен тоналдылықты талдау әдістері дәлірек және тиімді бола түсті. Қазіргі тәсілдерге қайталанатын нейрондық желілер (RNN), ұзын-қысқа мерзімді жады (LSTM) және трансформаторлар (мысалы, BERT) сияқты нейрондық желілерді қолдану кіреді. Бұл модельдер мәтіннің семантикасын тереңірек түсініп, контексті ескеруге қабілетті. Алайда, айтарлықтай жетістіктерге қарамастан, сөздердің көп мағыналылығы, ирония, сарказм және тілдің мәдени ерекшеліктері сияқты факторларға байланысты сентимент-анализ міндеті әлі де күрделі болып қала береді [1].

Әлемдік деңгейде сентимент-анализ қарқынды дамып келеді. АҚШ және Еуропа елдерінде Google, Meta, Microsoft сияқты ірі технологиялық компаниялар жасанды интеллект негізінде тілдік модельдерді жетілдірумен айналысады. Бұл елдерде ірі көлемдегі деректерге қолжетімділік және зерттеулерге инвестицияның жоғары деңгейі дәлдікті арттыруға ықпал етеді. Азия елдерінде, әсіресе Қытай мен Оңтүстік Кореяда, эмоцияны тану технологиялары әлеуметтік желілер мен коммерциялық платформаларда белсенді қолданылады.

Қазақстандық контексте мәтіндік деректерді өңдеу мен сентимент-анализді дамыту енді ғана қарқын алуда. Қазақ және орыс тілдеріндегі мәтіндерді автоматты өңдеу үшін қажетті аннотталған деректер жиынтығының жетіспеушілігі, морфологиялық күрделілік және тілдің ерекшеліктерін ескеретін жетілдірілген модельдердің болмауы негізгі қиындықтар болып табылады. Дегенмен, қазақстандық зерттеушілер бұл бағытта жұмыс істеп, қазақ тіліндегі мәтіндерді өңдеуге арналған модельдер мен алгоритмдерді дамытуда.

Бұл жұмыстың мақсаты – мәтіннің тоналдылығын талдаудың заманауи әдістерін зерттеу, олардың мүмкіндіктері мен шектеулерін қарастыру, сондай-ақ сентимент-талдаудың дәлдігі мен сенімділігін арттыруға мүмкіндік беретін тәсілді әзірлеу. Зерттеу аясында әртүрлі алгоритмдер мен модельдер қарастырылып, олар нақты деректерге тестілеуден өткізіледі, сондай-ақ талдау сапасын жақсарту бойынша ұсыныстар беріледі.

Зерттеудің өзектілігі мәтіндік деректерді автоматты түрде өңдеуге және талдауға арналған құралдарға сұраныстың артуына байланысты. Компаниялар мен ұйымдар өз клиенттерін жақсырақ түсінуге және кері байланыс негізінде стратегияларын бейімдеуге тырысатын жағдайларда, сентимент-анализ негізделген шешімдер қабылдаудың таптырмас құралына айналуда. Сонымен қатар, Қазақстанда мемлекеттік органдар мен бизнес құрылымдар үшін қоғамдық пікірді талдау қажеттілігі артып келеді, бұл сентимент-анализ технологияларының сұранысын арттырады.

Осылайша, бұл жұмыс мәтіннің тоналдылығын талдау әдістерін тереңірек түсінуге, оларды практикалық қолдануға және қызметтің әртүрлі салаларында мәтіндік ақпаратты өңдеудің тиімділігін арттыру үшін қолдануға болатын шешімдерді әзірлеуге бағытталған.

## Мәтіндердің эмоционалдық талдау ұғымы мен міндеттері

## Эмоционалдық талдауға әдістемелік тәсілдер

Мәтіндердің эмоционалдық талдауы, немесе пікірлерді талдау (Sentiment Analysis), табиғи тілдерді өңдеу (Natural Language Processing, NLP) саласында маңызды бағыттардың бірі болып табылады. Эмоционалдық талдаудың негізгі міндеті мәтіннің эмоциялық реңкін автоматты түрде анықтау болып табылады, яғни мәтіннің оң, теріс немесе нейтралды көзқарас білдіретінін анықтау [3]. Алайда, қазіргі заманғы әдістер қарапайым классификациядан тыс шығып, нақты эмоцияларды (мысалы, қуаныш, ашу, қайғы, таңдану, қорқыныш, жеккөрушілік) анықтау және аспектілік талдау (Aspect-Based Sentiment Analysis, ABSA) сияқты күрделірек міндеттерді қамтиды, мұнда мәтінде айтылған нақты объектілер немесе сипаттамаларға қатысты көзқарас бағаланады. Мысалы, мейрамхана туралы пікірде тамақ, қызмет көрсету және атмосфераға деген көзқарасты бөлек бағалауға болады.

Эмоционалдық талдау әртүрлі салаларда кеңінен қолданылады. Маркетингте өнімдер мен қызметтер туралы пікірлерді талдау үшін пайдаланылады, бұл компанияларға ұсыныстарын жақсартуға және клиенттердің қанағаттануын арттыруға мүмкіндік береді. Мысалы, Amazon мен Yelp сияқты компаниялар клиенттер пікірлерін талдау және өз қызметтерін жақсарту үшін эмоционалдық талдауды белсенді түрде пайдаланады. Саясатта эмоционалдық талдау қоғамдық пікірді және оқиғаларға немесе мәлімдемелерге реакцияны бақылауға көмектеседі. Мысалы, сайлау науқандары кезінде саяси аналитиктер кандидаттардың сөздеріне реакцияны бағалау үшін эмоционалдық талдауды қолданады. Қаржы саласында жаңалықтар мен әлеуметтік желілердегі хабарламалардың эмоционалдық талдауы нарықтық тенденцияларды болжауға көмектеседі. Мысалы, компания туралы жаңалықтардың эмоционалдық талдауы оның акцияларының құнының өзгерістерін болжауға мүмкіндік береді. Денсаулық сақтау саласында пациенттердің медициналық қызметтер немесе дәрілер туралы пікірлерін бағалау үшін пайдаланылады. Мысалы, фармацевтикалық компаниялар жаңа дәрілер туралы пікірлерді бақылау үшін эмоционалдық талдауды қолданады. Әлеуметтік ғылымдарда әлеуметтік желілердегі хабарламалардың эмоционалдық талдауы қоғамдық көңіл-күйлерді зерттеуге және трендтерді анықтауға көмектеседі. Мысалы, Twitter-дегі хабарламалардың эмоционалдық талдауы әртүрлі әлеуметтік мәселелер бойынша қоғамдық пікірді түсінуге көмектеседі.

Эмоционалдық талдаудың негізгі қиындықтарының бірі – табиғи тілді өңдеудің күрделілігі. Мәтіндер жиі көпмағыналы сөздерді, иронияны, сарказмды және басқа да формаларды қамтиды, олар алгоритмдермен дұрыс түсінбеуі мүмкін. Мысалы, «Әрине, керемет қызмет!» деген фраза оң пікір ретінде қабылдануы мүмкін, бірақ шын мәнінде сарказмды білдіреді. Сонымен қатар, контекст тональділікті анықтауда маңызды рөл атқарады, бұл семантика мен синтаксисті ескеретін күрделі модельдерді қолдануды талап етеді. Тағы бір проблема – әлеуметтік желілердегі хабарламалар немесе комментарийлер сияқты қысқа мәтіндерді өңдеу, олар көбінесе талдау үшін аз ақпарат қамтиды. Мысалы, «Бұл ужас болды» деген қысқа твит теріс ретінде түсіндірілсе де, контекстсіз нақты қандай реакция туғызғанын түсіну мүмкін емес.

Эмоционалдық талдау тарихы 2000 жылдардың басында машиналық оқытуды мәтіндерді классификациялау үшін қолдана бастаған кезеңнен басталады. Алғашқы әдістер қарапайым статистикалық модельдерге, мысалы, наивный байесовский классификатор мен опорлық векторлар әдісіне негізделді. Алайда, технологиялардың дамуы және деректер көлемінің артуымен эмоционалдық талдау әдістері күрделене және дәл бола түсті. Соңғы жылдары терең оқыту, әсіресе BERT және GPT негізіндегі трансформер модельдері эмоционалдық талдаудың сапасын айтарлықтай жақсартты. Бұл модельдер контекст пен мәтіннің семантикасын ескеру қабілетіне ие, бұл оларды ирония, сарказм немесе көпмағыналы сөздерді қамтитын күрделі мәтіндерді талдау үшін ерекше тиімді етеді [1].

Эмоционалдық талдаудың негізгі міндеттері эмоционалдық тонды классификациялау (мәтіннің жалпы эмоциялық тонусын анықтау), нақты эмоцияларды анықтау (мысалы, қуаныш, ашу, қайғы), аспектілік эмоционалдық талдау (негізгі объектілер немесе сипаттамаларға қатысты эмоционалдық тонды бағалау) және көптілді талдау (әртүрлі тілдердегі мәтіндермен жұмыс істеу үшін әдістерді бейімдеу) болып табылады. Әрбір міндеттің өзіндік ерекшеліктері бар және әртүрлі әдістер мен тәсілдерді қолдануды талап етеді. Мысалы, аспектілік эмоционалдық талдау мәтінді тереңірек түсінуді қажет етеді, себебі жалпы эмоциялық тонды анықтаумен қатар, мәтінде аталған нақты аспектілерді бөліп көрсету және олардың тонусын бағалау қажет.

Эмоционалдық талдаудың қолдану мысалдары маркетинг (өнімдер мен қызметтер туралы пікірлерді талдау, брендтің репутациясын бақылау), саясат (қоғамдық пікірді бақылау, саяси оқиғаларға немесе мәлімдемелерге реакцияны талдау), қаржы (жаңалықтар мен әлеуметтік желілердегі хабарламаларды талдау арқылы нарықтық тенденцияларды болжау), денсаулық сақтау (медициналық қызметтер немесе дәрілер туралы пациенттердің пікірлерін бағалау) және әлеуметтік ғылымдар (қоғамдық көңіл-күйлерді зерттеу, әлеуметтік медиа трендтерін анықтау) сияқты салаларды қамтиды. Бұл салалардың әрқайсысында эмоционалдық талдау адамдардың қажеттіліктері мен күтілімдерін жақсы түсінуге және негізделген шешімдер қабылдауға көмектеседі [2].

Эмоционалдық талдаудың проблемалары мен қиындықтары көпмағыналы сөздермен (бірдей сөздің контекстке байланысты түрлі эмоциялық реңктері болуы мүмкін), ирония мен сарказммен (иронияны қамтитын мәтіндер алгоритмдермен жиі дұрыс түсінілмейді), контекстік тәуелділікпен (тональділікті дәл анықтау үшін сөздердің қолданылу контекстін ескеру қажет), қысқа мәтіндерді өңдеумен (әлеуметтік желілердегі хабарламалар немесе комментарийлер жиі аз ақпарат қамтиды, бұл талдауды қиындатады) және көптілділікпен (әртүрлі тілдердегі мәтіндермен жұмыс істеу үшін әдістерді бейімдеу, соның ішінде орыс тіліндегі ерекшеліктерді ескеру) байланысты. Бұл мәселелер семантика мен синтаксисті ескере алатын күрделі модельдер мен тәсілдерді қолдануды талап етеді [3].

Эмоционалдық талдаудағы қазіргі тенденциялар терең оқытуды (трансформерлер негізіндегі модельдер, мысалы, BERT және GPT, эмоционалдық талдаудың сапасын айтарлықтай жақсартты), аспектілік эмоционалдық талдауды (нақты объектілер немесе сипаттамаларға қатысты эмоционалдық тонды бағалауға арналған зерттеулердің артуы), көптілді талдауды (әртүрлі тілдердегі мәтіндермен жұмыс істеуге арналған әдістерді дамыту) және модельдердің түсініктілігін (нақты әрі түсінікті модельдерді жасауға артықшылық беру) қамтиды. Бұл тенденциялар эмоционалдық талдауға қатысты міндеттердің күрделілігінің және әртүрлілігінің артуын көрсетеді, сондай-ақ неғұрлым дәл әрі әмбебап әдістерді дамыту қажеттілігін айғақтайды [1].

Эмоционалдық талдау саласындағы зерттеулер мен жетістіктердің мысалдарына Google әзірлеген BERT (Bidirectional Encoder Representations from Transformers) моделі кіреді, ол екі бағытты контекстті ескеру қабілеті арқасында эмоционалдық талдауда жаңа стандарттар орнатты, орыс тіліне бейімделген RuBERT моделі, ол орыс тіліндегі мәтіндер бойынша эмоционалдық талдау міндеттерінде жоғары дәлдік көрсетті, және OpenAI әзірлеген GPT (Generative Pre-trained Transformer) моделі, ол да эмоционалдық талдау мен мәтін генерациясы үшін пайдаланылады. Бұл модельдер эмоционалдық талдауда жоғары дәлдік пен тиімділікті көрсетіп, практикалық қолдану үшін ерекше пайдалы [1].

Эмоционалдық талдау әдістері мәтіндердегі адамдардың эмоцияларын, көзқарастарын және субъективті бағалауларын анықтауға бағытталған. Бұл саланың дамуы қазіргі заманғы цифрлық әлемде ақпаратты өңдеу мен талдаудың маңызды құралына айналды. Эмоционалдық талдаудың негізгі әдістері екі топқа бөлінеді: дәстүрлі тәсілдер мен заманауи машиналық оқыту әдістері. Дәстүрлі тәсілдерге лексикондық әдістер мен ережелерге негізделген әдістер жатады. Лексикондық әдістер алдын ала дайындалған сөздіктерді пайдаланады, онда әрбір сөзге оның эмоционалдық реңкі (мысалы, оң, теріс немесе нейтралды) белгіленген.[5] Мұндай сөздіктердің көпшілігі ағылшын тілінде дайындалған, бірақ қазіргі уақытта басқа тілдер үшін де ұқсас ресурстар құрастырылуда. Мысалы, SentiWordNet сөздігі WordNet лексикалық желісіне негізделген және әр сөз үшін үш баллдық жүйені (позитивті, негативті, нейтралды) қолданады [6]. AFINN сөздігі -5-тен +5-ке дейінгі шкала бойынша сөздерді бағалайды, ал VADER әсіресе әлеуметтік желілердегі мәтіндерді талдау үшін жасалған және смайлдарды, әріптердің үлкейтілуін, басқа да невербалды элементтерді есепке алады. Бұл әдістердің басты артықшылығы - оларды іске асырудың қарапайымдылығы және нәтижелерді түсінудің оңайлығы. Дегенмен, олардың дәлдігі шектеулі, әсіресе күрделі синтаксистік құрылымдарды, иронияны немесе сарказмді талдау кезінде. Мысалы, «Бұл фильм өте жақсы... жоқ!» деген сөйлемдегі ирония лексикондық әдістер арқылы дұрыс танылмайды, себебі олар контекстті толық ескермейді. Сонымен қатар, бір сөздің әртүрлі контексттерде әртүрлі мағыналары болуы мүмкін, мысалы, «қатал» сөзі «қатал ережелер» контекстінде теріс, ал «қатал жаттығу» контекстінде оң мағына береді [4].

Ережелерге негізделген әдістер мәтінді талдау үшін алдын ала анықталған лингвистикалық ережелер жиынтығын қолданады. Бұл әдістер грамматикалық құрылымдарды, сөз тіркестерін және басқа да синтаксистік элементтерді талдау арқылы эмоционалдық тонды анықтайды. Мысалы, «ешқашан» сияқты теріс сөздерді, «өте» сияқты күшейткіштерді немесе «бірақ» сияқты контрасттық қосылыстарды тану арқылы жалпы эмоционалдық бағаны есептеуге болады. Бұл әдістердің артықшылығы - оларды нақты тақырыптарға немесе домендерге бейімдеуге болады. Дегенмен, оларды әзірлеу көп еңбекті қажет етеді және көпмағыналы сөздермен, сөйлемнің күрделі құрылымдарымен жұмыс істеу кезінде тиімділігі төмендеуі мүмкін. Сонымен қатар, бұл әдістердің масштабталу қабілеті шектеулі, яғни олар үлкен көлемдегі мәтіндерді өңдеу кезінде өнімділігін жоғалтады.

Машиналық оқыту әдістері эмоционалдық талдау саласында революциялық өзгерістер әкелді [7]. Олар өз кезегінде классикалық машиналық оқыту алгоритмдері мен терең оқыту әдістеріне бөлінеді. Классикалық алгоритмдерге Naive Bayes, Support Vector Machine (SVM) және Random Forest жатады. Бұл алгоритмдер мәтіндерден мөлшерлік сипаттамаларды (мысалы, сөздердің жиілігін, n-граммаларды) шығарып алып, оларды белгіленген деректер жинағы бойынша үйретеді. Бұл әдістердің артықшылығы - олардың жалпы қолданылатындығы және салыстырмалы түрде аз есептеу ресурстарын қажет етуі. Дегенмен, олардың тиімділігі көбінесе деректердің сапасына және көлеміне тәуелді, ал күрделі лингвистикалық құбылыстарды (мысалы, контекстке тәуелді мағыналарды) талдауда шектеулері бар. Мысалы, ұзын мәтіндерді талдау кезінде олар сөйлемдер арасындағы семантикалық байланыстарды толық сақтай алмауы мүмкін [7].

