

Лабораториялық жұмыс 12: Сентименттік талдау үшін деректер жинау және дайындау

Мақсаты:

Бұл зертханалық жұмыста сіз казак немесе орыс тілдеріндегі пікірлерді интернеттегі ашық көздерден жинап, сентименттік талдауға арналған құрылымды деректер жиынын дайындайсыз.

Сіз келесі әрекеттерді орындауыңыз қажет:

1. Онлайн платформалардан пікірлер/шолуларды жинау
2. Деректерді тазарту және құрылымдау
3. Деректерді сақтап, талдауға ыңғайлы форматқа келтіру
4. Төмендегі негізгі бағандардың болуын қамтамасыз ету

Деректер жиыны келесі бағандардан тұруы керек:

1. `ID` — Әр шолу үшін бірегей идентификатор
2. `Атауы` — өнімнің, қызметтің, фильмнің, кітаптың немесе мейрамхананың атауы
3. `Күні` — Шолу қалдырылған күн
4. `Баға` — 1-ден 5-ке дейінгі рейтинг немесе баға
5. `Пікір` — Пайдаланушының нақты пікірі немесе шолуы

Тіл талабы: Барлық пікірлер қазақ немесе орыс тілінде болуы керек.

Дереккөз ретінде пайдалануға болатын платформалар:

Сіз келесі сайттардан дерек жинай аласыз (немесе ұқсас платформалардан):

- Kaspi Магазин (<https://kaspi.kz/>) — Өнімдерге арналған пікірлер
- 2GIS (<https://2gis.kz/>) — Мейрамхана немесе мекеме шолулары
- КиноПоиск немесе басқа кино сайттары
- Онлайн кітап дүкендері (пікірлері бар)
- Басқа жергілікті сайттар, пікір қалдыруға мүмкіндік беретін

Ұсынылатын құралдар мен әдістер:

- Python тілі
- Web scraping үшін: `requests`, `BeautifulSoup`, `Selenium`
- `pandas` кітапханасы — деректерді өңдеу және сақтау үшін