

INTELLEKTUAL MATN TAHLILI

Sabohat Yusupboyevna Allanazarova

O'zbekiston Milliy universiteti magistranti

allanazarovasabohat@gmail.com

Nilufar Zaynobiddin qizi Abduraxmonova

O'zbekiston Milliy universiteti dotsenti, DSc

abdurahmonova.1987@mail.ru

ANNOTATSIYA

Fan va texnikaning yuksak darajada rivojlanishi natijasida ijtimoiy saytlar orqali qabul qilinadigan ma'lumotlar hajmi boshqarib bo'lmaydigan tarzda keskin oshib ketdi. Ko'pchilik foydalanuvchilar duch keladigan asosiy muammo - vaqt yetishmasligi. Sababi, barcha ma'lumotlarni o'qishning iloji yo'q. Bu esa ushbu ma'lumotlarni saralash zaruratini tug'diradi. Shu sababli Tabiiy tillar jarayoni (NLP), Mashinaviy o'qitish (Machine learning), Ma'lumotlarga ishlov berish (Data Science) va Sun'iy intellekt kabi sohalarga qiziqishning o'sishi kuzatildi. XXI asr texnologiyalar va intellektual salohiyat asrida insonlarning fikrini o'rganish bo'yicha yangi metod paydo bo'ldi.

Kalit so'zlar: sentiment tahlili, tonallik, ijtimoiy tarmoq, ma'lumotlar bazasi, xizmatlarni baholash.

ABSTRACT

As a result of the high level of development of science and technology, the amount of information received through social sites has increased dramatically. The main problem that most users face is lack of time. Because not all data can be read. Which makes it necessary to sort this data. Therefore, there has been an increase in interest in areas such as the natural language process (NLP), machine learning, data science and artificial intelligence. The 21st century, in the age of technology and intellectual potential, a new method of studying people's opinions has emerged.

Keywords: sentiment analysis, tonality, social network, database, evaluation of services

KIRISH

Innovatsion texnologiyalar asrida hayotimizni ijtimoiy tarmoqlarsiz tasavvur qila olmaymiz, shu sababli ham kunlik



qabul qilinadigan ma'lumotlar hajmi keskin oshib ketadi. Internet ma'lumotlarining ko'pligi sababli, uni intellektual tahlil qilish zaruriyati va undan turli xil bilimlarni o'zlashtirish masalasi bugungi kunning talabidir.

ADABIYOTLAR TAHLILI VA METODOLOGIYA

Rivojlangan davlatlarda tabiiy tilni qayta ishlash texnologiyalari (NLP) va Informatsion texnologiyalar (IT) sohasida yaxshi natijalarga erishilmoqda. Ushbu ilmiy sohaning eng qiziqarli va samarali usullari orasida Sentiment tahlili ajralib turibdi. Sentiment tahlil qilish kompyuter lingvistikasi uchun samarali tadqiqot yo'nalishlaridan hisoblanadi. Dastlab Sentiment nima o'zi degan savolga javob bersak. Bu - matndagi emotsional rangdagi so'z boyligini avtomatlashtirilgan tarzda aniqlash uchun ishlatiladi. Sentiment tahlili yana quyidagicha nomlanadi:

- **Fikrlarni izlash** (Opinion extraction)
- **Fikrlarni o'rganish** (Opinion mining)
- **Hissiy tahlil qilish** (Sentiment analysis)
- **Subyektivlik tahlili** (Subjectivity analysis)

Hissiyotlarni tahlil qilish va o'rganish - bu yozma tildan odamlarning fikrlari, baholari, munosabati va hissiyotlarni tahlil qiladigan soha bo'lib, dunyo bo'ylab sentiment tahlil deb nomlanadi. Uning qo'llanilishi, biznes xizmatlaridan tortib, siyosiy kompaniyalargacha keng tarqalgan. Xususan:

- ❖ Media mahsulotlari;
- ❖ Fond bozorlari;
- ❖ Siyosat;
- ❖ Tabiiy ofatlar;
- ❖ Tibbiyot;
- ❖ Dasturiy ta'minot;
- ❖ Muhandislik;
- ❖ Xizmat ko'rsatish;
- ❖ Kiberhujum;

❖ **Ekstremistik** matnlarni **neytral** matnlardan farqlash va boshqa sohalarda foydalaniladi.

