

www.jpis.az

13 (1)
2022

Maşın təlimi əsasında Twitter məlumatlarının sentiment analizi

Məkrufə Ş. Hacırəhimova¹, Mərziyə İ. İsmayılova²

^{1,2}Azərbaycan Milli Elmlər Akademiyası, İnformasiya Texnologiyaları İnstitutu, B.Vahabzadə küç., 9A, AZ1141, Bakı, Azərbaycan

¹hmakrufa@gmail.com; ²imarziya@gmail.com

MƏQALƏ HAQQINDA

<http://doi.org/10.25045/jpis.v13.i1.07>

Məqalənin tarixi:

Təqdim olunub 21 iyul 2021

Rəy formasının alınması

27 sentyabr 2021

Qəbul olunub 29 dekabr 2022

Açar sözlər:

Sentiment analiz

Twitter

Mikrobloqlar

Maşın təlimi

Naive Bayes

Neural Network

Machine Learning-based sentiment analysis of Twitter data

Keyword:

Sentiment analysis

Twitter

Mikroblogging

Machine learning

Naive Bayes

Neural Network

Анализ тональности Twitter данных на основе машинного обучения

Ключевые слова:

Анализ тональности

Twitter

Микроблоги

Машинное обучение

Naive Bayes

Neural Network

X Ü L A S Ə

Məqalədə maşın təlimi alqoritmləri əsasında COVID-19 pandemiyası ilə bağlı müxtəlif ölkələrin Twitter sosial şəbəkə istifadəçilərinin fikirləri analiz edilmişdir. Sosial şəbəkə erasının gəlişi, mikrobloq əlavələrinin və forumların sürətlə yayılması ilə bağlı sentiment analizin rolu əhəmiyyətli dərəcədə artmışdır. Sosial şəbəkələr istifadəçilərin müxtəlif mövzularda fikir və düşüncələri haqqında məlumat toplamaq üçün əsas mənbələrdir. İnsanlar fikir və düşüncələrini başqaları ilə bölüşmək üçün hər gün saatlarla sosial şəbəkələrdə vaxt keçirirlər. Twitter kimi sosial şəbəkə platformalarında müzakirə olunan mövzulardan biri də COVID-19 pandemiyasıdır. Təqdim olunan məqalədə COVID-19 pandemiyası ilə bağlı tvitlərin emosional "rənginin" (müsbət, mənfi və neytral) analiz edilməsi üçün Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Neural Network (NN) kimi maşın təlimi metodundan istifadə edilmişdir. Tədqiqatda Python proqramlaşdırma mühitində scikit-learn kitabxanasından istifadə edərək təcrübələr aparılmışdır. Eksperimentlərin aparılması üçün Kaggle saytından götürülmüş COVID-19 pandemiyası ilə əlaqəli tvit verilənlər bazasından istifadə edilmişdir. Eksperimentlər zamanı RF klassifikatoru ən yüksək nəticə göstərmişdir.

The paper analyzes the views of Twitter users on the COVID-19 corona virus pandemic based on machine learning algorithms. The role of sentiment analysis increased with the advent of the social network era and the rapid spread of microblogging applications and forums. Social networks are the main sources for gathering information about users' thoughts on various themes. People spend more time on social media to share their thoughts with others. One of the themes discussed on social networking platforms Twitter is the COVID-19 corona virus pandemic. In the paper, machine learning methods as Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Neural Network (NN) are used to analyze the emotional "color" (positive, negative, and neutral) of tweets related to the COVID-19 corona virus pandemic. The experiments are conducted in Python programming using the scikit-learn library. A tweet database related to the COVID-19 corona virus pandemic from the Kaggle website is used for experiments. The RF classifier shows the highest performance in the experiments.

С наступлением эры социальных сетей и быстрым распространением приложений и форумов для микроблогов роль анализа тональности текста значительно возросла. Социальные сети являются основными источниками сбора информации о мыслях и мнениях пользователей по различным темам. Люди ежедневно часами проводят время в социальных сетях для того, чтобы поделиться своими мыслями и мнениями с другими пользователями. Одна из тем, обсуждаемых в социальных сетях, в частности в Twitter, – это пандемия COVID-19. В статье исследовано мнение пользователей Twitter из разных стран о пандемии COVID-19 на основе алгоритмов машинного обучения. Для анализа эмоциональной «окраски» (положительный, отрицательный и нейтральный) твитов о пандемии COVID-19 использованы четыре метода машинного обучения - Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF) и Neural Network (NN). Эксперименты проведены в программной среде Python с применением библиотеки scikit-learn. В экспериментах использовался набор данных твитов, связанных с пандемией COVID-19, с веб-сайта Kaggle. Сравнительный анализ полученных результатов показал наилучшую результативность классификатора RF.

1. Giriş

Sentiment analiz (SA) sosial şəbəkə veb-saytlarının sürətli inkişafı ilə son zamanlar çox populyar bir tədqiqat mövzusunə çevrilmişdir. SA, insanların müxtəlif məsələlər (məhsullar, hadisələr, təşkilatlar və s.) üzrə fikirlərini araşdıran bir tədqiqat sahəsidir [1, 2]. Sosial şəbəkələrdə, forumlarda, bloqlarda, mikrobloqlarda aparılan geniş müzakirələrlə bağlı SA-nın rolu əhəmiyyətli dərəcədə artmışdır. Bu gün, demək olar ki, hər bir veb-səhifədə istifadəçilərin məhsul və ya xidmətlər haqqında şərhlərini yerləşdirmələri və onları Facebook və ya Twitter-də dostları ilə bölüşmələri üçün bir bölmə vardır. Bu fikirlərin toplanması insanların kollektiv davranışlarını başa düşmək üçün informasiya verir və qiymətli kommersion marağı kəsb edir. Belə ki, getdikcə daha çox artan verilənlər toplusu göstərir ki, sosial mediadakı kontenti analiz edərək COVID-19 haqqında fikirləri və ya zaman-zaman işsizlik səviyyəsini proqnozlaşdırmaq olar.

Dünyanın ən çox istifadə etdiyi Twitter mikrobloqu qısa və sadə ifadə tərzinə görə son illərdə çoxsaylı istifadəçilərin: tələbələr, peşəkarlar, məşhurlar, şirkətlər və siyasətçilərin düşüncələrini bölüşdükleri bir məkana çevrilmişdir. Başqalarının rəyi (fikri) bəşəriyyət üçün həmişə əhəmiyyət kəsb etmişdir. İstər siyasət, istərsə də sadə bir seçim, məsələn, bir marketdə hər hansı bir məhsul seçmək olsun, hər zaman başqalarının edəcəyimiz seçim barədə nə düşündüyünü fikirləşirik. Bu baxımdan Twitter qısa müddətdə rəylərə təsir göstərən ən güclü sosial media platformalarından birinə çevrilmişdir [3].