Терең оқыту әдістері соңғы жылдары эмоционалдық талдау саласында елеулі жетістіктерге қол жеткізді. Оларға рекурренттік нейрондық желілер (RNN), ұзақ қысқа мерзімді жады желілері (LSTM) және трансформерлер жатады. Бұл әдістердің ерекшелігі - олар мәтіннен автоматты түрде сипаттамаларды шығарып алады және контекстті толық ескере алады. Мысалы, BERT (Bidirectional Encoder Representations from Transformers) модельі екібағытты контекстті талдауға мүмкіндік береді, яғни ол сөздің мағынасын оның алдындағы және кейінгі сөздерге қарай анықтайды. Бұл модельдердің көмегімен иронияны, сарказмды және басқа да күрделі лингвистикалық құбылыстарды анықтау әлдеқайда дәлдікпен жүзеге асырылады. Сонымен қатар, олар үлкен көлемдегі деректерді өңдеуге бейімделген, бұл әсіресе әлеуметтік желілердегі мәліметтерді талдау кезінде маңызды. Көптеген заманауи модельдер (мысалы, GPT, RoBERTa, XLM-R) көп тілділікті қолдайды, бұл оларды халықаралық деңгейде қолдануға мүмкіндік береді. Дегенмен, бұл әдістердің кемшіліктері де бар: оларды оқыту үшін үлкен есептеу қуаты қажет, ал модельдердің интерпретациялануы қиын болуы мүмкін [1].

Дәстүрлі және заманауи әдістерді салыстырғанда, әрқайсысының өз артықшылықтары мен кемшіліктері бар. Дәстүрлі әдістердің басты артықшылығы - олардың қарапайымдылығы және интерпретациялануының оңайлығы. Дегенмен, олар күрделі мәтіндерді талдауда дәлдігі төмен. Керісінше, терең оқыту әдістері жоғары дәлдікке қол жеткізеді, бірақ оларды іске асыру үшін көп ресурстар қажет және модельдердің жұмыс істеу принципін түсіну қиын. Қазіргі уақытта көптеген зерттеулер гибридтік тәсілдерді қолдануға бағытталған, яғни дәстүрлі лексикондық әдістер мен заманауи нейрондық желілерді біріктіру. Бұл тәсіл модельдердің дәлдігін арттыруға мүмкіндік береді, сонымен қатар олардың интерпретациялануын сақтайды. Эмоционалдық талдаудың болашағы осы екі тәсілдің үйлесімді қолданылуында, сондай-ақ жаңа технологиялардың, мысалы, жасанды интеллекттің дамуында жатыр.

## 

## Отандық және шетелдік зерттеулердің салыстырмалы талдауы

Шетелдік зерттеулер эмоционалдық талдау саласында ұзақ тарихқа ие және кең ауқымды әдістер мен қолданбаларды қамтиды. Негізгі назар терең оқыту модельдерін, мысалы, BERT, GPT және олардың нұсқаларын әзірлеу мен жетілдіруге бағытталған. Бұл модельдер мәтіндерді талдау мен эмоционалдық бағалауда, әсіресе ағылшын тілінде, өте жоғары дәлдікке жетті. Мысалы, BERT ағылшын тіліндегі мәтіндерді классификациялау мен эмоционалдық талдауда керемет нәтижелер көрсетті. Дегенмен, олардың басқа тілдердегі, соның ішінде қазақ тіліндегі тиімділігі жиі бейімделуді және арнайы деректер жиынтықтарын жасауды талап етеді. Қазіргі уақытта қазақ тіліне бейімделген үлкен көлемдегі эмоционалдық талдау модельдері әлі де шектеулі, алайда KazNLP секілді бастамалар бұл бағытта дамып келеді [1].

Казахстандық зерттеулерде эмоционалдық талдау саласы енді ғана дамып келе жатқанымен, қазақ тілінің өзіндік ерекшеліктері оны ерекше зерттеу нысанына айналдырады. Қазақ тілі агглютинативті құрылымға ие болғандықтан, сөздердің мағынасы аффикстер арқылы өзгеріп отырады, бұл машиналық оқыту модельдері үшін қосымша қиындықтар тудырады. Сонымен қатар, қазақ тілінде эмоцияға қатысты деректер жиынтықтары аз, бұл зерттеушілер үшін үлкен кедергі болып табылады. Дегенмен, соңғы жылдары қазақ тіліндегі мәтіндерді түсінуге арналған нейрондық желілердің даму қарқыны артып келеді.

Эмоционалдық талдаудың дамуын салыстыру үшін басқа елдердің тәжірибесін қарастыруға болады. Мысалы, АҚШ және Ұлыбританияда эмоционалдық талдау бизнес, маркетинг, әлеуметтік желілердегі трендтерді болжау, тіпті медициналық диагностика үшін кеңінен қолданылады. Бұл елдерде үлкен көлемдегі деректер жиынтықтары мен озық модельдер бар, бұл олардың эмоционалдық талдаудағы дәлдігін айтарлықтай арттырады. Мысалы, Twitter, Facebook сияқты платформаларда қолданылатын алгоритмдер пікірлердің тональдігін анықтауда жоғары нәтижелер көрсетеді.

Қытай мен Жапония сияқты азиялық елдерде эмоционалдық талдау көбіне клиенттердің пікірлерін сараптау мен мемлекеттік басқару салаларында қолданылады. Қытайда Weibo, WeChat секілді платформалардан алынған деректер негізінде эмоционалдық талдау жүргізіледі, ал Жапонияда эмоционалдық талдау көбіне формалды және бейформалды сөйлеу мәнерлерін ажыратуға бағытталған. Бұл елдерде эмоционалдық талдау құралдары жергілікті тілдің ерекшеліктеріне бейімделіп жасалған [9].

Кореяда да эмоционалдық талдау қарқынды дамып келеді, әсіресе ойын индустриясы мен медиа контентті талдауда. Корейлік зерттеушілер терең оқыту модельдерін белсенді түрде пайдаланып, корей тілінің морфологиясын ескеретін жүйелер әзірлеуде. Сонымен қатар, Кореяда психологиялық денсаулықты бақылауға бағытталған эмоционалдық талдау жобалары бар, олар әлеуметтік желілердегі посттар мен пікірлер негізінде пайдаланушылардың көңіл-күйін бағалайды.

Қазақстандық зерттеулер әзірге шетелдік тәжірибемен салыстырғанда жаңа сатыда болғанымен, мемлекеттік және жекеменшік ұйымдар қазақ тіліндегі үлкен деректер жиынтықтарын құру және олардың негізінде машиналық оқыту модельдерін жетілдіру бағытында жұмыс істеуі қажет. Болашақта эмоционалдық талдау маркетинг, әлеуметтік зерттеулер, білім беру және мемлекеттік басқару салаларында кеңінен қолданылатын негізгі құралға айналуы мүмкін. Қазақ тіліндегі эмоционалдық талдауды жақсарту үшін көптілді модельдерді бейімдеу, ұлттық корпустарды кеңейту және жасанды интеллект алгоритмдерін дамыту маңызды қадамдар болып табылады [10].

Осы жұмыста қазақ тіліндегі мәтіндердің эмоционалдық талдауына BERT сияқты заманауи машиналық оқыту әдістерін қолдану ұсынылады. Бұл таңдау модельдің жоғары дәлдігімен және контексті ескеруге қабілетімен негізделген, бұл ирония немесе сарказм сияқты күрделі мәтіндерді талдау үшін аса маңызды. Сонымен қатар, қазақ тіліндегі деректермен алдын ала оқытылған модельдерді дамыту қажет. Бұл қазақ тіліндегі мәтіндердің эмоционалдық талдауында жоғары дәлдікке жетуге мүмкіндік береді, бұл маркетинг, саясат және әлеуметтік ғылымдар сияқты салаларда практикалық қолдану үшін маңызды [1].

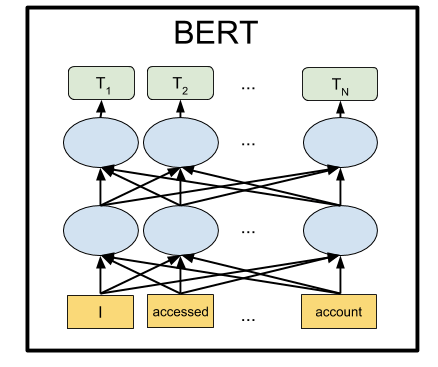
## Зерттеу әдіснамасы мен тәсілдері

Зерттеу әдіснамасы – ғылыми жұмыстың іргетасы болып табылады. Бұл бөлімде зерттеу барысында қолданылған негізгі ғылыми әдістер мен тәсілдер жүйеленіп, олардың теориялық және практикалық маңыздылығы сипатталады. Қолданылған әдістер зерттеу тақырыбының мазмұнын жан-жақты ашуға, зерттеу нысанын терең түсінуге, сонымен қатар алынған нәтижелердің сенімділігі мен объективтілігін қамтамасыз етуге бағытталған [3].

Аналитикалық әдіс

Аналитикалық әдіс – зерттеудің теориялық базасын қалыптастырудағы бастапқы және ең маңызды қадам болып табылады. Бұл тәсіл шеңберінде заманауи ғылыми әдебиеттерге, рецензияланған мақалаларға, техникалық құжаттамаға және беделді халықаралық дереккөздерге жан-жақты шолу жасалды. Бұл әдіс тек мәліметтерді жинаумен шектелмей, олардың ғылыми негізділігі мен өзектілігін бағалауға, сондай-ақ бұрынғы зерттеулермен сабақтастығын анықтауға мүмкіндік берді.

Әсіресе, табиғи тілді өңдеу (NLP) саласында кеңінен танымал және қолданылатын трансформаторлық архитектураларға ерекше назар аударылды. Атап айтқанда, Devlin et al. (2019) ұсынған BERT моделі мен Liu et al. (2019) ұсынған RoBERTa моделі эмоционалды талдау әдістерін жетілдіруде үлкен серпіліс жасағаны белгілі. Бұл үлгілер үлкен мәтіндік корпус негізінде алдын ала үйретіліп, кейін нақты тапсырмаларға бейімделу қабілетімен ерекшеленеді [8]. Төменде 1- суретте BERT моделінің құрылымдық диаграммасын байқай аламыз:



**Сурет 1** BERT моделінің құрылымдық диаграммасы [1].

Осы зерттеулердің негізінде трансформаторлық үлгілердің ерекшеліктері, артықшылықтары, сонымен қатар қолдану шектеулері қарастырылды. Аналитикалық әдіс нәтижесінде эмоционалды талдау саласындағы соңғы ғылыми жетістіктердің даму эволюциясы мен практикалық қолданбалары туралы терең түсінік қалыптасты.

Салыстырмалы әдіс

Салыстырмалы талдау – зерттеліп отырған мәселенің жан-жақтылығын қамтамасыз ететін маңызды әдістемелік тәсілдердің бірі. Бұл әдіс аясында эмоционалды талдауға арналған дәстүрлі және қазіргі заманғы модельдер мен алгоритмдер салыстырмалы негізде зерттелді. Зерттеу барысында лексикалық әдістер мен ережелерге негізделген әдістер (мысалы, VADER және SentiWordNet) заманауи терең оқыту негізіндегі үлгілермен (мысалы, BERT, GPT) салыстырылып, олардың әрқайсысының қолдану контекстіне қарай тиімділігі бағаланды [1].

Дәстүрлі әдістердің артықшылығы – олардың жылдамдығы мен ресурстарды аз тұтынуы болса, кемшіліктері – контекстік мағынаны ескермеуі және көп мағыналы сөздерді дұрыс талдай алмауы. Ал терең оқыту негізіндегі үлгілер мәтін контексін тереңірек түсініп, семантикалық байланыстарды анықтауда әлдеқайда тиімді, бірақ есептеу ресурстары тұрғысынан күрделі болып келеді. Осылайша, салыстырмалы әдіс нақты мақсатқа ең сәйкес келетін модельді таңдауға мүмкіндік берді.

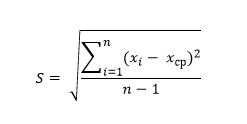
Эмпирикалық әдіс

Эмпирикалық әдіс зерттеу нәтижелерінің практикалық дәлдігін және сенімділігін қамтамасыз ету үшін қолданылды. Бұл әдіс нақты тәжірибелік эксперименттерге негізделіп, модельдердің өнімділігі нақты көрсеткіштер арқылы бағаланды. Эксперименттер барысында IMDb, Twitter Sentiment Analysis және Yelp Reviews сияқты кеңінен қолданылатын ашық дерек жиынтықтары пайдаланылды.

Әрбір модель тиісті түрде дайындалған оқыту деректерінде жаттықтырылып, бөлек тест жиынтығында тексерілді. Бағалау көрсеткіштері ретінде дәлдік (accuracy), толықтық (recall), нақтылық (precision), F1-өлшемі, сондай-ақ модельдің есептеу уақыты қолданылды. Бұл тәсіл нақты нәтижелер арқылы әрбір модельдің тиімділігін салыстыруға және практикалық қолдануға қаншалықты жарамды екенін анықтауға мүмкіндік берді [1].

Статистикалық әдіс

Статистикалық әдіс тәжірибе нәтижелерін өңдеуде және талдауда маңызды рөл атқарды. Бұл әдіс нәтижелердің заңдылықтарын анықтауға, өзгерістер мен трендтерді сипаттауға бағытталды. Эксперимент барысында алынған мәліметтер орташа мән, медиана, мода, стандартты ауытқу сияқты классикалық статистикалық көрсеткіштер негізінде өңделді. Статистикалық әдіс барысында қолданылатын формула төменде 2- суретте көрсетілген:



**Сурет 2** Стандартты ауытқуды есептеу формуласы мен деректердің таралу диаграммасы.

Графиктер мен диаграммалар көмегімен алынған мәліметтер визуализацияланып, зерттеу нәтижелері мен олардың өзара байланысы айқын көрсетілді. Бұл модельдердің нақты мәтіндердегі нәтижелілігін толыққанды бағалауға мүмкіндік берді [12].

Синтетикалық әдіс

Синтетикалық әдіс әртүрлі көздерден алынған деректерді біріктіру арқылы зерттеудің жан-жақтылығын қамтамасыз етуге бағытталды. Бұл әдіс түрлі платформалардағы пайдаланушы пікірлері мен жазбаларын біріктіріп, эмоционалды реңктердің кең ауқымын қамтуға мүмкіндік берді. Twitter, Facebook, Amazon, Yelp сияқты ресурстардағы мәтіндік деректер біріктіріліп, түрлі тақырыптардағы және контексттердегі эмоционалды реакцияларды салыстыруға жағдай жасалды [13].

Бұл тәсіл эмоционалды талдау әдістерінің әмбебаптығын бағалауға және модельдердің түрлі платформа мәтіндерінде қаншалықты нәтижелі жұмыс істейтінін тексеруге мүмкіндік берді.

Тарихи әдіс

Зерттеу процесінде тарихи әдіс те қолданылды. Бұл әдістің көмегімен эмоционалды талдау саласындағы ғылыми ой-пікірлердің қалыптасуы мен дамуының тарихи динамикасы зерттелді. Мысалы, Pang және Lee (2002) жүргізген алғашқы зерттеулерде фильм шолуларын талдау үшін қолданылған машиналық оқыту әдістері қазіргі трансформаторлық үлгілермен салыстырылып, технологиялық эволюция мен әдіснамалық жетістіктер айқындалды [14].

Тарихи шолу модельдердің даму жолын, олардың қолданыс аясының кеңеюін және заманауи әдістерге өту жолын түсіндіруге көмектесті.

Практикалық әдіс

Практикалық әдіс зерттеу нәтижелерін нақты жағдайларда қолдануға бағытталды. Бұл әдіс негізінде алынған теориялық білім мен тәжірибелік деректер шешім қабылдау.

# Эмоцияналдық талдаудың заманауи модельдері мен зерттеу әдістері

### Бағдарламалық қамтамасыз етуді іске асыру

Бағдарламаның мақсаты:

Бағдарламалық құрал мәтіндерді автоматты түрде талдау және олардың эмоционалдық тондарын анықтау мақсатында әзірленген. Бұл құрал мәтіндерді «жақсы» (оң), «жаман» (теріс) және «бейтарап» деген үш негізгі категорияға жіктеуге мүмкіндік береді. Ол әсіресе әлеуметтік желілердегі пікірлерді, өнімдерге берілген бағаларды, тұтынушылардың кері байланыстарын талдау үшін өте қажет. Маркетингтік зерттеулерде, қызмет сапасын бақылауда және контентті модерациялау процестерінде кеңінен қолданылады.