Xususan so'nggi yillarda ekstremizm g'oyalarning internet foydalanuvchilari o'rtasida tarqalishi, matnlarni tahlil qilishning nechog'lik muhim ekanligini isbotlamoqda. Ekstremistik ruhdagi matnlarni neytral matnlardan farqlash orqali yurtimizda sodir bolishi mumkin bo'lgan xavflarning oldi olinadi. Axborot urushlari paytida bu vazifa eng



muhim vazifalardan biridir Matn ohangini tahlil qilish, inglizcha fikrlarni tahlil qilish demakdir. Leksema yoki kommunikativ parcha darajasida ifodalangan hissiy tarkibiy qism leksik tonallik yoki leksik tuyg‘u deb nomlanadi.

Tonallik- bu muallifning matnda ifodalangan ba’zi bir obyektga, real hodisa, jarayonga nisbatan hissiy munosabatidir. Inson dunyoni birdaniga ko‘p miqyosda: **yaxshi-yomon, kuchli-zaif, katta-kichik, baxtli-baxtsiz, kulguli-g‘amgin, tez-sekin** sifatida baholaydi. Matnda ifodalangan hissiy bahoning tashuvchisi ham odatda aniq shaxs, umuman olganda u matn muallifidir. Agar matn muallifi birovning fikrini keltirsa, u holda hissiy bahoning tashuvchisiga aylanadi.

Nega hissiyotlar shunchalik muhim? Bu tahlilda asosan biror voqea-hodisaga, real vaziyat yoki mahsulot, ishlab chiqaruvchi korxonaga, xizmat ko‘rsatuvchi muassasalarga ijtimoiy tarmoq foydalanuvchilarining munosabati, e’tirofi, shaxsiy kechinmalari kabi norasmiy sharhlarga e’tibor qaratiladi.

Twitter, Facebook kabi ijtimoiy tarmoqlarda kishilar biznes va xizmatlar haqida baho beradilar va sharhlar yozadilar[6]. Ushbu sharhlar va reytinglar boshqa foydalanuvchilarga biznes yoki xizmatni baholashga va mahsulotni tanlashga yordam beradi. Restoran haqidagi sharhlarga e’tiboringizga qaratamiz:

Ijobiy: Qulay uy muhiti! Buvim yoqtirgani kabi! Oilaviy dam olish uchun ajoyib joy! Xizmat eng yuqori darajada!

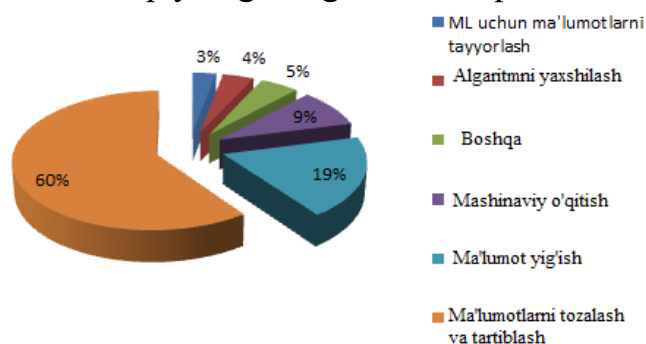
Betaraf: Odatiy. Bu oddiy kafega o‘xshaydi.

Salbiy: Xizmat ko‘rsatish juda yomon, narxlar juda baland, sho‘rva issiq emas, bezatilishigina yaxshi, men bu joyni tavsiya qilmayman!

Tasniflash uch toifada amalga oshiriladi: ijobiy yoki salbiy va betaraf. Biroq bu yondashuvlarda cheklov mavjud. Ba’zi tadqiqotchilar ishni besh guruh(ijobiy/o‘ta ijobiy, salbiy/o‘ta salbiy va betaraf)ga ajratgan holda amalga oshirishlari mumkin. Bu o‘rganilayotgan tilning nechog‘lik lingvistik resurslar va dasturiy modellarga boy ekanligiga bog‘liq. Tasniflashning eng oson usuli bu - ikkilik tasniflash bo‘lib, bu ish allaqachon o‘zbek tili uchun qilingan.