Rəy və ya fikirlərin analizi son zamanlarda Twitter kimi saytlar sayəsində təbii dilin emalı (Natural Language Processing, NLP), mətn analizi [4] və s. kimi sahələrdə ən çox araşdırılan və müzakirə olunan məsələlərdən birinə çevrilmişdir. Ancaq rəy toplamaq və ya onları analiz etmək o qədər də asan məsələ deyil. Son zamanlar mövcud informasiya mənbələrinin əlyətərliyinin artması səbəbindən bütün dünya şirkətləri real problemlərinin həlli və ya biznes məqsədlərinə çatmaq üçün verilənlərdən informasiya əldə etməyə cəhd edirlər. Generasiya olunan informasiyanın həcmi, sürəti və çeşidi görünməmiş bir həddə çatmışdır. Bu, yalnız böyük verilənlərin işlənməsi üçün Hadoop kimi yeni platformalar deyil, eyni zamanda verilənlərdən informasiya çıxarmaq üçün yeni maşın təlimi metodları və alqoritmləri də tələb edir [3].

Hazırda bütün dünya COVID-19 pandemiyası dövrünü yaşayır. Bu dövr ərzində bir çox insanlar həyatını itirmiş, bir çoxu isə bu xəstəliyi uğurla məğlub etmişdir. İlk olaraq 2019-cu ildə

koronavirus xəstəliyi olaraq tanınan COVID-19 11 Mart 2020-ci ildə Ümumdünya Səhiyyə Təşkilatı tərəfindən pandemiya elan edildi. COVID-19 pandemiyası sosial şəbəkələrdə müzakirə olunan əsas mövzuya çevrildi. Dünyanın hər yerində insanlar sosial şəbəkələrdə, o cümlədən Twitter-də COVID-19 pandemiyası haqqında öz fikirlərini ifadə etməyə başladılar. Bu toplanan tvitlər əsasında insanların xəstəliklə bağlı davranışları və əhval-ruhiyyələri təhlil edilmişdir.

Tədqiqat müxtəlif maraqlı tərəflər üçün faydalı ola bilər. Məsələn, hökumət insanların bu yeni gərginliyə necə reaksiya verdiyini, məsələn, qida çatışmazlığı, maska qıtlığı və s. kimi çətinliklərlə qarşılaşdıqlarını bildiyi üçün bu məlumatlardan siyasi məqsədlər üçün istifadə edə bilər. Müxtəlif kommersion təşkilatları sadalanan çətinliklərdən bəhs edən tvitlərdən müxtəlif fikirləri analiz edərək, zəruri məhsulların istehsalına başlaya və bununla da mənfəət əldə edə bilərlər. Qeyri-hökumət təşkilatları, insanları necə reabilitasiya etmə strategiyasına uyğun faktlar və məlumatlardan istifadə edərək qərar verə bilərlər və s. [5-9].

Twitter məlumatlarının SA sahəsində bir çox tədqiqatlarda müxtəlif yanaşmalar təqdim edilmişdir. Təklif olunan yanaşmalarda tvitlərdəki rəy sözlərini aşkarlamaq üçün leksikonəsaslı, qaydalara əsaslanan, hibrid yanaşmalar və maşın təlimi alqoritmləri istifadə olunmuşdur [10-13]. Leksikonəsaslı yanaşmanın çatışmazlığı ondan ibarətdir ki, lüğətə daxil edilən rəy sözlərinin sentiment qütbü predmet sahəsindən və kontekstdən asılıdır. Bu da hər bir sahənin öz lüğətinin yaradılmasını tələb edir. Qaydalara əsaslanan yanaşmanın çatışmazlığı ondadır ki, emosionallığın analizində yüksək dəqiqliyə nail olmaq üçün çoxlu sayda qaydalar təsvir etmək tələb olunur ki, bu da uzun və mürəkkəb bir prosesdir.

Məqalədə əsas məqsəd maşın təlimi alqoritmlərindən istifadə edərək COVID-19 pandemiyası ilə bağlı müxtəlif ölkələrin Twitter istifadəçilərinin məlumatları əsasında SA aparmaqdır.

Məqalənin sonrakı hissəsi aşağıdakı kimi təşkil edilmişdir. İkinci bölmədə əlaqəli işlər icmal olunmuşdur. Üçüncü bölmədə istifadə olunan klassifikatorlar təqdim olunmuşdur. Dördüncü bölmədə təklif olunan yanaşmanın metodologiyası təsvir edilmişdir. Beşinci bölmədə təklif edilmiş yanaşma eksperimental olaraq yoxlanılmış və ekperimentin nəticələri müzakirə olunmuşdur. Altıncı bölmədə tədqiqat üzrə nəticələr ümumiləşdirilmişdir. Sonuncu bölmədə istinad edilmiş ədəbiyyatların siyahısı təqdim olunmuşdur.

2. Əlaqəli tədqiqatlar

Son dövrlər Twitter mikrobloqunda müzakirə olunan müxtəlif mövzuların SA sahəsində çoxlu sayda tədqiqatlar aparılmışdır.

[11]-də Twitter məlumatlarında tvitlərdəki rəy sözlərinin semantik istiqamətini təyin etmək üçün həm korpusəsaslı, həm də lüğətəsaslı metodlardan istifadə edilərək yeni bir hibrid yanaşma təklif edilmişdir. Sentimenti aşkarlamaq üçün tvitlərdə olan rəy sözləri (feyllər və zərflərlə birlikdə sifətlərin birləşməsi) çıxarılır. Sifətlərin semantik istiqamətini tapmaq üçün korpusəsaslı metoddan, feil və zərflərin semantik istiqamətini tapmaq üçün isə lüğətəsaslı metoddan istifadə edilmişdir.

Başqa bir araşdırmada SA üçün iki yanaşma təklif edilmişdir: leksikonəsaslı və maşın təlimi metodu. Bu yanaşmaları həyata keçirmək üçün bir neçə üsul təsvir edilmiş və Twitter mesajlarının sentiment klassifikasiyası üçün onların necə istifadə edilə biləcəyi göstərilmişdir. Məqalədə müxtəlif lüğət birləşmələrinin müqayisəli analizi aparılmışdır. Müəlliflər göstərmişlər ki, sentiment leksiconları emosya ikonları (emojiconlar), abbreviaturalar və jarqon ifadələrlə gücləndirməklə Twitter üçün leksikonəsaslı klassifikasiyanın dəqiqliyini artırmaq olar. Sentiment klassifikasiyasında əlamət yaratma və əlamət seçmə prosesinin əhəmiyyəti analiz edilmişdir. SA metodlarının effektivliyini qiymətləndirmək üçün alqoritmlər SemEval 2013 (Semantic Evaluation Workshop) müsabiqəsinin Twitter verilənləri üzərində test edilmişdir. Nəticələr SVM və Naive Bayes klassifikatorlarına əsaslanan maşın təlimi metodlarının leksikon metodundan üstün olduğunu göstərmişdir [12].

[14]-də müəlliflər tərəfindən Twitter verilənlərinin SA üçün Sentimentor adlı veb-alət təqdim edilmişdir. Alət real zamanda Twitter verilənlərini pozitiv, neqativ və neytral olaraq klassifikasiya etmək üçün Naive Bayes klassifikatorundan istifadə edir.