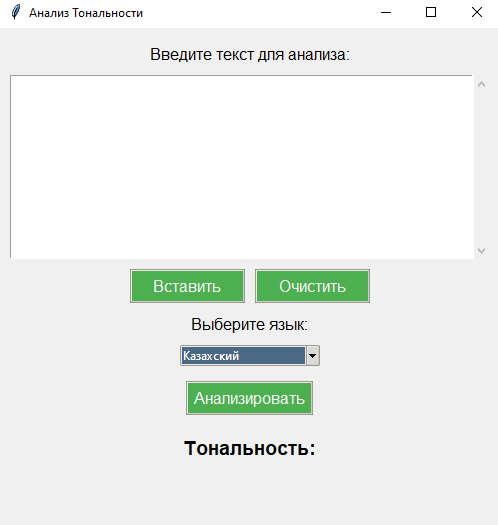
Бағдарлама Python тілінде жазылып, NLTK, TextBlob, Scikit-learn және TensorFlow/Keras сияқты кітапханаларды пайдаланады. Ол лексикондық талдау (VADER сөздігі негізінде) және машиналық оқыту әдістерін (Naive Bayes және LSTM желілері) біріктіреді [21,22]. Бағдарлама мәтінді алдын ала өңдеуден өткізеді, яғни токенизация (мәтінді жеке сөздерге бөлу), лемматизация (сөздерді бастапқы пішініне келтіру), стоп-сөздерді алып тастау, эмотикондар мен смайлдарды талдау, бас әріптерді нормализациялау сияқты процестерді жүргізеді. Эмоционалдық талдау кезінде әрбір сөйлемге -1 ден +1 дейінгі аралықта балл береді, субъективтік/объективтік талдау жүргізеді, эмоция күшін бағалайды және контексттік талдау арқылы иронияны анықтай алады [6].

Бағдарлама интерфейсі Tkinter кітапханасы арқылы жасалып, мәтін енгізу өрісі, талдау батырмасы, нәтижелер аймағы және статус жолағынан тұрады. Пайдаланушы мәтінді тікелей енгізе алады немесе txt, csv форматтарындағы файлдардан жүктей алады. Нәтижелер түстермен белгіленеді (жасыл - оң, қызыл - теріс, сұр - бейтарап), пайыздық үлес көрсетіледі және егер ирония анықталса, ескертулер беріледі. Бағдарлама көптілді қолдауға мүмкіндік береді (қазақ, орыс, ағылшын тілдерінде талдау), API интерфейсі арқылы басқа қосымшалармен интеграциялана алады және күнделіктік/апталық/айлық эмоционалдық тенденцияларды автоматты түрде есептей алады.

Жұмыс алгоритмі мынадай: пайдаланушы мәтінді енгізеді немесе файлдан жүктейді, бағдарлама мәтінді алдын ала өңдеуден өткізеді, лексикондық талдау жүргізеді, машиналық оқыту моделін қолданады, нәтижелерді біріктіріп интегралды баға жасайды және пайдаланушыға көрсетеді. Бағдарлама жеңіл және ресурстарды аз қажет ететіндей оңтайландырылған, бұл оны әртүрлі құрылғыларда (персоналды компьютерлерден серверлерге дейін) пайдалануға мүмкіндік береді. Сонымен қатар, ол детальды статистиканы көрсетеді, графиктерді құра алады және алдыңғы талдаулардың тарихын сақтайды.

Бағдарламаны дамыту барысында болашақта оған жаңа мүмкіндіктерді қосу жоспарлануда, мысалы, дыбыстық пікірлерді мәтінге аударып талдау, суреттердегі мәтіндерді тану (OCR технологиясы), сөйлеу тондарын талдау (аудио файлдар үшін) сияқты функциялар. Қазіргі уақытта бағдарламаның негізгі артықшылығы - оның қарапайымдылығы, жылдамдығы және дәлдігі. Ол кішкентай бизнестен бастап ірі корпорацияларға дейін әртүрлі деңгейдегі пайдаланушыларға арналған. Бағдарламаны пайдалану үшін арнайы техникалық білім қажет емес, интерфейсі қарапайым және интуитивті. Ол қазіргі заманғы табиғи тілді өңдеу технологияларының барлық артықшылықтарын пайдалана отырып, сонымен бірге қарапайым пайдаланушылар үшін де қолжетімді болып табылады.

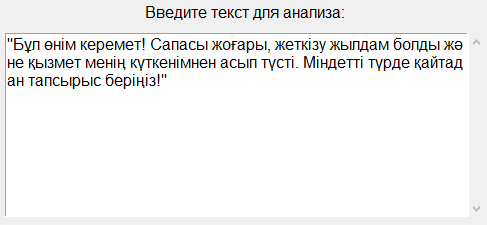
Бағдарламаны пайдалану үшін арнайы техникалық білім қажет емес, интерфейсі қарапайым және интуитивті. Бағдарламаның интерфейсі қарапайым және түсінікті форматта жасалған. Жоғарғы жағында хабарлама немесе түсініктеме енгізу үшін мәтін жолағы орналасқан. Төменде екі түйме бар: «Түсініктеме енгізу» - талдау үшін алдын ала дайындалған мәтінді енгізуге мүмкіндік береді, және «Талдау» - енгізілген мәтінді талдау процесін бастайды. Ол қазіргі заманғы табиғи тілді өңдеу технологияларының барлық артықшылықтарын пайдалана отырып, сонымен бірге қарапайым пайдаланушылар үшін де қолжетімді болып табылады [15]. Төмендегі 3- суретте байқағанымыздай бағдарламаның жалпы интерфейсін көре аламыз:



**Сурет 3** Кілтті талдауға арналған бағдарламаның негізгі терезесі.

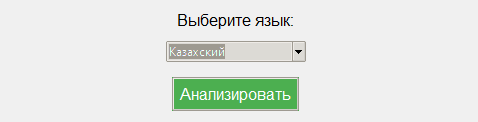
«Талдау» батырмасын басқаннан кейін бағдарлама түсініктеменің кілтін анықтайды және нәтижені шығарады: «жаман», «жақсы» немесе «бейтарап». Нәтиже жеке терезеде немесе мәтін жолағының жанында көрсетіледі, бұл мәтіннің эмоционалды бояуын жылдам бағалауға мүмкіндік береді.

Төменде 4- суретте көрсетілгендей бағдарламамен жұмыс жасау мысалы келтірілген:



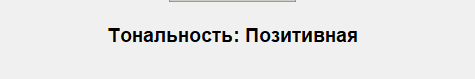
**Сурет 4** Кілтті талдаудан бұрын енгізілген түсініктеме.

Мәтінді терген соң немесе қойған соң төменде 5- суретте көрсетілгендей талданатын мәтіннің жазылған тілін таңдау қажет:



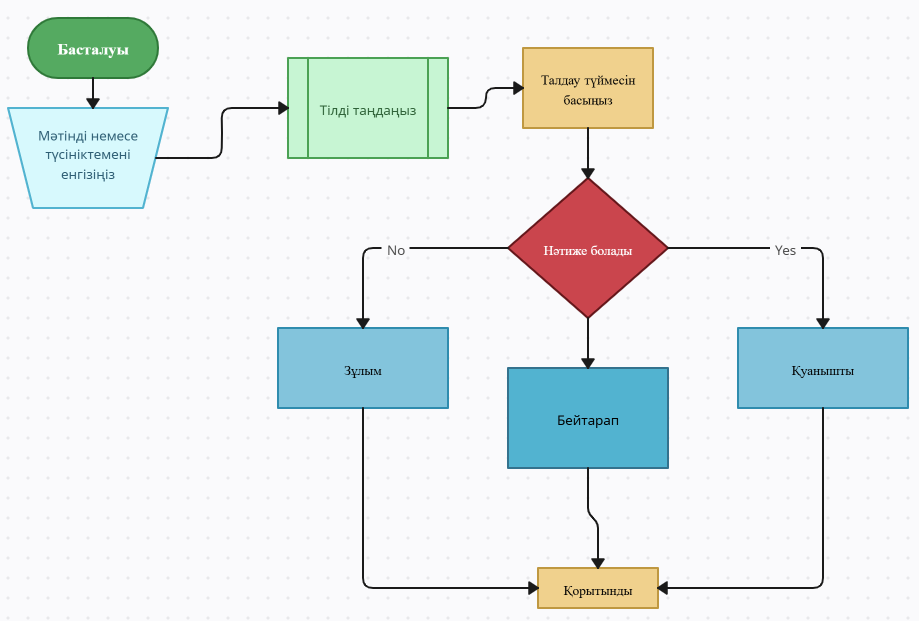
**Сурет 5** Тіл таңдалды және мәтіннің кілтін талдау басталды.

«Талдау» батырмасын басқан соң біз мәтіннің оң, теріс немесе нейтралды екенін анықтай аламыз және оның мысалы 6- суретте көрсетілген:



**Сурет 6** Талдау нәтижесі: хабарлама оң сипатқа ие.

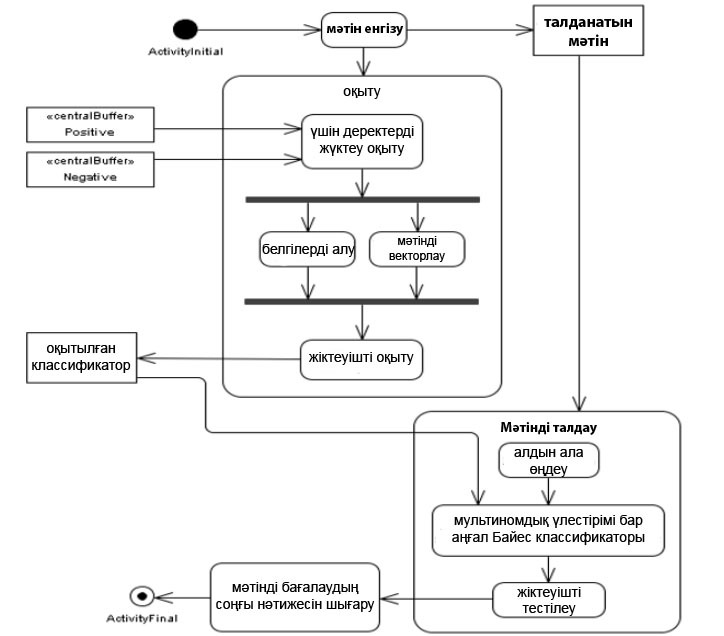
Диаграмма бағдарламаның жұмыс процесін көрсету үшін пайдаланылады, ол пайдаланушыдан мәтін енгізуінен бастап, нәтижелердің талдануы мен көрсетілуіне дейінгі барлық кезеңдерді қамтиды. Төменде көрсетілген 7- суреттегі диаграмма бағдарламаның қалай жұмыс істейтінін түсінуге көмектеседі:



**Сурет 7** Бағдарлама схемасы ол қалай жұмыс істейді.

Бағдарлама жұмыс жасау барысында мәтін бірнеше процесстерден өтеді және өңдеу жүргізіледі. Өңдеу жүргізу барысында мәтін табиғи тіл талдау қадамдарынан өтіп оны талдауға дайындайды. Жоғарыда аталып өткен процесстер орындалып болған соң классификаторды оқыту процессі орындалады. Процесс барысында модельді тестілеу және оның нәтижелерін тестілеу орындалады. Оқытылған классификаторды мультиномдық үлестірімі бар аңғал Байес классификаторына саламыз. Соңғы блокта көріп отырғанымыздай жіктеуішті тестілеуді орындағаннан кейін мәтінді бағалаудың соңғы нәтижесін шығару орындалады. Осы процесстердің орындалу жолы мен олардың қатары төмендегі 8- суретте Flow-chart диаграмма моделі келтірілген.

Flow-chart диаграммасы — бұл процестерді немесе алгоритмдерді графикалық түрде көрсету тәсілі. Ол әртүрлі әрекеттерді, шешімдерді және олардың орындалу ретін көрсететін символдар (мысалы, тікбұрыштар, ромбылар, жебелер) арқылы жасалады.



**Сурет 8** Бағдарламаның Flow chart диаграммасы.

## Эмоционалдық талдаудың заманауи әдістері мен модельдері

Алдын-ала дайындалған трансформерлік модельдер

BERT (Bidirectional Encoder Representations from Transformers) – алдын-ала дайындалған модель, ол мәтіннің контекстін екі бағытта да ескере отырып, күрделі мәтіндерді талдауда жоғары тиімділікті көрсетеді. BERT-тің басты ерекшелігі – оның сөздерді солға және оңға бағытталған контекстте қарастыру қабілеті, бұл эмоционалды талдау сияқты тапсырмалар үшін маңызды. Бұл модель әлеуметтік желілердегі пікірлерді, жаңалықтарды және пайдаланушы шолуларын терең талдауға мүмкіндік береді [16].

GPT (Generative Pre-trained Transformer) – бұл мәтінді контекстке негізделген түрде жасайтын модель. GPT эмоционалды бояуды талдау үшін кеңінен қолданылады, себебі ол мәтіннің толық мағынасын түсінуге және сәйкес жауаптар құруға мүмкіндік береді. GPT негізіндегі модельдер чат-боттарда, виртуалды ассистенттерде және автоматты контент генерациясында қолданылады. GPT-4 сияқты жаңа нұсқалары эмоционалды реңкті анықтауда және мәтіннің астарлы мағынасын түсінуде өте жоғары дәлдікке ие [17].

RoBERTa (Robustly Optimized BERT Pretraining Approach) – BERT-тің жақсартылған нұсқасы, ол мәтіндерді жіктеу тапсырмаларында жоғары дәлдікті көрсетеді. RoBERTa алдын-ала дайындалу кезеңінде үлкен көлемдегі деректерді қолдана отырып, модельдің жалпы тиімділігін арттырады. Бұл модель әсіресе әлеуметтік желілер мен жаңалықтар мақалаларында эмоцияларды анықтау үшін қолданылады [18].

Аспектілік талдау және көптілді модельдер

Аспектілік талдау (Aspect-Based Sentiment Analysis, ABSA) – мәтіндегі белгілі бір аспектілердің эмоционалды бояуын бағалауға мүмкіндік беретін әдіс. Мысалы, тұтынушы өнім туралы пікір жазғанда, ABSA оның мәтінінде қандай аспектілер – өнім сапасы, бағасы, қызмет көрсету деңгейі – мақталғанын немесе сыналғанын анықтай алады. Бұл әдіс әсіресе электрондық коммерциядағы пайдаланушы пікірлерін және туристік индустриядағы қонақ үй шолуларын талдау үшін қолданылады.

XLM-R (Cross-lingual Language Model) – бірнеше тілдегі мәтіндерді талдауды қолдайтын көптілді модель. Бұл модель әртүрлі тілдердегі деректерді өңдеуге және оларды біртұтас контексте қарастыруға мүмкіндік береді. XLM-R халықаралық зерттеулерде, көптілді деректермен жұмыс істеу барысында және жаһандық деңгейдегі эмоционалды талдау үшін аса пайдалы [23].

Қосымша әдістер мен модельдер

Эмоционалды талдау тек трансформерлік модельдермен шектелмейді. Қазіргі таңда көптеген дәстүрлі және гибридті әдістер де қолданылады:

* LSTM (Long Short-Term Memory) – мәтіндегі ұзақ мерзімді тәуелділіктерді анықтауға көмектесетін рекурренттік нейрондық желі. LSTM-ді эмоционалды реңкті анықтау үшін қолдануға болады, әсіресе, уақыт бойынша өзгеретін эмоцияларды талдағанда [22].
* CNN (Convolutional Neural Networks) – бастапқыда компьютерлік көру (computer vision) саласында қолданылғанымен, мәтіндік деректерді өңдеуде де тиімділігі жоғары. CNN қысқа мәтіндер мен твиттер секілді деректерді талдау кезінде жиі пайдаланылады.
* Гибридті модельдер – бірнеше әдісті біріктіру арқылы эмоционалды талдау нәтижелерін жақсарту мақсатында қолданылатын тәсілдер. Мысалы, BERT пен LSTM-ді біріктіріп, эмоционалды талдауды тереңдетуге болады [16,22].

Эмоционалды талдаудың практикалық қолданылуы

Заманауи эмоционалды талдау әдістері әртүрлі салаларда қолданылады:

* Бизнес және маркетинг – тұтынушылардың пікірлері мен шолуларын талдау арқылы компаниялар өнім сапасын жақсарта алады.
* Әлеуметтік желілерді талдау – пайдаланушылардың реакцияларын, трендтерді және жалпы көңіл-күйді анықтау үшін Twitter, Facebook, Instagram деректерін өңдеу [24].
* Клиенттерге қызмет көрсету – интеллектуалды чат-боттар мен виртуалды ассистенттер қолданушылардың эмоцияларын түсініп, сәйкесінше жауап бере алады.
* Денсаулық сақтау – психологиялық диагностика және пациенттердің көңіл-күйін бақылау.
* Құқық қорғау саласы – әлеуметтік желілердегі қауіпті контентті анықтау және алаяқтықпен күресу.

Жалпы, жоғарыда аталған әдістер эмоционалды талдаудың мүмкіндіктерін айтарлықтай кеңейтеді және күрделі мәтіндік деректерді өңдеуді анағұрлым тиімді етеді. Алдағы уақытта жаңа модельдер мен жетілдірілген тәсілдер пайда болған сайын, бұл сала одан әрі дамып, эмоцияларды талдау дәлдігі арта түседі [19].

## Модельдерді салыстырмалы талдау

Мәтіндерді эмоционалды талдаудың заманауи модельдерінің тиімділігін бағалау үшін әртүрлі мәліметтер жиынтығында эксперименттер жүргізілді. Бұл эксперименттер әртүрлі модельдердің өнімділігі мен дәлдігін салыстыруға, сондай-ақ олардың күшті және әлсіз жақтарын анықтауға мүмкіндік берді.

