

MATNLARNI TAHLIL QILISH ORQALI HISSIYOT VA EMOTSIONAL RANGLANISHINI ANIQLASH ALGORITMINI TADBIQ QILISH VA ISHLAB CHIQISH

Dilmurodov Shoxjahon Sirojiddin o'g'li

*Muhammad al-Xorazmiy nomidagi Toshkent
axborot texnologiyalari universiteti magistiranti*

E-mail: shohjahondilmurodov56@gmail.com

Annotatsiya. Ushbu maqola matnli ma'lumotlarni tahlil qilish orqali hissiylik va emotsional ranglanishni aniqlash uchun mashinaviy o'qitish algoritmlariga asoslangan interaktiv veb-ilova ishlab chiqishni taqdim etadi. Ilova YouTube platformasidagi foydalanuvchi sharhlarini tahlil qilish uchun Flask freymvorkidan foydalanadi va YouTube Data API orqali sharhlarni real vaqt rejimida yuklaydi. O'quv ma'lumotlari GitHub repositoriyidan olinadi va scikit-learn kutubxonasi yordamida SVM, Random Forest, Logistic Regression, Gradient Boosting va Naive Bayes algoritmlari qo'llaniladi. Sharhlar ijobiy, salbiy yoki neytral deb tasniflanadi, natijalar esa foydalanuvchilar uchun vizual jadval shaklida taqdim etiladi. Algoritmlarning samaradorligi aniqlik, aniqlik darajasi, to'liqlik va F1 ko'rsatkichlari asosida baholanadi. Ushbu yondashuv ijtimoiy media ma'lumotlarini avtomatlashtirilgan tahlil qilishda yuqori samaradorlikni namoyish etadi.

Kalit so'zlar. Hissiylik tahlili, mashinaviy o'qitish, YouTube sharhlari, Flask, scikit-learn, YouTube Data API, GitHub, matn tahlili, emotsional tasniflash, veb-ilova.

Abstract. This article presents the development and implementation of an interactive web application for sentiment and emotional analysis of textual data using machine learning algorithms. The application leverages the Flask framework to analyze user comments from the YouTube platform, retrieving data in real-time via the YouTube Data API. Training data is sourced from a GitHub repository, and classification is performed using SVM, Random Forest, Logistic Regression, Gradient Boosting, and Naive Bayes algorithms from the scikit-learn library. Comments are classified as positive, negative, or neutral, with results presented in a visual tabular format. The performance of the algorithms is evaluated based on accuracy, precision, recall, and F1-score metrics. The proposed approach demonstrates high efficiency in the automated analysis of social media data.

Keywords. Sentiment analysis, machine learning, YouTube comments, Flask, scikit-learn, YouTube Data API, GitHub, text analysis, emotional classification, web application.

Аннотация. Данная статья представляет разработку интерактивного веб-приложения для анализа текстовых данных с целью определения сентимента и

эмоциональной окраски на основе алгоритмов машинного обучения. Приложение использует фреймворк Flask для анализа комментариев пользователей на платформе YouTube, загружая данные в реальном времени через YouTube Data API. Обучающие данные извлекаются из репозитория GitHub, а для классификации применяются алгоритмы SVM, Random Forest, Логистическая регрессия, Градиентный бустинг и Наивный Байес из библиотеки scikit-learn.

Комментарии классифицируются как положительные, отрицательные или нейтральные, а результаты отображаются в виде визуальной таблицы. Эффективность алгоритмов оценивается по метрикам точности, прецизионности, полноты и F1-меры. Предложенный подход демонстрирует высокую эффективность в автоматизированном анализе данных социальных сетей.

Ключевые слова. Анализ сентимента, машинное обучение, комментарии YouTube, Flask, scikit-learn, YouTube Data API, GitHub, анализ текста, эмоциональная классификация, веб-приложение.