[15]-də müəlliflər seçkilərdə sentimentin tədqiq edilməsi üçün ən dəqiq yanaşmanı müəyyən etməkdən ötrü maşın təlimi alqoritmləri ilə müxtəlif sentiment analizatorların tətbiqinə baxmışlar. Leksikonəsaslı SA-da semantik istiqamət mətndə hesablanmış sözlər, ifadələr və ya cümlələrlə ifadə olunur. Leksikonəsaslı metoddakı qütüb, sözün semantik qiymətindən ibarət olan lüğət əsasında hesablanır. Mətni təsnif etmək üçün Naive Bayes və SVM alqoritmləri tətbiq edilmişdir. İstər sentiment leksikonlardan, istərsə də maşın təlimi yanaşmalarından istifadə edilərək SA sahəsində xeyli iş görülmüşdür. Lakin, bu tədqiqat W-WSD (Word Sequence Disambiguation), SentiWordNet, TextBlob

kimi üç sentiment lüğət arasında müqayisə aparmağa yönəldilmişdir. Bu üç sentiment analizator iki maşın təlimi alqoritmindən istifadə edilərək sınaqdan keçirilmişdir. Nəticədə rəylər üç analizatorla hesablanmışdır. Bundan başqa, nəticələr Naive Bayes və SVM kimi iki nəzarət edilən maşın təlimi klassifikatoru ilə yoxlanılmışdır. TextBlob lüğəti əsasında nəticələr nisbətən daha yaxşı olsa da, tvitləri analiz edərək ən yaxşı nəticə W-WSD ilə əldə edilmişdir.

Twitter-də istehzanın aşkarlanması və analizi son tendensiyalara və hadisələrə dair ictimai rəy barədə fikirlər verir. İstehzanın aşkarlanmasının köməyi ilə şirkətlər istifadəçilərin məhsullarına dair fikirlərini analiz edə bilirlər. Məlumatların ilkin emalı üçün TextBlob istifadə olunur. TextBlob NLTC-da (Natural Language Text Classification) quraşdırılmış bir paketdir. TextBlob qütüb və subyektivliyə inamı tapmaq üçün istifadə olunur. Tvitlərin qütübünü və subyektivliyini müəyyən etmək üçün RapidMiner istifadə olunur. SVM klassifikatoru və Naive Bayes klassifikatorunun göstərdiyi dəqiqliyi hesablamaq üçün Weka platformasından istifadə olunmuşdur. Cəmi 2250 tvit istifadə edilmişdir. Naive Bayes klassifikatorunun dəqiqliyi 65.2%, SVM-in dəqiqliyi isə 60.1% olmuşdur. Bu da Naive Bayes klassifikatorunun SVM klassifikatoru ilə müqayisədə daha çox dəqiqlik təmin etdiyini göstərir [16].

SA və ya fikirlərin əldə edilməsi son zamanlar bir çox tədqiqatçıların diqqət mərkəzinə çevrilmişdir. Belə ki, onlayn mətnlərin analizi bazarın araşdırılması, psixoloji və sosioloji nöqtəyi-nəzərdən elmi araşdırmalar, siyasi anketlər, biznes analitika, onlayn alış-veriş infrastrukturunu yaxşılaşdırmaq və s. üçün faydalıdır [17]. Twitter mesajlarının SA üçün yeni bir həll olan hibrid metod təqdim olunmuşdur. Metod, xətti SVM klassifikatorunun hazırlanması üçün yeni xüsusiyyətlər dəsti yaratmaq məqsədilə sentiment lüğətdən (SentiString) istifadə edir. Bu metodun unigram səviyyəsindən üstün olduğu göstərilmişdir.

Twitter insanların baxışlarını, fikirlərini və duyğularını sərbəst şəkildə ifadə etdikləri ən populyar sosial şəbəkə saytlarından biridir. Bu tvitlər Hindistanın Uri şəhəri yaxınlığında təhlükəsizlik qüvvələrinə edilən terror hücumu (Uri hücumu) ilə əlaqəli emosiyaları aşkarlamaq üçün analiz edilmişdir. Tvitlərdə emosiyaların və qütübün müəyyən olunması üçün intellektual mətn analizi üsullarından istifadə olunmuşdur. Tez-tez rast gəlinən sözlərdən verilənlər çoxluğu yaratmaq üçün təxminən 5000 tvit yenidən kodlaşdırılmış və ilkin emal edilmişdir. Emosiyaları və qütbləri əldə etmək üçün R proqramlaşdırma dilindən istifadə olunmuşdur. Eksperimentlərin nəticəsi insanların 94,3%-nin Uri hücumuna nifrət etdiyini göstər-

mişdir [18].

E.Evirgen öz tədqiqatında emosiyaları qısa və toplanmış bir formatda əks etdirən tvitlərin təhlili üçün bir şlüz (gateway) rolunu oynayan R proqramlaşdırma dilində struktur təklif etmişdir [3]. Hədəf tvitlərə emosiyaların qısa təsviri və səhv format və ya qrammatik quruluşda istifadə olunan sözlər daxildir. İşdə beş müxtəlif maşın təlimi metodları: SVM, Random Forest, Boosting, Maksimum Entropiya və Artificial Neural Networks müqayisə edilmişdir.

[19]-də məhsulun xüsusiyyətləri ilə bağlı rəyləri çıxarmaq üçün SA metodlarından istifadə edərək etibarlılıq modeli təklif etmişlər. Bu model, müştərilər üçün etibarlı məhsul xüsusiyyətləri haqqında faydalı məlumatlar təqdim edir. Model istifadəçilərin təqdim etdiyi mətn rəylərindən istifadə edir. Daha sonra hər bir məhsulun xüsusiyyəti və məhsulun özü üçün daha real bir etibarlılıq dəyəri yaratmaq üçün metod təqdim edilir. Məhsulun etibarlılığı hesablandıqda, onun bütün xüsusiyyətlərinə bərabər yanaşılır, çünki, bəzi xüsusiyyətlər müştərilər üçün digərlərindən daha vacibdir və nəticədə müştərilərin qərarlarına daha çox təsir göstərə bilər.

[20]-də sentiment klassifikasiyanın səmərəliliyini və dəqiqliyini yaxşılaşdırmaq məqsədi ilə baza təlimi klassifikatorlarını birləşdirərək ansambl klassifikatoru təklif etmişdir. Təklif olunan ansambl klassifikatoru ayrı-ayrı klassifikatordan daha yaxşı performans göstərmişdir.

[21]-də Twitter-dən götürülmüş rəyləri təsnif etmək üçün SA-nın köməyi ilə maşın təlimi metodlarından istifadə olunur. Əsas məqsəd multi sinif təsnifatı ilə çox sayda rəyləri emal etməkdir. Multi sinif SVM alqoritmi müsbət, mənfi, xoşbəxt, sevgi və əyləncə kimi müxtəlif sinifləri təsnif edə bilər. Təklif olunan alqoritm, *F*-ölçünün (*F*-measure) əldə olunmuş qiyməti baxımından Naive Bayes və SVM alqoritmindən daha yaxşı nəticələr vermişdir.

