Foydalanilgan adabiyotlar ro'yxati:

(1). Wiener M.J. (1990). Cryptanalysis of short RSA secret exponents. IEEE Transactions on Information Theory, 36(3), p. 553–558.

(2). Boneh D., Durfee G. (1999). Cryptanalysis of RSA with private key d less than N^0.292. EUROCRYPT 1999, LNCS 1592.

(3). Goldwasser S., Micali S. (1984). Probabilistic encryption. Journal of Computer and System Sciences, 28 (2), p. 270–299.

(4). Tsiounis Y., Yung M. (1998). On the security of ElGamal based encryption. PKC 1998, LNCS 1431.

(5). Boneh D. (1998). The decision Diffie-Hellman problem. ANTS III, LNCS 1423.

(6). Hankerson D., Menezes A., Vanstone S. (2004). Guide to Elliptic Curve Cryptography. Springer.

Kamolova Moxiniso Farmon qizi (Navoiy davlat universiteti talabasi, O'zbekiston;
kamolova.moxiniso@mail.ru; https://orcid.org/0009-0005-4052-4911)

NAIVE BAYES ALGORITMI ASOSIDA MATN HISSIYOTINI AVTOMATIK TAHLIL QILUVCHI DASTURIY VOSITANI ISHLAB CHIQISH TEXNOLOGIYASI

Annotatsiya. *Matnli ma'lumotlardan foydalanuvchilarning fikr va munosabatini avtomatik aniqlash (sentiment tahlil) tabiiy tilni qayta ishlashning dolzarb yo'nalishlaridan biridir. Ushbu maqolada Naive Bayes klassifikatori asosida o'zbek tilidagi qisqa matnlarning ijobiy yoki salbiy ekanligini avtomatik baholovchi dasturiy vosita ishlab chiqilgan. Dastur Python tilida, Tkinter grafik kutubxonasi yordamida yaratilgan bo'lib, tashqi kutubxonalarga bog'liq emas va internet aloqasini talab qilmaydi. Laplas tekislashidan foydalanish hisobiga noyob so'zlar uchun ham barqaror natija ta'minlanadi. Olib borilgan sinovlar dasturning amaliy va o'quv jarayonidagi qo'llanilishini ko'rsatadi.*

Kalit so'zlar: *sun'iy intellekt, mashinaviy o'rganish, tabiiy tilni qayta ishlash, sentiment tahlil, Naive Bayes, klassifikatsiya, Python, dasturiy ta'minot.*

Камолова Мохинисо Фармон кизи (Навоийский государственный университет, студент, Узбекистан;
kamolova.moxiniso@mail.ru; https://orcid.org/0009-0005-4052-4911)

ТЕХНОЛОГИЯ РАЗРАБОТКИ ПРОГРАММНОГО СРЕДСТВА ДЛЯ АВТОМАТИЧЕСКОГО АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА НА ОСНОВЕ АЛГОРИТМА NAIVE BAYES

Аннотация. *Автоматическое выявление мнений и отношения пользователей на основе текстовых данных (анализ тональности, сентимент-анализ) является одним из актуальных направлений обработки естественного языка. В данной статье разработано программное средство, которое на основе классификатора наивного Байеса автоматически оценивает положительную или отрицательную окраску коротких текстов на узбекском языке. Программа создана на языке Python с использованием графической библиотеки Tkinter; она не зависит от внешних библиотек и не требует подключения к сети Интернет. За счёт применения сглаживания Лапласа обеспечивается устойчивый результат и для редких (ранее не встречавшихся) слов. Проведённые испытания демонстрируют применимость программы в практической и учебной деятельности.*

Ключевые слова: *искусственный интеллект, машинное обучение, обработка естественного языка, анализ тональности (сентимент-анализ), наивный Байес, классификация, Python, программное обеспечение.*

Kamolova Moxiniso Farmon qizi (Navoi State University, student, Uzbekistan;
kamolova.moxiniso@mail.ru; https://orcid.org/0009-0005-4052-4911)

TECHNOLOGY FOR DEVELOPING A SOFTWARE TOOL FOR AUTOMATIC TEXT SENTIMENT ANALYSIS BASED ON THE NAIVE BAYES ALGORITHM

Annotation. *Automatic detection of users' opinions and attitudes from textual data (sentiment analysis) is one of the most relevant directions in natural language processing. This article presents a software tool that, based on the Naive Bayes classifier, automatically evaluates whether short texts in the Uzbek language are positive or negative. The program is developed in Python using the Tkinter graphical library; it does not depend on external libraries and does not require an internet connection. Through the use of Laplace smoothing, a stable result is ensured even for rare (previously unseen) words. The conducted experiments demonstrate the applicability of the program in both practical and educational settings.*

