

**BASHORATLI TAHLILLAR UCHUN MASHINALI O'QITISH  
ALGORITMLARI. QIYOSIY QARASHLAR*****Onarqulov Maqsad Karimberdiyevich****Farg'ona davlat universiteti f.m.f.b.f.d(PhD) o'qituvchisi****Yoqubjonov Alijon Mahamadamin o'g'li****Farg'ona davlat universiteti amaliy matematika yo'nalishi 2-kurs magistranti****Mahamadaminova Shahina Valijon qizi****Farg'ona davlat universiteti matematika fakulteti amaliy matematika yo'nalishi 3-kurs talabasi*

**Abstrakt.** Ma'lumotlar to'plamining borgan sari oshib borishi va ular asosida kelajakda sodir bo'ladigan analitik jarayonlarni o'rganish insoniyatga juda katta foyda keltirayotganligi tufayli optimal bashoratli tahlillar mashhurlikka erishdi va bunda mashinali o'qitish algoritmi muhim ro'l kasb etgan. Ko'pgina bashorat qilish uchun yaratilgan mashinali o'qitish algoritmlari odatda bashoratli tahlil uchun qo'llaniladi, ammo bu ma'lumotlar to'plamidagi o'rganilayotgan muammo uchun to'g'ri algoritmi tanlash hali ham qiyin vazifadir. Ushbu maqola bashoratli tahlil uchun eng mashhur mashinali o'qitish algoritmlari va ularni ma'lumotlar to'plamida qo'llash natijalari tahlili haqida umumiy ma'lumot beradi. Ushbu algoritmlar ishlash vaqti, aniqlik, sezgirlik, o'ziga xoslik va xatolarga yo'l qo'yish kabi qiyosiy ishlash ko'rsatkichlari bilan baholanadi va tahlil qilinadi.

Due to the increasing number of data sets and the fact that the study of analytical processes taking place in the future on their basis brings enormous benefits to humanity, optimal predictive analysis has gained popularity, and in this the machine learning algorithm has acquired an important role. Machine learning algorithms created for most predictions are usually used for predictive analysis, but choosing the right algorithm for the problem under study in this data set is still a difficult task. This article provides an overview of the most popular machine learning algorithms for predictive analysis and the analysis of the results of their application in a data set. These algorithms are evaluated and analyzed by comparative performance indicators such as run time, accuracy, sensitivity, specificity, and error-making.

Благодаря все большему количеству наборов данных и тому факту, что изучение аналитических процессов, происходящих на их основе в будущем, приносит огромную пользу человечеству, популярность приобрела оптимальная прогнозная аналитика, важную роль в которой сыграл алгоритм машинного обучения. Многие алгоритмы машинного обучения, предназначенные для прогнозирования, обычно используются для прогнозного анализа, но выбор правильного алгоритма для изучаемой проблемы в этом наборе данных по-прежнему является сложной задачей. В этой статье представлен обзор наиболее популярных алгоритмов машинного

обучения для прогнозного анализа и анализ результатов их применения в наборе данных. Эти алгоритмы оцениваются и анализируются с использованием сравнительных показателей производительности, таких как время безотказной работы, точность, чувствительность, специфичность и вероятность ошибок.

**Kalit so'zlar:** Mashinali o'qitish, Bashoratli tahlil, KNN-K eng yaqin qo'shni, SVM-Vektorlarni qo'llash mashinasi, Naive Bayes.

Machine learning, Predictive analysis, KNN- K nearest neighbor, SVM-Support vector machine, Naïve Bayes.

Машинное обучение, Прогнозный анализ, KNN-K-ближайших соседей, SVM- метод опорных векторов, Naïve Bayes.