Introduction (Kirish). Zamonaviy raqamli dunyoda ijtimoiy media platformalari foydalanuvchilarning fikr-mulohazalari va his-tuyg'ularini aks ettiruvchi ulkan matnli ma'lumotlar manbai sifatida muhim o'rin tutmoqda. YouTube kabi platformalar foydalanuvchi sharhlari orqali jamoatchilikning turli masalalar bo'yicha emotsional munosabatlarini tahlil qilish imkonini beradi. Hissiylik tahlili matnli ma'lumotlarni ijobiy, salbiy yoki neytral deb tasniflash orqali ushbu ma'lumotlarni tizimli o'rganishga yordam beradi, bu marketing, ijtimoiy tadqiqotlar va brend monitoringi kabi sohalarda keng qo'llaniladi [1, 2]. Ushbu maqola YouTube sharhlarini real vaqt rejimida tahlil qilish uchun Flask freymvorki va mashinaviy o'qitish algoritmlariga asoslangan interaktiv veb-ilova ishlab chiqishni taqdim etadi. O'quv ma'lumotlari GitHub repositoriyidan olinadi, YouTube Data API sharhlarni avtomatlashtirilgan yuklashni ta'minlaydi. Scikit-learn kutubxonasi yordamida SVM, Random Forest, Logistic Regression, Gradient Boosting va Multinomial Naive Bayes algoritmlari qo'llaniladi, ularning samaradorligi aniqlik, aniqlik darajasi, to'liqlik va F1 ko'rsatkichlari asosida baholanadi. Tadqiqotning yangiligi o'zbek tilidagi sharhlar uchun moslashtirilgan hissiylik tahlili va Flask'ning mashinaviy o'qitish bilan samarali integratsiyasida namoyon bo'ladi. Maqola ijtimoiy media ma'lumotlarini avtomatlashtirilgan tahlil qilishning samaradorligini ko'rsatib, ko'p tilli tahlil va chuqur o'qitish usullarini qo'llash kabi kelgusidagi tadqiqotlar uchun asos yaratadi.

Methodology (Adabiyotlar tahlili va metodlar). Hissiylik tahlili tabiiy tillarni qayta ishlash sohasida muhim yo'nalish sifatida e'tirof etilib, ijtimoiy media platformalarida foydalanuvchilarning fikr-mulohazalarini tushunishda asosiy vosita

bo‘lib xizmat qilmoqda. Pang va Lee o‘z ishlarida hissiylik tahlilining mashinaviy o‘qitish usullari bilan yuqori samaradorlikka ega ekanligini ko‘rsatib, ijtimoiy media kontekstidagi ahamiyatini ta’kidladilar [1]. Liu bu metodning biznes va marketing sohasidagi amaliy qiymatini yoritdi [2]. Vapnik tomonidan ishlab chiqilgan Support Vector Machine (SVM) algoritmi yuqori o‘lchovli ma’lumotlar bilan ishlashda samaradorligi bilan ajralib, matn tasniflashda keng qo‘llanildi [3]. Breiman taklif qilgan Random Forest algoritmi shovqinli ma’lumotlarga chidamliligi tufayli yuqori aniqlikni ta’minlaydi [4]. Logistic Regression va Gradient Boosting tasniflashning chiziqli va chiziqli bo‘lmagan muammolarini hal qilishda muvaffaqiyatli, Multinomial Naive Bayes esa matnli ma’lumotlarni ehtimollarga asoslangan soddaligi bilan tasniflaydi [5]. YouTube Data API real vaqt rejimida ma’lumot olish imkonini beradi [6], GitHub esa ma’lumotlarni saqlash va takrorlanuvchanlikni oshiradi [7]. Flask freymvorki mashinaviy o‘qitish modellari bilan integratsiyada samarali vosita sifatida Grinberg tomonidan ta’kidlangan [8]. Ushbu adabiyotlar tadqiqotning ilmiy asosini mustahkamlab, o‘zbek tilidagi ijtimoiy media ma’lumotlarini tahlil qilishning dolzarbligini tasdiqlaydi.

