

yordamida fuqarolik va infratuzilma muhandisligi, jild. 2021-yil, Wiley onlayn kutubxonasi.wiley.com

[21] K. Bai, X. Zhu, S. Wen, R. Zhang va boshqalar, "Broad learning based dynamic fuzzy inference system with adaptive structure and interpretable fuzzy rules," IEEE Transactions on Fuzzy Systems, 2021, [HTML](#)

[22] S. Chopra, G. Ziman, A. Sharma, M. Shabaz, "Zamonaviy muhandislik fanlarida adaptiv neyro-Noravshan xulosa chiqarish tizimining taksonomiyasi," Hisoblash intellekti va neyroiologiya, 2021, hindawi.com.hindawi.com

[23] Muminov Bahodir, Egamberdiyev Elyor VGG16 MODEL VA K-MEANS ALGORITMI ASOSIDA TASVIR MA'LUMOTLARINI KLASSTERLASH // Universum: texnicheskiye nauki. 2025. No1 (130). URL: <https://cyberleninka.ru/article/n/image-data-clustering-based-on-the-vgg16-model-and-the-k-means-algorithm> (data obrasheniya: 09.04.2025).

MA'LUMOTLARNI INTEKLEKTUAL TAHLIL QILISHDA CRISP-DM METODOLOGIYASINING FAZALARI (MA'LUMOTLARNI TAYYORLASH FAZASI)

Matchonov Shohruh Matkarim o'g'li

Jizzax shahridagi Qozon (Vo'lgabo'yi) federal universiteti filiali "Raqamli ta'lim texnologiyalari markazi" rahbari
shohruhmachonov@gmail.com

Annotatsiya: Mazkur ishda ma'lumotlarni intellektual tahlil qilishda CRISP-DM metodologiyasining o'quv ma'lumotlarini tayyorlash fazasiga etibor qaratilgan. Chunki bu faza modelni ishlab chiqishda vaqtning 80% gacha bo'lgan miqdorini o'z ichiga oladi. O'quv ma'lumotlarini tayyorlash fazasi bosqichlari rasmlar orqali tushuntirilgan, ma'lumotlarni modelga o'qitish yondoshuvlari ko'rsatilgan va mashinani o'qitish injinerlari va ma'lumotlar analitiklari uchun etibor berishlari kerak bo'lgan ma'lumotlar aytib o'tilgan.

Kalit so'zlar: CRISP-DM, Anamaliya, Dataset, Mashinali o'qitish, Ma'lumotlarni intellektual tahlil qilish, o'quv ma'lumotlari, ma'lumotlarni tozalash.

ЭТАПЫ ПРИМЕНЕНИЯ МЕТОДОЛОГИИ CRISP-DM В ИНТЕЛЛЕКТУАЛЬНОМ АНАЛИЗЕ ДАННЫХ (ЭТАП ПОДГОТОВКИ ДАННЫХ)

Аннотация: В данной работе внимание уделено этапу подготовки обучающих данных методологии CRISP-DM при интеллектуальном анализе данных. Потому что этот этап занимает до 80% времени разработки модели. Шаги этапа подготовки данных для обучения объясняются с помощью иллюстраций, представлены подходы к обучению данных в модели, а также упоминаются соображения для инженеров по машинному обучению и аналитиков данных.

Ключевые слова: CRISP-DM, Анамалия, Набор данных, Машинное обучение, Интеллектуальный анализ данных, данные обучения, очистка данных.

