

MASHINALI O'QITISHGA ASOSLANGAN MATNNI TASNIFLASH ALGORITMLARI

assistent *Choryorqulov G'iyos Husan o'g'li*

O'zbekiston Milliy universitetining Jizzax filiali, O'zbekiston

choryorqulov@jbnuu.uz

Annotatsiya. Ushbu maqolada matn va hujjatlarni tasniflash algoritmlarining cheklanishlari va afzalliklari tajribaviy tadqiqotlar asosida tahlil qilingan. Umuman olganda, ushbu bobda mashinali o'qitish algoritmlari matnni tasniflash uchun foydalaniladi, elektron hujjat aylanish tizimi esa elektron hujjatlarni tahlil qilish va kategoriyalash uchun ishlatiladi.

Kalit so'zlar. Mashinali o'qitish algoritmlari, chiziqli regressiya, logistik regressiya, dataset, xeshlash.

Tajribaviy tadqiqot uchun Toshkent axborot texnologiyalari universiteti Samarqand filialiga kelgan 4 ta kategoriyaga tegishli bo'lgan axborotlar olindi.

	ma'naviy	ilmiy	moliyaviy	o'quv ishlari
0	O'zbekiston Respublikasi Vazirlar Mahkamasini	yosh tadqiqotchi	moliyaviy mustaqillik	Fahriyat
1	Prezident	maqola	kredit modul tizimi	oily ta'lim muassasasi
2	ilmiy-iqtisodiy rivojlanish	tezis	El-yurt umid jamg'armasi	professor-o'qituvchi
3	mahalliy mehnat bozori	ilmiy tashkilotlar	Iqtisodiy taraqqiyot va kambag'allikni qisqart...	sodimlar
4	qiziqishlar	ilmiy ishlar va innovatsiyalar bo'yicha prorektor	Moliya vazirligi	kurslar
5	chora-tadbirlar	xalqaro ilmiy jurnal	Davlat soliq qo'mitasi	brinchi o'limbos
6	tanlov	qizilashirilgan patent muallifan	davlat byudjeti	kredit modul tizimi
7	ma'ruza tanlov o'tkazish	ilmiy jamoa	rag'batlantirish	ilova
8	mutaxassislarni xarida tayyorlash	dasturiy guvohnoma	Hisob palatasi	ta'lim jarayoni

1-rasm. Universitet hujjat ma'lumotlarining Datasetning ko'rinishi

Ushbu qismda kelgan axborotlar ma'naviy, ilmiy, moliyaviy va o'quv ishlari kategoriyalariga bo'lib chiqilib dataset holatiga keltirilgan [1]. (1-rasm)

	name	type
0	O'zbekiston Respublikasi Vazirlar Mahkamasi	ma'naviy
1	Prezident	ma'naviy
2	ijtimoiy-iqtisodiy rivojlantirish	ma'naviy
3	mahalliy mehnat bozori	ma'naviy
4	qiziqishlar	ma'naviy
...
42	yangi texnologiyalarni joriy etish	o'quv ishlari
43	o'quv metodik boshqarmasi	o'quv ishlari
44	kafolat xati	o'quv ishlari
45	ilmiy ishlar va innovatsiyalar bo'yicha prorektor	o'quv ishlari
46	yuristkonsult	o'quv ishlari

2-rasm. Datasetni nom va tur ustunlariga ajratish

Bunda datasetni 2 ta ustunga o'zgartirib olamiz. Nom va tur ustunlari mos ravishda berilgan ma'lumot va qaysi kategoriyaga tegishli ekanligini ko'rsatadi [2,3]. 2-rasmdan ko'rinib turibdiki, *type* 4 ta turdan iborat va biz uni raqamlar orqali ifodalashimiz mumkin. Buning uchun Mashinali o'qitishda *Labelencoder* metodidan foydalanamiz va quyidagi natijaga erishishimiz mumkin [4].(3-rasm)

	name	type
0	O'zbekiston Respublikasi Vazirlar Mahkamasi	1
1	Prezident	1
2	ijtimoiy-iqtisodiy rivojlantirish	1
3	mahalliy mehnat bozori	1
4	qiziqishlar	1
...
42	yangi texnologiyalarni joriy etish	3
43	o'quv metodik boshqarmasi	3
44	kafolat xati	3
45	ilmiy ishlar va innovatsiyalar bo'yicha prorektor	3
46	yuristkonsult	3

	name	type
40	barcha turdagi to'lovlar	2
20	Akademik litsey	3
27	Avtomobil yo'llari qo'mitasi	1
29	Axborot texnologiyalari va kommunikatsiyalarin...	3
17	COSCOM	1
...
42	yangi texnologiyalarni joriy etish	3
37	yig'im	2
0	yosh tadqiqotchi	0
46	yuristkonsult	3
15	Ўзбектелеком	1

