

FAN VA TEXNOLOGIYALAR TARAQQIYOTI

DEVELOPMENT OF SCIENCE AND TECHNOLOGI

1
2026

Tahririyat hay'ati raisi:
SIDDIQOVA S.G'. –
Buxoro davlat texnika universiteti rektori

Muovini:
NIZAMOV A.B. –
BuxDTU ilmiy ishlar va innovatsiyalar bo'yicha prorektori
Tahrir hay'ati:

MUQIMOV K.M. – O'zR FA akademigi (O'zMU)
JALILOV A.T. – O'zR FA akademigi (Toshkent kimyo-texnologiya ITI)
NEGMATOV S.N. – O'zR FA akademigi ("Fan va taraqqiyot" DUK)
BAHODIROV G'.A. – t.f.d., professor, O'zR FA bosh ilmiy kotibi
XAMIDOV O.X. – iqtisod fanlari doktori, professor (BuxDU)
JALILOV T.K. – iqtisod fanlari doktori (DSc), professor (TKTI)
PARDAYEVA M.D. – BuxDTU yoshlar masalalari va ma'naviy-ma'rifiy ishlar bo'yicha birinchi prorektori, falsafa fanlari doktori (DSc)
XOJIYEV A.X. – o'quv ishlari bo'yicha prorektor, texnika f.f.d. (PhD)
SAIDOV S.B. – Buxoro DTU moliya va iqtisod ishlari bo'yicha prorektori
QURBONOV J.M. – texnika fanlari doktori, professor (Samarqand ISI)
ADIZOV B.Z. – texnika fanlari doktori (DSc), pprofessor, O'zRFA UNKI
ASTANOV S.X. – fizika-matematika fanlari doktori, professor
RAXMONOV X.Q. – texnika fanlari doktori, professor
VOXIDOV M.M. – texnika fanlari doktori, professor
JO'RAYEV X.F. – texnika fanlari doktori, professor
SADULLAYEV N.N. – texnika fanlari doktori (DSc), professor
MAJIDOV Q.X. – texnika fanlari doktori, professor
FOZILOV S.F. – texnika fanlari doktori, professor
ISABAYEV I.B. – texnika fanlari doktori, professor
ABDURAHMONOV O.R. – texnika fanlari doktori, professor
GAFUROV K.X. – texnika fanlari doktori (DSc), professor
XAYDAROV A.A. – texnika fanlari doktori (DSc), dotsent
JO'RAYEV F.O'. – texnika fanlari doktori (DSc), professor
MURADOVA F.R. – pedagogika fanlari doktori (DSc), professor
JUMAYEV M.R. – fizika-matematika fanlari doktori (DSc), professor
YUNUSOVA G.S. – falsafa fanlari doktori (DSc), professor
BOBOYEV A.Ch. – iqtisodiyot fanlari nomzodi, professor
TO'XTAYEVA Z.Sh. – texnika fanlari doktori (DSc), professor
MAXMUDOV M.J. – texnika fanlari doktori (DSc), professor
HAYITOV R.R. – texnika fanlari doktori (DSc), professor
BOZOROV G'.R. – texnika fanlari doktori (DSc), professor
BOLTAYEV Z.I. – fizika-matematika fanlari doktori (DSc), professor
OLTIYEV A.T. – texnika fanlari doktori, (DSc)
JALILOV R.B. – texnika fanlari doktori (DSc), professor
MAXMUDOV M.I. – texnika fanlari doktori (DSc), professor
MAJIDOVA N.Q. – texnika fanlari doktori (DSc), professor
AXMEDOV V.N. – texnika fanlari doktori (DSc), professor
MAXMUDOV R.A. – texnika fanlari doktori (DSc), professor
PULATOVA M.I. – fizika-matematika fanlari nomzodi, professor
RAHMATOV Sh.A. – pedagogika fanlari bo'yicha falsafa doktori (PhD)
OCHILOV A.R. – texnika fanlari doktori (DSc), dotsent
O'RINOV U.A. – pedagogika fanlari doktori (DSc), professor
PO'LATOVA S.U. – texnika fanlari doktori (DSc), professor
SAMIYEVA Sh.X. – pedagogika fanlari doktori (DSc), professor
TESHAYEV M.X. – fizika-matematika fanlari doktori (DSc), professor
XAITOV V.U. – iqtisodiyot fanlari doktori (DSc), dotsent
XOJIYEV Sh.M. – texnika fanlari doktori (DSc), dotsent
XAYITOV Sh.N. – iqtisodiyot fanlari doktori (DSc), dotsent
ZOIROV E.X. – falsafa fanlari doktori (DSc), dotsent
NARZIYEV M.S. – texnika fanlari doktori (DSc), dotsent
NAMAZOVA N.J. – iqtisodiyot fanlari b.f.d. (PhD), dotsent

Bosh muharrir: DO'STOV H.B. – kimyo fanlari doktori, professor

Muharrirlar: Artikova M.M., Istamova G.X.
Musahhih: Barakayeva D.F.

