



DATA MININGDA CRISP-DM METODOLIGIYASI TASNIFI

Muxtorov Doston Naim o'g'li

E-mail: dmuxtorov062@gmail.com

O'zbekiston Milliy Universiteti Jizzax filiali, Talaba

Sanayev Mashrab Eshquvvat o'g'li

E-mail: sanayevmashrab@gmail.com

O'zbekiston - Finlandiya Pedagogika Instituti, Assistent

Narzikulov Nodirbek Xamrobek o'g'li

E-mail: narzikulovnodir@gmail.com

O'zbekiston Milliy Universiteti Jizzax filiali, Talaba

Annotatsiya

"CRISP-DM" (Cross-Industry Standard Process for Data Mining) – bu ma'lumotlar chiqarishda amal qilingan tizimli usulni tashkil etuvchi metodologiya. Bu metodologiya tashkil etish, tahlil qilish va ma'lumotlar chiqarish jarayonlari uchun yagona standartni taqdim etadi

Kalit so'zlar: Modeling, Namuna olish, Ma'lumotlarni aniqlash, Ma'lumotlar Modelini Baholash,

KIRISH

CRISP-DM (Cross-Industry Standard Process for Data Mining), ma'lumotlar do'koni tahlilini amalga oshirish uchun ishlab chiqilgan bo'lgan standart metodologiya tizimidir. Bu metodologiya, ma'lumotlar tahlili loyihalari uchun boshqaruvning har bir bosqichini belgilab chiqadi. CRISP-DM yoki ma'lumotlar do'koni tahlili jarayoni quyidagi besh bosqichdan iborat:

Ma'lumotlar Tahlilining Tushunchalari (Business Understanding): Bu bosqichda loyihaning mohiyati, maqsadi, ma'lumotlar tahlilini olib chiqishga qaratilgan maqsadlar va boshqa biznes va ma'lumotlar bilan bog'liq tushunchalar aniqlanadi.

Ma'lumotlarni aniqlash (Data Understanding): Bu bosqichda mavjud ma'lumotlar, ularning manzili, shakli, qanday olinishlari, omborlanish shakllari va xususiyatlari to'g'risida tushunchalar, olingan ma'lumotlarni tahlil qilish uchun qo'llaniladigan muhim ma'lumotlarni aniqlash shart.

Ma'lumotlar Tuzilishi (Data Preparation): Bu bosqichda olingan ma'lumotlarni tahlil uchun tayyorlash jarayoni o'tkaziladi. Bu, ma'lumotlarni



to'plangan, to'plangan ma'lumotlarni o'chirish, ma'lumotlarni ishlash, tuzatish, birlashtirish va boshqa amallarni o'z ichiga oladi.

Ma'lumotlar Modelini Tuzilishi (Modeling): Bu bosqichda ma'lumotlar analitik modellarga o'zgaradi. Bu modellar ma'lumotlarni tahlil etishda va maqsadga muvofiq topilgan ma'lumotlarni qisqa muddat ichida tushunarli natijaga olib chiqishda qo'llaniladi.

Ma'lumotlar Modelini Baholash (Evaluation): Bu bosqichda olingan modelni sinovdan o'tkazish, natijalarni baholash va aniqlangan natijalarning maqsadga muvofiqligini tekshirish jarayoni o'tkaziladi. Agar natijalar maqsadlarni bajarib chiqmasa, modelni qayta tuzatish va baholash lozim bo'ladi.

Ma'lumotlar Modelini Amalga Oshirish (Deployment): Agar ma'lumotlar modeli muvofiq baholangan bo'lsa, uni amalga oshirish bosqichi o'tkaziladi. Model odatda real hayotda ishlash uchun boshqa ilovalar yoki tizimlar bilan integratsiya qilinadi. CRISP-DM metodologiyasi tahlilchi va modelni tuzish bosqichlari orasida amaliy ishlarni tashkil etish, muvofiq miqdordagi ma'lumotlarni olish va ularni tuzatish, modelni o'rganish va tuzilish, natijalarni baholash, va so'ng modelni amalga oshirish bosqichlarini tizimga olib kiradi. "Ma'lumotlar Tahlilining Tushunchalari" (Business Understanding) - bu CRISP-DM metodologiyasining birinchi bosqichi. Bu bosqichda, ma'lumotlar tahlili loyihasi uchun asosiy maqsadlar va talablar aniqlanadi. Asosiy mazmuni quyidagicha: Biznesda muammolar va talablar: Loyiha boshlashdan oldin, shuningdek, o'tkazmoqchi bo'lgan biznes muammolarini tushunish lozim. Bu muammolar qanday ma'lumotlardan foydalanilishi orqali yechilishi mumkinligini bilishni talab qiladi. Maqsadlar va o'zgarishlarni aniqlash: Loyiha uchun aniq maqsadlar belgilanadi. Bu maqsadlar biznes muammolari yoki biznes bosqichlari yechishga qaratilgan. Ma'lumotlar tahlili loyihasi bu maqsadlarni qanday bajarishi kerakligini aniqlaydi.