3.3-rasm. Matnli ustunni LabelEncoder yordamida sonli ko'rnishga o'tkazish

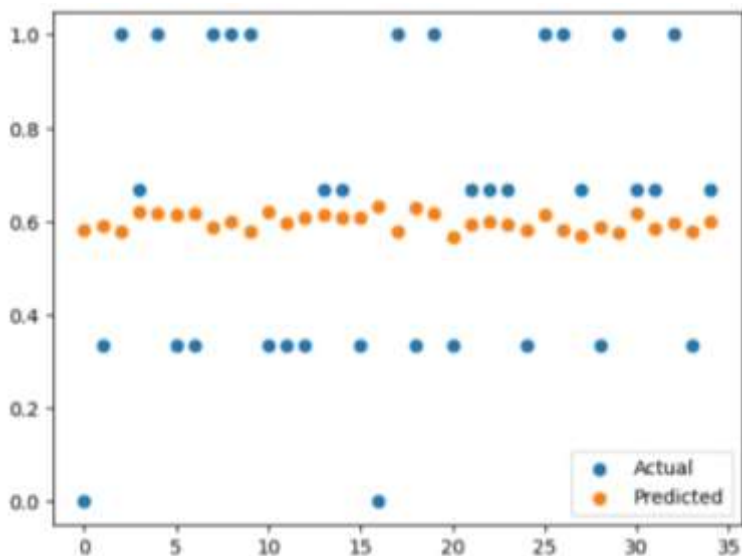
LabelEncoder yorliqlarni normallashtirishga yordam beradigan yordamchi sinf bo'lib, ular 0 va 1 orasidagi qiymatlarni o'z ichiga oladi. Bundan tashqari, raqamli bo'lmagan yorliqlarni (agar ular xeshlangan va solishtirish mumkin bo'lsa) raqamli teglarga aylantirish uchun ham foydalanish mumkin [5].

	name	type	hash
0	barcha turdagi to'lovlar	2	1932670463124744049
1	Akademik litsey	3	-4452531108823077823
2	Avtomobil yo'llari qo'mitasi	1	3359624676329283642
3	Axborot texnologiyalari va kommunikatsiyalarin...	3	2823616649310933661
4	COSCOM	1	-3536471060794299275
...
169	yangi texnologiyalarni joriy etish	3	4809138968845019490
170	yig'im	2	7883120098521571565
171	yosh tadqiqotchi	0	-3087103009392467657

3.4-rasm. Xeshlash jarayoni

Bu yerda xeshlash jarayonini ko'rish mumkin (4-rasm). Xeshlash bu ma'lumot turini o'zgartirish hisoblanadi. Mashinaga ma'lumotni o'qitish jaroyonida matnli ma'lumotlarni o'qitish murakkab jarayon hisoblanadi. Shuning uchun matnli ma'lumotlar xeshlanadi [6].

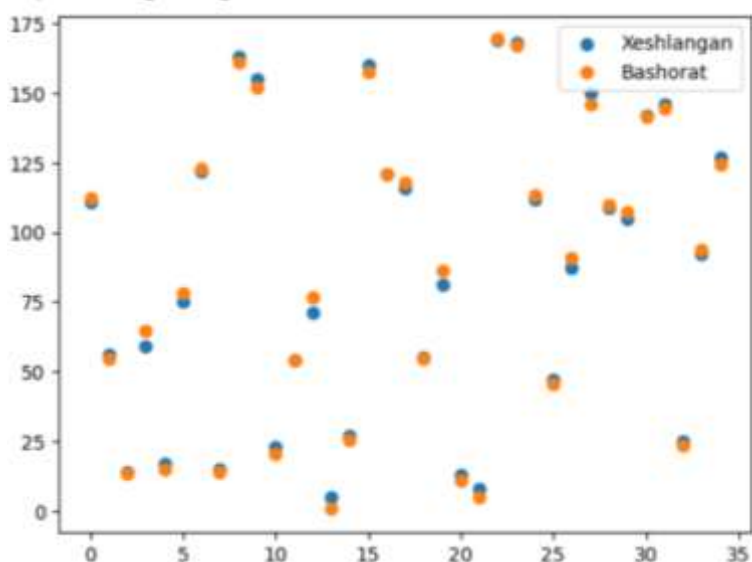
Endi dataset ustida bema'lol mashinali o'qitish algoritmlarini qo'llashimiz mumkin. Chiziqli regressiya bashoratli tahlilning asosiy va keng tarqalgan turi hisoblanadi [7,8]. Regressiyaning umumiy g'oyasi ikkita narsani tekshirishdan iborat: (1) bashorat qiluvchi o'zgaruvchilar to'plami natija o'zgaruvchini bashorat qilishda yaxshi ish qiladimi? (2) Qaysi o'zgaruvchilar, xususan, natija o'zgaruvchisining muhim prognozi hisoblanadi va ular beta baholarning kattaligi va belgisi bilan ko'rsatilgan - natija o'zgaruvchisiga qanday ta'sir qiladi? Ushbu regressiya baholari bir bog'liq o'zgaruvchi va bir yoki bir nechta mustaqil o'zgaruvchilar o'rtasidagi munosabatni tushuntirish uchun ishlatiladi [9].