[22]-də müəlliflər qütblülük leksikonuna (Polarity Lexicon, PL) və süni intellektə (Artificial Intelligence, AI) əsaslanan iki üsul tətbiq edərək, türk tvitlərindəki sentimentini proqnozlaşdırmaq üçün mexanizm təqdim etmişlər. PL metodu sözlərdən ibarət lüğət təqdim edir və tvitlər onlarla uyğunlaşdırılır. Tvitlər lüğətdəki sözlərlə uyğunlaşdıqdan sonra əldə edilən nəticəyə əsasən pozitiv, neqativ və ya neytral olaraq təsnif edilir. AI metodu tvitləri pozitiv, neqativ və ya neytral təsnif etmək üçün SVM və RF klassifikatorlarından istifadə edir. Eksperimentlər nəticəsində SVM qeyd olunmuş verilənlər üzərində 76% dəqiqlik göstərmişdir. RF isə xam verilənlər üzərində 88% dəqiqliklə daha yaxşı performans göstərmişdir. PL meto-

dunun performansı xam verilənlər baza veriləni ilə əvəz edildikdə davamlı olaraq 45%-dən 57%-ə qədər yüksəlmişdir.

Baxdığımız tədqiqatlardan da görüldüyü kimi, sosial şəbəkə məlumatlarının, xüsusilə də Twitter məlumatlarının SA sayəsində tədqiqatların sayı durmadan artmaqdadır [5, 6, 8, 9]. Son dövrlərdə sosial şəbəkələrdə müzakirə olunan əsas mövzularlardan biri də COVID-19 pandemiyasıdır. [5]-də COVID-19 pandemiyası ilə əlaqədar rəylər təhlil edilmişdir. İnsanların bu xəstəliklə bağlı fikirlərini analiz etmək üçün Twitter API istifadə edərək koronavirus ilə əlaqəli tvitlər əldə edilir və bu tvitlər maşın təlimi metodları və vasitələrindən istifadə edilərək pozitiv, neqativ və neytral olaraq analiz edilir. Python proqramlaşdırma mühitində tvitlərin ilkin emalı üçün istifadə olunan NLTC kitabxanasından istifadə etməklə müxtəlif tvitlər üzrə eksperimentlər aparılmışdır. Daha sonra tvit verilənlər dəsti Textblob paketi vasitəsilə analiz edilmiş və pozitiv, neqativ və neytral rəylər müxtəlif təsvirlərlə təqdim olunmuşdur.

[6]-də COVID-19 haqqında 5977653 fərddən toplanan 19298967 tvitdən ibarət verilənlər dəsti təqdim etmişdir. Bu tvitlər 'koronavirus', 'kovid' və 'maska' sorğu terminlərindən istifadə edilərək 2020-ci ilin mart-iyul ayları üçün toplanmışdır. Müəlliflər COVID-19 ilə əlaqəli toplanmış tvitləri və tvitlərin coğrafi məkanını təsvir etmək üçün mövzu modelləşdirmə, SA və təsviri statistikadan istifadə etmişlər.

[7]-də heç bir tibbi qeydlər olmadan Twitter mesajlarının köməyi ilə COVID-19 pandemiyasına yoluxmuş xəstələrin aşkarlanması üçün yanaşma təklif etmişlər. Twitter mesajlarında COVID-19 pandemiyasını aşkar etmək üçün TF-IDF-əlamət çıxarma texnikasının köməyi ilə dəstək vektor maşını, logistik reqressiya, Naive Bayes, təsadüfi meşə və qərar ağacı kimi maşın təlimi yanaşmalarından istifadə edən intellektual model təklif edilmişdir. Təklif olunan intellektual model Twitter mesajlarını önlər, sağalanlar və şübhəli kimi üç kateqoriya üzrə klassifikasiya etmişdir. Eksperimental analiz üçün, maşın təlimi yanaşmalarının nəticələrini qiymətləndirmək məqsədi ilə COVID-19 pandemiyası ilə bağlı tvit məlumatları təhlil edilmişdir. Eksperimentlər nəticəsində məlum olmuşdur ki, Twitter mesajlarında COVID-19 pandemiyasının doğruluğu 70–80% arasında dəyişir. Təklif olunan yanaşma performansı precision, recall, F-measure meyarları və xətalər matrisi üsulları ilə qiymətləndirilmişdir.

COVID-19-un qarşısının alınması tədbirləri milyonlarla insanın gündəlik fəaliyyətlərinə mane

olur, həm də onların psixi sağlamlıqlarına təsir edir. Belə profilaktik tədbirlərdən biri də sosial uzaqlaşmadır. İstifadəçilər Twitter sosial media platformalarını vasitəsi ilə fikirlərini sərbəst şəkildə ifadə edirlər. [8]-də əsas məqsəd cəmiyyətin sosial uzaqlaşma ilə bağlı ictimai fikirlərini analiz etməkdir. Kanadaya xas Twitter məlumatlarının tvit mətnlərinin sentiment qütübünü artırmaq üçün SentiStrength alətindən istifadə edilmiş və sentiment klassifikasiya üçün SVM alqoritmi tətbiq edilmişdir. Performansın qiymətləndirilməsi xətarlar matrisi, precision, recall, F-measure meyarı ilə ölçülmüşdür. Araşdırmada cəmi 629 tvit mətni çıxarılmışdır. Bunlardan 40%-i neytral, 35%-i neqativ və yalnız 25%-i sosial uzaqlaşma ilə bağlı pozitiv fikirləri ifadə etmişdir. SVM alqoritmi verilənlər dəstinin 80%-i təlim və 20%-i test məlumatlarına bölünərək tətbiq edilmişdir. Performans qiymətləndirməsi 71% dəqiqliklə nəticələnmişdir. Yalnız müsbət və mənfə qütbə malik Twitter mətnləri istifadə edildikdə, dəqiqlik 81%-ə yüksəlmişdir. Test məlumatlarının 10% azaldılması ilə dəqiqliyin 87%-ə çatdığı müşahidə edilmişdir. Nəticələr göstərmişdir ki, təlim məlumatlarının artması alqoritmin performansını artırmışdır.

[9]-da koronavirus pandemiyası ilə bağlı mövzuları çıxara bilən və sonra 2020-ci ilin iki fərqli dövründə: mart - aprel və sentyabr-oktyabr aylarında bu mövzuların SA-nı yerinə yetirən bir model hazırlamışlar. Bu araşdırma üçün, COVID-19 haqqında ingilis dilli tvitlərin məlumat bazası seçilmişdir. 567064 tvit işlənib analiz edilmişdir. İşin reallaşdırılması ilkin emaldan başlayaraq optimal modelin alınmasına qədər bir neçə mərhələdə həyata keçirilmişdir. Bu iş, ən yaxşı nəticəni əldə etmək üçün sözlərin kökünün (stemming) ayrılması və lemmatizasiya (lüğət və marfoloji təhlildən istifadə etməklə sözün əsas və ya lüğət formasını əldə etmək) prosesini yerinə yetirmişdir. Bundan başqa bu tədqiqatda mövzuları tapmaq üçün LDA (The Latent Dirichlet Allocation) modelindən istifadə edilmişdir və koronavirusla əlaqəli ən vacib səkkiz mövzu aşkarlanmışdır. Bu model NLP əsasında işləyir və məlumat bazasının 80%-i təlim verilənləri, 20%-i isə test verilənləridir.

Bundan əlavə, bu məqalədə aşkarlanan mövzuların əksəriyyətinə əsaslanaraq, insanların hissələrini təsnif etmək üçün leksikaya əsaslanan yanaşmalardan istifadə edərək toplanan tvitlərin SA təqdim edilmişdir.

Bu araşdırma çərçivəsində çoxlu sayda tvitlərin analizi və emalını yaxşılaşdırmaq üçün eksperimentlər Spark platformasında python proqramlaşdırma dilində aparılmışdır.