Keywords: *artificial intelligence, machine learning, natural language processing, sentiment analysis, Naive Bayes, classification, Python, software.*

Kirish. So‘nggi yillarda ijtimoiy tarmoqlar, elektron tijorat platformalari va ta’lim tizimlarida hosil bo‘layotgan matnli ma’lumotlar hajmi keskin ortib bormoqda. Bunday ma’lumotlardan foydalanuvchilarning biror mahsulot, xizmat yoki hodisaga bo‘lgan munosabatini avtomatik ravishda aniqlash sentiment tahlil (matn tonalligini tahlil qilish) deb ataladi va u tabiiy tilni qayta ishlash (Natural Language Processing) hamda sun’iy intellektning muhim amaliy masalalaridan biri hisoblanadi [1,2].

Matnni tasniflashda mashinaviy o‘rganish usullari keng qo‘llaniladi. Pang, Li va Vaithyanathan o‘zlarining 2002-yildagi tadqiqotida ushbu masalaga ilk bor mashinaviy o‘rganishni qo‘llab, Naive Bayes, maksimal entropiya va tayanch vektorlar mashinasi (SVM) algoritmlarini taqqoslab chiqdi [1]. Keyingi tadqiqotlar sodda tuzilishiga qaramay, Naive Bayes klassifikatori kichik hajmli ma’lumotlarda ham raqobatbardosh natija berishini va hisoblash jihatidan tejamkorligini ko‘rsatdi [3,4].

O‘zbek tili kabi resurslari cheklangan tillar uchun esa tayyor korpuslar va lingvistik vositalarning yetishmasligi qo‘shimcha qiyinchilik tug‘diradi [6,7]. Shu sababli ko‘p hollarda transformer asosidagi zamonaviy modellar yuqori aniqlik bersa-da, ular katta hisoblash resurslari va o‘qitilgan korpuslarni talab qilgani bois oddiy shaxsiy kompyuterda, internetsiz ishlovchi yengil yechimlar hamon dolzarbligicha qolmoqda.

Tadqiqotda Naive Bayes algoritmi asosida ishlovchi, tashqi kutubxonalarga bog‘liq bo‘lmagan, mustaqil .exe fayl sifatida tarqatiladigan va o‘zbekcha qisqa matnlarning foiz aniqligida baholovchi dasturiy vositani loyihalash va amalga oshirishdan texnologiyasi yoritilgan.

Adabiyotlar tahlili. Sentiment tahlil sohasidagi tadqiqotlar shartli ravishda uch yo‘nalishga ajratiladi: lug‘atga (leksikonga) asoslangan, mashinaviy o‘rganishga asoslangan va chuqur o‘rganishga (neyron tarmoqlarga) asoslangan yondashuvlar. Pang va Li tomonidan amalga oshirilgan dastlabki tajribalarda Naive Bayes klassifikatori IMDb kinotanqid sharhlarida 82,9% aniqlik ko‘rsatgan [1]. Deyv va hammualliflari mahsulot sharhlari korpusida Naive Bayesga Laplas tekislashini qo‘llab, aniqlikni 87% gacha oshirishga erishgan [3]. Rish o‘zining empirik tadqiqotida ushbu algoritmning belgilar o‘zaro bog‘liq bo‘lmagan hollarda ham yuqori barqarorlik ko‘rsatishini ta’kidlagan [4], Makkallum va Nigam esa matnni tasniflashda multinomial Naive Bayes modeli ko‘pgina holatlarda eng samarali ekanligini asoslab bergan [5].

O‘zbek tili uchun bu sohadagi ishlar nisbatan yangi bo‘lib, Quriyozov va Matlatipov o‘zbek tili uchun birinchi sentiment tahlil ma’lumotlar to‘plamini shakllantirib, bazaviy modellarni taqdim etishgan [6]. Keyinchalik Matlatipov va hammualliflari mahalliy restoran sharhlari asosida kengaytirilgan korpus va tasniflash modellarini ishlab chiqdi [7]. Rabbimov va hammualliflari esa kinosharhlar misolida emoji belgilarining tasniflash sifatiga ta’sirini LSTM neyron tarmog‘i yordamida o‘rgangan [8].