FAN VA TEXNOLOGIYALAR
TARAQQIYOTI
ILMIY-TEXNIKAVIY JURNAL

DEVELOPMENT OF SCIENCE
AND TECHNOLOGY
SCIENTIFIC AND TECHNICAL JOURNAL

Jurnal O'zbekiston matbuot va axborot agentligi Buxoro viloyati boshqarmasida 2014 yil 22-sentyabrda № 05-066-sonli guvohnoma bilan ro'yxatga olingan

Muassis:
Buxoro davlat texnika universiteti

Jurnal O'zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi OAK Rayosatining 2017 yil 29-martdagi №239/5-sonli qarori bilan dissertatsiyalar asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlar ro'yxatiga kiritilgan. 2019 yilda O'zbekiston Respublikasi OAK Rayosatining qarorlari bilan qayta ro'yxatdan o'tkazilgan.

Tahririyat manzili:
200117, Buxoro shahri, Q. Murtazoyev ko'chasi, 15-uy, Buxoro davlat texnika universiteti
Tel: 0(365) 223-92-40
Faks: 0(365) 223-78-84
E-mail: fannt_jurnal@umail.uz

Jurnalning to'liq elektron varianti bilan <http://journal.bstu.uz> sayti orqali tanishish mumkin.

Ushbu jurnalda chop etilgan materiallar tahririyatning yozma ruxsatisiz to'liq yoki qisman chop etilishi mumkin emas. Tahririyatning fikri mualliflar fikri bilan har doim ham mos tushmasligi mumkin. Jurnalda yoritilgan materiallarning haqqoniyligi uchun maqolalarning mualliflari va reklama beruvchilar mas'uldirlar.

MUNDARIJA – СОДЕРЖАНИЕ – CONTENT

TEXNIKA, TEXNOLOGIYA VA JIHOZLAR	
Normayev Q.H., Pardayeva Sh.S., Pirnazarov Sh.V., Ashirova M.R., Jalg'asova G.S. Burg'ilash mashinalari va ularning mexanizmlarini samarali qayta tiklash texnologiyalari.	5
Azamjonov R.S. Eksploatatsiyadagi avtoyo'l ko'priklarining yuk ko'tarish qobiliyatini oshirish usulini takomillashtirish	11
Xolmuxammadiyev A.M., Toshev Sh.O. Burg'ulash eritmasidan markazdan qochma separator yordamida yengillashtirilgan eritma tayyorlash texnologiyasi	15
Джураев Х.Ф., Ўктамова Ш.Х. Ковок уруғи қобиғининг синдириш кучларини аниқлаш методикаси	20
Mansurov O.A. Improving the drying process of fruits using convective method.	24
Jo'rayev F.O'., Jo'rayev A.A., Saidova G.K. Uneversal qurilmaning yumshatuvchi ish oraganining parametrlarini uning ish ko'rsitkichlariga ta'siri.	27
Авлиякулов Н.Н. Достижение качества измерительных процессов при метрологическом обеспечении производства	32
Do'stov H.B., Rahimov B.R., Savriyev F.A. Suvlangan gaz quduqlarini qayta ishga tushirish texnologiyasini takomillashtirish tahlili.	36
KIMYO VA KIMYOVIY TEXNOLOGIYALAR	
Toshboyev S.O'., Panoyev E.R., Hasanov Z.Z. Gazlarni tozalashda qo'llaniluvchi absorbentlarning tavsifi va unda hosil bo'ladigan ko'piklanish sabablari.	43
Умаров С.Х., Рустамов В.Дж., Халлоков Ф.К., Ходжаев У.О., Нарзуллаева З.М. Механизм отрицательной фотопроводимости в TiGaSe ₂	47
Фозилов Х.С., Туробжонов С.М., Фозилов С.Ф. Маҳаллий иккиламчи хом ашё – куйи молекулали полиэтиленни оксидлаб синтетик юқори ёғ кислоталари олиш ва уларнинг физик-кимёвий хоссалари.	52
Абдирахимов И.Э. Способы предотвращения образования гидрата при переработке газов.	57
Alinazarov A.X., Shariboyev N.Y., Jo'raxanov D.M. Anaerob parchalanish jarayonida harorat barqarorligining metanogen bakteriyalar faolligiga ta'siri va uni quyoshli issiqlik tizimi orqali ta'minlash.	61
Abdirakhimov I.E. Anticorrosion properties of copper phosphide.	66
Norqulov J.F., Ibodullayev M.X., Abduraxmonov O.R., Kodirov O.Sh. Amin eritmalarini issiqlikka bardoshli tuzlardan tozalash jarayonini modellashtirish.	72
Абдирахимов И.Э., Буранов Ф.Э., Курбанов А.Т. Исследование технологических параметров, рекомендованных для синтеза этилена из метана.	79
Rahimov B.R. Yuqori qatronli neftlarning qovushqoqligini pasaytiruvchi depresantlar kompozitsiyasini ishlab chiqish.	86
Эшдавлатова Г.Э., Камолов Л.С. Определение концентрации диэтанолamina титриметрическим методом	91
Rahimov B.R. Qoldiq neftlarni olishda chuqurlik nasos qurilmasini takomillashtirishning dolzarb muamolari.	96
Yuldashev T.R., Buronov F.E., Hafizov S.Sh. Tabiiy gazning tarkibidagi zaharli komponentlarni aminlar yordamida tozalash jarayonlari.	100