STAGES OF APPLICATION OF CRISP-DM METHODOLOGY IN DATA MINING (DATA PREPARATION STAGE)

Abstract: This paper focuses on the training data preparation stage of the CRISP-DM methodology in data mining. Because this stage takes up to 80% of the model development time, the steps of the training data preparation stage are explained with illustrations, approaches to

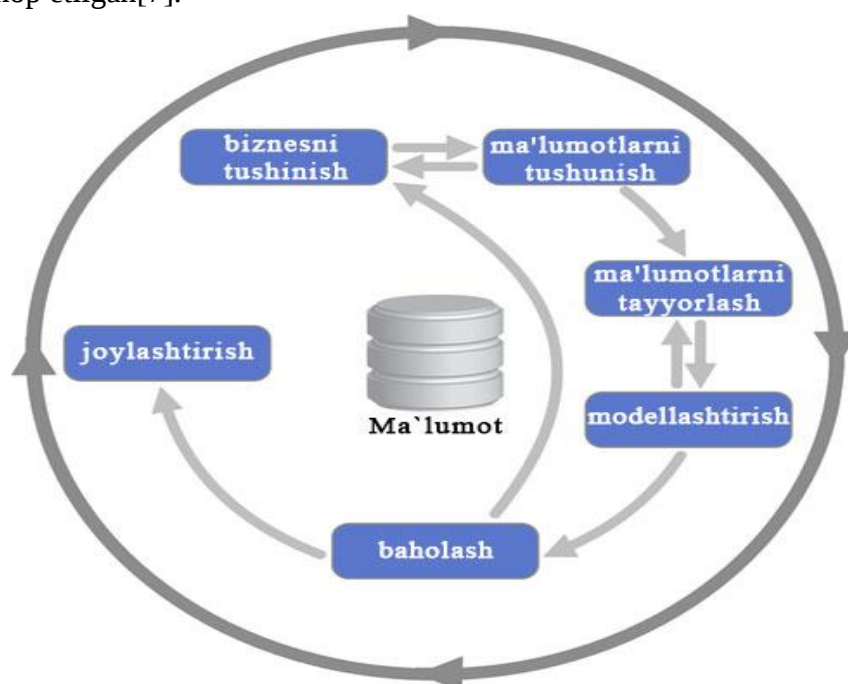
training data in the model are presented, and considerations for machine learning engineers and data analysts are mentioned.

Keywords: CRISP-DM, Anomaly, Dataset, Machine learning, Data mining, Training data, Data cleaning.

Kirish. CRISP-DM [1]- ma'lumotlarni intellektual tahlil qilishda eng ko'p qo'llaniladigan metodologiya hisoblanadi.[2,3,4].

Metodologiyaning ma'lumotlarni intellektual tahlil qilish hayotiy tsikli olti fazadan iborat bo'lib, yo'nalishlar fazalar orasidagi umumiy bog'liqliklarni ko'rsatadi. Ushbu bosqichlarning ketma-ketligi aniq belgilanmagan. Ko'pgina loyihalarda avvalgi bosqichlarga qaytish va keyin yana oldinga siljish zaruriyati paydo bo'ladi[5]

Ushbu metodologiyaning birinchi versiyasi 1999 yil mart oyida Bryusselda bo'lib o'tgan to'rtinchi CRISP-DM SIG seminarida taqdim etilgan[6] va bosqichma-bosqich ko'rsatmalar bir yildan so'ng chop etilgan[7].



1-rasm. Ma'lumotlarni intellektual tahlil qilish hayotiy tsikli

Fazalar ketma-ketligi qat'iy emas va har doim talabga qarab fazalar o'rtasida oldinga va orqaga harakat qilinadi. Jarayon diagrammasidagi o'qlar fazalar orasidagi muhim bog'liqliklarni ko'rsatadi. Diagrammadagi tashqi doira ma'lumotlar qazib olishning tsiklik xususiyatini anglatadi. Yechim o'rnatilgandan keyin ma'lumotlarni yig'ish jarayoni davom etadi. Jarayon davomida olingan bilimlar yangi, ko'pincha yo'naltirilgan biznes savollarini yaratishi mumkin va keyingi ma'lumotlarni qidirish jarayonlari avvalgilarining o'rganishlaridan foydalanadi, bu ma'lumotlarni qidirish jarayoni sifatini oshiradi.

Modelni ishlab chiqishda vaqtning 80% gacha o'quv ma'lumotlarini tayyorlashga sarflanadi. O'quv ma'lumotlarini tayyorlash jarayoni ko'p bosqichni o'z ichiga oladi. Bularga quyidagilarni misol qilish mumkin 2-rasm.