Mavjud ishlarning aksariyati ilmiy korpuslar va modellar yaratishga qaratilgan bo‘lib, oxirgi foydalanuvchi uchun mo‘ljallangan, mustaqil ishlaydigan va o‘rnatishni talab qilmaydigan tayyor dasturiy mahsulotlar deyarli uchramaydi. Bu esa nazariy natijalar bilan amaliy qo‘llanish o‘rtasida ma’lum bo‘shliqni yuzaga keltiradi. Ushbu ishda aynan shu bo‘shliqni to‘ldirishga Naive Bayes algoritmini sodda, yengil va ko‘chma (.exe) dasturiy vosita ko‘rinishida amalga oshirishga harakat qilingan. Algoritmning nazariy asoslari va dasturlash texnikasi bo‘yicha [9,10] manbalardan foydalanildi.

Tadqiqot metodologiyasi. Dasturning matematik asosini multinomial Naive Bayes klassifikatori tashkil etadi. Bayes teoremasiga ko‘ra, berilgan matn (d) uchun sinf (c) ehtimoli quyidagicha baholanadi:

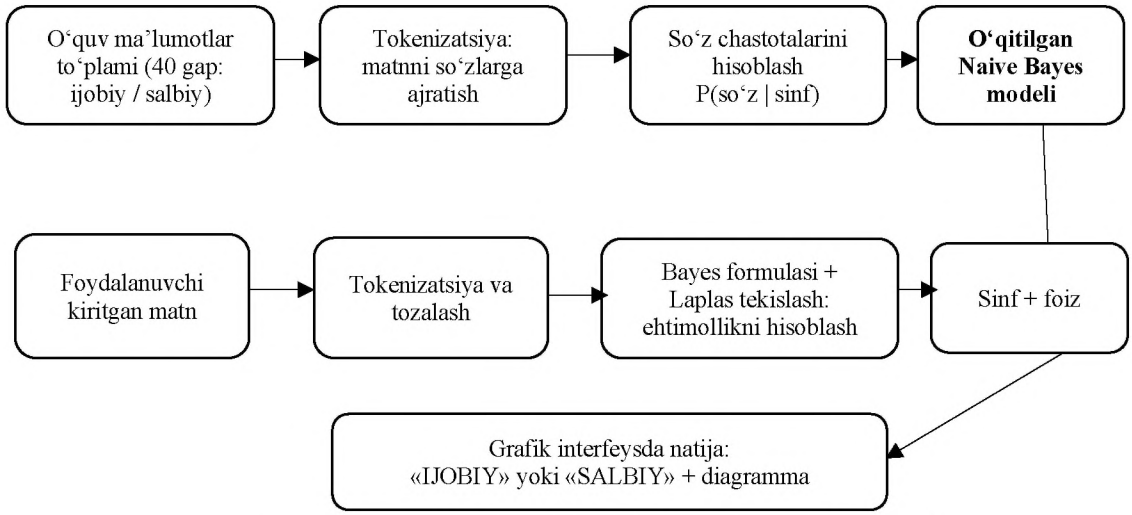
$$P(c \mid d) \propto P(c) \cdot \prod P(w_i \mid c)$$

bu yerda: P(c) – sinfning boshlang‘ich ehtimoli, $P(w_i \mid c)$ – w_i so‘zining shu sinfda uchrash ehtimoli. So‘z ehtimoli noyob so‘zlar uchun nolga aylanib qolmasligi maqsadida Laplas tekislashi qo‘llaniladi:

$$P(w \mid c) = \left(n(w, c) + 1\right) / \left(N(c) + |V|\right)$$

bu yerda: n(w, c) – so‘zning c sinfdagi chastotasi, N(c) – sinfdagi jami so‘zlar soni, |V| – lug‘at hajmi. Hisoblash jarayonida sonlar juda kichik qiymatlarga aylanmasligi uchun ehtimolliklar logarifmik shaklda jamlanadi va so‘ngra normallashtiriladi.