MASHINASOZLIK VA ENERGETIKA	
Bafoyev D.X. Detallar sirtini plastik deformatsiyalash va bir tekis mikrorelefnı shakllantirish	107
Majitov J.A. Bioreaktor devorining turli sirtlarida issiqlik berish koeffitsiyentlarini hisoblash va tahlil etish.	112
Azamov S.S. Uch fazali tok o'zgartkichlari yordamida asinxron motorlarni himoya qilish tizimi tatqiq etish.	118
Maxmudov M.I., Qo'shshayeva M.R., Nurov S.S., Timirov H.N., Sayfiyev H.O. Quyosh panellarining samaradorligiga changlanganlikning ta'sirini o'rganish bo'yicha ilmiy tadqiqotlar tahlili.	122
Атауллаев А.О. Исследование и разработка виброустойчивых инклинометрических преобразователей для систем контроля и управления.	128
Tovboyev A.N., Tog'ayev I.B. Tojlanish jarayonining yuqori kuchlanishli havo liniyalaridagi quvvat yo'qotishlariga ta'sirini baholash.	135
Murodov K.J. Ko'p qavatli binolarning vertikal chiqindi quvurlaridagi suyuqlik oqimlaridan elektr energiya hosil qilish.	141
Xamroyev X.X. Mexanik ishlov berishda sirt qatlamining mexanik xususiyatlari.	145
Tojimurodov D.D. Asinxron motorlarni nosimmetrik rejimlari.	149
Sadullayev N.N., Nematov Sh.N., To'xtayev Sh.B., Soliyeva Z.N., Murtazoyev F.F. O'zbekistonda quyosh sovitish tizimlari va ularning texnik imkoniyatlari.	154
INFORMATIKA VA AXBOROT – KOMMUNIKATSION TIZIMLAR	
Babomuradov O.J., Kuyliyeva F.A. Statistical analysis of social network messages.	163
Madaminov U.A., Qodirov D.R. Katta hajmli ma'lumotni parallel qayta ishlashda Hadoop MapReduce tizimini qo'llash.	168
Каххоров М.М. Моделирование ветровой турбины типа Дарье и обоснование её параметров в системе Ansys Fluent.	173
Maxmudov M.Sh. Relief morfometrik ko'rsatkichlari asosida marshrutlashning model va algoritmlari.	179
OZIQ-OVQAT SANOATI TEXNOLOGIYALARI	
Ravshanov S.S., Botirov M.Sh., Ergashev A.M., Mirzaev J.D., Shodmonqulova Q. Quruq iqlimda yetishtirilgan bug'doy donini ochiq omborlarda saqlashning mahsulot xossalariga ta'siri.	184
Sharipov N.Z., Qo'ldosheva F.S. Gilos danagidan moy olish jarayonini infraqizil nurlar yordamida jadallashtirish.	187
Шарипов Н.З. Интенсификация процесса получения масла из ядра косточек вишни в поле сверхвысокочастотного тока.	190
Баракаев Н.Р., Шукуров Ю.У. Мева ва сабзавотларни сублимация йўли билан қуритишнинг илмий асослари.	194
Ibragimov R.R., Sharipov N.Z. Bodringni konservalash jarayonida noan'anaviy usulda issiqlik ishlov berish jarayonini nazariy asoslari.	199
Kuliyev N.Sh. Meva va sabzavot sharbatlarini ko'pirtirishda ko'pik hosil qilish va emulgirlash xususiyatlarining dinamikasi.	203
Мухамедова М.Э. Разработка рецептуры диабетических сдобных сухарей с использованием синергии растительных компонентов.	209
Хужакулов У.К., Мажидова Н.К., Мажидов К.Х. Исследование показателей качества местных сортов овощной продукции длительного хранения.	220