- Takrorlanishlarni olib tashlash
- Imlo xatolarni to'g'irlash
- Anamaliyalarni oldini olish
- Bo'sh joylarni qiymat bilan to'ldirish [9]

2-rasm. O'quv ma'lumotlaridagi kamchiliklar

Ism	Familiya	Yosh	Bo'y	Mashina	
Dilrabo	Fazilova	45	160	Jentra	Takrorlanish
Dilrabo	Fazilova	45	160	Jentra	
Axrор	Burxonov	20	170	Jettra	Xatolik
Timur	Asatov	145	170	Tracker	
Anvar	Tavashev	47		BYD	

Anamaliya Bo'shliq

3-rasm. Ma'lumotlarni modelga o'qitish jarayoni

Ma'lumotlar idal holatda tayyor bo'lgach barcha ma'lumotlar datasetda tartiblangan jadval holatiga olib kelinadi va modelga qo'llaniladi 3-rasm.

Datasetda tartiblangan jadvallarda matnlar shaklda ma'lumotlar joylashtiriladi. Mashinani o'qitish injinerlari datasetdagi ma'lumotlarni modelga mos qilib dasturlash tilida mashinani o'qitish bibliotekalari yordamida ma'lumotlarni tahlil qilish orqali qayta ishlashadi. Qayta ishlangan ma'lumotlar qaysidir xususiyatiga ega bo'ladi. Bu xususiyatlar ko'pincha naqishlar, gipotezalar bo'lishi mumkin. Ma'lumotlar to'g'ri xususiyatga ega bo'lishi uchun maksimal ravishda ma'lumotlarni tozalash va ma'lumotlarni qayta ishlash talab qilinadi.[8] Biz qanday masalani yechishimizdan qat'iy nazar, aytaylik avtasalonda avtamabillarni sotish biznesi bo'ladimi, haridorning kredit limitini hisoblash bo'ladimi yoki rasmdagi tirik mavjudotlarni tanib olish maslasi bo'ladimi siz kiruvchi ma'lumotlarini qayta ishlamasdan masalani yecholmaysiz. Yaxshi tayyorlangan ma'lumotlar modelning yaxshi ishlashini ta'minlaydi.

Xulosa. Mazkur ishda biz CRISP-DM metodologiyasining eng ko'p vaqt oladigan va asosiy ma'lumotlarni tayyorlash fazasiga qaratdik. Ma'lumotlarni tayyorlash fazasida asosiy bosqichlar bilan tanishdik, va tahlil qilish uchun olingan ma'lumotlarning sifatini yaxshilash uchun mashinani o'qitish injinerlari va ma'lumotlar analitiklari qayerlarga diqqat qilishlari kerakligi haqida yoritishga harakat qildik.

Adabiyotlar ro'yxati

1. Shearer C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13—22.
2. Poll: What main methodology are you using for data mining? www.kdnuggets.com. (2016)
3. Poll: Data Mining Methodology. www.kdnuggets.com. (2016)
4. . What main methodology are you using for your analytics, data mining, or data science projects? Poll. www.kdnuggets.com. (2016)
5. Óscar Marbán, Gonzalo Mariscal and Javier Segovia (2009); A Data Mining & Knowledge Discovery Process Model
6. Pete Chapman (1999); The CRISP-DM User Guide
7. Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth (2000); CRISP-DM 1.0 Step-by-step data mining guide
8. Matchonov, S., Asatov, T. and Burhonov, A., 2024. BI TIZIMLARINING AFZALLIKLARINI BAHOLASH UCHUN MASHINALI O'QITISH MODELLARI. International Journal of scientific and Applied Research, 1(3), pp.142-145.
9. Бабамурадов, О., & Матчонов, III. (2022). Bi tizimining ishlash prinspi va elektron jadvallar bilan ozaro bogliqligi. Современные инновационные исследования актуальные проблемы и развитие тенденции: решения и перспективы, 1(1), 11–15. извлечено от <https://inlibrary.uz/index.php/zitdmrt/article/view/5078>