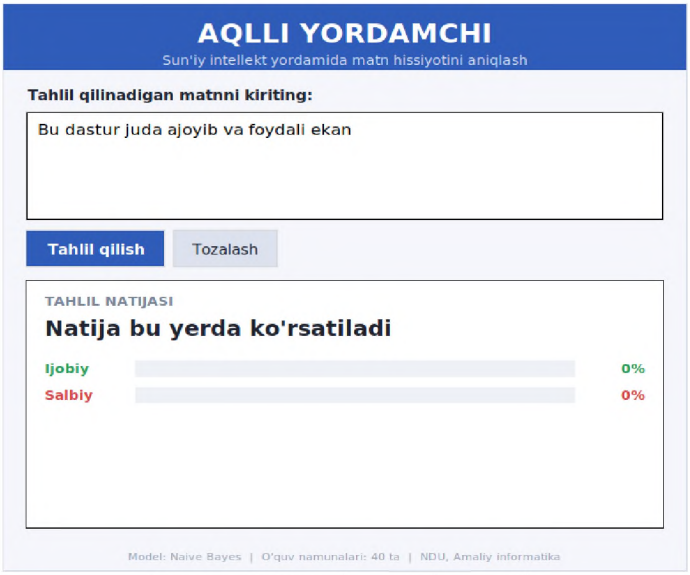
Dasturning umumiy ishlash sxemasi 1-rasmda keltirilgan. U ikki bosqichdan iborat: o‘qitish (training) bosqichida belgilangan namunaviy gaplar tokenlarga ajratilib, so‘z chastotalari hisoblanadi va model shakllantiriladi; tahlil bosqichida esa foydalanuvchi kiritgan matn xuddi shu tarzda qayta ishlanib, eng yuqori ehtimollikka ega sinf aniqlanadi (1-rasm).



1-rasm. Dasturning ishlash algoritmi (umumiy arxitekturasi).

Dastur Python 3 tilida amalga oshirilgan. Naive Bayes klassifikatori tashqi kutubxonalarsiz, faqat standart vositalar yordamida yozilgan; bu exe faylning hajmini kichik saqlash va ortiqcha bog‘liqliklardan xoli bo‘lishini ta’minlaydi. Grafik interfeys Tkinter kutubxonasida qurilgan. O‘qitish to‘plami sifatida ijobiy va salbiy deb belgilangan 40 ta o‘zbekcha gap olingan bo‘lib, ular asosida 115 ta noyob so‘zdan iborat lug‘at hosil bo‘ladi. Matnni dastlabki qayta ishlash bosqichida barcha harflar kichik registrga keltiriladi va matn so‘zlarga ajratiladi.

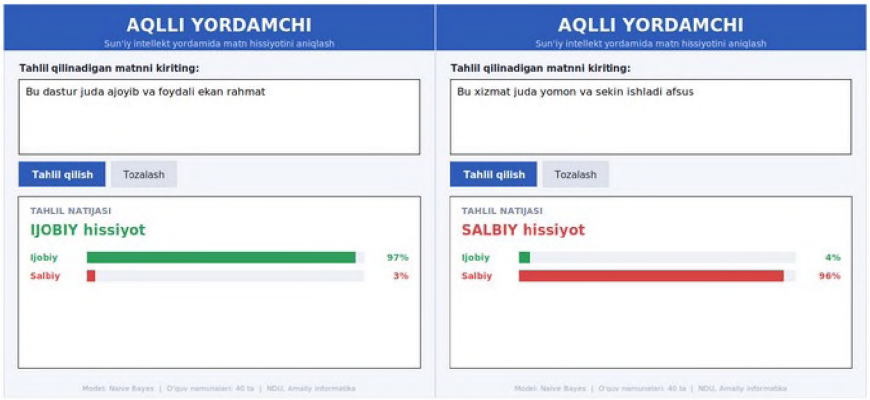
Natijalar va tahlil. Ishga tushirilgandan so‘ng dasturning asosiy oynasi ochiladi (2-rasm). Oyna tarkibida matn kiritish maydoni, “Tahlil qilish” va “Tozalash” boshqaruv tugmalari hamda natijani aks ettiruvchi panel mavjud.



2-rasm. Dasturning asosiy oynasi.

Ishlab chiqilgan dasturiy vositaning asosiy imkoniyatlari quyidagilardan iborat: matnni real vaqt rejimida tahlil qilish; natijani nafaqat sinf nomi (“IJOBIY” yoki “SALBIY”), balki har bir sinf uchun foizli ehtimollik va ko‘rgazmali diagramma ko‘rinishida taqdim etish; internet aloqasi va tashqi kutubxonalarsiz, to‘liq avtonom ishlash; kichik hajm va past resurs talabi; mustaqil .exe fayl sifatida boshqa kompyuterlarga ko‘chirib ishlatish; shuningdek, o‘quv to‘plamini kengaytirish orqali tasnif sifatini oshirish imkoniyati.