Файзиев А.А., Ахмедов В.Н., Кедельбаев Б.Ш. Технология приготовления натурального мясного полуфабриката ферментированного бифштекса	227
TO‘QIMACHILIK VA YENGIL SANOAT TEXNOLOGIYALARI	
Ramazonov S.S., Ismoyilov F.B. Tola sifatini yaxshilash va tozalash maqsadida jinlash mashinasini ishchi qismlarini takomillashtirish.	233
Muxitdinova M.A., Bebutova N.N. Milliylik va zamonaviylik uyg‘unligi: O‘zbek libos dizaynida yangi yo‘nalishlar.	237
Nematova L.X., Qodirova D.X. Yoshi katta ayollar fe‘l atvori belgilari bo‘yicha iste’molchilarning tipologik xususiyatlarini aniqlash.	242
Saidova A.S., Saidova G.Sh., Fayzilloyeva N.B. Tikuv mashinasining tahlama qurilmadagi taxlam hosil qilishda takomillashgan tahlam berishdagi tishlarini turli ko‘rinishda choklarga ta’siri.	246
Совутов М.Э., Мусаев Н.М., Холиқов К.М., Мукимов М.М. Иссиқлик сақлаш хусусияти юқори бўлган трикотаж тўқималарининг физик-механик кўрсаткичлари тадқиқи	251
Shodmonova M.S. Elastikligi yuqori bo‘lgan polimer kompozitsiya olish usulini ishlab chiqish	258
Садуллаева Д.А. Современные материалы: Хан-атлас – традиция и инновация.	261
Джурраев А., Росулов Р.Х., Пардаев Б.Ч., Имомназаров М.С. Пахтани майда ифлосликлардан тозалагичнинг самарали конструкцияси.	267
Aliyev Sh.B. Tikuv mashinasining tishli reykasini lazer nuri yordamida termik ishlov berishning matematik modeli.	270
Musayev S.S., Qodirov T.J. Reseptura ko‘rsatkichlarining poliolefin va etilenpropilen kau-chugi asosidagi poyabzal kompozitsiyalarining iste’mol va texnologik xususiyatlariga ta’siri.	275
Amonov A.R., Muxammedjanov M.M. Tikuv mashinasi qavariq rezinali vtulkali qayishqoq tayanchlarga o‘rnatilgan bosh vali tebranish amplitudasi tahlili.	281
Berdimuratov U.T., Rosulov R.X., Pardayev B.Ch. Arrali jinlar konstruksiyasi, asosiy ishchi organlari, ishlash jarayoni.	284
Muxammedova M.O., Axmedov J.J. O‘smirlar uchun soddalashtirilgan ortopedik poyabzal elementlarining konstruktiv yechimlarini ishlab chiqish.	286
Temirova G.I., Shodmonova M.S. Polimer kompozitsiya asosida ko‘p qatlamli mo‘ynali paketlar tayyorlash texnologiyasi.	291
Saidova X.X., Murodova Z., Ulugova N. 3D va ko‘p o‘lchovli geometrik yondashuvlar asosida bolalar kiyimini konstruktiv loyihalash.	295
Kuliyeva D.R. Bazalt mato qo‘llanilgan maxsus kiyim turini tanlash va konstruksiyasini ishlab chiqish	299
Sharipov J.O., Sayitqulov S.O., Begmurodov A.F. Detallarni yeyilishi va korroziyalanishiga asosiy sabablar hamda ulardan himoyalaniş.	306
Toshpulotov L.I., Xamrayeva S.A. Turli o‘rilishli ikki qatlamli ko‘ylakbop matolarning mexanik va deformatsion xususiyatlarining o‘zgarishi.	309
Sayitqulov S.O., Razzoqov R.R. Paxtani mayda iflosliklardan tozalash mashinasining yangi konstruksiyasini ishlab chiqish.	314
Qodirova S.X. Ayollar ustki liboslari bo‘yicha istiqbolli moda yo‘nalishlarini tahlil qilish	318
Каримова Н.Х. Оценка свойств двухслойных трикотажных тканей на основе структурного и механического анализа противостойчивости к трению и разрывного удлинения	322
Maxmudova X.M. Paxta quritishning ahamiyatliligi va undagi texnik muammolar.	326
Behbudov Sh.H., Samadova M.O. Ip va matoga ta’sir etuvchi tashqi kuchlar asosida tikuv mashinasi bo‘g‘inlarining harakat tenglamalarini tahlil qilish.	332

