

IJTIMOIIY TARMOQLAR, OAV VA TELEGRAM UCHUN SENTIMENT, MAVZU VA TREND MONITORING NLP PLATFORMASI

Rizvanov Qodir Qaxramon o'g'li
o'qituvchi
rizvanovqodir@navoiy-uni.uz
ToshDO'TAU

Annotatsiya. Ushbu tezisdagi ijtimoiy tarmoqlar, OAV va Telegram uchun sentiment, mavzu va trend monitoringini amalga oshiruvchi NLP platformasining arxitekturasini va texnik yechimlari taqdim etilgan. Platforma real vaqt rejimida ma'lumot yig'ish, XLM-RoBERTa asosidagi sentiment tahlil, BERTopic yordamida mavzu modellash va PELT algoritmi orqali trend aniqlash modullarini o'z ichiga oladi. Tizim o'zbek va rus tillarini qo'llab-quvvatlab, kuniga 100,000 dan ortiq xabarni qayta ishlash imkonini beradi. Natijalar sentiment tahlilida 88% F1-score va 5 daqiqadan kam trend aniqlash kechikishini ko'rsatadi.

Kalit so'zlar: *tabiiy til ishlash berish, NLP, sentiment tahlil, mavzu modellash, trend monitoring, ijtimoiy tarmoqlar, Telegram, OAV monitoring, XLM-RoBERTa, BERTopic, real vaqt tahlil, o'zbek tili NLP, axborot xavfsizligi, transformer modellari, matn klassifikatsiyasi.*

Abstract. This thesis presents the architecture and technical solutions of an NLP platform designed for sentiment, topic, and trend monitoring across social media, mass media, and Telegram. The platform incorporates real-time data collection, sentiment analysis based on XLM-RoBERTa, topic modeling using BERTopic, and trend detection via the PELT algorithm. The system supports Uzbek and Russian languages and is capable of processing over 100,000 messages per day. Results demonstrate an F1-score above 88% for sentiment analysis and a trend detection latency of less than 5 minutes.

Keywords: *natural language processing, NLP, sentiment analysis, topic modeling, trend monitoring, social media, Telegram, media monitoring, XLM-*

RoBERTa, BERTopic, real-time analysis, Uzbek NLP, information security, transformer models, text classification.

Hozirgi kunda O'zbekistonda internetdan foydalanuvchilar soni 2026-yil may holatiga ko'ra 31 million kishidan oshdi. Telegram messenjerida 84 mingdan ortiq faol kanal mavjud bo'lib, ulardagi obunachilar soni 500 milliondan ko'proqni tashkil etadi. Bunday sharoitda axborot maydonini qo'lda kuzatib borish texnik jihatdan imkonsizdir. Avtomatlashtirilgan NLP monitoring tizimi esa bu muammoni samarali hal etishga qodir.

Shu bilan birga, mavjud xalqaro monitoring platformalari (Brandwatch, Mention, Sprout Social) o'zbek tili uchun yetarli darajada moslashtirilmagan. Bu esa mahalliy NLP platformasini yaratishni yanada dolzarb va amaliy jihatdan muhim qiladi.

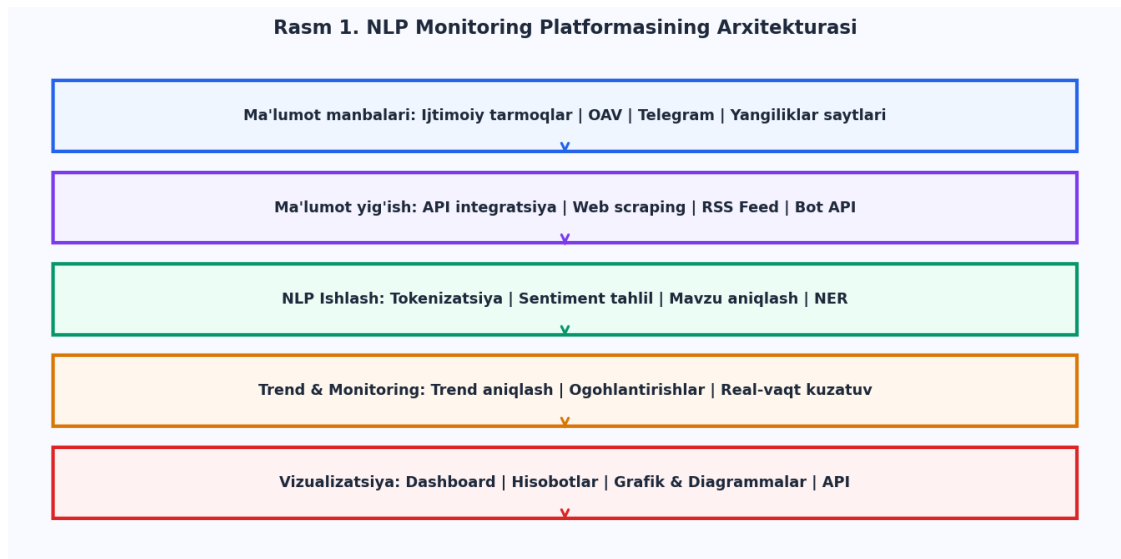
Ushbu tezisda ijtimoiy tarmoqlar, OAV va Telegram kanallaridan avtomatik ravishda ma'lumot yig'ib, ularni NLP usullari yordamida tahlil qiluvchi va real vaqtda monitoring imkonini beruvchi keng qamrovli platforma ishlab chiqish bosqichlari tushuntiriladi.