Dasturning tahlil natijalari 3-rasmda namoyish etilgan. Ijobiy shakldagi matn kiritilganda model uni 97% ishonch bilan ijobiy deb baholadi (3-rasm, a), salbiy matnda esa salbiy sinf ulushi 96%ni tashkil etdi (3-rasm, b).



a) b)

3-rasm. Tahlil natijalari: a) ijobiy matn; b) salbiy matn.

Dasturning ishlash sifatini baholash maqsadida o‘qitishda ishtirok etmagan bir qator nazorat gaplari tahlil qilindi. Natijalar 1-jadvalda keltirilgan.

1-jadval

Kiritilgan matn	Aniqlangan sinf	Ijobiy, %	Salbiy, %
Mahsulot sifatli va arzon ekan, rahmat	ijobiy	86,0	14,0
Xizmat juda yomon va sekin edi	salbiy	3,2	96,8
Yangi interfeys qulay va zamonaviy	ijobiy	92,5	7,5
Natija past chiqdi, afsuslandim	salbiy	10,2	89,8
Ovqat mazali va issiq edi	ijobiy	60,5	39,5
Dastur qotib qoldi va xato berdi	salbiy	14,6	85,4

Jadvaldagi natijalardan ko‘rinib turibdiki, model matndagi kalit so‘zlarga (“sifatli”, “qulay”, “yomon”, “xato” kabi) tayanib, hissiy bo‘yoqni ishonchli aniqlaydi. Hissiy jihatdan aniq ifodalangan matnlarda ishonch darajasi 85–96%ni tashkil etadi, neytralga yaqin yoki lug‘atda kam uchragan so‘zli matnlarda esa bu ko‘rsatkich pasayadi (masalan, beshinchi qatordagi 60,5%). Bu holat o‘quv to‘plami hajmining cheklanganligi bilan izohlanadi: korpus kengaytirilgan sari tasniflash sifati ortib boradi.

Xulosa. Ishda Naive Bayes algoritmi asosida o‘zbek tilidagi qisqa matnlarning hissiy bo‘yog‘ini avtomatik aniqlovchi dasturiy vosita ishlab chiqildi. Dasturning sodda va tushunarli tuzilishi, tashqi kutubxonalarga bog‘liq emasligi hamda mustaqil .exe fayl sifatida tarqatilishi uni nafaqat amaliy masalalarda, balki o‘quv jarayonida – talabalarga sun’iy intellekt va tabiiy tilni qayta ishlash asoslarini o‘rgatishda ko‘rgazmali vosita sifatida foydalanish imkonini beradi.

Tadqiqotning kelgusi bosqichlarida olingan natijalarni transformer asosidagi zamonaviy modellar bilan taqqoslash rejalashtirilgan.

Foydalanilgan adabiyotlar ro‘yxati:

(1). Pang B., Lee L., Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of EMNLP. Philadelphia: ACL, 2002, p. 79–86.

(2). Pang B., Lee L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval. 2008, vol. 2, № 1–2, p. 1–135.

(3). Dave K., Lawrence S., Pennock D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th International Conference on World Wide Web (WWW). 2003, p. 519–528.

(4). Rish I. An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. 2001, vol. 3, № 22, p. 41–46.

(5). McCallum A., Nigam K. A comparison of event models for naive Bayes text classification. AAAI-98 Workshop on Learning for Text Categorization. 1998, p. 41–48.

(6). Kuriyozov E., Matlatipov S. Building a new sentiment analysis dataset for the Uzbek language and creating baseline models. Proceedings (MDPI). 2019, vol. 21, № 1, p. 37.

(7). Matlatipov S., Rahimboeva H., Rajabov J., Kuriyozov E. Uzbek sentiment analysis based on local restaurant reviews. CEUR Workshop Proceedings. 2022, vol. 3315, p. 126–136.

(8). Rabbimov I., Mporas I., Simaki V., Kobilov S. Investigating the effect of emoji in opinion classification of Uzbek movie review comments. Speech and Computer (SPECOM 2020). Cham: Springer, 2020, p. 435–445.

(9). Jurafsky D., Martin J.H. Speech and Language Processing. 3rd ed. Pearson, 2023, p. 1024.

(10). Bird S., Klein E., Loper E. Natural Language Processing with Python. Sebastopol: O‘Reilly Media, 2009, p. 504.