Шопулатов А.А., Райимкулов Ж.К., Бердимуратов У.Т., Росулов Р.Х. Пахтани майда ифлосликлардан тозалашда қозикчали-планкали барабан планкаларининг таъсири	336
Sultonov M.M., Aloxodjayev A.M., Mirzaaxmedova D.Yu. Takomillashtirilgan tozalash qurilmasining texnologik parametrlarini tadqiq qilish.	340
Azimova G.A., Shokirov L.B. Mahalliy xomashyolarga asoslangan yarim jun to‘qimalar ishlab chiqishning dolzarb masalalari.	350
Agzamova S.M., Gulyayeva G.X., Mukimov M.M., Karimova N.H. Turli xomashyoli ikki qatlamli trikotaj to‘qimalarining sifat ko‘ratkichlarini kompleks baholash.	354
ANIQ VA IJTIMOIIY-IQTISODIY FANLAR	
Mannonov Y.O. Kitobxonlik madaniyatini oshirishning ayrim masalalari.	360
Мухамадиев Б.Т, Мухамадиева З.Б., Джамалов С.Ш. Значение сбалансированного кишечного микробиома в укреплении физического и психологического здоровья	363
Xakimova M.Y. Elektron kutubxonani yaratishda zamonaviy dasturiy-texnik ta’minot	372
Шарипов М.З., Назаров Э.С., Зокирова З.М., Иброхимова С.Ш., Муминов Ш.Х. Структура и свойства одномерных микромагнитных конфигураций в ферромагнитных кристаллах.	375
EKOLOGIYA VA ATROF MUHIT MUHOFAZASI	
Мамедов Р.А., Жамолов Ж.Ж. Оценка низконапорного гидроэнергетического потенциала водозаборных сооружений и ирригационных каналов.	381
Мамедов Р.А. Исследование зависимости выхода водорода от термохимических характеристик электролита в процессе электролиза.	391

УДК 004.93

STATISTICAL ANALYSIS OF SOCIAL NETWORK MESSAGES

¹Babomuradov O.J., ²Kuyliyeva F.A.

¹*Jizzakh branch of kazan federal university*, ²*Tashkent state agrarian university*.

Abstract. This article highlights the theoretical and practical aspects of the intellectual analysis of social network communications using statistical methods. It examines the possibilities of discovering hidden knowledge in texts through techniques such as text digitization, word frequency, TF-IDF, N-gram, clustering, and classification methods. The study also demonstrates, based on scientific sources, that these approaches have been successfully applied in various fields in recent years, particularly in social, legal, biomedical, and media domains.

Keywords: text mining, social network, statistical analysis, TF-IDF, N-gram, clustering, classification.

IJTIMOY TARMOQ XABARLARINING STATISTIK TAHLILI

¹Babomuradov O.J., ²Qoyliyeva F.A.

¹*Kazan Federal Universiteti Jizzax filiali*, ²*Toshkent davlat agrar universiteti*.

Annotatsiya. Ushbu maqolada ijtimoiy tarmoqdagi yozishmalarni intellektual tahlil qilishning nazariy va amaliy jihatlari yoritilgan. Matnlarni raqamlashtirish, soʻz chastotasi, TF-IDF, N-gram, klasterlash va klassifikatsiya kabi usullar orqali matnlarda yashirin bilimlarni aniqlash imkoniyatlari koʻrib chiqilgan. Tadqiqot ilmiy manbalar asosida ushbu yondashuvlarning soʻnggi yillarda ijtimoiy, huquqiy, biotibbiy va ommaviy axborot sohalarida muvaffaqiyatli qoʻllanilganini koʻrsatadi.

Kalit soʻzlar: matn qazib olish (text mining), ijtimoiy tarmoq, statistik tahlil, TF-IDF, N-gram, klasterlash, klassifikatsiya.

Introduction. In recent years, the rapid expansion of digital technologies and online communication platforms has led to an unprecedented growth in textual data. As a result, text mining – also known as text analytics – has become one of the most important tools in both scientific research and real-world applications. A significant portion of human-generated information today is expressed in the form of unstructured text, including news articles, blogs, social media posts, comments, reviews, and official documents. This massive volume of text data contains valuable insights about human behavior, opinions, and trends, which can be discovered only through systematic analysis.

The main objective of text mining is to extract meaningful and previously unknown knowledge from large collections of unstructured texts. It involves identifying hidden patterns, semantic relationships, and statistical regularities that can assist in decision-making processes, policy development, and predictive modeling. By integrating techniques from natural language processing (NLP), machine learning, and data mining, text mining allows researchers to automatically process and interpret huge text corpora that would be impossible to analyze manually.

With the growth of social networks such as Telegram, Twitter, and YouTube, the importance of text mining has expanded even further. These platforms generate millions of user-generated messages daily, making it crucial to develop algorithms capable of filtering, categorizing, and evaluating their content. Text mining methods help identify toxic, extremist, or harmful messages, detect sentiment polarity, and reveal communication trends within online communities.

Therefore, text mining today serves as a bridge between linguistic research and computational analysis, providing valuable tools for understanding digital communication in modern society. It supports government agencies, media analysts, and academic researchers in discovering knowledge from text, ensuring that large-scale communication data can be effectively used for social, political, and technological insights. [1].

The statistical method converts text into digital form and analyzes it through mathematical and statistical calculations. The advantage of this approach is the ability to process large volumes of text quickly. The main methods are as follows:

- Word Frequency – counting the most frequent words;
- TF-IDF (Term Frequency – Inverse Document Frequency) – identifying important words in a corpus of texts;
- N-gram analysis – studying the repetition system of word combinations;

- Clustering – grouping similar texts;
- Classification – dividing texts into predefined categories [2–4].