3. Təklif olunan yanaşma

Təklif olunan yanaşma aşağıdakı addımlardan ibarətdir.

1-ci addım. Verilənlərin toplanması. SA-da verilənlərin toplanması ilk addımdır. Twitter verilənlər bazası Kaggle saytından götürülmüşdür (<https://www.kaggle.com>). İşin reallaşdırılması Python mühitində həyata keçirilmişdir. Bu verilənlər bazası 6 (UserName, ScreenName, Location, TweetAt, OriginalTweet, Sentiment) atributdan ibarətdir. Analiz üçün biz yalnız 12726 yazıdan, iki atributdan (OriginalTweet, Sentiment) istifadə edərək onları üç sinif (pozitiv, neqativ, neytral) üzrə klassifikasiya etmişik. Eksperimentlər üçün 10180 tvit təlim verilənləri və 2546 tvit test verilənlərindən istifadə olunmuşdur.

2-ci addım. Verilənlərin ilkin emalı. Bu addım SA prosesinin ən vacib mərhələsidir. Twitter məlumatları sözlər, simvollar, emosiya ikonaları, URL, @ simvolu ilə başlayan istifadəçi adları, heşteqlər və istifadəçilərə istinadlardan ibarətdir. Sözlər, həmçinin səhv yazılmış sözlərdən, çoxlu nöqtələrdən və bir çox təkrarlanan hərflərdən düzələn sözlərin qarışığıdır. Buna görə məlumatları standart formaya gətirmək üçün onları ilkin emal etmək lazımdır. Verilənlərin ilkin emalı küyün azaldılması və maşın təlimi alqoritmlərinin yaxşılaşdırılması üçün verilənlər bazasının əhəmiyyətsiz və lazımsız məlumatlardan təmizlənməsi prosesidir. Müxtəlif mənbələrdən daxil olan emal edilməmiş verilənlərin çox zaman analizdən əvvəl ilkin emalı tələb olunur. Onlayn mətnlər adətən çox sayda küy və HTML etiketləri, skriptlər və reklamlar kimi məlumatsız hissələrdən ibarət olur. Twitterdən götürülmüş emal edilməmiş tvitlər çox zaman küylü verilənlər toplusu olur. Tvitlərin istifadəçi istinadları, emosiya ifadələri və s. kimi xüsusi əlamətləri var. Emal edilməmiş Twitter verilənləri müxtəlif klassifikatorlar vasitəsilə asanlıqla öyrənilə biləcək bir verilənlər toplusu yaratmaq üçün normallaşdırılmalıdır. Tvit verilənlərini standart formaya gətirmək və ölçüsünü azaltmaq üçün çox sayda ilkin emal addımları tətbiq etmək lazım gəlir. İlkin emala aşağıdakı mərhələlər daxildir [10, 23, 24]:

- Tvitlər kiçik hərflərə çevrilir.
- 2 və ya daha çox nöqtə boş sahə ilə əvəz olunur.
- Tvitlərin sonunda durğu işarələri, rəqəmlər və xüsusi simvollar silinir.
- 2 və ya daha çox boş sahə bir boş sahə ilə əvəz olunur.
- Tvitdəki bütün URL-lər, heşteqlər (#topic), istifadəçi adları (@username) silinir.

- “RT” ilə başlayan retvitlər silinir.
- Qısa sözlər silinir.
- Stop words-lər və ya lazımsız sözlər tvitdən silinir.
- Stemming, yəni sözün kökü saxlanıb şəkilçiləri atılır.
- Tokenizasiya prosesi həyata keçirilir.

Aşağıda ilkin emal mərhələlərinin kodu verilmişdir:

```
#new column with removed @user
import re
df['Tweet'] = np.vectorize(remove_pattern)(df['OriginalTweet'], '@[\w]*')
df.head(2)
import re
df['Tweet'] = df['Tweet'].apply(lambda x: re.split('https://\./.*', str(x))[0])
df.head(3)
# remove special characters, numbers,
punctuations
df['Tweet'] = df['Tweet'].str.replace('[^a-zA-Z#]+', '')
df.head(5)
# remove short words
df['Tweet'] = df['Tweet'].apply(lambda x: ' '.join([w for w in x.split() if len(w) > 2]))
df.head(2)
```

```
# create new variable tokenized tweet
tokenized_tweet = df['Tweet'].apply(lambda x: x.split())
df.head(2)
from nltk.stem.porter import *
stemmer = PorterStemmer()
# apply stemmer for tokenized_tweet
tokenized_tweet = tokenized_tweet.apply(lambda x: [stemmer.stem(i) for i in x])
df.head(2)
# join tokens into one sentence
for i in range(len(tokenized_tweet)):
    tokenized_tweet[i] = ' '.join(tokenized_tweet[i])
# change df['Tweet'] to tokenized_tweet
df['Tweet'] = tokenized_tweet
df.head(2)
def hashtag_extract(x):
    hashtags = []
    for i in x: ht = re.findall(r'#(\w+)', i)
```

İlkin emal məsələləri üçün ümumi istifadəyə açıq olan bəzi alətlər haqqında məlumatı [25]-dən əldə etmək olar.

4. İstifadə olunan klassifikatorlar

Məqalədə çoxsınıflı təsnifatı aparmaq üçün Naive Bayes, SVM, RF və NN maşın təlimi alqoritmləri istifadə olunmuşdur.

Naive Bayes. Naive Bayes klassifikatoru Bayes teoreminə əsaslanır. Naive Bayes mətnlərin klassifikasiyası üçün istifadə edilə bilən ən sadə modeldir. Alqoritm mətnlərin sentiment “rəngini” müəyyənləşdirmək üçün istifadə edilir. Bayes teoremindən istifadə edərək siniflər üçün şərti ehtimal aşağıdakı kimi hesablanır.

$$P(c|t) = \frac{P(c)P(t|c)}{P(t)} \quad (1)$$

$$c = \underset{c}{\operatorname{argmax}} P(c|t) \quad (2)$$

$$P(c|t) \propto P(c) \prod_{i=1}^n P(f_i|c) \quad (3)$$

Yuxarıdakı düsturda f_i n əlamət arasından i -ci əlaməti işarə edir. $P(c)$ və $P(f_i|c)$ isə maksimal həqiqətə oxşarlıq qiymətləri ilə hesablanır [25,26].

Dayaq vektor maşını. Dayaq vektor maşını verilənlərin klassifikasiyası üçün hiperüstəvidən istifadə edən və nəzarət edilən təlim alqoritmidir. Hiperüstəvi müxtəlif siniflərin yaxın nöqtələrini ayıran səhələrin tapılması yolu ilə qurulur. Sərhəddən kənarında yerləşən nöqtələr dayaq vektorları təşkil edir. Sadə məsələlər üçün hiperüstəvi sinifləri asanlıqla ayıra bilər. Hiperüstəvi aşağıdakı kimi təyin olunur [24, 27-30]:

$$h(x) = w^T x \quad (4)$$

Burada w - $h(x) = 0$ boyunca hiperüstəvidə yerləşən nöqtələrin çəki vektorudur.