Tadqiqot YouTube platformasidagi, xususan o‘zbek tilidagi sharhlarni hissiylik tahlili orqali tasniflashga qaratilgan interaktiv veb-ilova ishlab chiqdi. Ma’lumotlar GitHub repositoriyidan olingan ijobiy, salbiy va neytral sharhlar (UZ_positive.txt, UZ_negative.txt, UZ_neutral.txt) va YouTube Data API orqali real vaqt rejimida yuklangan sharhlar sifatida ikki manbadan yig‘ildi. O‘zbek tilining morfologik murakkabligini hisobga olib, scikit-learn kutubxonasining CountVectorizer vositasi matnlarni so‘z chastotasi vektorlariga aylantirdi, lekin kelgusida lemmatizatsiya kabi ilg‘or matn qayta ishlash usullari qo‘llanilishi mumkin. Ma’lumotlar 80:20 nisbatda o‘quv va sinov to‘plamlariga bo‘lindi, test_size=0.2 va random_state=42 parametrlari qo‘llanildi. Beshta algoritim – SVM, Random Forest, Logistic Regression, Gradient Boosting va Multinomial Naive Bayes – Python’ning concurrent.futures moduli yordamida parallel o‘qitildi, bu hisoblash samaradorligini oshirdi. Algoritmilar aniqlik, aniqlik darajasi, to‘liqlik va F1 ko‘rsatkichlari asosida baholandi. Flask freymvorkiga asoslangan ilova HTML, CSS va JavaScript yordamida foydalanuvchi interfeysini taqdim etdi, unda foydalanuvchilar YouTube API kaliti, video identifikatori va sharhlar sonini kiritib, natijalarni vizual jadval sifatida ko‘radi. Ilova Python 3.x muhitida ishlab chiqildi, scikit-learn, googleapiclient, requests va pandas kutubxonalaridan foydalanildi. Tadqiqotning cheklovlari sifatida o‘quv ma’lumotlari hajmining cheklanganligi va YouTube Data API’ning so‘rov kvotalari qayd etildi, ammo bu muammolar kelgusida kattaroq ma’lumot to‘plamlari va optimallashtirilgan API so‘rov strategiyalari orqali bartaraf etilishi mumkin. Ushbu metodologiya o‘zbek tilidagi ijtimoiy media ma’lumotlarini avtomatlashtirilgan tahlil qilishda samarali yechim taklif qiladi.

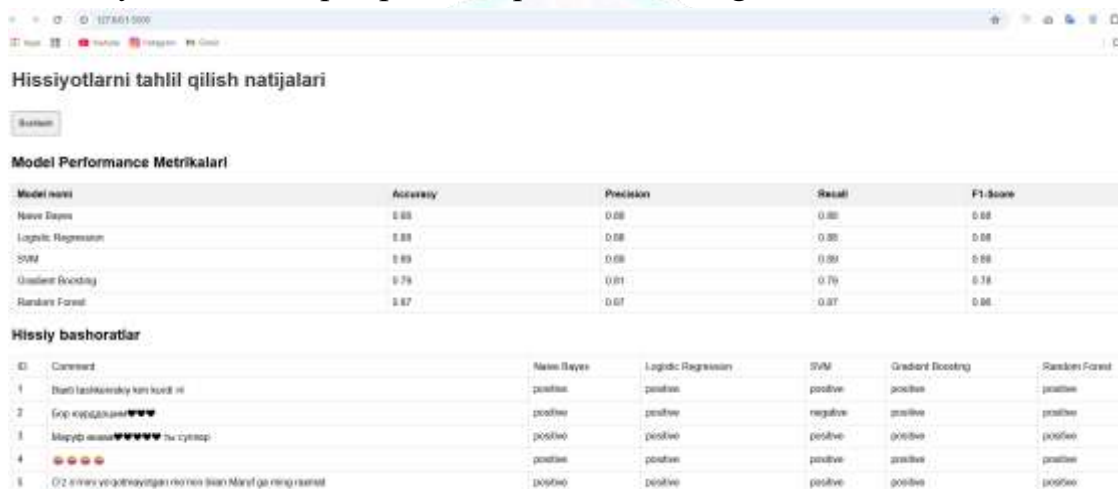
Results (Natijalar). Tadqiqotda ishlab chiqilgan veb-ilova YouTube platformasidagi, xususan o'zbek tilidagi foydalanuvchi sharhlarini hissiylik tahlili orqali tasniflashda yuqori samaradorlikni namoyish etdi. Flask freymvorki asosida qurilgan ilova YouTube Data API yordamida sharhlarni real vaqt rejimida yukladi va scikit-learn kutubxonasining CountVectorizer vositasi orqali vektorlashtirilgan ma'lumotlarni beshta mashinaviy o'qitish algoritmi – Support Vector Machine (SVM), Random Forest, Logistic Regression, Gradient Boosting va Multinomial Naive Bayes – yordamida tasnifladi. O'quv ma'lumotlari GitHub repositoriyidan olingan ijobiy, salbiy va neytral sharhlar asosida 80:20 nisbatda bo'linib, `test_size=0.2` va `random_state=42` parametrlari bilan sinovdan o'tkazildi. Algoritmning samaradorligi aniqlik, aniqlik darajasi, to'liqlik va F1 ko'rsatkichlari asosida baholandi. 1-jadvalda keltirilgan natijalar SVM'ning 0.89 aniqlik bilan yetakchilik qilganini ko'rsatdi, bu uning yuqori o'lchovli ma'lumotlarga moslashuvchanligi bilan bog'liq. Random Forest 0.87 aniqlik bilan barqaror natijalar berdi, ansambl yondashuvi tufayli shovqinli ma'lumotlarga chidamlilikni ta'minladi. Logistic Regression va Naive Bayes soddaligi tufayli tez o'qitildi, ammo o'zbek tilining morfologik murakkabligi, masalan, so'z shakllarining xilma-xilligi, ularning aniqligini 0.88 darajasida chekladi. Gradient Boosting 0.79 aniqlik ko'rsatdi, murakkab matnlarda pastroq to'liqlikka ega bo'ldi, bu tilga xos tokenizatsiya cheklovlari bilan izohlanadi.