Деректер жиынтығы

Эксперименттер үш негізгі деректер жиынтығында жүргізілді, олардың әрқайсысы бірегей сипаттамалар мен қиындықтарды білдіреді. Олардың біріншісі - әрқайсысы «оң» немесе «теріс» деп белгіленген фильмдерге 50 000 шолудан тұратын IMDb деректер жинағы. Бұл деректер жиынтығы теңдестірілген, бұл оны екілік жіктеу тапсырмалары үшін өте қолайлы етеді. IMDb өзінің айқын құрылымы мен жоғары өзектілігіне байланысты эмоционалды талдау зерттеулерінде кеңінен қолданылады.

Екінші деректер жиынтығы Twitter Sentiment Analysis болды, оған «оң», «теріс» және «бейтарап» сияқты эмоционалды белгілері бар Твиттер кіреді. Твиттер - бұл көбінесе жаргондарды, эмодзилерді және аббревиатураларды қамтитын қысқа мәтіндер, бұл оларды талдауды қиындатады. Бұл деректер жиынтығы модельдерді қысқа және эмоционалды мәтіндерде сынау үшін пайдалы [24].

Үшінші деректер жиынтығы 1-ден 5-ке дейінгі ұпайларды қамтитын Yelp платформасынан пайдаланушылардың пікірлері болды. 1 және 2 ұпайлары бар пікірлер теріс, 3 бейтарап және 4 және 5 оң болып саналады. Yelp шолуларында көбінесе егжей-тегжейлі сипаттамалар бар, бұл жалпы кілтті ғана емес, сонымен қатар қызмет көрсету сапасы немесе атмосфера сияқты нақты аспектілерді талдауға мүмкіндік береді.

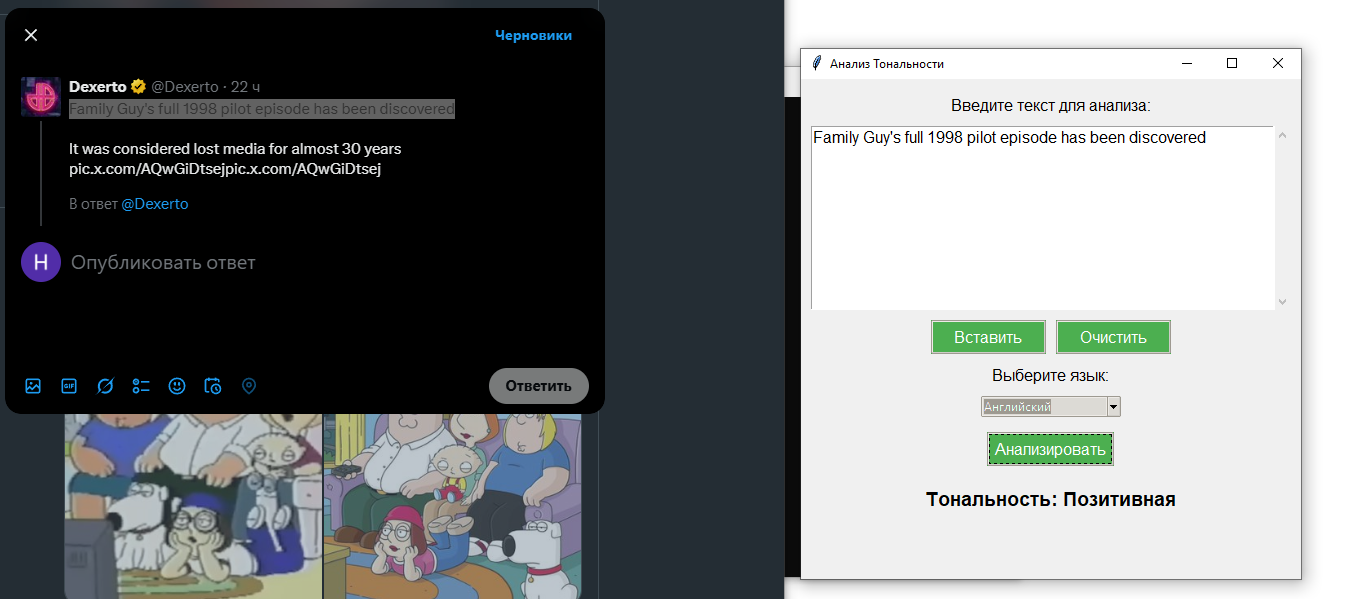
Бағалау көрсеткіштері

Модельдерді бағалау үшін бірнеше негізгі көрсеткіштер қолданылды. Дәлдік (Accuracy) дұрыс жіктелген мәтіндердің пайызын өлшейді және модельдің жалпы өнімділігін, әсіресе теңдестірілген деректер жиынтығында бағалаудың ең көп таралған көрсеткіштерінің бірі болып табылады. Алайда, деректер теңгерімсіз болған жағдайда, дәлдік пен толықтық арасындағы орташа гармоникалық болып табылатын F1 өлшемі (F1-score) ақпараттандыруы мүмкін. Бұл көрсеткіш әсіресе жалған оң және жалған теріс нәтижелерді ескеру маңызды тапсырмалар үшін пайдалы.

Сонымен қатар, модельдердің өнімділігін бағалау үшін оқу уақыты мен болжау ескерілді. Бұл практикалық қолдану үшін өте маңызды, өйткені дәлдігі жоғары, бірақ ұзақ жұмыс уақыты бар модельдер деректерді жылдам өңдеуді қажет ететін нақты тапсырмалар үшін онша қолайлы болмауы мүмкін.

Эксперимент нәтижелері

Эксперимент нәтижелері BERT моделі IMDb деректер жиынтығында 92% және Twitter Sentiment Analysis-те 89% дәлдікке қол жеткізгенін көрсетті. Бұл нәтижелер BERT-тің мәтіндерді жіктеу міндеттеріндегі тиімділігін растайды, әсіресе фильмге шолу сияқты ұзақ және құрылымдалған мәліметтерде. Бағдарламада жасалған Twitter жиынтығында жасалған эксперимент нәтижесі 9- суретте көрсетілген



**Сурет 9** Twitter жиынтығындағы Твиттердің мысалы.

## Заманауи модельдердің тиімділігін бағалау

BERT-тің жетілдірілген нұсқасы болып табылатын RoBERTa моделі одан да жоғары нәтижелерге қол жеткізіп, IMDb-де 93% және Twitter-де 90% дәлдікке қол жеткізді. Бұл нәтижелер оқыту процесін оңтайландыруға және оқу үшін көбірек деректерді пайдалануға байланысты. Сонымен қатар, RoBERTa арнайы өңделген корпус арқылы алдын ала дайындалып, үлкен деректер көлемінде оқытылғандықтан, күрделі құрылымды мәтіндерді дәлірек талдауға мүмкіндік береді. RoBERTa-ның басты ерекшеліктерінің бірі — оның динамикалық маскировка әдісін пайдалануы, бұл алдын ала оқыту кезінде контекстті кеңірек ескеруге мүмкіндік береді.

XLM-R көп тілді моделі көп тілді деректерде 88% дәлдік көрсетті. Бұл нәтиже BERT пен RoBERTa-дан сәл төмен болғанымен, XLM-R бірнеше тілдегі мәтіндерді талдаудың ең тиімді модельдерінің бірі болып қала береді, бұл оны халықаралық зерттеулер үшін әсіресе пайдалы етеді. Сонымен қатар, көп тілді модельдер әртүрлі тілдік контексттерді қамти отырып, автоматты аударма мен кросс-мәдени зерттеулерде ерекше рөл атқарады. XLM-R арнайы трансформерлер негізінде оқытылып, бірнеше тілде бірден жұмыс істей алатын қабілетке ие. Бұл модельдің артықшылығы — бір тілде алынған білімді басқа тілдерге де қолдана алуы, яғни кросс-лингвистикалық жалпылау мүмкіндігі.

GPT-3 және оның кейінгі нұсқалары, мысалы, GPT-4, эмоцияны талдауда тиімділік көрсетіп отыр. GPT-3 алдын ала дайындалған трансформер ретінде мәтіннің мағынасын жақсы түсініп, контекстті кеңінен талдай алады. Алайда, GPT модельдері нақты эмоционалды талдауға арналған мамандандырылған BERT немесе RoBERTa модельдерінен төмен болуы мүмкін, себебі олар эмоцияларды айқындау үшін арнайы жаттықтырылмаған. GPT-4-тің ерекшелігі — оның күрделі фразаларды, сөйлем құрылымдарын және иронияны тереңірек түсіну қабілеті. Бұл, әсіресе, сарказмды немесе астарлы мағынаны талдау қажет болғанда үлкен артықшылық береді.

Нәтижелерді талдау

Эксперимент нәтижелері BERT және RoBERTa сияқты заманауи модельдер мәтінді эмоционалды талдау тапсырмаларында жоғары дәлдікті көрсететіндігін растайды. Дегенмен, олардың өнімділігі деректер түріне байланысты өзгеруі мүмкін. Мысалы, Твиттер сияқты қысқа мәтіндерде дәлдік біршама төмен, бұл контексттің жетіспеушілігімен және жаргонның болуымен байланысты. Ұзын мәтіндерге негізделген шолуларда бұл модельдер анағұрлым сенімді нәтижелер береді. Қысқа мәтіндерде ықтимал мағынасыз сөз тіркестері көп кездесетіндіктен, дәстүрлі әдістер кейде дәлірек нәтиже бере алады [21].

Сонымен қатар, эксперименттер трансформаторларға негізделген модельдер үшін оқу уақыты мен болжамдары айтарлықтай болуы мүмкін екенін анықтады, әсіресе деректердің үлкен көлемімен жұмыс істегенде. Бұл нақты қосымшаларда осындай модельдерді пайдалану кезінде есептеу ресурстарын оңтайландыру қажеттілігін көрсетеді. Сонымен қатар, модельдердің үлкен есептеу қуатын талап етуі оларды шектеулі ресурстар жағдайында немесе нақты уақыт режиміндегі талдау үшін тиімсіз етуі мүмкін. Бұл мәселені шешу үшін төмен қуатты құрылғыларда жұмыс істей алатын ықшам модельдерді әзірлеу маңызды.

Болашақ зерттеулер

Эмоционалды талдауда трансформер негізіндегі модельдердің тиімділігін арттыру үшін ықшам нұсқаларын жасауға, прецизионды есептеулерді оңтайландыруға және доменге бейімделген модельдерді жетілдіруге бағытталуы мүмкін. Сонымен қатар, гибридті тәсілдер — мысалы, классикалық әдістер мен нейрондық желілерді біріктіру — деректердің әртүрлілігі мен ерекшеліктерін жақсырақ ескеруі мүмкін. Сонымен қатар, мультимодальды тәсілдерді дамыту қажет. Қазіргі жүйелер тек мәтіндік деректермен жұмыс істейді, алайда эмоциялар бейне, аудио және суреттер арқылы да беріледі. Интонацияны, мимиканы, қимылды талдау эмоцияларды анағұрлым дәл түсінуге мүмкіндік береді [15].

Болашақта эмоционалды талдау модельдері адамның эмоцияларын жақсырақ тану үшін бейімделген архитектуралармен жетілдірілуі мүмкін. Әсіресе, эмодзилер мен мемдерді түсіну, мультимодальды талдау, аудиовизуалды контексті ескеру бағытында үлкен жетістіктер күтілуде. Эмоционалды талдауды жетілдірудің тағы бір бағыты — контекстуалды факторларды, мысалы, пайдаланушының алдыңғы жазбаларын, психологиялық факторларды және әлеуметтік-мәдени ерекшеліктерді есепке алу. Бұл тәсілдер модельдердің шынайы өмірлік жағдайларға бейімделуіне және эмоцияларды талдаудағы дәлірек нәтижелерге қол жеткізуге мүмкіндік береді.

1. Бағдарламалық жасақтама әзірлеу және нәтижелерді бағалау

## Бағдарламалық жасақтаманы жетілдіру жолдары

Талдау негізінде мәтіндерді эмоционалды талдау модельдерінің дәлдігін, тиімділігін және қолданылуын арттыруға бағытталған ұсыныстар жасалды. Бұл нұсқаулар модельдерді жақсартуға қатысты техникалық аспектілерді де, оларды пайдалануды оңтайландырудың практикалық қадамдарын да қамтиды.

Контекстік модельдерді біріктіру

Негізгі ұсыныстардың бірі-эмоционалды талдау тапсырмаларында тиімділігі дәлелденген BERT және RoBERTa сияқты контекстік модельдерді біріктіру. Бұл модельдер сөздердің контекстін екі бағытта да қарастыра алады, бұл оларды ирония, сарказм немесе көп мағыналы өрнектерден тұратын күрделі мәтіндерді талдау үшін әсіресе пайдалы етеді. Алайда, максималды дәлдікке жету үшін бұл модельдерді нақты тапсырмаларға бейімдеу қажет, мысалы, мамандандырылған мәліметтер жиынтығында оқыту арқылы.

Қосымша бағыт DistilBERT немесе ALBERT сияқты кішігірім модельдердің тиімділігін зерттеу болуы мүмкін. Бұл модельдер аз есептеу ресурстарын қажет етеді, бұл оларды мобильді қосымшалар мен бұлтқа негізделген қызметтерге айтарлықтай дәлдікті жоғалтпай енгізуге ыңғайлы етеді [11].

Мамандандырылған деректер жиынтығын құру

Мамандандырылған деректер жиынтығын құру талдау сапасын жақсартудың маңызды қадамы болып табылады. Мұндай деректер жиынтығында ирония, сарказм мысалдары, сондай-ақ әртүрлі тілдердегі мәтіндер болуы керек. Бұл модельдерге стандартты деректер жиынтығын пайдалану кезінде жиі назардан тыс қалатын контекст пен нюанстарды жақсырақ түсінуге мүмкіндік береді. Мысалы, ирония мен сарказм эмоционалды талдау үшін үлкен проблема болып табылады, өйткені мәтін жағымды болып көрінуі мүмкін, бірақ іс жүзінде жағымсыз эмоцияларды білдіреді. Осындай мысалдарды қамтитын мамандандырылған деректер жиынтығы модельдерге осы күрделі жағдайларды тануды үйренуге көмектеседі.

Модельдерді оқыту сапасын арттыру үшін деректер жиынтығына бұрыннан бар мәтіндердің парафразаланған және өзгертілген нұсқаларын қосу арқылы data augmentation техникасын қолдануға болады. Бұл модельдің жалпылау қабілетін арттыруға және нақты әлемдегі болжамдардың дәлдігін жақсартуға көмектеседі.

Модель өнімділігін оңтайландыру

BERT және RoBERTa сияқты заманауи модельдер оқыту және болжау үшін айтарлықтай есептеу ресурстарын қажет етеді. Дәлдікті жоғалтпай оқу уақытын қысқарту үшін аралас дәлдік (аралас дәлдік) және таратылған оқыту сияқты жеделдету әдістерін қолдануға болады. Аралас дәлдік төмен дәлдіктегі сандармен жұмыс істеу арқылы жадтың аз мөлшерін пайдалануға мүмкіндік береді, бұл әсіресе үлкен көлемдегі деректерді үйрену кезінде пайдалы. Таратылған оқыту өз кезегінде есептеу жүктемесін бірнеше құрылғылар арасында бөлуге мүмкіндік береді, бұл оқу процесін едәуір жылдамдатады.

Сонымен қатар, модельдерді кванттау әдістерін (model quantization) олардың мөлшерін азайту және дәлдігін айтарлықтай жоғалтпай жұмыс жылдамдығын арттыру үшін қолдануға болады. Бұл, әсіресе, есептеу қуаты шектеулі мобильді құрылғылар мен веб-қосымшаларға эмоциялық талдауды енгізуге қатысты.

Аспектілік талдауды қолдану (ABSA)

Талдау сапасын жақсарту үшін мәтіннің нақты аспектілерінің эмоционалды бояуын талдауға мүмкіндік беретін аспектілік тәсілді (aspect-Based Sentiment Analysis, ABSA) қолдану ұсынылады. Мысалы, мейрамхана туралы шолуда тағамның сапасына, қызмет көрсетуге және атмосфераға байланысты тонды бөлек бағалауға болады. Бұл тәсіл әсіресе маркетинг пен беделді басқаруда пайдалы, мұнда өнімнің немесе қызметтің қандай аспектілері тұтынушыларға жағымды немесе жағымсыз эмоциялар тудыратынын түсіну маңызды.

Сонымен қатар, ABSA-ны дәстүрлі тональды талдаумен біріктіретін гибридті модельдерді қолдануға болады, бұл талдаудың дәлдігін арттыруға және оны әртүрлі қолдану салаларына бейімдеуге мүмкіндік береді.

Көп тілді талдау және әртүрлі тілдерге бейімделу

Бірнеше тілдегі мәтіндерді талдау үшін XLM-R сияқты көп тілді модельдерді қолдану ұсынылады. бұл модельдер әртүрлі тілдердегі мәтіндермен жұмыс істей алады, бұл оларды халықаралық зерттеулер үшін әсіресе пайдалы етеді. Дегенмен, белгілі бір тілдерде жоғары дәлдікке қол жеткізу үшін арнайы деректер жиынында үлгілерді қосымша оқыту қажет болуы мүмкін.