Ma'lumotlardan olingan foyda: Ma'lumotlar tahlili natijasida olinadigan foyda va farqni tushunish. Bu, ma'lumotlar tahlili natijalari orqali nima o'rganish va qanday qadamda ishlaydigan bo'lishi kabi mavzularni o'z ichiga oladi. Resurslar va cheklovlar: Ma'lumotlar tahlili loyihasi uchun kerakli resurslar va cheklovlar (finans, odam resurslari, vaqt, ma'lumotlar resurslari) belgilanadi. Bu, loyihaning amalga oshirilishi uchun kerakli imkoniyatlarni va qo'llaniladigan voqealarni o'rganishni o'z ichiga oladi. Boshqa shartlar: Boshqa talablar va shartlar, masalan, ma'lumotlar xavfsizligi, qonunchilikni saqlash, yashirin ma'lumotlarni saqlash shartlari va boshqalar kabi muhim ko'nikmalarni qamrab oladi.



"Ma'lumotlarni aniqlash" (Data Understanding) - bu CRISP-DM metodologiyasining ikkinchi bosqichi. Bu bosqichda tahlil qilish uchun kerakli ma'lumotlarni tuzish, tahlil qilish va natijalarni olish uchun ma'lumotlar bilan tanishish jarayoni o'tkaziladi. Quyidagi nuqtalarni o'z ichiga oladi: Ma'lumotlar tuzilishi: Bu qismatda loyihaning maqsadlariga qarab ma'lumotlar tuzilishi va ko'rsatkichlari tavsiflanadi. Ma'lumotlar qanday formatda saqlanmoqda, ular qanday olinmoqda, va ularning umumiy miqdori qancha ekanligi haqida to'liq ma'lumot beriladi. Ma'lumotlarni tahlil qilish uchun tayyorlash: Bu bosqichda, kerakli ma'lumotlarni olish, yig'ish, va o'rganish jarayonlari tavsiflanadi. Bu jarayonlar orqali tahlilchi, tahlil qilish uchun ma'lumotlarni tayyorlashga va ulardan olingan ma'lumotlarni aniqlashga tayyorlanadi.

FOYDALANILGAN ADABIYOTLAR:

1. Naim o'g'li M. D., Abdishukur o'g'li S. A. THE NUMPY LIBRARY OF THE PYTHON PROGRAMMING LANGUAGE IS AN OPTIMAL SOLUTION FOR WORKING WITH ARRAYS //JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH. – 2022. – T. 2. – №. 13. – C. 195-197.
2. Naim o'g'li M. D., Baxtiyor o'g'li E. S. DATA SCIENCE METHODOLOGY IN LEARNING PROGRAMMING //JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH. – 2022. – T. 2. – №. 13. – C. 207-210.
3. Amanbayevna A. S., Naim o'g'li M. D. GEOMETRIC MODELING AND VISUALIZATION OF SELF-SIMILAR STRUCTURES BASED ON FRACTAL THEORY //JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH. – 2022. – T. 2. – №. 13. – C. 187-188.
4. Naim o'g'li M. D., Baxtiyor o'g'li E. S. KATTA HAJMDAGI MA'LUMOTLARINI TAHLIL QILISHDA CRISP-DM USULLARINI QO'LLASH //JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH. – 2022. – T. 2. – №. 13. – C. 211-216.
5. Baxtiyor o'g'li E. S., Naim o'g'li M. D. YAGONA INTERAKTIV DAVLAT XIZMATLARI PORTALNING TRANSPORT BO'LIMINI IDF0, DFD VA IDF3 SATANDART DIAGRAMMARI YORDAMIDA YARATILGAN LOYIHASI //JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH. – 2022. – T. 2. – №. 13. – C. 198-206.
6. Doston M., Abdulatif S. SUN'IY INTELLEKT TEXNOLOGIYALARI VA ULARNI SOHALARDA QO'LLANILISHI //Educational Research in Universal Sciences. – 2022. – T. 1. – №. 5. – C. 144-147.
7. 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. – T. 19. – №. 6. – C. 352-357.



