СӨЖ 10

Тақырыбы: Табиғи тілдерді өңдеу NLP

Университет студенттердің пән туралы қысқа пікірлерін автоматты түрде оң, бейтарап немесе теріс санатқа бөлетін NLP жүйесін әзірлеп жатыр.

Жүйеге түскен пікірлердің мысалдары:

Көрсеткіш	Мән
1-пікір	Материал түсінікті және мысалдар пайдалы болды.
2-пікір	Тапсырма көп, бірақ бағалау талаптары анық.
3-пікір	Жүйе жиі істемей қалады, файлды жүктей алмадым.
4-пікір	Дәріс жаман емес.
Тілдер	Қазақша және орысша
Мақсат	Пікірдің тональдігін анықтау

Тапсырма:

1. Мәтінді өңдеудің tokenization, қалыпқа келтіру және қажетсіз таңбаларды тазалау кезеңдерін көрсетіңіз.

2. 1–4 пікірлерге болжамды тональдік белгісін беріп, таңдауыңызды негіздеңіз.

3. «Дәріс жаман емес» сөйлемі модель үшін неге күрделі болуы мүмкін екенін түсіндіріңіз.

4. Мәтінді сандық түрде ұсынудың TF–IDF және embedding тәсілдерін салыстырыңыз.

5. Оқыту жиынында қазақша пікір аз болса, қандай мәселе пайда болатынын және оны қалай азайтуға болатынын жазыңыз.

6. Жүйе сапасын бағалау үшін confusion matrix, precision, recall және F1-score көрсеткіштерін қалай қолданатыныңызды түсіндіріңіз.

Бұл тапсырмада мәтіннің мағынасы тек жеке сөздерден құралмайтынын ескеріңіз. Тіл, контекст, терістеу және деректер теңгерімі модель шешіміне бірге әсер етеді.