Tutaq ki, iki sınıflı klassifikasiya məsələləri üçün təlim verilənləri toplusu (x_i, y_i) -dir. Burada $x_i \in R$ -əlamətlər vektoru, y_i isə ona uyğun sinif nişanıdır, $y_i \in \{+1, -1\}$. Elə hiper-üstəvi tapmaq lazımdır ki, o $y_i = 1$ və $y_i = -1$ nöqtələrini ayırsın, təlim çoxluğunun ən yaxın nöqtələrindən maksimal məsafədə keçsin. Hiperüstəvinin tənliyi $w \cdot x - b = 0$ ilə verilir. Əgər w vektorunun x_i ilə skalyar hasil b -nin icazə verilən qiymətindən böyükdürsə, $w \cdot x_i > b \Rightarrow y_i = 1$, onda yeni nöqtə birinci kateqoriyaya, əgər azdırsa, $w \cdot x_i < b \Rightarrow y_i = -1$, ikinciyə aiddir.

Təsadüfi meşə alqoritm. Təsadüfi meşə alqoritm qərar ağacı ansamblından istifadə edən maşın təlimi alqoritmidir. RF çoxlu sayda klassifikasiyaların və reqressiyanın vahid qərar ağacını təqdim edir. Alqoritm Bagging

yanaşmasına və əlamətlərin təsadüfi altcoxlugunun seçilməsinə əsaslanır. Tutaq ki, x_1, x_2, \dots, x_n tvitlər çoxluğu y_1, y_2, \dots, y_n isə onların müvafiq sentiment əlamətləridir. Bagging tətbiq edilməklə əvəzləmə yolu ilə təsadüfi (X_i, Y_i) cütü seçilir. Hər bir klassifikasiya ağacı f_k təsadüfi (X_k, Y_k) istifadə edərək öyrədilir. Burada $k \in \{-1, K\}$. Alqoritm klassifikasiyanın yüksək dəqiqliyinə nail olmağa imkan verir. Qərar ansamblı bir-birindən asılı olmadan qurulur. Sonda K sayda qərar ağacının yekun qərarı majoritar səsvermə yolu ilə müəyyən edilir [23,25].

Neyron şəbəkə. Neyron şəbəkə bioloji neyron şəbəkələrin – canlı orqanizmin sinir hüceyrələrinin şəbəkələrinin təşkili və işləməsi prinsipi əsasında qurulmuş riyazi model, eləcə də onun proqram təminatı və ya aparat təminatıdır. Bioloji neyronun əlaqələri süni neyron şəbəkələrdə düyünlər arasında çəkilər kimi modelləşdirilir. Müsbət çəki həyəcanverici əlaqəni, mənfi çəki isə ləngidilməni əks etdirir. Bütün daxilolmalar çəkiyə görə modifikasiya olunur və cəmlənir. Bu fəaliyyət xətti birləşmə adlanır. Nəticədə, aktivləşdirmə funksiyası çıxış siqnalının amplitudunu idarə edir. Məsələn, məqbul çıxış diapazonu adətən 0-dan 1-ə qədərdir və ya -1-dən 1-ə qədər ola bilər [31].

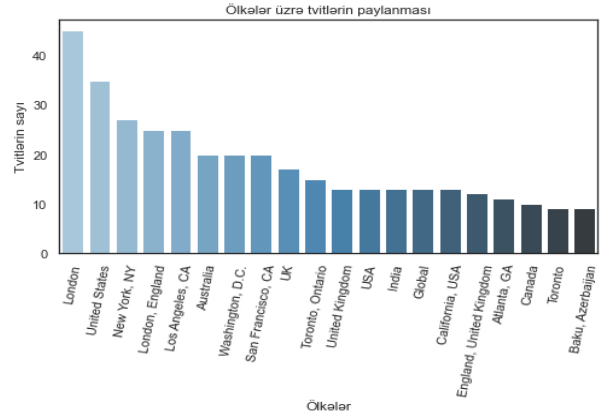
$$\vec{h}_i = f(W_x \vec{x}_i + b_x) \quad (5)$$

Burada \vec{h}_i gizli xüsusiyyətlər; $f(z)$ aktivasiya funksiyası; W_x, b_x çəki və meyllilik parametrləri; x_i giriş vektorudur.

5. Eksperimentlər və diskussiya

Yuxarıda təsvir olunmuş müxtəlif klassifikatorlar Python mühitində scikit-learn kitabxanasından istifadə edilməklə reallaşdırılıb və <https://www.kaggle.com> saytıdan götürülmüş məlumatlar üzərində eksperimentlər aparılmışdır. Şək.1-də bir neçə ölkənin Twitter istifadəçilərinin COVID-19 pandemiyasına münasibəti göstərilmişdir. Seçilmiş (Naive Bayes, SVM, RF və NN) alqoritmlər götürülmüş təlim verilənləri üzərində öyrədilmişdir. Daha sonra eksperimentlər test verilənləri üzərində aparılmışdır. Eksperimentdə istifadə olunan hər bir klassifikatorun test verilənləri üzərindəki nəticələri cədvəl 1-də təqdim olunmuşdur.

Klassifikatorların Precision, Recall, F-measure meyarları üzrə göstərdiyi nəticələr müəyyən olunmuşdur. Tvitlərin sentiment qütbləri neytral, neqativ, pozitiv olaraq təsnifatlandırılmışdır.



Şək. 1. COVID-19 pandemiyasına görə tvitlərin ölkələr üzrə paylanması.

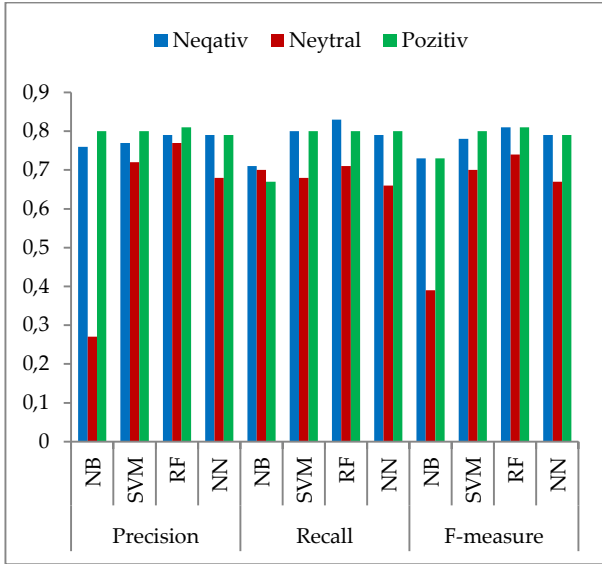
Cədvəl 1. Klassifikatorların meyarlar üzrə nəticələri

Sentiment	Precision				Recall				F-measure			
	NB	SVM	RF	NN	NB	SVM	RF	NN	NB	SVM	RF	NN
Neqativ	0.76	0.77	0.79	0.79	0.71	0.80	0.83	0.79	0.73	0.78	0.81	0.79
Neytral	0.27	0.72	0.77	0.68	0.70	0.68	0.71	0.66	0.39	0.70	0.74	0.67
Pozitiv	0.80	0.80	0.81	0.79	0.67	0.80	0.80	0.80	0.73	0.80	0.81	0.79