8. Naim o'g'li M. D., Xasan o'g'li A. M. CLASSIFICATION OF CRISP-DM METHODOLOGY IN DATA ANALYSIS //MODELS AND METHODS FOR INCREASING THE EFFICIENCY OF INNOVATIVE RESEARCH. – 2023. – Т. 2. – №. 19. – С. 220-222.
9. Naim o'g'li, Muxtorov Doston, and Qudratov Akbar Akmal o'g'li. "ROLAP VA DATA MINING INTEGRATSIYASI." *Yangi O'zbekiston taraqqiyotida tadqiqotlarni o'rni va rivojlanish omillari* 1.1 (2023): 233-236.
10. Naim o'g'li, Muxtorov Doston, and RASHIDOV BOYSARI TO'RA O'G. "STREAM DATA MINING VA REAL VAQT TAHLILI." *Yangi O'zbekiston taraqqiyotida tadqiqotlarni o'rni va rivojlanish omillari* 1.1 (2023): 217-222.
11. Naim o'g'li, Muxtorov Doston, ALISHEROV BOBURJON ANVAR O'G'LI, and SAFAROV ULUG'BEK ANORBOY O'G. "OLAP VA DATA MINING INTEGRATSIYASI." *Yangi O'zbekiston taraqqiyotida tadqiqotlarni o'rni va rivojlanish omillari* 1.1 (2023): 229-232.
12. Naim o'g'li, Muxtorov Doston, SAFAROV ULUG'BEK ANORBOY O'G, and G'ANIYEV SOBIR QODIR O'G. "DATA MINING KLASSIFIKATSIYA VA KLAUSTERIZATSIYA." *Yangi O'zbekiston taraqqiyotida tadqiqotlarni o'rni va rivojlanish omillari* 1.1 (2023): 223-228.
13. Naim o'g'li, Muxtorov Doston, and Shermuhammedov Abdulatif Abdishukur o'g'li. "THE NUMPY LIBRARY OF THE PYTHON PROGRAMMING LANGUAGE IS AN OPTIMAL SOLUTION FOR WORKING WITH ARRAYS." *JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH* 5.4 (2022): 195-197.
14. Naim o'g'li, Muxtorov Doston, and Ergashev Sirojiddin Baxtiyor o'g'li. "KATTA HAJMDAGI MA'LUMOTLARINI TAHLIL QILISHDA CRISP-DM USULLARINI QO'LLASH." *JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH* 5.4 (2022): 211-216.
15. Baxtiyor o'g'li, Ergashev Sirojiddin, and Muxtorov Doston Naim o'g'li. "YAGONA INTERAKTIV DAVLAT XIZMATLARI PORTALNING TRANSPORT BO'LIMINI IDF0, DFD VA IDF3 SATANDART DIAGRAMMARI YORDAMIDA YARATILGAN LOYIHASI." *JOURNAL OF INNOVATIONS IN SCIENTIFIC AND EDUCATIONAL RESEARCH* 5.4 (2022): 198-206.
16. Yusupovich X. J. BEMORLARNING SHIFOKOR YOZGAN RETSEPTI BOYICHA DORILARNI QABUL QILGANLIK DARAJASINI ANIQLASH AVTOMATIK TIZIMNI YARATISH //MODELS AND METHODS FOR INCREASING THE EFFICIENCY OF INNOVATIVE RESEARCH. – 2023. – Т. 2. – №. 19. – С. 223-234.
17. Бурнашев В. Ф., Холматов Ж. Ю. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГИДРОДИНАМИЧЕСКИХ ПРОЦЕССОВ МНОГОФАЗНОЙ ФИЛЬТРАЦИИ В НЕФТЯНОМ ПЛАСТЕ ПРИ ЕГО ЗАВОДНЕНИИ //RESEARCH AND EDUCATION. – 2022. – Т. 1. – №. 1. – С. 137-154.
18. Kholmatov Javlon, & Mustafoyev Erali. (2023). STRUCTURE AND PRINCIPLE OF OPERATION OF FULLY CONNECTED NEURAL NETWORKS. *International Journal of Contemporary Scientific and Technical Research*, 136–141.