Word Frequency and Zipf's Law

In studies related to word frequency, two main concepts are used:

- Absolute frequency (how many times a word occurs in a text)
- Relative frequency (the proportion of a word relative to the entire corpus).

Based on frequency-based statistical observations, Zipf's law plays a significant role in linguistics and text mining. It states that in any natural language corpus, the frequency of a word is inversely proportional to its rank in the frequency table. In other words, the most common words occur extremely frequently, while rare words appear only a few times. For example, in English, words such as “the,” “and,” “of,” and “to” dominate most texts, whereas words describing specific concepts, emotions, or events occur infrequently.

The implications of Zipf's law for text mining are profound. It suggests that a relatively small number of high-frequency words carry a large portion of the overall word occurrences, while a vast number of low-frequency words contribute to the richness and diversity of language. This insight allows researchers to compress and process large corpora efficiently by focusing on statistically significant terms while filtering out extremely common or extremely rare words that add noise rather than meaning.

However, this efficiency comes with a limitation. Although statistical models like word frequency analysis and TF-IDF enable rapid and automatic processing of large amounts of text, they cannot fully capture the semantic depth, contextual meaning, or pragmatic intent behind words and phrases. For example, the same word may convey different meanings depending on context (“bank” as a financial institution or river edge), which pure frequency-based models cannot differentiate. [5–6].

Recent studies (2023–2025) clearly demonstrate the effectiveness of this approach. For example, in [5], the method was effectively applied in various fields: in media discourse, word frequency was calculated for over 10,000 news headlines to identify the main topics. The results showed that technological terms such as “Big Tech,” “AI,” and “ChatGPT” were dominant, while social issues such as “jobs,” “diversity,” and “ethics” appeared in only about 6 percent of the material. In another field, [6] attempted to detect traces of LLMs (Large Language Models) in biomedical texts using word frequency analysis. By examining the frequency of certain stylistic units in more than 15 million scientific abstracts, it was concluded that at least 13.5 percent of papers written in 2024 might have been generated with the help of artificial intelligence.

TF-IDF Method. The TF-IDF approach is one of the most widely used statistical techniques in text mining and information retrieval. Its primary goal is to identify and quantify how important a particular word is within a document relative to a collection of documents (the corpus). It does so by combining two distinct but complementary measures: Term Frequency (TF) and Inverse Document Frequency (IDF).

The Term Frequency (TF) component measures how often a word appears in a single document. Words that occur more frequently within a document are considered more representative of that document's topic. However, very common words such as “the,” “and,” or “is” often appear in nearly every text, and thus do not provide meaningful distinctions between documents.

As a result, words that occur frequently in a document but rarely across the corpus receive a higher weight. The TF-IDF method combines local and global importance of words. Today, TF-IDF still retains its importance in text analysis, and despite the emergence of complex models, it remains a fundamental method due to its simplicity, speed, and interpretability [7–8].

For example, in a study by Chen et al. (2023), comments in medical forums were analyzed using TF-IDF, and the sentiment model achieved 9–12% higher accuracy [7]. In [8], 200,000 tweets about COVID-19 were analyzed using TF-IDF, and logistic regression and SVM models achieved

fast and resource-efficient results. Similarly, in [11], the TF-IDF method was used to automatically classify court decisions with 87% accuracy.

N-gram. N-gram analysis in text mining examines not only individual words but also the sequential relationships between them. An N-gram represents a contiguous sequence of N linguistic units — which may be words, characters, or syllables — extracted from a text. For example, a bigram (N=2) consists of two consecutive words such as “artificial intelligence”, while a trigram (N=3) might be “text mining analysis”. The choice of N determines the level of contextual detail captured: smaller N-grams (e.g., unigrams, bigrams) represent local word co-occurrence, whereas larger N-grams (e.g., trigrams or 4-grams) capture more complex syntactic and semantic patterns.

N-gram models play a central role in language modeling — estimating the probability of a word given its preceding context. This capability makes them foundational for many natural language processing (NLP) tasks, including machine translation, speech recognition, text classification, sentiment analysis, spam detection, and even neurolinguistic research. By examining frequent or rare N-grams, researchers can identify stylistic tendencies, domain-specific expressions, or emotional tones that are not apparent from single-word frequency analysis alone.

In practical applications, N-gram analysis helps quantify contextual meaning and identify collocations — words that commonly occur together, forming stable linguistic units such as idioms or technical terms. For instance, the bigram “machine learning” or the trigram “deep neural network” provides far more meaningful information than analyzing “machine,” “learning,” or “deep” separately. This ability to capture contextual dependencies makes N-gram modeling an essential bridge between simple statistical approaches (like TF-IDF) and more sophisticated deep learning-based language models. [9–10].