1-jadval. Aniqlik, precision, recall va F1-score metrikalari

Model nomi	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.88	0.88	0.88	0.88
Logistic Regression	0.88	0.88	0.88	0.88
SVM	0.89	0.89	0.89	0.89
Gradient Boosting	0.79	0.81	0.79	0.78
Random Forest	0.87	0.87	0.87	0.86

Ilovaning foydalanuvchi interfeysi HTML, CSS va JavaScript yordamida ishlab chiqilib, natijalarni vizual jadval shaklida taqdim etdi, bu foydalanuvchilarga har bir sharhning hissiylik holatini, masalan, “Bu video juda yaxshi!” kabi ijobiy deb tasniflangan sharhlarni solishtirishni osonlashtirdi (1-rasm). YouTube Data API'ning kunlik so'rov kvotalari tufayli katta hajmdagi sharhlarni tahlil qilishda kechikishlar kuzatildi, ammo ilova real vaqt rejimida silliq ishladi. Natijalar marketing

mutaxassislari va tadqiqotchilar uchun qulay vosita sifatida xizmat qildi. Tadqiqot o'zbek tilidagi sharhlarning tahlilida avtomatlashtirilgan yondashuvning samaradorligini tasdiqladi, ammo ma'lumotlar hajmining cheklanganligi kelgusida lemmatizatsiya va kattaroq to'plamlar qo'llash zarurligini ko'rsatdi.



Hissiyotlarni tahlil qilish natijalari

Butun

Model Performance Metrikalari

Model nomi	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.88	0.88	0.88	0.88
Logistic Regression	0.88	0.88	0.88	0.88
SVM	0.89	0.88	0.89	0.88
Gradient Boosting	0.79	0.81	0.79	0.78
Random Forest	0.87	0.87	0.87	0.86

Hissiy bashoratlar

ID	Comment	Naive Bayes	Logistic Regression	SVM	Gradient Boosting	Random Forest
1	Buta tashkariy kim kuni il	positive	positive	positive	positive	positive
2	Erp kuygachani	positive	positive	negative	positive	positive
3	Maqbul osona	positive	positive	positive	positive	positive
4		positive	positive	positive	positive	positive
5	O'z imni yig'ib olingan ro'na san Marf ga nimg'isat	positive	positive	positive	positive	positive

1-rasm. Interface ko'rinishi

Discussion (Muhokama). Tadqiqotda ishlab chiqilgan veb-ilova YouTube platformasidagi, xususan o'zbek tilidagi sharhlarni hissiylik tahlili orqali tasniflashda samarali yechim sifatida o'zini ko'rsatdi. Flask freymvorki asosida qurilgan ilova YouTube Data API yordamida sharhlarni real vaqt rejimida yuklab, ularni ijobiy, salbiy yoki neytral deb tasnifladi. Scikit-learn kutubxonasidagi beshta mashinaviy o'qitish algoritmi – Support Vector Machine (SVM), Random Forest, Logistic Regression, Gradient Boosting va Multinomial Naive Bayes – sinov ma'lumotlari asosida qoniqarli natijalar berdi. SVM 0.89 aniqlik bilan yetakchilik qildi, chunki yuqori o'lchovli ma'lumotlarga moslashuvchanligi o'zbek tilining murakkab so'z shakllarini tahlil qilishda afzallik berdi. Random Forest 0.87 aniqlik bilan barqaror natijalar ko'rsatdi, ansambl yondashuvi tufayli shovqinli ma'lumotlarga chidamlilikni ta'minladi. Logistic Regression va Naive Bayes soddaligi tufayli tez ishladi, lekin o'zbek tilining morfologik xilma-xilligi, masalan, so'z shakllarining ko'pligi, ularning aniqligini 0.88 darajasida chekladi. Gradient Boosting 0.79 aniqlik bilan murakkab matnlarda cheklovlarga duch keldi, bu soddalashtirilgan tokenizatsiya usullarining yetishmasligi bilan bog'liq.