Ma’lumotlar bilan ishlash. Sentiment tahlilida ma’lumotlar bazasini tuzish o‘zbek tili uchun nisbatan murakkab va ancha vaqtni oladi. Buning uchun bot tuzish va ma’lum mavzuga aloqador so‘rovnoma o‘tkazish kerak bo‘ladi yoki ijtimoiy saytlardagi fikrlarni yuklash uchun maxsus kod tuzish mumkin. Meta kompaniyasining xavfsizlik tizimi tufayli sharhlarni yuklashda qiyinchilik yuzaga kelishi mumkin.

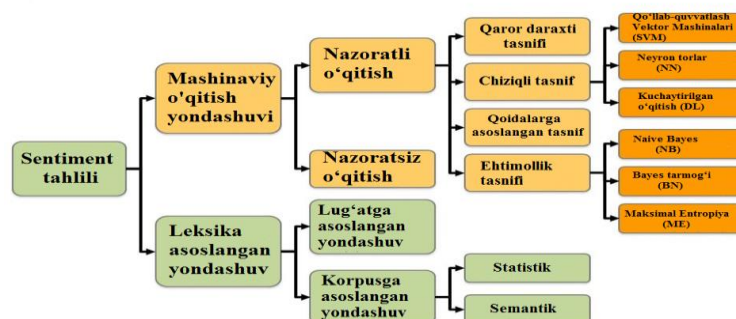
Avvalo, ma'lumotlarga ishlov berish jarayoni qanchalik muhim va asosiy qism ekanligini ta'kidlash zarur. Buni quyidagi diagramma orqali ko'rishimiz mumkin:



1-rasm. AI muhandislarining 80% dan ortiq vaqti aynan ma'lumotlarga ishlov berishdan iborat bo'ladi. Faqatgina 9% vaqt bevosita mashinaviy o'qitish va 4% ML algoritmlarni sozlashga ketadi.

Ikkinchi katta muammo so'zlarning o'zagini topish (stemming yoki lemmatizasiya) ya'ni buni avtomatlashtirish.

Mazkur mavzu bo'yicha ko'plab tadqiqot ishlari mavjud. Ushbu bobda tasniflash vazifasi uchun qo'llanilishi mumkin bo'lgan algoritmlarning qisqacha izohlari keltirilgan.



2-rasm. Sentiment tahlili texnikasi

Hissiyotlarni avtomatik aniqlash muammosini hal qilishning ikkita asosiy usuli mavjud:

1. Statistik uslub. Buning uchun oldindan tonallik bilan belgilangan matnlar to'plami (korpus) kerak bo'ladi. Ular bo'yicha model o'qitiladi, uning yordamida matn tasniflanadi;

2. Lug'atlar va qoidalarga asoslangan uslub. Buning uchun ijobiy va salbiy so'zlar va iboralarning lug'ati tuziladi.

Ba'zida aralash (gibrid) uslub ham qo'llaniladi.

MUHOKAMA

Matn tahlilidagi asosiy qiyinchilik inson tillarining noaniqligidadir[4]. O'zbek tilidagi matnlar uchun sentiment tahlili dasturini ishlab chiqishda mazkur tilning xususiyati hisobga olinishi zarur. O'zbek tilida berilgan matnli ma'lumotlardan inson fikrini ifodalashni avtomatik keltirib chiqarishni shakllantiradigan modellar qurish asosiy vazifa bo'lib turibdi. Buning uchun ingliz tilida bo'lgan **SentiWordNet**, **SentiNet** kabi hissiy tahlil uchun korpus qurish kerak. SentiNet qanday ishlaydi? Hamma fikrlar ham hamisha faqat salbiy yoki ijobiy bo'lavermaydi. Yarmi salbiy, yarmi ijobiy bo'lishi mumkin. Bunda gapning chuqur ma'nosini aniqlash uchun kalit so'zlardan tashkil topgan baza kerak. SentiNet gapning jihati bo'yicha tahlil qiladi.

Ba'zi yomon narsalar yuz berdi, xafa bo'ldim, lekin endi o'zimni yaxshi his qilyapman.

Ijobiy	Salbiy
Yaxshi	Yomon
	Xafa

↓
Ijobiy matn

Yuqoridagi misolda ikki xil fikr ifodalangan. Garchi salbiy so'zlar soni ko'p bo'lsa ham, tahlil gapni ijobiy deb topadi. Sababi ammo, lekin, biroq kabi zidlov bog'lovchilari qatnashgan gaplar tarkibidagi qism tashlab ketiladi. Bu gapning chuqur ma'nosi (aspect based) bo'yicha tasniflashdir.