N-gram analysis has been widely used as a main analytical tool in many studies. For example, [9] used bigram and trigram features to detect hate speech in social networks. Based on more than 50,000 Twitter posts, it was noted that when trigram features were added, accuracy increased by 6 percent, showing that N-gram expanded the context and improved semantic precision. Similarly, [10] analyzed large German newspaper archives using N-gram analysis to observe how key terms in political discourse changed between 2000–2020. The combinations of “Klimawandel” (climate change) bigram and trigram sharply increased since 2015 and were more often used with “Energiewende” (energy transition) in political texts. This demonstrates the effectiveness of this approach in studying historical discourse. N-gram also plays an important role in psycholinguistics.

Clustering. In text mining, not only the frequency or statistical weight of words is important, but also the degree of similarity between documents. Understanding these relationships allows researchers to organize large text collections into meaningful groups and uncover hidden thematic structures. For this purpose, clustering methods are applied. Clustering represents an unsupervised learning approach, meaning that it groups documents or textual units based on similarity without relying on predefined categories or labels.

Theoretically, clustering algorithms first represent each document as a vector in a high-dimensional feature space. These feature vectors can be constructed using various text representations such as TF-IDF, Word2Vec, or BERT embeddings, depending on the desired balance between interpretability and semantic richness. Once the vector representations are obtained, clustering is performed using similarity metrics like cosine similarity, Euclidean distance, or Manhattan distance. These metrics measure how close or distant documents are in the multidimensional space, allowing similar texts to be grouped together and dissimilar ones separated into different clusters.

Among the numerous clustering algorithms, several stand out as the most commonly used in text mining. The K-Means algorithm partitions data into a predefined number of clusters by minimizing intra-cluster variance. It is efficient and scalable, making it suitable for large-scale text datasets such as social media messages. Hierarchical clustering, on the other hand, builds a tree-like structure (dendrogram) that represents nested groupings of documents, enabling analysis at various

levels of granularity. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is particularly useful for identifying clusters of arbitrary shape and handling outliers or “noise” texts that do not belong to any group. Gaussian Mixture Models (GMM) offer a probabilistic interpretation of clusters, assuming that data points are generated from a mixture of Gaussian distributions. Meanwhile, Spectral Clustering uses graph-based methods and eigenvalue decomposition to separate data based on similarity graphs, often yielding more flexible results for complex semantic patterns [11–12].

Clustering is widely used in research and practical projects. For example, in [11], hierarchical clustering was applied to a large news article archive, analyzing 120,000 articles from 2010–2022 and identifying political, economic, cultural, and sports topics as natural clusters. Hierarchical clustering provided higher semantic accuracy compared to K-means because its tree structure allowed the identification of smaller subtopics. In [12], the DBSCAN algorithm was used to detect plagiarism in academic papers. By analyzing 30,000 student works, suspiciously similar documents were identified through clustering, determining the degree of plagiarism. DBSCAN was noted to be more effective than traditional similarity metrics in detecting outlier documents.

Classification. Classification is one of the most fundamental and widely used approaches in text mining and natural language processing. The main purpose of classification is to automatically assign each text document to one or more predefined categories based on its content. Theoretically, classification represents a supervised learning paradigm, meaning that it requires a labeled dataset — a collection of examples where the correct category (or “label”) for each text is already known. Using these examples, the model learns to recognize patterns and linguistic cues that distinguish one class from another, allowing it to predict the appropriate class for new, unseen texts.

Text classification has a broad range of practical applications, including spam filtering, sentiment analysis, topic detection, fake news identification, and harmful content detection in social networks. Depending on the size, complexity, and nature of the dataset, different algorithms can be applied, ranging from traditional statistical models to advanced deep learning architectures [13–14].

Classification methods are effectively applied in various fields. In [13], several classification algorithms were compared to detect hate speech in social networks. On a dataset of 120,000 tweets, Logistic Regression and SVM performed well with TF-IDF, while the BERT-based classifier achieved the highest result with 92% accuracy. In [14], 500,000 articles were collected and classified using Naive Bayes, Random Forest, and Transformer models. The results showed that the Transformer-based RoBERTa model achieved 95% accuracy, while Naive Bayes remained effective in low-resource conditions.

Conclusion. Statistical approaches — such as word frequency analysis, TF-IDF weighting, N-gram modeling, clustering, and classification — provide a robust foundation for the quantitative analysis of textual data. These methods are especially effective for exploring patterns within social network messages, where data are often massive in scale yet limited in semantic richness. By focusing on numerical relationships and probabilistic distributions, statistical models make it possible to identify trends, detect topic clusters, and classify messages according to their semantic or emotional tone with relatively low computational cost.

Moreover, statistical techniques offer a high level of interpretability. Their outputs — frequencies, similarity measures, and cluster boundaries — are transparent and easily validated, which is crucial for research areas such as social network analysis and digital communication studies. In comparison to deep learning approaches, which often operate as “black boxes,” statistical methods remain valuable for preliminary text analysis, feature extraction, and for projects where explainability is a priority.