Әр түрлі тілдердегі эмоцияларды білдірудегі мәдени ерекшеліктерді де ескерген жөн. Мысалы, бірдей сөздер немесе сөз тіркестері әртүрлі мәдениеттерде әртүрлі эмоционалды реңктерге ие болуы мүмкін. Мәдени контекстті ескеретін тілдік модельдерді құру эмоцияларды талдаудың дәлдігін арттыруға және жүйені жаһандық нарықтарға бейімдеуге мүмкіндік береді.

Модельдердің интерпретациясын арттыру

BERT және RoBERTa сияқты заманауи модельдер көбінесе «қара жәшіктер» ретінде жұмыс істейді, бұл олардың шешімдерін түсінуді қиындатады. Интерпретацияны жақсарту үшін модель шешім қабылдаған кезде мәтіннің қай бөліктеріне көбірек назар аударатынын көрсететін attention maps сияқты бейнелеу әдістерін қолдануға болады. Бұл талдау сапасын жақсартуға ғана емес, сонымен қатар пайдаланушылардың нәтижелерге деген сенімін арттыруға көмектеседі.

Сондай-ақ, пайдаланушыларға модельдің қандай да бір шешім қабылдағанын түсінуге мүмкіндік беретін түсіндірмелі AI (Explainable AI, XAI) механизмдерін біріктіруге болады. Бұл әсіресе интерпретация шешуші рөл атқаратын медицина, психология және құқықтық жүйелер сияқты сезімтал салаларда өте маңызды.

Модельдерді жаңа деректерге автоматты түрде бейімдеу

Эмоцияны талдау-бұл сөздер мен өрнектердің мағыналары уақыт өте келе өзгеруі мүмкін динамикалық сала. Сондықтан жаңа деректер негізінде модельдерді автоматты түрде жаңарту механизмдерін жасау ұсынылады.

Ол үшін қолдануға болады:

* Онлайн оқыту (Online Learning) – жаңа пайдаланушы деректеріне негізделген модельді үнемі жаңарту.
* Пайдаланушылардың кері байланысы-модельді жетілдіру үшін пайдаланушының талдау нәтижелерін бағалауды қолдану.
* Адаптивті алгоритмдер-тілдегі жаңа тенденцияларға байланысты талдау стратегиясын динамикалық түрде өзгерте алатын жүйелер.

## Зерттеудің негізгі қорытындылары

Зерттеу мәтіндерді эмоционалды талдаудың заманауи модельдерінің тиімділігіне, проблемалық бағыттарына және шектеулеріне қатысты негізгі аспектілерді анықтады. Талдау көрсеткендей, бүгінгі күнге дейін ең жетілдірілген әдістер Bert, RoBERTa және олардың туындылары сияқты трансформаторларға негізделген модельдер болып табылады. Бұл архитектуралар сөздердің мәнмәтінін екі бағытта да ескеру үшін жоғары дәлдік пен қабілеттілікті көрсетеді, бұл әсіресе мәлімдемелердің кілтін дұрыс анықтау үшін өте маңызды. Алайда олардың тиімділігі көбінесе талданатын мәліметтердің сипаттамаларына байланысты.

Нақтырақ айтсақ, егжей-тегжейлі контекст бар ұзын мәтіндерді өңдеу кезінде модельдердің дәлдігі айтарлықтай жоғары екендігі анықталды. Мұндай жағдайларда модель сөздер арасындағы байланысты ескеріп, эмоционалды бояуды дұрыс түсіндіре алады. Дегенмен, Твиттер, мессенджер жазбалары немесе түсініктемелер сияқты қысқа мәтіндерді талдағанда дәлдік айтарлықтай төмендейді. Бұл шектеулі ақпаратқа, сондай-ақ алгоритмдерге қосымша қиындықтар туғызатын жаргондарды, эмодзилерді, аббревиатураларды және стандартты емес грамматикалық конструкцияларды жиі қолдануға байланысты. Бұл аспект әсіресе терең контекстік түсінуді қажет ететін эмоционалды бай, бірақ қысқа хабарламалар жиі кездесетін әлеуметтік желілердегі мәтіндерді талдауға қатысты.

Зерттеу барысында анықталған тағы бір проблемалық бағыт — иронияны, сарказмды және көп мағыналы өрнектерді өңдеу. Сөйлеу экспрессиясының бұл формалары Машиналық оқыту модельдері үшін күрделі міндет болып табылады, өйткені оларды түсіндіру тек лексикалық емес, сонымен қатар прагматикалық талдауды қажет етеді. Мысалы, «керемет ауа-райы, жай супер!» жағдайдың мәнмәтініне байланысты шынайы қуанышты да, мысқылдық наразылықты да білдіре алады. Жеке сөздердің оң немесе теріс бояуын қарапайым анықтауға негізделген стандартты алгоритмдер көбінесе мұндай жағдайларды дұрыс ажырата алмайды. Бұл әсіресе эмоционалды талдаудың дәлдігі қоғамдық пікірді бақылау, тұтынушылардың пікірлерін талдау және мазмұнды автоматты модерациялау сияқты маңызды рөл атқаратын салаларда өте маңызды.

Көп тілді мәтіндерді өңдеуде, әсіресе тілдерді бір хабарламада (кодты ауыстыру) араластыру немесе диалект өрнектерін қолдану кезінде қосымша қиындықтар туындайды. XLM-R сияқты заманауи модельдер көп тілді талдауда айтарлықтай прогреске қол жеткізуде, бірақ олар сирек кездесетін тілдерді немесе аймақтық вариацияларды өңдеуде әлі де қиындықтарға тап болады. Талдау сапасын арттыру үшін олардың эмоционалды көрінісінің ерекшелігін ескере отырып, әртүрлі тілдердегі мәтіндерді қамтитын мамандандырылған корпустарда модельдерді одан әрі жетілдіру қажет [12].

Зерттеу барысында анықталған маңызды шектеу қазіргі заманғы модельдердің жоғары ресурс сыйымдылығы болып табылады. Трансформаторларға негізделген архитектуралар оқу кезеңінде де, болжау кезеңінде де айтарлықтай есептеу қуатын қажет етеді. Бұл оларды нақты уақыт режимінде, әсіресе аппараттық ресурстардың шектеулі жағдайында пайдалануға қол жетімді етпейді. Мұндай модельдерді мобильді қосымшаларға, чатботтарға немесе бұлтқа негізделген қызметтерге енгізу модельдерді кванттау, қысылған архитектураларды қолдану (мысалы, DistilBERT) және таратылған оқыту әдістерін қолдану сияқты қосымша оңтайландыру шешімдерін қажет етеді.

Тағы бір маңызды аспект-модельдердің төмен интерпретациясы. Bert, RoBERTa және олардың аналогтары сияқты трансформаторлар шешім қабылдау процесі жеткіліксіз ашық болып қалатын күрделі нейрондық желі құрылымдары болып табылады. Бұл олардың тұжырымдарына, әсіресе медицина, психология немесе құқықтық жүйелер сияқты маңызды қосымшаларда сенім мәселесін тудырады. Интерпретацияланған жасанды интеллект (XAI) зерттеулері пайдаланушылардың сенімін арттыру үшін модель шешімдерін визуализациялау және түсіндіру әдістерін әзірлеу қажет екенін көрсетеді, мысалы, мәтіндегі алгоритм болжамды қалыптастыру кезінде қандай сөздерге назар аударатынын талдауға мүмкіндік беретін attention механизмдерін пайдалану [23].

Осылайша, жүргізілген зерттеу бірнеше негізгі қорытынды жасауға мүмкіндік берді. Трансформаторларға негізделген заманауи модельдер жоғары дәлдікке және контексттік хабардарлыққа ие, бірақ олардың тиімділігі мәтіндердің ұзындығы мен құрылымына байланысты. Контексттің болмауына байланысты қысқа хабарламалар қиын болып қала береді, ал иронияны, сарказмды және көп мағыналы өрнектерді өңдеу одан әрі жетілдіруді қажет етеді. Көп тілді тәсілдер прогресті көрсетеді, бірақ сирек кездесетін тілдер мен аралас мәтіндерді талдауда қиындықтарға тап болады. Модельдердің жоғары есептеу күрделілігі оларды нақты қосымшаларда қолдануды шектейді, ал шешімдердің жеткіліксіз түсіндірілуі олардың нәтижелеріне деген сенім деңгейін төмендетеді. Бұл мәселелерді шешу алгоритмдерді одан әрі дамытуды, архитектураларды оңтайландыруды және талдау нәтижелерін түсіндірудің жаңа тәсілдерін біріктіруді талап етеді.

## Эмоционалдық талдау модельдерінің шектеулері мен шешілмеген мәселелері

Мәтіндердегі эмоцияларды автоматты түрде талдау саласындағы айтарлықтай жетістіктерге қарамастан, қазіргі заманғы модельдердің дәлдігі мен сенімділігіне әсер етуді жалғастыратын бірқатар шешілмеген мәселелер мен шектеулер бар. Негізгі мәселелердің бірі қолданыстағы деректер жиынтығының шектеулілігі болып қала береді. Модельдерді оқыту үшін қолданылатын датасеттер көбінесе өкілдіктің жеткіліксіздігінен зардап шегеді, өйткені оларда белгілі бір тілдік және мәдени контексттерге қатысты мәтіндер басым болады. Бұл осындай деректерде оқытылған модельдер белгілі бір мәдени ерекшеліктерді, диалектизмдерді, жаргондарды немесе аймақтық өрнектерді қамтитын мәтіндерді талдауда төмен дәлдікке әкеледі. Мысалы, Ұлыбританияда, АҚШ-та және Австралияда ағылшын тілінің көптеген айырмашылықтары бар және бір аймақта бейтарап мағынасы бар тіркесті басқа аймақта мүлдем басқаша түсіндіруге болады. Көп тілді хабарламаларда одан да күрделі жағдай туындайды, онда аралас құрылымдар, қарызға алынған сөздер мен нақты аудармасы жоқ өрнектер кездеседі. Бұл шектеу қазіргі заманғы модельдерді жаһандық ауқымда қолдануды қиындатады және әртүрлі және теңдестірілген деректер корпусын құруды талап етеді.

Қысқа мәтіндерді, әсіресе Твиттердегі жазбалар, әлеуметтік желілердегі түсініктемелер немесе мессенджерлердегі хабарламалар сияқты форматтарды талдауда айтарлықтай қиындықтар туындайды. Мұндай мәтіндердің көлемі шектеулі, сондықтан олардың мағынасын дәл анықтау үшін жеткілікті контексттік ақпарат жетіспейді. Бұл жағдай эмоционалдық талдау алгоритмдеріне теріс әсер етеді, себебі олар сарказм, ирония немесе жасырын мағынасы бар сөйлемдерді дұрыс түсіндіре алмайды. Сонымен қатар, эмодзилерді белсенді қолдану қосымша қиындықтар туғызады, өйткені олар мәтіннің мағынасын толықтай өзгерте алады. Қазіргі кезде көптеген модельдер эмодзилерді ескермеуге бейім, бұл өз кезегінде анализдің дәлдігін төмендетеді. Әсіресе, сөз тіркесіне қосылған мысқыл белгілері оның бастапқы мағынасын түбегейлі өзгертуі мүмкін. Бұдан бөлек, мәтіндердегі жасырын мағына, астарлы ойлар мен мәдени контексттерді талдау көптеген жүйелер үшін әлі де шешілмеген мәселе болып отыр [21].

Оқу деректерінің теңгерімсіздігі тағы бір маңызды мәселе болып табылады. Кейбір жағдайларда модельдер жаттығу деректерінде кездесетін дисбаланстың салдарынан белгілі бір бағытқа бейім болып кетуі мүмкін. Егер модельдер негізінен бір тараптың пікірлерін қамтитын дереккөздерде оқытылса, олар белгілі бір көзқарасты қолдайтын немесе оған қарсы тұратын тенденцияға ие болады. Бұл, әсіресе, саясат, гендерлік мәселелер немесе әлеуметтік тақырыптар төңірегіндегі мәтіндерді талдауда байқалады. Егер эмоцияларды талдайтын жүйе біржақты деректер негізінде үйретілсе, ол объективтілікті сақтай алмай, белгілі бір пікірлерді әдейі күшейтіп немесе әлсіретіп көрсетуі мүмкін. Мысалы, егер модель негізінен жағымды пікірлерден үйренсе, онда ол теріс пікірлерді әлдеқайда жұмсартып көрсетуге бейім болуы мүмкін, ал шынайы бейтарап пікірлерді де жағымды немесе теріс деп қате жіктеуі ықтимал. Бұл жағдай, әсіресе, әлеуметтік медиа контентін модерациялау, тұтынушы пікірлерін талдау немесе нарықтық зерттеулер жүргізу кезінде айтарлықтай қиындықтар туғызады.

Қазіргі эмоционалды талдау модельдерінің тағы бір шектеуі – мультимодальды талдаудың болмауы. Көптеген жүйелер тек мәтіндік ақпаратпен шектеліп, суреттер, аудио немесе бейне сияқты қосымша дереккөздерді елемейді. Алайда, адамдар арасындағы қарым-қатынаста эмоциялар тек мәтін арқылы ғана емес, сонымен қатар интонация, мимика, қимылдар мен басқа да вербалды емес элементтер арқылы жеткізіледі. Әлеуметтік желілерде пайдаланушылар өз хабарламаларын эмоцияны күшейту үшін фотосуреттермен, гифтермен немесе бейнежазбалармен толықтырады. Осы факторларды ескермеу талдау нәтижелерінің дәлдігіне айтарлықтай әсер етеді. Мысалы, «Мен өте бақыттымын!» деген сөйлем тек мәтін бойынша қарастырылғанда қуанышты көңіл-күйді білдіруі мүмкін, бірақ оған қосылған сурет немесе бейне контекстке байланысты мағынаны мүлдем өзгертіп жіберуі мүмкін. Сондықтан, эмоцияларды дәлірек анықтау үшін мәтіндік, визуалды және аудио ақпараттарды біріктіретін мультимодальды модельдерді әзірлеу аса маңызды.

Осы мәселелерді шешу үшін эмоционалды талдау модельдерінің дамуын бірнеше бағытта жетілдіру қажет. Біріншіден, мәдени әртүрлілікті ескеретін кеңейтілген деректер жиынтықтарын құру маңызды. Әртүрлі тілдер мен диалектілерде жазылған мәтіндерді қамтитын мәліметтер базасын жасау модельдердің әмбебаптығын арттырады. Екіншіден, қысқа мәтіндерді талдау сапасын жақсарту үшін контекстуалды ақпаратты ескеретін жетілдірілген алгоритмдерді дамыту қажет. Бұл үшін трансформер негізіндегі нейрондық желілер мен өзара байланысты мәтіндердің үлкен массивтерін талдай алатын үлгілерді қолдануға болады. Үшіншіден, оқыту процесінде бейтараптықты қамтамасыз ететін әдістерді енгізу керек. Бұл үшін теңдестірілген деректерді пайдалану, сондай-ақ әділеттілік пен объективтілікті бақылайтын арнайы механизмдер әзірлеу қажет. Төртіншіден, эмоцияларды дәлірек анықтау үшін мультимодальды тәсілді қолдану аса маңызды. Мұнда мәтіндік, аудио және визуалды деректерді бірлесіп талдай алатын жүйелерді құру керек. Бұл әсіресе әлеуметтік желілердегі контентті талдау кезінде өзекті болмақ.

Қорытындылай келе, мәтіндердегі эмоцияларды талдау саласы қарқынды дамып келе жатқанына қарамастан, шешімін таппаған көптеген мәселелер әлі де бар. Деректер жиынтықтарының шектеулілігі, қысқа мәтіндерді талдаудағы қиындықтар, алгоритмдердің біржақтылығы және мультимодальды тәсілдің болмауы – бұл бағыттағы негізгі кедергілердің бірі болып қала береді. Осы мәселелерді шешу үшін әртүрлі тілдік және мәдени контексттерді қамтитын кеңейтілген деректер жиынтықтарын әзірлеу, контекстуалды ақпаратты ескеретін жетілдірілген алгоритмдерді қолдану, бейтараптықты қамтамасыз ететін әдістерді енгізу және мультимодальды талдау тәсілін дамыту қажет. Тек осы бағыттарда жетілдірулер енгізу арқылы ғана эмоцияларды талдаудың дәлдігі мен сенімділігін арттыруға болады, бұл өз кезегінде оны түрлі салаларда, оның ішінде әлеуметтік желілерді мониторингтеу, клиенттерге қызмет көрсету, психологиялық диагностика және автоматтандырылған модерация жүйелерінде кеңінен қолдануға мүмкіндік береді.