Sentiment tahlilida matn ohangini aniqlashda so'rovlar muhim ahamiyatga ega. Masalan: aynan bir matnning o'ziga har xil savol berish orqali hukm o'zgarib ketadi:

- 1. Bu tadbirdan sizga nimalar yoqdi?**
- 2. Bu tadbirdan sizga nimalar yoqmadi?**

↓
Hammasi

↓
Hech qaysi

Guvohi bo'lganimizdek, birinchi javob ijobiy, ikkinchi javob esa salbiy bahoni ifodalagan. Yana boshqa misolni ko'rsak:

- 1. Sumka qizil**
- 2. Sumka yaxshi**

Bir qarashda birinchi izoh betaraf, ikkinchi izoh esa, ijobiy fikr ifodalagandek tuyuladi. Aslida ikkala izoh ham ijobiy fikr bo'lib, yaxshi so'zida subyektiv baho kuchli bo'lganligi sababli darrov e'tiborni tortadi.

O'zbek tilida ibora, kinoya piching, kesatirlardan keng foydalaniladi. Bu esa mashinaga tahlil qilish jarayonida

qiyinchilik tugʻdiradi. Masalan: **Qirq marta oʻlchab, bir marta kesmoq** – extiyotkor maʼnosida qoʻllaniluvchi ushbu iborani kompyuter alohida-alohida soʻzlar sifatida qabul qiladi. Xususan, kinoya haqida gap ketganda, odamlar oʻzlarining salbiy his-tuygʻularini ijobiy soʻzlardan foydalangan holda ifoda etadilar. Kinoyaviy qoʻllanilgan soʻz ogʻzaki nutqda oʻziga xos, farqli intonatsiya bilan aytilsa, yozuvda koʻpincha qoʻshirnoq bilan ijratiladi: Bugun geografiya oʻqituvchisi darsga kelmadi. Soʻridan yiqilib, oyogʻini sindiribdi. Bu **“xushxabar”** dan keyin butun sinf boʻshadi-qoldi. **“Shumxabar”** tarzidagi bu kinoyani mashina tushunadigan dastur ishlab chiqishni talab qiladi Bu esa mashinaga tahlil qilish jarayonida qiyinchilik tugʻdiradi[3]. Masalan: **“Elmurod bu sohada eng mohiri”**. Kompyuter bu jumladagi teskari maʼnoni tushunmaydi. Natijada fikrni ijobiy deb baholaydi. Hozirgacha oʻzbek tilida qilingan ishlarda paremiologik birliklar tashlab ketilgan.

XULOSA

Maʼlumki, biror mahsulot yoki muammoga qiziqish dastlab uni oʻrganishdan boshlanadi. Ijtimoiy tarmoqlarda esa istalgancha izlash imkoni mavjud. Axborotlar shunchalik koʻpki, bu maʼlumotlarni toʻliq oʻrganib chiqish uchun bir kishining 24 soat vaqti kifoya qilmaydi. Sentiment tahlilining maqsadi ijtimoiy tarmoqlarning kuchidan turli xil sohalardagi kayfiyatni oʻrganishdan iborat.

REFERENCES

1. Abduraxmonova N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018).
2. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
3. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
4. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019;6(1-2019):131-7.
5. Abdurakhmonov N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine



- Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.
6. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
 7. Aripov M., Sharipbay A., Abdurakhmonova N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018)
 8. Go A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12).
 9. Kubedinova L. Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020) .2020/11: 90-101
 10. Kuriyozov E., Matlatipov S. Building a new Sentiment Analysis Dataset for Uzbek Language and Creating Baseline Models.// Multidisciplinary Digital Publishing Institute Proceedings. 2019.-№1. Pages 37.
 11. Rabbimov I., Mporas I., Kobilov S. Investigating the effect of emoji in opinion classification of uzbek movie review comments. International Conference on Speech and Computer Science. -2020.-P.435-445.
 12. Chen Yanqing and Steven Skiena, 2014. Building sentiment lexicons for all major languages. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Vol. 2: Short Papers). Baltimore, Maryland: Association for Computational Linguistics. Chollet, Francois et al., 2015. Keras. <https://github.com/fchollet/keras>.