Therefore, despite the growing popularity of transformer-based and neural models, statistical methods continue to play an important role in natural language processing. Their balance of speed, interpretability, and computational efficiency ensures their relevance in both academic research and applied data analysis of social media communications. Combining these methods with modern

machine learning or deep learning models can further enhance the precision and insightfulness of text analytics in the context of social network message evaluation.

References

1. Aggarwal C.C., Zhai C. Mining Text Data. – Springer, 2012. – 454 б.
2. Zare D., Fernandez-Sanz L., Pospelova V., Lopez-Balominos I. A Comprehensive Survey on Text Mining: From Theory to Practice. – ResearchGate, 2025. – 38 б.
3. Alebqbie H.J., Sfoq M.S., Albeer R.A. A Review of Text Mining Techniques: Trends and Applications in Various Domains. – ResearchGate, 2024. – 42 б.
4. A comprehensive survey of text classification techniques. – Journal of Applied Logic. – 2024. – ScienceDirect. – № 3. – Б. 45–78.
5. Karanouh M. Mapping ChatGPT in Mainstream Media to Unravel Jobs and Diversity Challenges. – arXiv, 2023. – 15 б.
6. Kobak D., González-Márquez M., Horvát E.-Á., Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. – arXiv, 2024. – 21 б.
7. Chen Y. ва бошқалар. Using TF-IDF features for sentiment analysis in medical forums. – Journal of Biomedical Informatics, 2023. – Т. 145. – Б. 112–124.
8. Almeida R., Costa F. TF-IDF based approaches for misinformation detection in Twitter data. – Social Network Analysis and Mining. – 2024. – Т. 13. – № 2. – Б. 65–79.
9. Rahman M., Karim S. Bigram and trigram features for hate speech detection on social media. – Computers in Human Behavior Reports. – 2023. – Т. 10. – № 1. – Б. 44–56.
10. Müller T., Schmid H. N-gram analysis of German political discourse (2000–2020). – Journal of Language and Politics. – 2024. – Т. 23. – № 3. – Б. 212–234.
11. Petrov I., Smirnova A. Hierarchical clustering of news articles for topic discovery. – Journal of Computational Linguistics. – 2024. – № 5. – Б. 99–115.
12. Rahimi M., Ahmad N. Using DBSCAN clustering for plagiarism detection in student essays. – Education and Information Technologies. – 2024. – Т. 29. – № 4. – Б. 180–193.
13. Ahmed S., Qureshi M. Classification of hate speech on Twitter using TF-IDF and deep learning methods. – Computers in Human Behavior Reports. – 2023. – Т. 9. – № 2. – Б. 140–155.
14. Li Y., Zhao H. Automated classification of large-scale news articles: A comparison of traditional and transformer-based approaches. – Journal of Information Science. – 2024. – Т. 50. – № 1. – Б. 37–52.

Babomuradov Ozod Jo'rayevich – D.Sc., Executive Director of the Jizzakh Branch of Kazan Federal University, Tel.: +998 97 403 76 11 E-mail: babomuradovozod@gmail.com

Qo'liyeva Feruzakhon Alisher qizi – Assistant Lecturer at the Department of Information Systems and Technologies, Tashkent state agrarian university, Tel.: +998 88 333 33 66 E-mail: qoyliyevaferuza@gmail.com

**O‘ZBEKISTON RESPUBLIKASI OLIY TA’LIM,
FAN VA INNOVASIYALAR VAZIRLIGI
BUXORO DAVLAT TEXNIKA UNIVERSITETI**

Buxoro davlat texnika universiteti nashri
“FAN VA TEKNOLOGIYALAR TARAQQIYOTI”

Ilmiy-texnikaviy jurnal

2026 yil, 1-son. ISSN 1281-8193

Jurnal tahririyat kompyuterida sahifalandi.

Manzil: 200117, Buxoro sh., Q.Murtazoev ko‘chasi 15.

Tel: (8-365) 223-92-40, E-mail: fantt_jurnal@umail.uz

<http://journal.bstu.uz>

БУХАРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Издатель: Бухарский государственный технический университет

«РАЗВИТИЕ НАУКИ И ТЕХНОЛОГИЙ»

Научно-технический журнал

Выпуск 1, 2026. ISSN 1281-8193

Отредактирована на компьютере редакции.

Адрес: 200117, г. Бухара, ул. К. Муртазаева, 15.

Тел.: (8-365) 223-92-40, E-mail: fantt_jurnal@umail.uz

<http://journal.bstu.uz>

BUKHARA STATE TECHNICAL UNIVERSITY

Publisher: Bukhara State Technical University

DEVELOPMENT OF SCIENCE AND TECHNOLOGY

Scientific and Technical Journal

Issue 1, 2026. ISSN 1281-8193

Yedited on the yeditorial computer.

Address: 200117, Bukhara, K. Murtazaev Street, 15

Tel.: (8-365) 223-92-40, E-mail: fantt_jurnal@umail.uz

<http://journal.bstu.uz>