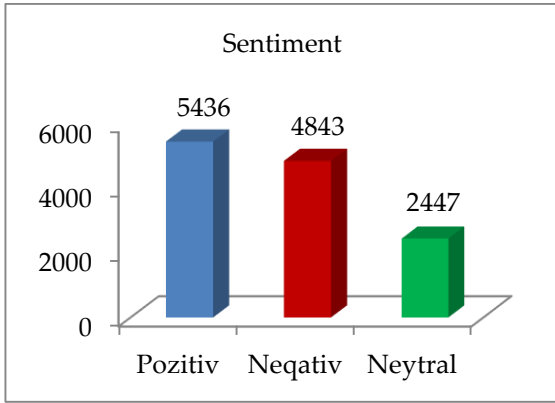
Cədvəl 1-dən görüldüyü kimi, Precision meyarı üzrə neytral əhval-ruhiyyəyə görə, ən aşağı nəticəni NB klassifikatoru və ən yüksək nəticəni RF klassifikatoru, neqativ əhval-ruhiyyəyə görə, ən aşağı nəticəni NB klassifikatoru və ən yüksək nəticəni RF klassifikatoru, pozitiv əhval-ruhiyyəyə görə, ən aşağı nəticəni NN klassifikatoru, ən yüksək nəticəni isə RF klassifikatoru göstərmişdir. Recall meyarı üzrə neytral əhval-ruhiyyəyə görə, ən aşağı nəticə NN klassifikatoru və ən yüksək nəticə RF klassifikatoru, neqativ əhval-ruhiyyəyə görə, ən aşağı nəticə NB klassifikatoru və ən yüksək nəticə RF klassifikatoru, pozitiv əhval-ruhiyyəyə görə ən aşağı nəticə NB klassifikatoru, ən yüksək nəticə SVM, RF və NN klassifikatorları ilə əldə edilmişdir. F-measure meyarı üzrə neytral əhval-ruhiyyəyə görə, ən aşağı nəticəni NB klassifikatoru və ən yüksək nəticəni RF klassifikatoru, neqativ əhval-ruhiyyəyə görə, ən aşağı nəticəni NB klassifikatoru və ən yüksək nəticəni RF klassifikatoru, pozitiv əhval-ruhiyyəyə görə, ən aşağı nəticəni NB klassifikatoru, ən yüksək nəticəni isə RF klassifikatoru göstərmişdir.

Şək. 2-də test verilənləri üzərində aparılmış sınaqlarda klassifikatorların meyarlar üzrə nəticələrinin vizual görüntüsü və şək. 3-də isə test verilənləri üzrə tvitlərin sentiment qütblərinin paylanması təqdim olunmuşdur.

Şək.2-dəki qrafikdən görüldüyü kimi, meyarlar üzrə neytral əhval-ruhiyyəyə görə, ən aşağı nəticəni NB klassifikatoru, pozitiv və neqativ əhval-ruhiyyəyə görə, ən yüksək nəticəni RF klassifikatoru göstərmişdir.



Şək. 2. Klassifikatorların meyarlar üzrə göstərdiyi nəticələr.



Şək. 3. Tvitlərin sentiment qütblər üzrə paylanması.

Şək. 3-dəki qrafikə əsasən, 5436 pozitiv əhval-ruhiyyəli tvit, 4843 neqativ əhval-ruhiyyəli tvit və 2447 neytral əhval-ruhiyyəli tvit olduğu müşahidə olunur.

Alqoritmlərin effektivliyini qiymətləndirmək üçün istifadə olunan metrikalar aşağıdakı düsturlarla ifadə olunur [32,33]:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F - measure = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (8)$$

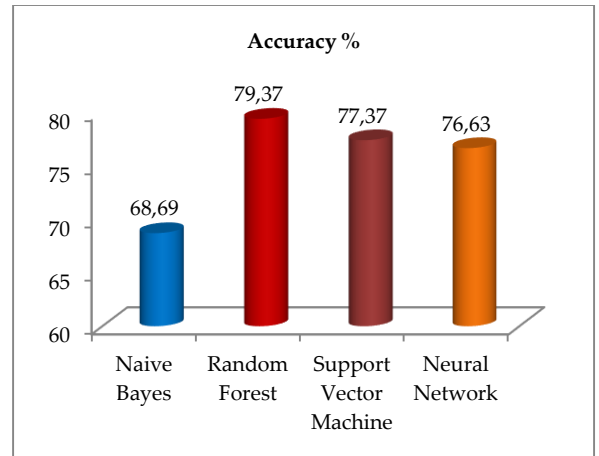
$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (9)$$

Cədvəl 2-də klassifikatorların doğruluq testi və şəx. 4-də doğruluq testinin qrafiki təqdim edilmişdir.

Cədvəl 2. Klassifikatorların doğruluq testi.

Klassifikatorlar	Doğruluq testi %
Naive Bayes	68.69
Random Forest	79.37
Support Vector Machine	77.37
Neural Network	76.63

Cədvəl 2-dən görüldüyü kimi, klassifikatorların doğruluq testində ən yüksək nəticəni RF klassifikatoru göstərmişdir.



Şək. 4. Klassifikatorların meyarlar üzrə doğruluq testinin qrafiki.

Şək. 4-dəki qrafikdən görüldüyü kimi, ən aşağı nəticəni NB klassifikatoru, ən yüksək nəticəni isə RF klassifikatoru göstərmişdir. SVM və NN klassifikatorları isə eyni nəticə göstərmişdir.

6. Nəticə

COVID-19 gündən-günə bütün dünyada sürətlə yayılaraq milyonlarla insanların həyatına təsir etdiyi üçün bir sıra ölkələr onun şiddətini yoxlamaq məqsədilə karantin elan etmişdilər. Karantin dövründə, insanlar sosial şəbəkələr vasitəsilə duyğularını ifadə etdikləri üçün Twitter COVID-19 pandemiyası haqqında məlumatı bütün dünyada yaymaqda mühüm rol oynamışdır. Bu fəlakətli vəziyyəti nəzərə alaraq, Twitterdə insanların bu pandemiya ilə əlaqədar tvitlərinə analiz etmək üçün yanaşma təklif olunur. Bu məqalə, Coronavirus və ya COVID-19 pandemiyası ilə əlaqədar tvitlər haqqında SA-nı əks etdirir. Məqalədə maşın təlimi metodlarından (Naive Bayes, Dayaq Vektor Maşını, Neyron Şəbəkə və Təsadüfi Meşə) istifadə edərək COVID-19 pandemiyası ilə bağlı müxtəlif ölkələrin Twitter istifadəçilərinin məlumatları analiz edilmiş, on-

ların üç sinif üzrə (neytral, pozitiv və neqativ) klassifikasiyası məsələsinə baxılmışdır. Tədqiqatda dörd klassifikatordan (NB, RF, SVM və NN) istifadə olunmuşdur. Klassifikatorlardan Random Forest klassifikatoru ən yüksək nəticə göstərmişdir.