## Зерттеу нәтижелерінің қорытындылары мен болашақтағы зерттеулер

Мәтіндердегі эмоцияларды талдаудың дамуы жасанды интеллект пен табиғи тілді өңдеудегі маңызды зерттеу саласы болып қала береді. Қазіргі заманғы модельдер дәлдіктің жоғары деңгейін көрсетеді, бірақ одан әрі жетілдіруді қажет ететін шешілмеген мәселелер қалады. Ең перспективалы бағыттардың бірі-мәдени ерекшеліктерді ескере отырып, әртүрлі тілдердегі мәтіндерді тиімді талдай алатын көп тілді модельдерді дамыту. Қазіргі уақытта XLM-R және mBERT сияқты модельдер көптеген тілдердегі мәтіндерді өңдеуде айтарлықтай жетістіктерге қол жеткізуде, бірақ олар сирек тілдермен, диалектілермен және аралас тілдік құрылымдармен жұмыс жасауда қиындықтарға тап болады. Әрі қарайғы зерттеулердің маңызды бағыты-тілдік вариациялар мен мәдени контексттердің кең ауқымын қамтитын теңдестірілген деректер корпусын құру арқылы тілдік қолдауды кеңейту. Бұл әсіресе жаһандану жағдайында, әлеуметтік желілерді, форумдарды және мессенджерлерді қолданушылар аралас тілдерді, қарызға алынған сөздерді және ерекше идиомалық өрнектерді жиі қолданады. Талдаудың дәлдігін қамтамасыз ету үшін модельдерді сөйлеудің жергілікті ерекшеліктеріне бейімдеу тетіктерін әзірлеу, сондай-ақ мәдени контексттерді ескеруге мүмкіндік беретін мәтіндерді аудару және түрлендіру әдістерін енгізу қажет.

Тағы бір перспективалық бағыт-хабарламалардың эмоционалды бояуын түсінудің дәлдігі мен тереңдігін едәуір арттыратын мультимодальды талдауды біріктіру. Қазіргі уақытта модельдердің көпшілігі суреттер, аудио және бейне сияқты басқа маңызды әдістерді елемей, тек мәтіндік деректермен жұмыс істейді. Алайда, табиғи қарым-қатынастағы эмоциялар тек мәтін арқылы ғана емес, сонымен қатар мимика, дауыс интонациясы, жест-ишара және визуалды контекст арқылы беріледі. Мысалы, сарказм немесе ирония көбінесе белгілі бір мимикалық өрнектермен немесе интонациялық ерекшеліктермен бірге жүреді, олар таза мәтіндік модельдер үшін қол жетімсіз болып қалады. Мультимодальды тәсілдерді біріктіру модельдерге мәтіннің мазмұнын ғана емес, сонымен бірге ілеспе ақпаратты да талдауға мүмкіндік береді, бұл болжамдарды дәлірек және адамның табиғи қабылдауына жақындатады. Ол үшін әр түрлі типтегі деректерді бір уақытта өңдей алатын нейрондық желілік архитектураларды әзірлеу, сондай-ақ пайдаланушылардың нақты өзара әрекеттесуінің мысалдарын қамтитын үлкен мультимодальды деректер корпустарын құру қажет. Мұндай технологиялардың дамуы әлеуметтік медианы талдауда, мазмұнды автоматты модерациялауда, цифрлық көмекшілерде және қоғамдық пікірді бақылау жүйелерінде қолдануға болатын күрделі және интеллектуалды жүйелерді құруға мүмкіндік береді.

Қазіргі заманғы модельдердің негізгі мәселелерінің бірі олардың төмен түсіндірілуі болып қала береді, бұл қабылданған шешімдерді түсіндіруде қиындықтар туғызады және олардың нәтижелеріне деген сенімділікті төмендетеді. Қазіргі уақытта трансформаторларға негізделген модельдердің көпшілігі «қара жәшіктер» ретінде жұмыс істейді, бұл олардың жұмысын талдауды және ықтимал қателерді анықтауды қиындатады. Бұл әсіресе медицина, заң жүйелері және автоматтандырылған шешім қабылдау жүйелері сияқты алгоритмдердің ашықтығы маңызды салаларда өте маңызды. Бұл мәселені шешудің бір бағыты-модельдің шешім қабылдау процесін визуализациялауға және болжамды қалыптастыру кезінде мәтіннің қай бөліктеріне назар аударатынын талдауға мүмкіндік беретін түсіндірмелі жасанды интеллект (XAI) әдістерін енгізу. Attention механизмдерін, маңыздылықты кері тарату әдістерін және интерпретацияланған деректерді ұсынуды пайдалану алгоритмдердің ашықтығын арттыруға және оларды маңызды жүйелерге біріктіруді жеңілдетуге көмектеседі. Болашақта XAI әдістерін дамыту модельдердің жұмысын түсінуді жақсартып қана қоймайды, сонымен қатар әлсіз жерлерді анықтау және күрделі мысалдармен қосымша оқыту арқылы олардың бейімделуін арттырады.

Тағы бір маңызды перспектива-модельдердің өнімділігін оңтайландыру, өйткені қазіргі заманғы трансформаторлық архитектуралар айтарлықтай есептеу ресурстарын қажет етеді, бұл оларды нақты мәселелерде қолдануды шектейді. BERT және GPT сияқты қуатты нейрондық желілерді іске қосу арнайы жабдықты қажет етеді, бұл оларды шағын компаниялар, стартаптар және Мобильді қосымшаларды жасаушылар үшін қол жетімді етпейді. Бұл мәселені шешу үшін модельдерді кванттау, білімді дистилляциялау және үлестірілген есептеу сияқты оңтайландыру әдістері белсенді түрде әзірленуде. Кванттау модельдің өлшемін кішірейтуге және оның орындалуын жылдамдатуға мүмкіндік береді, бұл таразылар мен белсендірулердің биттілігін азайту арқылы оны мобильді құрылғыларда жұмыс істеуге ыңғайлы етеді [15]. Білімді дистилляциялау неғұрлым күрделі және қуатты алгоритмдерге негізделгенныхам модельдерді оқытудан тұрады, бұл болжау сапасының айтарлықтай нашарлауынсыз есептеу жүктемесін айтарлықтай төмендетуге мүмкіндік береді. Таратылған есептеулерді пайдалану бұлтты ортада модельдерді іске қосуға мүмкіндік береді, жүктемені бірнеше серверлер арасында бөледі және оларды жоғары жүктеме жағдайында тиімді пайдалануды қамтамасыз етеді. Бұл технологиялар мобильді қосымшаларға, чатботтарға және нақты уақыттағы деректер жүйелеріне эмоционалды талдау модельдерін енгізудің жаңа мүмкіндіктерін ашады.

Болашақта осы бағыттардың дамуы мәтіндердегі эмоцияларды тиімді талдай алатын, мультимодальды контекстті ескеретін және шектеулі есептеу ресурстарында жұмыс істей алатын дәлірек, жылдам және түсіндірілетін модельдерді құруға әкеледі. Бұл эмоционалды талдау технологияларын тұтынушыларды қолдаудың автоматтандырылған жүйелерінен бастап интеллектуалды ұсыныс жүйелеріне дейін қолданушылардың жекелендірілген және бейімделгіш тәжірибесін қамтамасыз ететін көптеген қосымшаларға біріктіруге мүмкіндік береді.

## Мәтіндерді эмоционалдық талдау сапасын арттыру әдістері

Мәтіндерді эмоционалды талдау модельдерін жетілдіру күрделі және көп қырлы міндет болып табылады, ол кешенді тәсілді қажет етеді. Ең маңызды бағыттардың бірі-ирония, сарказм және мәдени-спецификалық өрнектер сияқты күрделі тілдік құрылымдардың мысалдарын қамтитын мамандандырылған мәліметтер жиынтығын құру. Модельдерді оқыту үшін қолданылатын стандартты деректер жиындары көбінесе нақты тілдің барлық нюанстарын көрсетпейді, бұл күрделі жағдайларда болжау дәлдігінің төмендеуіне әкеледі. Талдаудың сапасын арттыру үшін нақты эмоционалды бояуы бар мәтіндерден тұратын мамандандырылған деректер корпусын әзірлеу және жинау, сондай-ақ тілдің аймақтық ерекшеліктерін ескеру қажет. Мұндай мәліметтер жиынтығына контекстке байланысты әр түрлі түсіндірілуі мүмкін көп мәнді өрнектердің мысалдарын қосу маңызды. Мысалы, әртүрлі интонациямен немесе әртүрлі әлеуметтік топта айтылған бір сөз тіркесінің қарама-қарсы мағыналары болуы мүмкін. Сондықтан модельдерді әзірлеу кезінде мәтіннің лексикалық мазмұнын ғана емес, оның прагматикалық контекстін де ескеру қажет.

Оқыту деректерінің сапасын жақсартудың тиімді әдістерінің бірі-мәтіннің әртүрлі модификациялары арқылы оқу үлгісінің көлемін жасанды түрде арттыруға мүмкіндік беретін күшейту. Бұл әдістерге сөйлемдерді парафразалау, сөздерді синонимдермен ауыстыру, орфографиялық қателерді енгізу және сөздердің ретін өзгерту жатады. Күшейту модельдерге тілдің өзгергіштігіне жақсы бейімделуге көмектеседі және шектеулі деректерде қайта оқыту мүмкіндігін азайтады. Сонымен қатар, мәтінді басқа тілге аудару және кері аудару (back translation) әдістерін қолдануға болады, бұл деректерді одан әрі байытуға және олардың әртүрлілігін арттыруға мүмкіндік береді [16]. Бұл тәсіл әсіресе әлеуметтік медианы өңдеуде пайдалы, мұнда пайдаланушылар жиі орфографияны, қысқартуларды және жаргондарды пайдаланады.

Жетілдірудің тағы бір маңызды бағыты-заманауи нейрондық желілік архитектураларды дәстүрлі Машиналық оқыту әдістерімен біріктіретін гибридті тәсілдерді дамыту. BERT және RoBERTa сияқты терең оқыту модельдері жоғары нәтижелерге қол жеткізгенімен, олар жоғары интерпретацияны немесе аз көлемдегі деректермен жұмыс істеуді қажет ететін есептердегі қарапайым алгоритмдерден әрқашан асып түсе бермейді. Нейрондық желілерді логистикалық регрессия, кездейсоқ ормандар және градиентті күшейту сияқты классикалық Машиналық оқыту әдістерімен біріктіру болжау дәлдігінің жоғарылауына әкелуі мүмкін. Мысалы, мәтіннен белгілерді алу үшін трансформаторларды қолдануға болады, содан кейін нәтижелерді түсіндіру үшін дәстүрлі жіктеу әдістерін қолдануға болады. Бұл тәсіл деректердің ерекшеліктерін жақсырақ қарастыруға мүмкіндік береді және үлкен нейрондық желілік архитектураларға тәуелділікті азайтады.

Талдаудың дәлдігін жақсарту үшін модельдік ансамбльдерді қолдану тиімді, мұнда бірнеше Алгоритмдер бірлесіп жұмыс істейді және олардың болжамдарын біріктіреді. Ансамбльдер әртүрлі принциптерге негізделуі мүмкін: кейбір модельдер жалпы ақпаратты өңдей алады, басқалары мәтіннің белгілі бір аспектілеріне назар аударады, мысалы, жеке сөздердің эмоционалды бояуы немесе сөйлем құрылымы. Біріктірілген тәсіл көбірек параметрлерді ескеруге мүмкіндік береді және қателіктердің ықтималдығын азайтады, әсіресе бір модель дұрыс емес болжам жасай алатын қиын жағдайларда. Сонымен қатар, модельдер ансамбльдері тұрақты және сенімді нәтижелерді қамтамасыз ете отырып, жеке алгоритмдердің кемшіліктерін тегістеуге мүмкіндік береді [17].

Есептеу ресурстарын оңтайландыру эмоционалды талдау модельдерін практикалық қолдануда, әсіресе оларды өнімді жүйелерге орналастыруда шешуші рөл атқарады. Қазіргі заманғы трансформаторлық модельдер айтарлықтай есептеу қуатын қажет етеді, бұл оларды шектеулі ресурстар жағдайында пайдалануды қиындатады. Бұл мәселенің шешімдерінің бірі-кванттауды қолдану, бұл модельдің көлемін кішірейтуге және аз биттік сандарды қолдану арқылы оның жұмысын тездетуге мүмкіндік береді. Квантталған модельдер жедел жадты аз пайдаланады және дәлдіктің қолайлы деңгейін сақтай отырып, тезірек жұмыс істей алады. Қосымша жеделдетуді білімді дистилляциялау арқылы алуға болады, мұнданаяам модель неғұрлым күрделі және қуатты архитектурадан үйренеді [27]. Бұл болжам ретінде айтарлықтай шығынсыз есептеу жүктемесін азайтуға мүмкіндік береді.

Бұлтты есептеулерді пайдалану бірнеше серверлер арасында есептеу жүктемесін бөлуге мүмкіндік беретін өнімділікті оңтайландырудың тағы бір әдісі болып табылады. Бұл, әсіресе, деректердің үлкен көлемін жылдам өңдеуді қажет ететін чатботтар немесе қоғамдық пікірді бақылау платформалары сияқты нақты уақыттағы жүйелерге қатысты. Модельдерді бұлтқа орналастыру оларды жаңартуды және оқуды жеңілдетеді, бұл үнемі өзгеріп отыратын тілдік конструкциялармен және пайдаланушылардың сөйлеуіндегі жаңа тенденциялармен жұмыс істеу үшін маңызды.

Модельдердің интерпретациясын арттыру маңызды мәселе болып қала береді, өйткені көптеген заманауи Алгоритмдер, әсіресе терең нейрондық желілер «қара жәшіктер» ретінде жұмыс істейді. Бұл дегеніміз, пайдаланушы нәтиже алады, бірақ модельдің қандай да бір қорытындыға қалай келгенін нақты түсіне алмайды. Бұл мәселені шешу үшін LIME және SHAP сияқты түсіндірілетін жасанды интеллект (Explainable AI, XAI) әдістерін қолдануға болады. Бұл әдістер модельді болжауға қай сөздер немесе сөз тіркестері көбірек әсер еткенін визуализациялауға мүмкіндік береді, бұл талдау процесін айқынырақ етеді. Сонымен қатар, пайдаланушыларға модельдің нәтижелерін түзетуге және оны жаңа мәліметтерге үйретуге мүмкіндік беретін кері байланыс механизмдерін енгізуге болады [19].

Тағы бір маңызды аспект-эмоционалды талдау модельдерін қолданудың этикалық жағы. Қазіргі алгоритмдердің негізгі мәселелерінің бірі олардың теңгерімсіз оқу деректерінен туындауы мүмкін біржақтылығы болып қала береді. Мысалы, егер модель негізінен белгілі бір әлеуметтік топтың мәтіндерінде оқытылған болса, ол басқа топтардың өкілдеріне бейімділік танытуы мүмкін, бұл нәтижелердің бұрмалануына әкеледі. Мұндай мәселелерді азайту үшін модельдерді пайдаланушылардың әртүрлі санаттарындағы мәтіндерден тұратын теңдестірілген деректер жиынтығында оқыту керек, сонымен қатар алгоритмдерді біржақтылыққа үнемі тексеріп отыру керек. Сонымен қатар, модельдерге кері байланысты ескеруге және жүйелік қателерді анықтау кезінде болжамдарын өзгертуге мүмкіндік беретін Өзін-өзі түзету механизмдерін жасауға болады.

Жоғарыда аталған барлық ұсыныстарды енгізу мәтіндерді эмоционалды талдау модельдерінің дәлдігін, тиімділігі мен ашықтығын едәуір арттырады. Мамандандырылған деректер жиынтығын құру, гибридті тәсілдерді біріктіру, есептеу ресурстарын оңтайландыру, интерпретацияны арттыру және біржақтылықты жою мұндай модельдерді нақты өмірде сенімдірек және қолдануға мүмкіндік береді. Ұзақ мерзімді перспективада бұл эмоцияларды дәлірек тануға, контекстті ескеруге және маркетинг, пікір аналитикасы, психолингвистикалық зерттеулер және тұтынушыларды қолдаудың автоматтандырылған жүйелерін қоса алғанда, қолданбалардың кең ауқымында талдаудың жоғары дәлдігін қамтамасыз етуге қабілетті интеллектуалды жүйелердің дамуына әкеледі.

## Бағдарламалық жасақтаманың негізгі функциялары

Мәтіндердің эмоционалды талдауын дамыту күрделі міндет болып табылады, ол кешенді тәсілді және әртүрлі заманауи технологияларды біріктіруді талап етеді. Негізгі бағыттардың бірі-трансформаторлық архитектуралардың, мультимодальды талдаудың және түсіндірілетін жасанды интеллект әдістерінің артықшылықтарын біріктіруге қабілетті әмбебап модельді әзірлеу. BERT, RoBERTa және олардың туындылары сияқты қазіргі модельдер мәтіндерді талдауда жоғары дәлдікті көрсетеді, бірақ олар бірқатар шектеулерге тап болады, соның ішінде ресурстардың жоғары сыйымдылығы, ирония мен сарказм сияқты күрделі лингвистикалық құрылымдарды өңдеудің күрделілігі және болжамдардың әлсіз интерпретациясы. Осы қиындықтарды тиімді жеңе алатын әмбебап модель құру әрі қарайғы зерттеулердің перспективалы бағыттарының бірі болып табылады.