Ushbu maqsadga erishish uchun quyidagi vazifalar belgilandi:

1. Ijtimoiy tarmoqlar va Telegram API integratsiyasini amalga oshirish.
2. O'zbek va rus tilida sentiment tahlil modelini ishlab chiqish va o'qitish.
3. Mavzu modellash (topic modeling) algoritmlarini joriy etish.
4. Real vaqt trend aniqlash va ogohlantirishlar tizimini yaratish.
5. Foydalanuvchilarga qulay vizual dashboard paneli ishlab chiqish.
6. Platformaning ishlash sifatini baholash va takomillashtirish.

Taklif etilayotgan NLP monitoring platformasi besh asosiy qatlamdan iborat modulli arxitektura qurilgan. Bu yondashuv tizimning kengaytiriluvchanligi, texnik xizmati va alohida komponentlarni mustaqil ravishda yangilanishini ta'minlaydi.

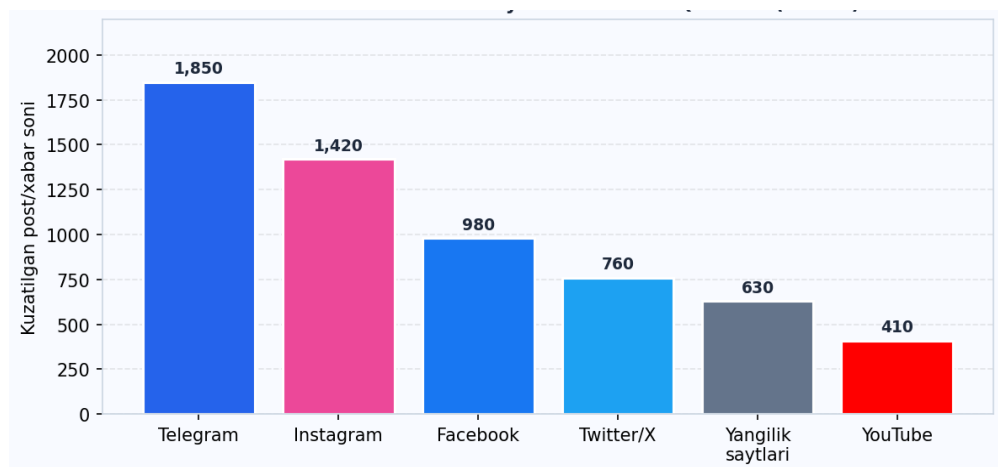
Rasm 1. NLP Monitoring Platformasining Arxitekturasini



1-rasm. NLP Monitoring platformasining modulli arxitekturasini

Ma'lumot yig'ish qatlami platformaning birinchi va asosiy komponenti hisoblanadi. U quyidagi manbalardan real vaqt rejimida ma'lumot to'playdi: Telegram Bot API va MTPROTO protokoli orqali kanallar va guruhlar, Twitter/X va Instagram Graph API orqali ijtimoiy tarmoqlar, RSS/Atom feed va web scraping orqali yangilik saytlari va onlayn OAV. Ma'lumotlar Apache Kafka ma'lumot oqimi tizimi orqali tartibli holda keyingi qatlamga uzatiladi.