**Bashoratli tahlillar uchun mashinali o'qitish algoritmlari. qiyosiy qarashlar.** Bashoratli tahlil qilish - bu kelajakdagi voqealarni yoki kutilmagan vaziyatlarni sezilarli darajada aniqlik bilan bashorat qilish uchun joriy va avvaldan mavjud bo'lgan ma'lumotlarni tahlil qilishning eng keng tarqalgan va keng qo'llaniladigan usulidir. Bu ma'lumotlar mayningi, statistik tahlil va mashinali o'qitish kabi keng ko'lamli analitik usullarni o'z ichiga oladi. Mashinali o'qitish esa - bu tizimlarning ma'lumotlar asosida o'qitilishi, modellarni tanish va minimal inson ishtiroki bilan bashorat qilish imkonini beruvchi statistik va analitik modellarni yaratishni avtomatlashtiradigan sun'iy intellektning bir tarmog'idir. Bashoratli tahlillar uchun eng mashhur va keng qo'llaniladigan mashinali o'qitish algoritmlariga Logistik regressiya, KNN, SVM, qaror daraxtlari va tasodifiy o'rmon algoritmlarini misol qilib keltirish mumkin.

Mashinali o'qitishning bashorat qilish usullari va tahlillini talab qiladigan sog'liqni saqlash, sanoat, moliya, qishloq xo'jaligi, ta'lim, ijtimoiy media, kiberxavfsizlik va boshqa ko'plab sohalarni o'z ichiga oladi.

Birinchi navbatda turli tadqiqotchilar tomonidan olib borilgan tegishli tadqiqot ishlarini muhokama qilsak va ularning eng dolzarb topilmalari qayd etsak. Bu tegishli bashorat qilish algoritmlarni ko'rib chiqish uchun tanlovni yo'naltiradi. Shundan so'ng, ular asosida tanlangan strategiyalarni afzalliklari va kamchiliklariga oydinlik kiritiladi. Keyin esa amalga oshirish uchun ko'rib chiqilgan ma'lumotlar to'plamining qisqacha ko'rinishi beriladi.

Bu kungacha ko'plab tadqiqotchilar maqsadli iste'molchilarni qidirish uchun yanada samarali marketing va ijtimoiy media tahlillarining marketing strategiyalarini tanlashda bashorat qilish usullari va tahlillarining imkoniyatlari va muammolari bo'yicha tadqiqotlarni taqdim etdilar. Yana bir tadqiqod natijalariga ko'ra "intraoperative" bashorat tahlillarini sog'liqni saqlash tizimining elektron bazalardagi mavjud ma'lumotlarga asoslanib mashinali o'qitish usullaridan foydalangan holda amalga oshiradi. Tadqiqodchilar post-induksion gipotenzivni bashorat qilish vazifasi uchun sakkizta eng yaxshi mashinani o'rganish algoritmlarini amalga oshirdilar va qo'shimcha parametrlarni sozlash test to'plamiga qo'llanilishidan oldin ular orasida eng yaxshi bashoratchi yordamida amalga oshiriladi. Eksperimental natijalar shuni ko'rsatdiki, o'quv jarayonida ham,

test jarayonida ham gradientni kuchaytirish boshqa barcha yondashuvlarga nisbatan aniq diskriminatsiyani ko'rsatdi.

2017 yilda Intellektual hisoblash va boshqarish tizimlari (ICICCS) bo'yicha xalqaro konferentsiyada sog'liqni saqlash sohasida mashinani o'rganish algoritmlari yordamida samarali hal qilinishi mumkin bo'lgan prognozlash muammolari ko'rib chiqiladi. Bunda yurak-qon tomir kasalliklari bashoratlari, diabet bashoratlari, gepatit va saraton bashoratlari va boshqalarni keng ko'lamda muhokama qilishadi. Tadqiqodchilar ushbu bashorat qilish vazifalari uchun ishlatiladigan CART, Naïve Bayes, RBF, SVM, Simulated Annealing (SA) kabi umumiy mashinali o'qitish algoritmlarini aniqladilar. Kelgusi ishlarda ko'rib chiqilishi mumkin bo'lgan ba'zi cheklovlar, masalan, analitik modellashtirishning katta ma'lumotlar tizimlari bilan to'liq integratsiyalashuvi va quvvat sarfidan tashqari ko'plab boshqa muhim barqarorlik natijalariga joriy etishni kengaytirish kabilar muhokama qilindi.