Бұл тәсіл тек мәтінді ғана емес, сонымен қатар аудио және кескін сияқты деректердің басқа түрлерін талдауға мүмкіндік беретін мультимодальды талдауды енгізуді қажет етеді. Мысалы, мәтінді ғана емес, сонымен қатар эмодзи немесе суреттер сияқты онымен бірге жүретін визуалды элементтерді, сондай-ақ аудио хабарламалар жағдайында дауыстың акустикалық сипаттамаларын ескере отырып, хабарламаның кілтін талдау айтарлықтай дәлірек болуы мүмкін. Деректердің бірнеше түрін бірден ескеретін біріктірілген модельдер контекстті тереңірек түсіну арқылы талдау сапасын айтарлықтай арттыруға мүмкіндік береді. Бұл бағытта белсенді зерттеулер жүргізілуде және болашақ әзірлемелер мультимодальды деректерді жалпы талдау процесіне біріктіру мүмкіндігін ескеруі керек.

Даму стратегиясының тағы бір маңызды аспектісі-модельдерді практикалық қолдануға бағдарлау. Эмоционалды талдау саласындағы зерттеулердің жоғары теориялық маңыздылығына қарамастан, бұл әзірлемелерді нақты қолданбаларға енгізу негізгі міндет болып қала береді. Қазіргі заманғы бизнес-процестер мәтіндік деректерді тиімді талдай алатын автоматтандырылған құралдарды қажет етеді, мейлі ол пайдаланушылардың пікірлері, әлеуметтік желілердегі түсініктемелер немесе клиенттердің сұраныстары болсын. Ең перспективалы бағыттардың бірі-CRM жүйелеріне эмоционалды талдауды біріктіру, бұл компанияларға клиенттермен тиімді қарым-қатынас жасауға, олардың көңіл-күйін талдауға және жекелендірілген шешімдерді ұсынуға мүмкіндік береді [11].

Сонымен қатар, мазмұнды автоматты түрде модерациялау құралдарын жасауға ерекше назар аудару керек. Интернеттегі пайдаланушы жасаған мазмұн көлемінің қарқынды өсуі жағдайында компаниялар улы лексика, қорлау немесе экстремистік сөздер бар хабарламаларды бақылау және сүзу мәселесіне тап болады. Қазіргі заманғы модельдер мұндай мәтіндерді жоғары дәлдікпен анықтай алады, бірақ оларды қолдану нақты платформалардың ерекшеліктеріне қосымша бейімделуді қажет етеді. Болашақ зерттеулер нақты уақыттағы пікірлерді автоматты түрде модерациялай алатын, жалған позитивтерді азайтатын және жоғары дәлдікті қамтамасыз ететін шешімдер жасауға бағытталуы керек.

Осы саланы табысты дамыту үшін ғылыми қоғамдастықпен белсенді ынтымақтастық, мәтіндерді талдауға арналған халықаралық жобалар мен конкурстарға қатысу қажет. Зерттеу топтарымен үнемі өзара әрекеттесу және білім алмасу бар әдістерді жақсартуға, жаңа тәсілдерді сынауға және нақты тапсырмалар үшін оңтайлы шешімдерді табуға мүмкіндік береді. SemEval немесе Kaggle сияқты халықаралық конференциялар мен ғылыми жарыстар жаңа модельдерді сынау және әлемдік мамандар қауымдастығынан кері байланыс алу үшін тамаша алаң болып табылады. Әртүрлі елдердің зерттеушілерінің бірлескен жұмысы әртүрлі елдер мен аймақтарда қолданылатын әмбебап үлгілерді әзірлеу кезінде әсіресе маңызды болып табылатын әртүрлі тілдер мен мәдени контексттерді ескеруге мүмкіндік береді.

Мәтіндерді эмоционалды талдау саласында мамандар даярлауға бағытталған білім беру бағдарламаларын әзірлеу де маңызды бағыт болып табылады. Қазіргі уақытта табиғи тілді өңдеу (NLP) құзыреттілігі бар мамандарға сұраныс ұсыныстан едәуір асып түседі және бұл алшақтық өсуде. Бұл теңгерімсіздікті жою үшін мамандарға озық технологияларды игеруге және оларды практикада қолдануға мүмкіндік беретін мамандандырылған курстар мен оқыту бағдарламаларын әзірлеу қажет. Осы курстар шеңберінде трансформаторлық модельдердің жұмыс принциптері, мәтінді өңдеу әдістері, Машиналық оқыту алгоритмдері және нейрондық желілік архитектуралар, сондай-ақ олардың хабарламалардың эмоционалды бояуын талдауда практикалық қолданылуы сияқты тақырыптарды қарастыруға болады [22].

Сонымен қатар, хакатондар мен зертханаларды ұйымдастыруға болады, онда студенттер мен жаңадан бастаған зерттеушілер алған білімдерін іс жүзінде қолдана алады, өз модельдерін жасап, оларды нақты жағдайда сынай алады. Мұндай бастамалар пәндік саланы тереңірек түсінуге және студенттердің салада сұранысқа ие дағдыларын қалыптастыруға ықпал етеді.

Осылайша, мәтіндерді эмоционалды талдауды дамыту стратегиясы бірнеше негізгі бағыттарды қамтуы керек: эмоцияларды дәл және түсіндірілетін болжауға бағытталған әмбебап мультимодальды модельдерді әзірлеу, тұтынушылардың пікірлерін талдауды және мазмұнды модерациялауды автоматтандыру үшін бизнес-процестерге әзірленген шешімдерді енгізу, жаңа әдістерді сынау және жетілдіру үшін халықаралық ғылыми қауымдастықпен белсенді ынтымақтастық және осы салада жоғары білікті мамандарды даярлауға бағытталған білім беру бастамаларын қалыптастыру. Осы бағыттардың кешенді дамуы қолданыстағы алгоритмдерді едәуір жақсартып қана қоймай, сонымен қатар мәтіндердің эмоционалды талдауын Болашақтың цифрлық экожүйесінің ажырамас бөлігіне айналдыра отырып, оларды іс жүзінде енгізуді жеделдетуге мүмкіндік береді.

## Мәселелер мен шешімдердің таксономиясы

Кесте 1

Мәселелер мен ұсынылған шешімдерді жүйелеу

|  |  |
| --- | --- |
| Мәселе | Шешім |
| Қысқа мәтіндердегі төмен дәлдік | Мәтінмәндік модельдерді (BERT, RoBERTa) және деректерді күшейтуді қолдану. |
| Деректер жиынтығының шектеулілігі | Мәдени ерекшеліктерді ескере отырып, мамандандырылған деректер жиынтығын құру. |
| Модельдердің ресурс сыйымдылығы | Кванттауды, білімді дистилляциялауды және бұлтты есептеулерді қолдану. |
| Төмен интерпретация | Түсіндірілетін AI (LIME, SHOP) әдістерін енгізу және attention maps визуализациясы. |
| Модельдік бейімділік | Теңдестірілген деректер бойынша оқыту және біржақтылықты үнемі тестілеу. |
| Ескерту – кесте дереккөз негізінде жасалған [19] | |

Зерттеу мәтінді эмоционалды талдау қоғамдық пікірді талдаудың, беделді басқарудың және тұтынушыларға қызмет көрсетуді жақсартудың маңызды құралы екенін растады. Дегенмен, бұл саланы одан әрі дамыту үшін деректер жиынының шектеулілігі, модельдердің ресурс сыйымдылығы және төмен интерпретация сияқты бірқатар мәселелерді шешу қажет.

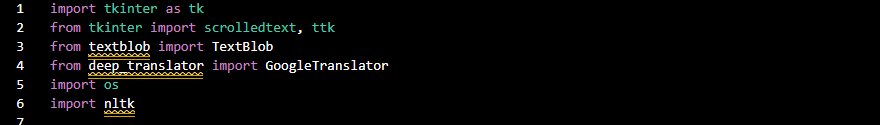
Ұсынылған нұсқаулар, соның ішінде арнайы деректер жиынтығын құру, өнімділікті оңтайландыру және интерпретацияны жақсарту мәтінді эмоционалды талдаудың дәлдігі мен тиімділігін жақсартуға көмектеседі. Бұл модельдерді маркетинг, беделді басқару және қоғамдық пікірді талдау сияқты нақты міндеттерге қолдануға жаңа мүмкіндіктер ашады.

## Бағдарлама кодының сипаттамасы

Бұл бөлімде проблемалардың таксономиясында сипатталған ұсынылған шешімдердің бағдарламалық жасақтамасы ұсынылған. Бағдарлама мәтіндік түсініктемелерді автоматты түрде талдауға және олардың эмоционалды бояуын анықтауға арналған. Бағдарламаның негізгі міндеті — түсініктемелерді үш санатқа бөлу: «жаман» (теріс), «жақсы» (оң) және «бейтарап».

Код-мәтіннің кілтін талдауға арналған графикалық интерфейс (GUI) бағдарламасы.

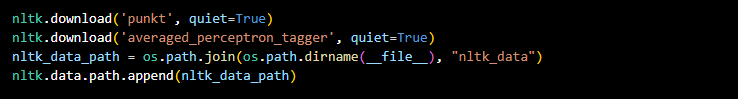
Python-дағы қажетті кітапханаларды жүктеу мысалы 10- суретте көрсетілген:



**Сурет 10** Python-дағы қажетті кітапханаларды жүктеу.

Python-дағы бағдарлама tkinter көмегімен графикалық интерфейсті жасайды, оған айналдыру мәтін өрісі және мәтінді аудару батырмасы кіреді. Енгізілген мәтін Textblob кітапханасы арқылы талданады, ал аударма Google Translator арқылы жүзеге асырылады.

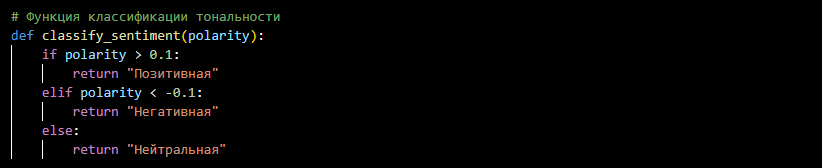
Интерфейсте пайдаланушы мәтін енгізетін жылжымалы мәтін өрісі бар. Түймені басқан кезде мәтін өңделеді және аударылады. Бағдарлама жұмыс істеуі үшін tkinter, text blob, deep\_translator модульдері, сондай-ақ тілді талдау үшін nltk қолданылады. Модульдерді жүктеу коды 11- суретте көрсетілген:



**Сурет 11** Бұл код nltk кітапханасының жұмыс істеуі үшін қажетті деректерді жүктейді және оларды сақтау жолын көрсетеді.

Алдымен мәтінді сөйлемдер мен сөздерге бөлу үшін қолданылатын punkt модулі жүктеледі, содан кейін averaged\_perceptron\_tagger сөйлеу бөліктерінің теггері әр сөздің қай бөлікке жататынын анықтауға мүмкіндік береді. Әрі қарай, nltk\_data қалтасына жол жасалады, ол сценарийдің өзі сияқты каталогта болады, содан кейін бұл жол nltk жүктелген модельдер мен деректерді іздейтін жолдар тізіміне қосылады. Бұл бағдарламаға стандартты nltk каталогтарына шектеулі қол жетімділікпен де қажетті ресурстарды пайдалануға мүмкіндік береді.

Classify\_sentiment функциясының жұмыс істеу принципі код жүзінде 12- суретте көрсетілген:

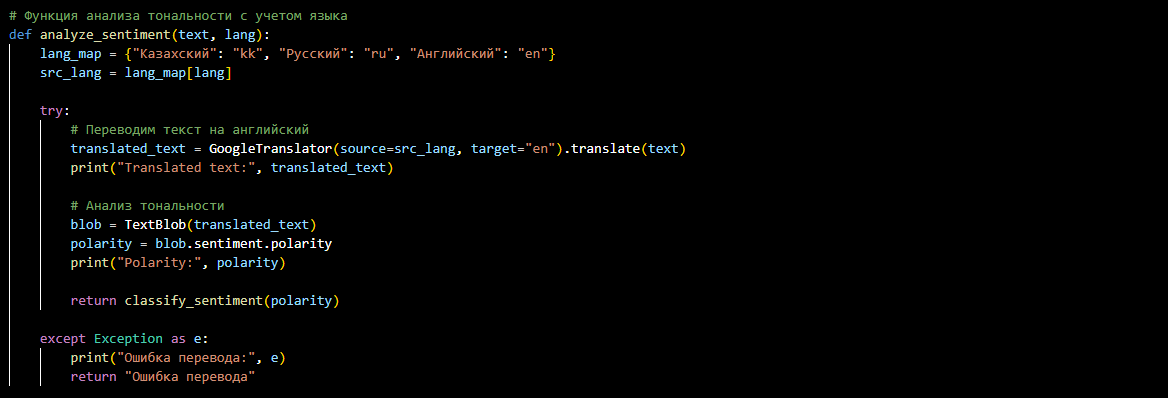


**Сурет 12** Бұл код мәтіннің тоналдылығын полярлық мәніне қарай жіктейтін classify\_sentiment(polarity) функциясын анықтайды.

Функция мәтіннің эмоционалды бояуын көрсететін polarity санын қабылдайды: оң мән оң тонды, теріс теріс және нөлге жуық мән бейтарапты көрсетеді.

Егер полярлық 0.1-ден үлкен болса, «оң» жолы қайтарылады, егер -0.1-ден аз болса, онда «теріс», ал егер -0.1-ден 0.1-ге дейін болса, нәтиже «бейтарап» болады. Бұл мәтінді талдауға және оның эмоционалды бояуын анықтауға мүмкіндік береді.

Analyze\_sentiment функциясының құрылымы 13- суретте көрсетілген:



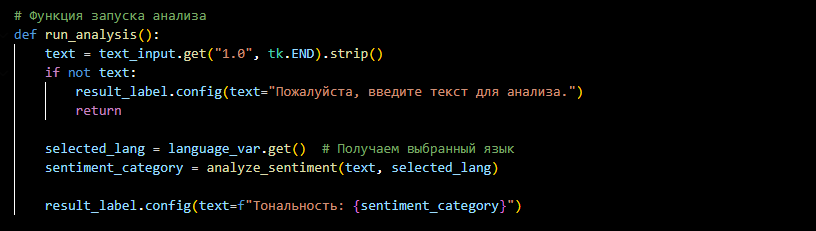
**Сурет 13** Бұл код мәтіннің тоналдылығын (эмоционалды бояуын) оның полярлығына қарай талдайтын analyze\_sentiment (text, lang) функциясын анықтайды.

Алдымен Тілдерді «Қазақ», «Орыс», «Ағылшын» олардың кодтарымен (kk, ru, en) байланыстыратын lang\_map сөздігі жасалады. Содан кейін берілген lang параметріне негізделген мәтіннің бастапқы тілі (src\_lang) анықталады.

Мүмкіндік дұрыс талдауды қамтамасыз ету үшін GoogleTranslator көмегімен мәтінді ағылшын тіліне аударуға тырысады, өйткені textblob кітапханасы ағылшын тілімен жақсы жұмыс істейді. Аудармадан кейін textblob нысаны жасалады, ол мәтінді талдайды және оның полярлығын есептейді — -1 (теріс кілт) мен 1 (оң) арасындағы сан.

Алынған мән classify\_sentiment() функциясына беріледі, ол кілттер санатын қайтарады: «оң», «теріс» немесе «бейтарап».

Егер мәтінді аудару кезінде қате пайда болса (мысалы, API қосылымына қатысты мәселе), ерекшелік өңделеді, «аударма қатесі» хабары көрсетіледі және функция сәйкес нәтижені қайтарады. 14- суретте көрсетілгендей Run\_analysis () функциясы енгізілген мәтіннің кілтін талдауды бастайды:



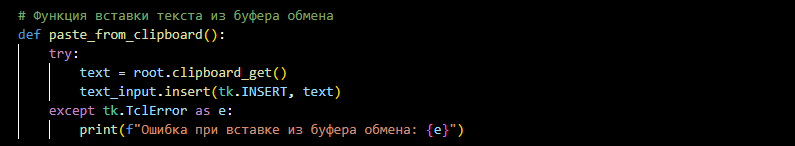
**Сурет 14** Run\_analysis () функциясы енгізілген мәтіннің кілтін талдауды бастайды.

Алдымен ол мәтінді `мәтін\_инпут` мәтін жолағынан алады, оның барлық мазмұнын оқиды (`get("1.0", tk.IND)`) және қосымша бос орындарды жою (`strip()`). Егер мәтін бос болса, «талдау үшін мәтінді енгізіңіз» деген хабарлама көрсетіледі және функция аяқталады.

Содан кейін пайдаланушы таңдаған тіл `language\_var` айнымалысынан анықталады. Алынған мәліметтер `analyze\_sentiment()` функциясына жіберіледі, ол мәтінді талдайды және кілттер санатын қайтарады: «оң», «теріс» немесе «бейтарап».

Нәтиже `result\_label`-де пайда болады, оның мәтінін табылған кілтпен жаңартады.