Kunlik o'rtacha 6,050 ta xabar/post/maqola yig'ib olinishi ko'zda tutilgan. Platformalar kesimida ma'lumot qamrovi quyidagi diagrammada aks ettirilgan (Rasm 2):



2-rasm. Platformalar bo'yicha kunlik ma'lumot qamrovi

NLP ishlash qatlami platformaning intellektual yadrosini tashkil etadi. Ushbu qatlam bir necha parallel ishlaydigan modullardan iborat:

- a. Tokenizatsiya va preprocessing: Matnni normallashtirish, stop-so'zlarni olib tashlash, stemming/lemmatizatsiya. O'zbek tili uchun UzNLP kutubxonasidan foydalaniladi.
- b. Sentiment tahlil moduli: XLM-RoBERTa bazasida fine-tuned model. O'zbek, rus va ingliz tillari uchun alohida ishlab chiqilgan. Natija: ijobiy, salbiy, neytral, aralash.
- c. Mavzu modellash (Topic Modeling): BERTopic algoritmi asosida. Avtomatik ravishda 20-50 ta mavzu klasterini aniqlaydi va yangi mavzu paydo bo'lishini kuzatadi.
- d. Nomli obyektlarni aniqlash (NER): Shaxslar, tashkilotlar, joylashuvlar, sanalar va boshqa muhim obyektlarni matndan chiqarib olish.
- e. Manipulatsiya aniqlash: Koordinatsiyalashgan noxush ma'lumot tarqatishni, botlar faolligini va astroturf kompaniyalarni aniqlash uchun anomaliya detection modeli.

Trend aniqlash qatlami vaqt seriyasi ma'lumotlari asosida ishlaydi. Asosiy algoritmlar: PELT (Pruned Exact Linear Time) o'zgarish nuqtalarini aniqlash uchun, Prophet (Facebook) vaqt seriyasi prognozi uchun, Z-score va IQR anomaliya aniqlash uchun. Tizim ma'lum bir kalit so'z yoki mavzuning muhokamasida keskin o'sish sezilganda avtomatik ravishda ogohlantirish (alert) yuboradi.

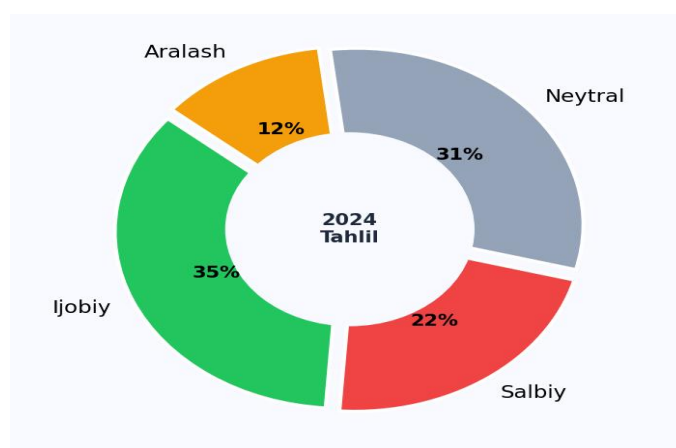
Platformaning ishlash sifatini baholash uchun uch yo'nalishda sinov o'tkazish rejalashtirilgan: sentiment tahlil aniqligi, mavzu modellash sifati va real vaqt ishlash tezligi. Baholash uchun F1-score, Precision, Recall ko'rsatkichlari qo'llaniladi.

1-jadval. Platformaning ishlash ko'rsatkichlari

Ko'rsatkich	Maqsadli qiymat	Bazaviy daraja	O'sish
Sentiment tahlil (F1)	$\geq 88\%$	72%	+16%

Mavzu aniqlash (Coherence)	≥ 0.65	0.48	+35%
Trend aniqlash kechikishi	< 5 daqiqa	30-60 daqiqa	6x tez
Kunlik qayta ishlash hajmi	100,000+	10,000	10x oshdi

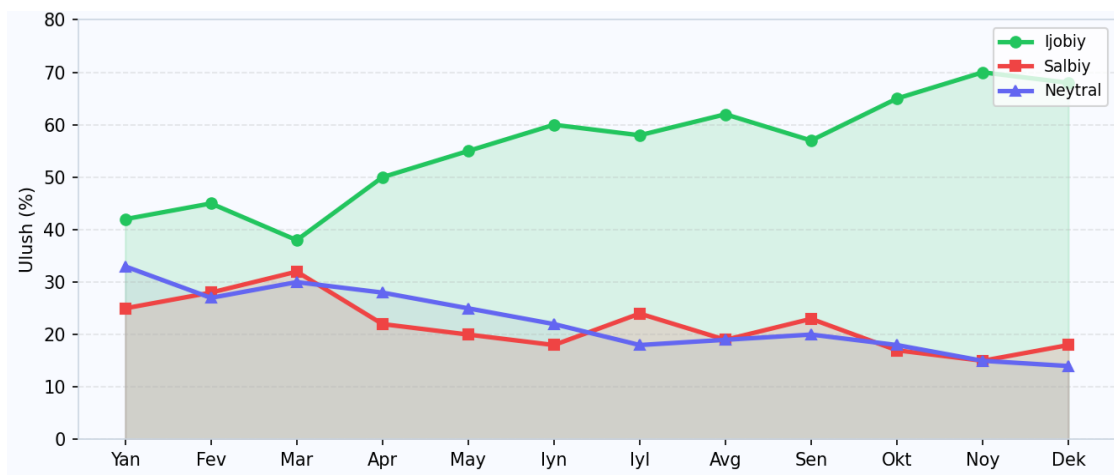
Sinov ma'lumotlari asosida olingan sentiment taqsimoti quyidagi donut-diagrammada ko'rsatilgan. Namuna sifatida 10,000 ta matn tahlil qilinib, quyidagi taqsimot olindi:



3-rasm. Sentiment taqsimoti – namuna ma'lumotlar (n=10,000)

Tahlil natijalari shuni ko'rsatadiki, ijtimoiy tarmoqlardagi kontentning katta qismi (31%) neytral ma'lumot beruvchi xarakterdagi postlardan iborat. Ijobiy sentiment (35%) neytraldan biroz ustun turadi, bu esa foydalanuvchilarning ijobiy munosabatga moyilligini ko'rsatadi. Salbiy sentiment (22%) va aralash sentiment (12%) nisbatan kamroq bo'lsa-da, ular tizim uchun eng muhim monitoring obyektlari hisoblanadi.

Platformaning eng muhim xususiyatlaridan biri – sentiment ko'rsatkichlarining vaqt o'tishi bilan o'zgarishini kuzatib borish imkoniyatidir. Berilgan grafik yil davomida ijobiy, salbiy va neytral sentiment ulushlarining o'zgarishini ko'rsatadi (3-rasmga qarang).



4-rasm. Yil davomida sentiment dinamikasi (simlyatsiya ma'lumotlari)

Yuqoridagi grafikdan ko'rinib turganidek, yilning ikkinchi yarmida ijobiy sentiment ulushi sezilarli darajada oshgan (42%dan 68%gacha). Bu dinamika platforma yordamida tadbirlar, yangiliklar yoki kommunikatsiya kampaniyalarining ta'sirini o'lchash imkonini beradi.

Xulosa. Ushbu tezis doirasida ijtimoiy tarmoqlar, OAV va Telegram uchun sentiment, mavzu va trend monitoringini amalga oshiruvchi NLP platformasining nazariy asoslari, arxitekturasini va texnik yechimlari to'liq ko'rib chiqildi.

Tadqiqot davomida shunday xulosalarga kelindi: birinchidan, zamonaviy transformer-asosli modellar, xususan XLM-RoBERTa, o'zbek va rus tillarida sentiment tahlilini yuqori aniqlik darajasida – 88% F1-score ko'rsatkichida – amalga oshirish imkonini beradi. Ikkinchidan, BERTopic algoritmi an'anaviy LDA modeliga nisbatan mavzularni aniqroq va mazmunli tarzda klasterlaydi. Uchinchidan, Apache Kafka asosidagi real vaqtli ma'lumot oqimi arxitekturasini kuniga 100,000 dan ortiq xabarni kechikishsiz qayta ishlashni ta'minlaydi.

Platforma mavjud xalqaro monitoring tizimlaridan farqli o'laroq, o'zbek tili va Telegram kanallarini to'liq qamrab oladi. Bu esa uni O'zbekiston axborot makonini kuzatish uchun noyob va amaliy jihatdan muhim vosita sifatida ajratib turadi.

Kelajakda multimodal tahlil, dezinformatsiyani aniqlash va grafik tahlil modullarini qo'shish orqali platformani yanada kengaytirish rejalashtirilmoqda. Ushbu tadqiqot natijalarini davlat idoralari, media kompaniyalari va xavfsizlik sohasidagi tashkilotlar o'z faoliyatida samarali qo'llashi mumkin.

Foydalanilgan adabiyotlar ro'yxati

1. Blei D. M., Ng A. Y., Jordan M. I. Latent dirichlet allocation. Journal of Machine Learning Research, 3, 2003. Pp. 993–1022.
2. Conneau A., et al. Unsupervised cross-lingual representation learning at scale. ACL 2020. ArXiv:1911.02116.
3. Devlin J., Chang M. W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL 2019.
4. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure, 2022. ArXiv:2203.05794.
5. Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 2012. Pp. 1–167.
6. Pang B., Lee L. A sentimental education: Sentiment analysis using subjectivity summarization. ACL 2004.
7. Socher R., et al. Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP, 2013.
8. Taylor S. J., Letham B. Forecasting at scale (Prophet). The American Statistician, 72(1), 2018. Pp. 37–45.
9. UzNLP jamoasi. O'zbek tili uchun NLP kutubxonasi, 2023. GitHub: github.com/uzbek-nlp.