Ushbu maqolada bashorat qilish usullari va tahlillari uchun keng qo'llaniladigan mashinali o'qitishning "Naive Bayes", "SVM" va "KNN" algoritmlari asosida olib borilgan tadqiqod natijalarini e'tiboringizga havola qilinadi. Tegishli ravishda bashorat qilish usullari va tahlillar bilan ular qisqacha o'rganiladi.

**Naive Bayes (NB)**- bu eng oddiy Bayes tarmog'i modeli ostida *supervised* mashinali o'qitish algoritmi sifatida tanilgan "ehtimoliy tasniflagichlar" bo'lib, Bayes teoremasini atributlar orasidagi erklilik haqidagi sodda taxminlar bilan amalga oshiradi. Bayes teoremasiga yanada soddaroq nazar tashlaydigan bo'lsak boshqa ehtimollarni bilgan xolda ehtimollikni toppish usuli deb atasak ham bo'ladi. Naiv Bayes klassifikatorlarining afzalligi yuqori darajada masshtabli bo'lib, o'quv muammosidagi atributlar sonini bir qator parametrlarga simmetrik bo'lishiga imkon beradi. U birinchi marta matnni qayta ishlashga kiritilgan va matn va hujjatlarni, masalan, firibgar yoki qonuniy ekanligini atribut sifatida so'z chastotalari bilan turkumlash uchun asosiy vosita hisoblangan, shuningdek, avtomatlashtirilgan tibbiy diagnostikada ham keng qo'llanilgan. Naïve Bayes algoritmlarini qo'llagan xolda analitik modellar yaratilgan va ular asosida chiqarilgan natijalar klinik ma'lumotlarga to'g'ri kelganligi sifatida o'zini oqlagan.

**SVM**- bu tasniflash va regressiya tahlili muammolari uchun ishlatiladigan shunga o'xshash *supervised* algoritmlari to'plamidir. U chiziqli klassifikatorlar oilasiga mansub bo'lib, Tixonov tartibga solishning alohida holati sifatida ham ko'rib chiqilishi mumkin. SVMning o'ziga xos xususiyati shundaki, empirik tasniflash xatosi doimiy ravishda kamayadi va bo'shliq oshadi, shuning uchun usul ***maksimal bo'shliqni tasniflash usuli*** sifatida ham tanilgan.

Usulning asosiy g'oyasi asl vektorlarni yuqori o'lchamli fazoga aylantirish va bu bo'shliqda eng katta bo'shliqqa ega bo'lgan ajratuvchi giperplanani qidirishdir. Sinflarni ajratib turadigan giperplanning ikkala tomonida ikkita parallel giperplanlar qurilgan. *Ajratuvchi giperplane* ikkita parallel giper tekislikgacha eng katta masofani yaratadigan giper tekislik bo'ladi. Algoritm ushbu parallel

giperplanlar orasidagi farq yoki masofa qanchalik katta bo'lsa, o'rtacha klassifikator xatosi shunchalik kichik bo'ladi degan taxminga asoslanadi.

**KNN** oddiy parametrik bo'lmagan *supervised* mashinali o'qitish algoritmi bo'lib, tasniflash muammolari va regressiya muammolarini hal qilish uchun ishlatiladi. Bu shunchaki ma'lumotlarni aniqlash uchun xayoliy chegara yaratadi va yangi ma'lumotlar nuqtalari uchun eng yaqin chegara chizig'ini bashorat qilishga harakat qiladi, ammo ma'lumotlarning hajmi oshgani sayin jarayon keskin sekinlashib boradi.