15- суретте Paste\_from\_clipboard () функциясы алмасу буферінен мәтінді мәтін жолағына қобды орындайды:

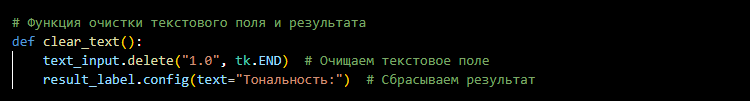


**Сурет 15** Paste\_from\_clipboard () функциясы алмасу буферінен мәтінді мәтін жолағына енгізеді.

Алдымен ол алмасу буферінің мазмұнын түбірімен алуға тырысады.clipboard\_get(). Содан кейін алынған мәтінді ағымдағы курсордың (tk) орнына text\_input ішіне енгізеді.INSERT).

Егер алмасу буферінен деректерді алу кезінде қате пайда болса (мысалы, буфер бос болса немесе Қолдау көрсетілмейтін деректер болса), TK ерекшелік.TclError ұсталып, консольде қате туралы хабарлама пайда болады.

Clear\_text() функциясы `text\_input.delete("1.0", tk.IND)` қолданады, бұл бірінші жолдан, бірінші таңбадан соңына дейін барлық мәтінді жою үшін. Содан кейін мәтінді `result\_label`-де жаңартады, оны алдыңғы талдау нәтижесін тазарту үшін «кілт:» күйіне орнатады. Оның бағдарламалық коды 16- суретте көрсетілген:



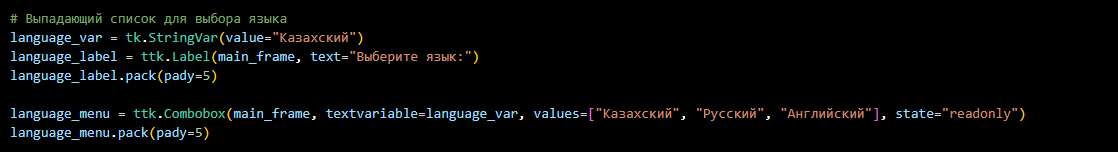
**Сурет 16** Clear\_text () функциясы мәтін өрісін тазартады және талдау нәтижесін қалпына келтіреді.

Tkinter ктапханасы бағдарлама интерфейсін жасау барысында алдымен `tk` типті `language\_var` айнымалысы жасалады, ол ағымдағы таңдалған тілді сақтайтын `StringVar`. Әдепкі бойынша, оның мәні «қазақша» деп орнатылады.

Содан кейін `ttk.Label` белгісі жасалады, ол «Тілді таңдау» мәтінін көрсетеді және ол `main\_frame`-де кішкене шегініспен (pady=5) орналастырылады.

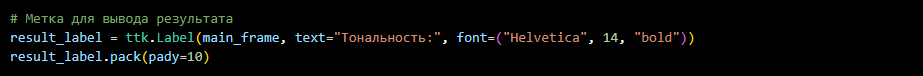
Осыдан кейін ашылмалы тізім (ttk.Combobox) жасалады, ол `language\_var`-мен байланысты. Бұл тізімде үш тілдің нұсқасы бар: «Қазақ», «Орыс», «Ағылшын». Тізім "read only" режиміне орнатылған, бұл пайдаланушыға еркін мәндерді енгізуге жол бермейді. Ашылмалы тізім `main\_frame`-де шегініспен орналастырылады.

Код Tkinter кітапханасын қолдану арқылы жазылған. Tkinter кітапханасы python тілінде жазылған бағдарламаларға олардың интерфейстерін құруға, олардың элеменнтерін басқаруға мүмкіндік береді. Бұл интерфейс элементі пайдаланушыға мәтін енгізілген тілді таңдауға мүмкіндік береді, ал кейіннен ол кілтті талдау үшін қолданылады және оның бағдарламалық коды 17- суретте көрсетілген:



**Сурет 17** Бұл код Tkinter кітапханасын пайдаланып тілді таңдау үшін интерфейс жасайды.

Бұл белгі іске қосу түймесін басқаннан кейін мәтінді талдау нәтижесін шығару үшін қолданылады және оның орындау принципі 18- суретте көрсетілген:



**Сурет 18** Нәтиже шығару жүйесі

Бұл код main\_frame ішінде мәтіндік белгіні (Label) жасайды, ол бастапқыда "кілтті:"көрсетеді. Қаріп Helvetica 12-ге орнатылған, ал тік шегініс (pady=10) интерфейсті оқуға ыңғайлы етеді.

# Қорытынды

Мәтінді эмоционалды талдау табиғи тілді өңдеудегі (NLP) маңызды міндет болып табылады және маркетингте, беделді басқаруда және қоғамдық пікірді талдауда кеңінен қолданылады. Осы жұмыс аясында Bert және RoBERTa сияқты эмоционалды талдаудың заманауи әдістері мен модельдері зерттелді, олар мәтіндерді жіктеу міндеттерінде тиімді екендігі дәлелденді. Алайда, олардың Твиттер сияқты қысқа мәтіндердегі өнімділігі контексттің болмауына және жаргонның болуына байланысты шектеулі болып қалады.

Тәжірибелер көрсеткендей, BERT IMDb деректер жиынтығында 92% және Twitter Sentiment Analysis — те 89% дәлдікке жетеді[16], ал Роберта одан да жоғары нәтижелерге қол жеткізеді-сәйкесінше 93% және 90%. XLM-R көп тілді моделі біршама төмен дәлдікті көрсетсе де (88%)[20], бірнеше тілдегі мәтіндерді талдауда ең тиімді болып қала береді.[18]

Зерттеу барысында анықталған негізгі мәселелердің бірі-иронияны, сарказмды және көп мағыналы өрнектерді талдаудың күрделілігі [28]. Бұл құбылыстар көбінесе жіктеу қателіктеріне әкеледі, өйткені мәтін жағымды болып көрінуі мүмкін, бірақ іс жүзінде жағымсыз эмоцияларды білдіреді. Сондай-ақ, заманауи модельдердің жоғары ресурс сыйымдылығы атап өтілді, бұл олардың деректерді жылдам өңдеуді қажет ететін нақты міндеттерде қолданылуын шектейді.

Осы мәселелерді шешу үшін арнайы деректер жиынтығын құру, модельдердің өнімділігін оңтайландыру және олардың түсіндірілуін жақсарту бойынша ұсыныстар ұсынылды. Атап айтқанда, модельдердің жұмысын жеделдету үшін білімді кванттау және айдау әдістерін қолдану, сондай-ақ олардың шешімдерінің ашықтығын арттыру үшін түсіндірілетін AI (XAI) әдістерін енгізу ұсынылады.

Мәтінді эмоционалды талдауды одан әрі дамытудың болашағы мәдени және тілдік ерекшеліктерді ескеретін көп тілді модельдерді құруды, сонымен қатар мәтіндік деректерді ғана емес, сонымен қатар суреттерді, аудио және бейнелерді де ескеретін мультимодальды талдауды біріктіруді қамтиды. Бұл модельдерді тұтынушылардың пікірлерін талдау, мазмұнды модерациялау және беделді басқару сияқты нақты міндеттерге қолданудың жаңа мүмкіндіктерін ашады.

Қорытындылай келе, мәтінді эмоционалды Талдау қоғамдық пікірді талдаудың және тұтынушыларға қызмет көрсету сапасын жақсартудың маңызды құралы деп айтуға болады. Дегенмен, бұл саланы одан әрі дамыту үшін деректер жиынының шектеулілігі, модельдердің ресурс сыйымдылығы, төмен интерпретация мәселелерін шешу қажет. Ұсынылған ұсыныстар мен стратегиялар мәтіндерді эмоционалды талдаудың дәлдігі мен тиімділігін жақсартуға көмектеседі, бұл оны әр түрлі қызмет салаларында қолдануға мүмкіндік береді.

# **Пайдаланылған әдебиеттер тізімі**

1. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool. [Электрондық ресурс]. – URL: <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
2. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. [Электрондық ресурс].–URL: [https://www.sciencedirect.com/science/article/pii/ S2090447914000550?via%3Dihub](https://www.sciencedirect.com/science/article/pii/%20S2090447914000550?via%3Dihub)
3. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,8(4). [Электрондық ресурс] - URL: [https://wires.onlinelibrary.wiley.com/doi/10.1002 /widm 1253](https://wires.onlinelibrary.wiley.com/doi/10.1002%20/widm%201253)
4. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, 28(2), 15–21. [Электрондық ресурс] - URL:<https://ieeexplore.ieee.org/document/6468032>
5. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O’Reilly Media. [Электрондық ресурс] - URL: <https://www.nltk.org/book/>
6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1810.04805>
7. Socher, R., Perelygin, A., et al. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. [Электрондық ресурс] - URL: <https://aclanthology.org/D13-1170.pdf>
8. Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1907.11692>]
9. Araque, O., Gatti, L., & Iglesias, C. A. (2019). A modular approach to multilingual and multi-domain sentiment analysis. Expert Systems with Applications, 115, 136–155. [Электрондық ресурс] - URL: https:// www.sciencedirect.com /science/article /abs/ pii/S0957417418304950?via%3Dihub
10. Feldman, R. (2013). Techniques and Applications for Sentiment Analysis. Communications of the ACM, 56(4), 82–89. [Электрондық ресурс] - URL: <https://dl.acm.org/doi/10.1145/2436256.2436274>
11. Mohammad, S. M. (2016). Challenges in Sentiment Analysis. [Электрондық ресурс]-URL:[https://www.researchgate.net/publication/316049100\_Challenges\_in\_ Sentiment\_ Analysis](https://www.researchgate.net/publication/316049100_Challenges_in_%20Sentiment_%20Analysis)
12. Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment Analysis is a Big Suitcase. IEEE Intelligent Systems, 32(6), 74–80. [Электрондық ресурс] - URL: [<https://ieeexplore.ieee.org/document/8267597>]
13. Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding Bag-of-Words Model: A. Statistical. Framework. [Электрондық ресурс] - URL: [https://www.researchgate.net/ publication/226525014\_Understanding\_bag-of-words\_model\_A\_statistical\_framework](https://www.researchgate.net/%20publication/226525014_Understanding_bag-of-words_model_A_statistical_framework)
14. Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2020). Multi-level multi-task learning for emotion detection in conversations. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1905.05739>
15. Schuller, B., & Batliner, A. (2013). Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. Wiley. [Электрондық ресурс] - URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118706664>
16. OpenAI. (2023). OpenAI API Documentation – Sentiment Classification Examples. [Электрондық ресурс] - URL: <https://platform.openai.com/examples>
17. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135. [Электрондық ресурс] - URL: <https://www.nowpublishers.com/article/Details/INR-011>
18. Yao, L., Mao, C., & Luo, Y. (2019). Graph Neural Networks for Text Classification. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1905.13129>
19. Howard, J., & Gugger, S. (2020). Deep Learning for Coders with Fastai and PyTorch. O’Reilly Media. [Электрондық ресурс] - URL: [https://github.com/fastai/ fastbook](https://github.com/fastai/%20fastbook)
20. Sun, C., Qiu, X., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? [Электрондық ресурс] - URL: <https://arxiv.org/abs/1905.05583>
21. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1906.08237>
22. Raffel, C., Shazeer, N., Roberts, A., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1910.10683>
23. Zhou, P., Shi, W., Tian, J., et al. (2016). Attention-Based Bidirectional LSTM for Text Classification. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1607.06450>
24. Peters, M. E., Neumann, M., Iyyer, M., et al. (2018). Deep contextualized word representations. [Электрондық ресурс] - URL: <https://aclanthology.org/N18-1202/>
25. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. [Электрондық ресурс] - URL: <https://arxiv.org/abs/1310.4546>
26. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. [Электрондық ресурс] - URL: <https://nlp.stanford.edu/pubs/glove.pdf>
27. Al-Smadi, M., Talafha, B., & Al-Ayyoub, M. (2019). Using deep learning for aspect-based sentiment analysis: A survey. Journal of Information Science. [Электрондық ресурс] - URL: [https://www.researchgate.net/publication/ 311113413\_An\_Enhanced\_Framework\_for\_Aspect-Based\_Sentiment\_Analysis\_of\_ Hotels'\_Reviews\_Arabic\_Reviews\_Case\_Study](https://www.researchgate.net/publication/%20%20311113413_An_Enhanced_Framework_for_Aspect-Based_Sentiment_Analysis_of_%20Hotels'_Reviews_Arabic_Reviews_Case_Study)
28. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), 267–307. [Электрондық ресурс] - URL: <https://direct.mit.edu/coli/article/37/2/267/2105/Lexicon-Based-Methods-for-Sentiment-Analysis>
29. Thelwall, M., Buckley, K., Paltoglou, G., et al. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544–2558. [Электрондық ресурс] - URL: <https://onlinelibrary.wiley.com/doi/10.1002/asi.21416>
30. Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM). [Электрондық ресурс] - URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>

# Қосымша

Мәтінді сентименталды талдауға бағытталған бағдарлама листингі

import tkinter as tk

from tkinter import scrolledtext, ttk

from textblob import TextBlob

from deep\_translator import GoogleTranslator

import os

import nltk

# NLTK деректерін жүктеу

nltk.download('punkt', quiet=True)

nltk.download('averaged\_perceptron\_tagger', quiet=True)

nltk\_data\_path = os.path.join(os.path.dirname(\_\_file\_\_), "nltk\_data")

nltk.data.path.append(nltk\_data\_path)

# Тональдылықты анықтайтын функция

def classify\_sentiment(polarity):

if polarity > 0.1:

return "Позитивті"

elif polarity < -0.1:

return "Негативті"

else:

return "Бейтарап"

# Тональдылықты тілге байланысты талдайтын функция

def analyze\_sentiment(text, lang):

lang\_map = {"Қазақша": "kk", "Орысша": "ru", "Ағылшынша": "en"}

src\_lang = lang\_map[lang]

try:

# Мәтінді ағылшын тіліне аудару

translated\_text = GoogleTranslator(source=src\_lang, target="en").translate(text)

print("Аударылған мәтін:", translated\_text)

# Тональдылықты талдау

blob = TextBlob(translated\_text)

polarity = blob.sentiment.polarity

print("Полярлық көрсеткіш:", polarity)

return classify\_sentiment(polarity)

except Exception as e:

print("Аудару қатесі:", e)

return "Аудару қатесі"

# Анализді бастайтын функция

def run\_analysis():

text = text\_input.get("1.0", tk.END).strip()

if not text:

result\_label.config(text="Мәтінді енгізіңіз.")

return

selected\_lang = language\_var.get() # Таңдалған тілді алу

sentiment\_category = analyze\_sentiment(text, selected\_lang)

result\_label.config(text=f"Тональдылық: {sentiment\_category}")

# Буферден мәтінді қою функциясы

def paste\_from\_clipboard():

try:

text = root.clipboard\_get()

text\_input.insert(tk.INSERT, text)

except tk.TclError as e:

print(f"Буферден қою қатесі: {e}")

# Мәтінді тазалау функциясы

def clear\_text():

text\_input.delete("1.0", tk.END) # Мәтін өрісін тазалау

result\_label.config(text="Тональдылық:") # Нәтижені тазалау

# Графикалық интерфейсті жасау

root = tk.Tk()

root.title("Тональдылықты талдау")

root.geometry("500x500") # Терезе өлшемін орнату

# Стильдерді баптау

style = ttk.Style()

style.theme\_use("clam") # Қазіргі заманғы стильді қолдану

# Қаріптер мен түстерді баптау

style.configure("TLabel", font=("Helvetica", 12), background="#f0f0f0")

style.configure("TButton", font=("Helvetica", 12), background="#4CAF50", foreground="white")

style.configure("TCombobox", font=("Helvetica", 12))

style.configure("TFrame", background="#f0f0f0")

# Негізгі жақтау

main\_frame = ttk.Frame(root, padding="10")

main\_frame.pack(fill=tk.BOTH, expand=True)

# Нұсқаулық жазуы

instruction = ttk.Label(main\_frame, text="Талдау үшін мәтінді енгізіңіз:")

instruction.pack(pady=5)

# Мәтін енгізу өрісі

text\_input = scrolledtext.ScrolledText(main\_frame, width=60, height=10, font=("Helvetica", 12))

text\_input.pack(pady=5)

# Батырмалар үшін жақтау

button\_frame = ttk.Frame(main\_frame)

button\_frame.pack(pady=5)

# "Қою" батырмасы

paste\_button = ttk.Button(button\_frame, text="Қою", command=paste\_from\_clipboard)

paste\_button.pack(side=tk.LEFT, padx=5)

# "Тазалау" батырмасы

clear\_button = ttk.Button(button\_frame, text="Тазалау", command=clear\_text)

clear\_button.pack(side=tk.LEFT, padx=5)

# Тіл таңдау мәзірі

language\_var = tk.StringVar(value="Қазақша")

language\_label = ttk.Label(main\_frame, text="Тілді таңдаңыз:")

language\_label.pack(pady=5)

language\_menu = ttk.Combobox(main\_frame, textvariable=language\_var, values=["Қазақша", "Орысша", "Ағылшынша"], state="readonly")

language\_menu.pack(pady=5)

# "Анализ" батырмасы

analyze\_button = ttk.Button(main\_frame, text="Анализ жасау", command=run\_analysis)

analyze\_button.pack(pady=10)

# Нәтиже шығару белгісі

result\_label = ttk.Label(main\_frame, text="Тональдылық:", font=("Helvetica", 14, "bold"))

result\_label.pack(pady=10)

# Бағдарламаны іске қосу

root.mainloop()