Ədəbiyyat

- Liu, B. (2012). Sentiment analysis and opinion mining. Toronto: Morgan & Claypool Publishers. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Hacırahimova, M.Ş., İsmayilova, M.İ. (2020). Sentiment analiz: problemləri və həlli yolları. İnformasiya texnologiyaları problemləri, 2, 111-123. <https://doi.org/10.25045/jpit.v11.i2.11>
- Evirgen, E. (2016). Sentiment analysis of turkish tweets. Master Thesis. Turkey: Bahcesehir University. https://www.academia.edu/38504379/IJMET_10_01_094_pdf?from=cover_page
- Alguliev, R., Aliguliyev, R., Hajirahimova, M. (2010). Multi-document summarization model based on integer linear programming. Intelligent Control and Automation, 1(2), 105-111. <https://doi.org/10.4236/ica.2010.12012>
- Kaur, C., Sharma, A. (2020). Twitter Sentiment Analysis on Coronavirus using Textblob. EasyChair Preprint, 2974, 1-10. <https://easychair.org/publications/preprint/Fd5m>
- Dashtian, H., Murthy, D. (2021). CML-COVID: A large-scale covid-19 Twitter dataset with latent topics, sentiment and location information, Academia Letters, 1-9. <https://doi.org/10.20935/AL314>
- Samina, A. et al. (2021). Machine Learning Approach for COVID-19 Detection on Twitter. Computers, Materials & Continua, 68(2), 2231-2247. <https://doi.org/10.32604/cmc.2021.016896>
- Shofiya, C., Samina, A. (2021). Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data. International Journal of Environmental Research and Public Health, 18(11), 1-10. <https://doi.org/10.3390/ijerph18115993>
- Abdulaziz, M. et al. (2021). Topic based Sentiment Analysis for COVID-19 Tweets. International Journal of Advanced Computer Science and Applications, 12 (1), 626-636. <https://doi.org/10.14569/IJACSA.2021.0120172>
- Vohra, S. M., Teraiya, J. B. (2013). A comparative study of sentiment Analysis techniques. International Journal of Information, knowledge and research in Computer engineering, 2(2), 313-317.
- Kumar, A., Sebastian, T. M. (2012). Sentiment Analysis on Twitter. International Journal of Computer Science, 9(4), 372-378. <http://www.ijcsi.org/>
- Kolchyna, O. et al. (2015). Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination, Cornell university, 1-32. [arXiv.org](https://arxiv.org)
- Borele, P., Borikar, D.A. (2016). An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques. Journal of Computer Engineering (IOSR-JCE), 18(2), 64-69.
- Spencer, J., Uchyigit, G. (2012). Sentimentor: Sentiment Analysis of Twitter Data. In The 1st International Workshop on Sentiment Discovery from Affective Data (SDAD), Bristol, UK, 28 September (pp.56-66). <http://www.iosrjournals.org/>
- Hasan, A., Moin, S., Karim, A., Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. International Journal of Mathematical and Computational Applications, 23(11), 1-15. <https://doi.org/10.3390/MCA23010011>
- Yadav, S.J., Ranjan, P. (2017). Proposed Approach for Sarcasm Detection in Twitter Shubhodip. Indian Journal of Science and Technology, 10(25), 1-8. <https://doi.org/10.17485/ijst/2017/v10i25/114443>
- Bahrainian, S.A., Denge, A. (2013). Sentiment Analysis using Sentiment Features. Proc. of IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), Washington, USA, November 2013, pp. 26-29. <https://doi.org/10.1109/WI-IAT.2013.145>
- Kawade, D.R., Oza, K.S. (2017). Sentiment Analysis: Machine Learning Approach. International Journal of Engineering and Technology, 9(3), pp.2183-2186. <https://doi.org/10.21817/IJET/2017/V9I3/1709030151>
- Hafez, A.A., Xu, Y., Tjondronegoro, D. (2012). Product Reputation Model: An Opinion Mining Based Approach. Proc. of the 1st International Workshop on Sentiment Discovery from Affective Data, London, UK, January 2012 (pp.16-27). <http://ceur-ws.org/Vol-917/>
- Saleena, A.N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. Proc. of the International Conference on Computational Intelligence and Data Science (ICCID), Gurugram, India, January 2018 (pp. 937-946). <https://doi.org/10.1016/j.procs.2018.05.109>
- Sumathy, Dr.P., Muthukumari, S.M. (2018). Sentiment Analysis of Twitter Data Using Multi Class Semantic Approach. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 3(6), 262-269. <https://doi.org/10.32628/CSEIT183654>
- Shehu, H.A., Tokat, S., Sharif, Md. H., Uyaver, S. (2019). Sentiment Analysis of Turkish Twitter Data. Proc of the AIP Conference, 2183(1), AIP Publishing (pp. 1-4). <https://doi.org/10.1063/1.5136197>
- Hacırahimova, M., İmamverdiyeva, A. (2018) Twitter məlumatlarının sentiment analizi. "İnformasiya təhlükəsizliyinin aktual multidissiplinar elmi-praktiki problemləri" IV respublika konfransı, Bakı, Azərbaycan, 14 dekabr 2018 (s. 245-248). <https://doi.org/10.25045/NCInfoSec.2018.59>
- Kumar, R. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46. <https://doi.org/10.1016/j.knosys.2015.06.015>
- Kharde, V.A., Sonawane, S.S. Sentiment Analysis of Twitter Data: A (2016). Survey of Techniques Vishal A. International Journal of Computer Applications, 139(11), 5-15. <https://doi.org/10.5120/ijca2016908625>
- Tang, H., Tan, S., Cheng, X. (2009). A survey on sentiment detection of reviews. International Journal Expert Systems with Applications, 36(7), 10760-10773. <https://doi.org/10.1016/j.eswa.2009.02.063>
- Medhat, W., Hassan, A., Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Vinodhini, G., Chandrasekaran, R.M. (2012). Sentiment Analysis and Opinion Mining: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering, 2(6), 282-292. <https://doi.org/10.5121/ijasuc.2013.4102>


29. Wawre, S.V., Deshmukh, S.N. (2016). Sentiment Classification using Machine Learning Techniques. *International Journal of Science and Research*, 5(4), 819-821.
30. Padmaja, S., Fatima, S.S. (2013). Opinion Mining and Sentiment Analysis –An Assessment of Peoples’ Belief: A Survey. *International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC)*, 4(1), 21-33.
<http://www.ijsr.net>
31. Pranali, B., Dilipkumar, A.B. (2016). An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques. *Journal of Computer Engineering*, 18(2), 64-69.
<https://doi.org/10.9790/0661-1802056469>
32. Chena, L-S, Liub, C-H, Chiu, H-J. (2011). A neural network based approach for sentiment classification in the blogosphere. *Journal of Informetrics*, 5(2), 313–322.
<https://doi.org/10.1016/j.joi.2011.01.003>
33. Mathur, R., Bandil, D., Pathak, V. (2018). Analyzing Sentiment of Twitter Data using Machine Learning Algorithm. *GADL Journal of Inventions in Computer Science and Communication Technology (JICSCT)*, 4(2), 1-7.

Makrufa Sh.Hajirahimova ¹, Marziya I. Ismaylova ²

^{1,2}Azerbaijan National Academy of Sciences, Institute of Information Technology, B.Vahabzade str., 9A, AZ1141 Baku, Azerbaijan

Макруфа Ш. Гаджирагимова ¹, Марзия И. Исмаилова ²

^{1,2} Национальная Академия Наук Азербайджана, Институт Информационных Технологий, ул. Б.Вахабзаде, 9А, AZ1141, Баку, Азербайджан

 ¹ 0000-0003-0786-5974; ² 0000-0002-3080-0952