1-jadval. Mashinali o'qitish usullari

Strategiya	Afzalliklar	Cheklovlar
Naïve Bayes	Agar erklilik prezumpsiyasi mavjud bo'lsa, u boshqa algoritmlarga qaraganda samaraliroq ishlaydi. Yuqori darajada kengaytiriladigan va doimiy va diskret ma'lumotlarni qo'llab-quvvatlashi mumkin	O'zgaruvchilar orasidagi assotsiatsiyani integrasiya qilib bo'lmaydi. Yana bir kamchilik – bu real dunyoda mutlaqo mustaqil bashoratchilar to'plamini olish deyarli amaliy emasligi haqidagi mustaqil bashoratchilarning prezumpsiyasi
Support Vector machine	Kategoriyalar orasidagi farqlanishning aniq chegarasi mavjud bo'lganda, u umumlashtirishda yaxshi ko'rsatkichga ega. Bundan tashqari, yuqori o'lchamli bo'shliqlar uchun yanada samarali.	Bu katta ma'lumotlar to'plamlari uchun maqbul emas va maqsadli sinflar bir-biriga mos kelganda unchalik yaxshi ishlaymaydi
K-nearest neighbor	Amalga oshirish va tushunish oson. Algoritm qidiruv maydonida mustahkam va hisoblashda kamroq vaqt talab etadi	Ma'lumotlar to'plami hajmi kattalashganda yomon ishlaydi. Bu sekin o'rganadi, shuning uchun hisoblash uchun ko'proq vaqt talab etiladi.

1-jadvalda ko'rinib turibdiki har bir algoritmnining afzalliklari va kamchiliklari bor. Ushbu algoritmlarni moliya sektoridan olingan kredit karta ma'lumotlar to'plamiga qo'llab ko'ramiz. Quidagi jadvalda kredit kartalarning ma'lumotlari to'plamining tavsifi va xususiyatlari keltirilgan keltirilgan:

2-jadval. Ma'lumotlar to'plami tavsifi.

Ma'lumotlar to'plami	Turi	Xususiyatlar	Na'munalar	Sinflar	Nisbat(holatlardan hususiyatlarga)	Domen
Kredit karta	Ikkilik tur	24	30000	2	1250,0	Moliya
Xodimlarni saqlash	Ikkilik tur	10	14998	2	1499,0	Sanoat
Uy joy	Ikkilik	18	497	2	27.61	Moliya

ma'lumotlari	tur					
--------------	-----	--	--	--	--	--

Standart to'lov uchun ( $H_0/Y_0 = 1/0$ ) 24 ta atributlar orasidagi qaram o'zgaruvchidir, qolgan 23 tasi esa bashorat qiluvchi o'zgaruvchilardir.

2-jadvalda tasvirlangan kredit karta ma'lumotlar to'plamida NB, SVM va KNN ning amalga oshirish natijalari va qiyosiy tahlilini ko'rib chiqamiz. Bu foydalanuvchilarning moliyaviy tranzaksiya shakllari va tarixiga asoslangan sukut to'lovlari bashorat qilishga qaratilgan. Bu bashoratli tahlil to'rt bosqichdan iborat:

- dastlabki ishlov berish bosqichi, unda xomaki ma'lumot yig'iladi va oldindan qayta ishlanadi;

- oldindan qayta ishlangan ma'lumotni osongina boshqarilishi mumkin bo'lgan shaklga (muammo talabiga ko'ra eng mos ma'lumot yoki atributlarni tanlash orqali) aylantirish uchun to'g'ri mashinali o'qitish algoritmi qo'llaniladigan ma'lumotlarni o'zgartirish yoki xususiyatni tanlash bosqichi;

- o'quv modeli o'zgartirilgan ma'lumotlardan foydalangan holda tuzilgan o'quv bosqichi;

- hisobot berish uchun avval yaratilgan o'quv modelidan foydalanadigan sinov yoki bashorat bosqichi.