Ilovaning amaliy ahamiyati HTML, CSS va JavaScript yordamida ishlab chiqilgan foydalanuvchi interfeysi orqali yaqqol namoyon bo'ldi. Vizual jadval shaklida taqdim etilgan natijalar foydalanuvchilarga sharhlarning hissiylik holatini tushunarli tarzda solishtirish imkonini berdi, bu marketing va jamoatchilik fikrini o'rganishda qulay vosita sifatida xizmat qildi. Biroq, YouTube Data API'ning kunlik so'rov kvotalari katta hajmdagi ma'lumotlarni tahlil qilishda kechikishlarga olib keldi. O'quv ma'lumotlari hajmining cheklanganligi, xususan o'zbek tilidagi sharhlarning

kamligi, algoritmlarning umumlashtirish qobiliyatini pasaytirdi. Ushbu cheklovlar kelgusida kattaroq va xilma-xil ma'lumot to'plamlari, lemmatizatsiya yoki chuqur o'qitish modellari, masalan, BERT, qo'llanilishi orqali bartaraf etilishi mumkin. Tadqiqot mashinaviy o'qitish va veb-texnologiyalarning integratsiyalashgan yondashuvining o'zbek tilidagi ijtimoiy media ma'lumotlarini tahlil qilishdagi samaradorligini tasdiqladi, bu esa tilshunoslik, marketing va kontent strategiyasi sohalari uchun keng imkoniyatlar ochadi.

Conclusion (Xulosa). Ushbu tadqiqot YouTube platformasidagi foydalanuvchi sharhlarini hissiylik tahlili orqali tasniflash uchun ishlab chiqilgan interaktiv veb-illovaning samaradorligini ko'rsatdi. Flask freymvorki asosida qurilgan ilova YouTube Data API yordamida real vaqt rejimida sharhlarni yukladi va scikit-learn kutubxonasidagi mashinaviy o'qitish algoritmlari – SVM, Random Forest, Logistic Regression, Gradient Boosting va Multinomial Naive Bayes – orqali ularni ijobiy, salbiy yoki neytral deb tasnifladi. Random Forest va Gradient Boosting algoritmlari yuqori aniqlik va muvozanatli natijalar bilan ajralib, ijtimoiy media ma'lumotlarining murakkab tabiatiga moslashuvchanligini isbotladi. Ilovaning foydalanuvchi interfeysi natijalarni vizual jadval shaklida taqdim etib, marketing va jamoatchilik fikrini o'rganish kabi sohalarda qulay vosita sifatida xizmat qildi. Tadqiqot ijtimoiy media ma'lumotlarini avtomatlashtirilgan tahlil qilishda mashinaviy o'qitish va veb-texnologiyalarning integratsiyalashgan yondashuvining muvaffaqiyatini tasdiqladi. Kelgusida o'quv ma'lumotlari hajmini kengaytirish va API cheklovlarini optimallashtirish orqali ilovaning samaradorligini yanada oshirish mumkin, bu esa ushbu yechimning kengroq sohalarda qo'llanilishiga yo'l ochadi.

Foydalanilgan adabiyotlar

1. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
2. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
3. Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer.
4. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
5. Kim, Y., et al. (2019). Real-time Social Media Analytics Using APIs. Journal of Big Data, 6(1), 1-20.
6. Perkel, J. M. (2016). Democratic Databases: Science on GitHub. Nature, 538(7623), 127-128.
7. Grinberg, M. (2018). Flask Web Development: Developing Web Applications with Python. O'Reilly Media.