5-rasm. Chiziqli regressiya algoritmining natijasi

Demak birinchi algoritm natijasi yaxshi emas. Endi logistik regressiya algoritmidan foydalanib ko'ramiz. Logistik regressiya mashinali o'qitishning eng mashhur algoritmlaridan biri bo'lib, u Supervised Learning algoritmi hisoblanadi. U ma'lum bir mustaqil o'zgaruvchilar to'plamidan foydalangan holda toifali bog'liq o'zgaruvchini bashorat qilish uchun ishlatiladi [10].

hash	Bashorat
60	111 112.135252
44	56 54.677374
127	14 13.589371
120	59 64.584458
143	17 14.925465
22	75 78.406055
9	122 122.667821
145	15 13.932334
146	163 161.192801
170	155 151.743434
19	23 20.286299
17	54 53.950825



6-rasm. Logistik regressiya algoritmining natijasi.

Bu natijadan ko'rish mumkinki, logistik regressiya algoritmidan bu yerda foydalanish yaxshi natija beradi. Jadvalga qaraydigan bo'lsak, xesh va bashorat qilingan qiymatlar bir-biriga mutanosib. Ba'zi qiymatlar uzoqroq kelgan bo'lishi mumkin. Natija grafikda yaqqol tasvirlangan. Ko'k rang datasetdagi xeshlangan qiymatlar, sariq esa bashorat qiluvchi qiymatlar hisoblanadi. E'tibor bersangiz, ko'k va sariq nuqtalar bir-birga deyarli mos keladi. Ba'zi sariq nuqtalar ko'k rangga ustma ust tushgan. Bu jadvaldagi eng yaqin kelgan bashorat qiymatlarini anglatadi.

Foydalanilgan adabiyotlar:

1. Nizomiddin N. et al. TA'LIMDA DASTURLASH JARAYONINI BAHOLASHGA ASOSLANGAN AVTOMATLASHTIRILGAN TIZIMNI TADBIQ ETISH //International Journal of Contemporary Scientific and Technical Research. – 2023. – С. 24-28.

2. Choryorqulov G. H., & Qosimov NS (2023) //ELEKTRON JADVAL MODELINING TAVSIFLANISHI. PEDAGOGS Jurnal. – T. 30. – №. 3. – С. 67-73.

3. Чорркулов Г., Норматов Н., Мамараимов А. Роль анализа текстовых связей в электронных документах в информационной безопасности //Информатика и инженерные технологии. – 2023. – Т. 1. – №. 1. – С. 67-71.

4. Норматов Н., Мамараимов А. Та'lim tizimida baholash tizimini avtomatlashtirishni joriy etish jarayonlari va foydalanish metodlari //Информатика и инженерные технологии. – 2023. – Т. 1. – №. 2. – С. 356-359.

5. Мамараимов А., Чорёркулов Г., Норматов Н. Tanib olish modullarini dasturiy amalga oshirish //Информатика и инженерные технологии. – 2023. – Т. 1. – №. 2. – С. 38-44.

6. Sanoqul o'g'li Q. N. et al. ELEKTRON HUJJAT ALMASHINUVINI AVTOMATLASHTIRISH MODELINI ANALITIK TAHLILI //Лучшие интеллектуальные исследования. – 2023. – Т. 10. – №. 5. – С. 89-100.

7. Javlon, K., & Erali, M. (2023). STRUCTURE AND PRINCIPLE OF OPERATION OF FULLY CONNECTED NEURAL NETWORKS. International Journal of Contemporary Scientific and Technical Research, 136-141.

8. Мустафоев, Е., & Холматов, Ж. (2023). Brayl matn tasviri sifatini oshirish usullari. Информатика и инженерные технологии, 1(2), 23-27.

9. Obid o'g A. S. J. et al. Numpy Library Capabilities. Vectorized Calculation In Numpy Va Type Of Information //Eurasian Research Bulletin. – 2022. – Т. 15. – С. 132-137.

10. Javlon X. et al. Классификатор движения рук с использованием биомиметического распознавания образов с помощью сверточных нейронных сетей с методом динамического порога для извлечения движения с использованием датчиков EF //Journal of new century innovations. – 2022. – Т. 19. – №. 6. – С. 352-357.