Ushbu bosqichlardagi qadamlar sosidagai tajribalar ko'plab dasturlash tilari orasida mashhur bo'lgan Python dasturlash muhitida o'tkazildi. Bu ko'plab oldindan ishlov berish jarayonlarini va ma'lumotlarni tahlil qilishni osonlashtirdi. Ayniqsa python dasturlash tilidagi transformatsiya kutubxonalarini tushunarsiz bo'lgan ma'lumotlar to'plamini qayta ishlashni amalga oshiruvchi jarayonlarni yuqori aniqlikda ko'rsatib bera oldi.

Tegishli turli xil ishlash ko'rsatkichlari bilan solishtirganda bashorat qiluvchi algoritmlarni amalga oshirish natijalari aniqlik balliga ko'ra SVM boshqa algoritmlardan ustundir. Biroq, SVM algoritmi uchun talab qilinadigan vaqt boshqa algoritmlarga qaraganda ko'proq sarflandi. Naïve Bayes algoritmi esa tajribalar vaqtida KNN bilan bir xil aniqlik natijalarini ko'rsatdi.

#### Foydalanilgan adabiyotlar

1. Onarqulov, M., Yaqubjonov, A., & Yusupov, M. (2022). COMPUTER NETWORKS AND LEARNING FROM THEM OPPORTUNITIES TO USE. Models and methods in modern science, 1(13), 59-62.
2. Онаркулов Максад. (2021). Influence of the temperature and cyclic deformations of  $(\text{Bi}_x\text{Sb}_{1-x})_2\text{Te}_3$  films on their resistance. *Инженерно-Физический журнал. Беларусь*, 94 (5), 1403-1408.
3. Абдулазиз угли, Ю. М., Каримбердиевич, О. М., & Махамедин угли, Ё. А. (2022). АЛГОРИТМЫ РАСПОЗНОВАНИЯ РЕЧИ И КЛАССИФИКАЦИЯ МЕТОДОВ РАСПОЗНОВАНИЯ РЕЧИ. CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES, 3(10), 15-19.
4. Rajeshkanna, A., Preetha, V. & Arunesh, K. (2019 ). Mobil tarmoq provayderlarini tasniflash vazifasida mashinali o'qitish algoritmlarini eksperimental tahlil qilish. Elektron biznes va telekommunikatsiyalar

- bo'yicha xalqaro konferentsiya (335-343).
5. Kendale, S., Kulkarni, P., Rosenberg, AD, & Wang, J. (2018). Postinduksion gipotenziyani bashorat qilish uchun nazorat qilinadigan mashinani o'rganish bashoratli tahlili. *Anesteziologiya*, 129 (4), 675-688.
  6. Онаркулов, М. К., Отажонов, С. М., Ботиров, К. А., Юнусов, Н., Мамаджонов, У. М., & Каххорова, Б. А. (2020). УСТРОЙСТВО ДЛЯ ИЗУЧЕНИЯ ТЕНЗОЧУВСТВИТЕЛЬНОСТИ В ФОТОЧУВСТВИТЕЛЬНЫХ ПОЛУПРОВОДНИКОВЫХ ПЛЕНКАХ. *Главный редактор: Ахметов Сайранбек Махсутович, д-р техн. наук; Заместитель главного редактора: Ахмеднабиев Расул Магомедович, канд. техн. наук; Члены редакционной коллегии*, 55.
  7. Nithya, B. va Pango, V. (2017). Mashinani o'rganish vositalari va usullaridan foydalangan holda sog'liqni saqlash sohasida bashoratli tahlil. 2017 yilda Intellektual hisoblash va boshqarish tizimlari (ICICCS) bo'yicha xalqaro konferentsiya (492-499-betlar). IEEE.
  8. Lin, J. va Kolcz, A. (2012). Twitterda keng ko'lamli mashinani o'rganish. Ma'lumotlarni boshqarish bo'yicha 2012 yilgi ACM SIGMOD xalqaro konferensiyasi materiallari (793-804).