

DATA SCIENCE JARAYONLARI

Onarqulov Maqsadjon Karimberdiyevich

*Farg'ona davlat universiteti amaliy matematika va
informatika kafedrası dotsenti*

maxmaqsad@gmail.com

Yusupov Mirsaid Abdulaziz o'g'li

*Farg'ona davlat universiteti amaliy matematika va
informatika kafedrası o'qituvchisi*

mirsaidbeky@gmail.com

Nuritdinova Nargizaxon Muxtorali qizi

Farg'ona davlat univesiteti 3-kurs talabasi

aikas32ew@gmail.com

Annotatsiya: *Data science ko'p tarmoqli soha bo'lib, u tartiblangan va tartiblanmagan ma'lumotlardan tushuncha va bilimlarni olish uchun ilmiy usullar va jarayonlar, algoritmlar va tizimlardan foydalanadi. U murakkab ma'lumotlar to'plamlarini tahlil qilish va sharhlash uchun turli sohalaridagi tajribalarni, jumladan, statistika, informatika, matematika va ma'lum bir sohaga oid bilimlarni birlashtiradi. Data science har bir sohada qarorlar qabul qilishda va yangi bilimlarni kashf etishda yordam beradigan muhim bir soha hisoblanadi. Ma'lumotlar ilm-fani kompaniyalar va tashkilotlar uchun muhim qarorlar qabul qilishda, jarayonlarni optimallashtirishda, marketing strategiyalarini ishlab chiqishda va mijozlar ehtiyojlarini tushunishda yordam beradi.*

Kalit so'zlar: *ma'lumotlarni yig'ish, tozalash, tahlil qilish, modellash, va natijalarni taqdim etish, ma'lumotlarni vizualizatsiya qilish, metodologik yondashuv.*

Annotation: *Data science is a multidisciplinary field that uses scientific methods and processes, algorithms and systems to extract insights and knowledge from structured and unstructured data. It combines expertise from a variety of fields, including statistics, computer science, mathematics, and domain-specific knowledge, to analyze and interpret complex data sets. Data science is an important field that helps in making decisions and discovering new knowledge in every industry. Data science helps companies and organizations make important decisions, optimize processes, develop marketing strategies, and understand customer needs.*

Keywords: *data collection, cleaning, analysis, modeling, and presentation of results, data visualization, methodological approach.*

Аннотация: *Наука о данных — это междисциплинарная область, которая использует научные методы и процессы, алгоритмы и системы для извлечения идей и знаний из структурированных и неструктурированных данных. Он сочетает в себе опыт из различных областей, включая статистику, информатику, математику и отраслевые знания, для анализа и интерпретации сложных наборов данных. Наука о*

TA'LIM, TARBIYA VA INNOVATSIYALAR

I son, Dekabr

данных — важная область, которая помогает принимать решения и открывать новые знания в каждой отрасли. Наука о данных помогает компаниям и организациям принимать важные решения, оптимизировать процессы, разрабатывать маркетинговые стратегии и понимать потребности клиентов.

Ключевые слова: *сбор данных, очистка, анализ, моделирование и представление результатов, визуализация данных, методический подход.*

Kirish

Data science jarayonlari odatda bir nechta asosiy bosqichlarga bo'linadi: ma'lumotlarni yig'ish, tozalash, tahlil qilish, modellash, va natijalarni taqdim etish.

○ Data collection (Ma'lumotlarni yig'ish):

Har qanday Data science loyihasida birinchi qadam tegishli ma'lumotlarni yig'ishdir. Bu ma'lumotlar bazalari, kompaniya tizimlaridan, veb-sahifalar, sensorlar, ijtimoiy media va boshqalar kabi turli manbalardan olingan ma'lumotlarni o'z ichiga oladi. To'plangan ma'lumotlarsifatini baholash zarur. Yig'ilgan ma'lumotlar keyinchalik tahlil qilinishiga tayyor bo'lishi kerak.

○ Data cleaning and preprocessing (Ma'lumotlarni tozalash va qayta ishlash):

Daslabki yig'ilgan ma'lumotlar ko'pincha tartibsiz, to'liq va mukammal bo'lmaydi, unda yetishmayotgan qiymatlar, o'zgarishlar yoki xatolar bo'lishi mumkin. Data science mutaxassislari aniqlik va izchillikni ta'minlash uchun ma'lumotlarni tozalaydi va oldindan qayta ishlaydi. Bu yetishmayotgan qiymatlarni qayta ishlash, o'zgaruvchan qiymatlarni olib tashlash va o'zgaruvchilarni o'zgartirishni o'z ichiga oladi.

○ Exploratory data analysis (Ma'lumotlarni tahlil qilish):

Tadqiqot ma'lumotlarini tahlil qilish tendensiyalar va o'zaro bog'liqliklarni tushunish uchun ma'lumotlarni vizual va statistik jihatdan o'rganishni o'z ichiga oladi. Ma'lumotlarning asosiy tuzilishi haqida tushunchaga ega bo'lish uchun tavsiflovchi statistika, ma'lumotlarni vizualizatsiya qilish va boshqa usullardan foydalaniladi.

○ Feature engineering (Xususiyat muhandisligi):

Feature engineering – Machine learning modellarining ishlashini yaxshilash uchun yangi xususiyatlarni yaratish yoki mavjudlarini o'zgartirish jarayoni. Ushbu qadam maxsus bilim va ijodkorlikni talab qiladi.

○ Data visualization and reporting (Natijalarni taqdim etish):

Data science jarayonining bu bosqichi natijalarni taqdim etishdir. Bu bosqichda, tahlil va modelning natijalari foydalanuvchilar yoki qaror qabul qiluvchilar uchun aniq va tushunarli tarzda taqdim etiladi. Vizualizatsiya, interaktiv hisobotlar, grafikalar, diagrammalar va boshqalar yordamida natijalar samarali tarzda taqdim qilinadi. Bu bosqichda natijalar maqsadga muvofiq ravishda taqdim etilishi kerak, shunda ular amaliy qarorlar qabul qilishda yordam beradi.

Xulosa. Data science jarayonlari samarali va to'g'ri amalga oshirilsa, ular kompaniyalar va tashkilotlarga katta foyda keltirishi mumkin. Har bir bosqichda aniqlik, diqqat va

metodologik yondashuv muhim rol o'ynaydi. Har bir jarayon, ma'lumotlar tahlilidan boshlanib, natijalarni taqdim etishgacha bo'lgan barcha bosqichlarda, maqsadga erishishga yordam beradi. Data science sohasining rivojlanishi va yangilanishi, yangi texnologiyalar va metodlarni o'z ichiga olgan holda, kelajakda yanada samarali va kuchli yechimlar yaratishga imkon beradi.

FOYDALANILGAN ADABIYOTLAR:

1. "Data Science from Scratch: First Principles with Python" - Joel Grus
2. "Python for Data Analysis" - Wes McKinney
3. "An Introduction to Statistical Learning: with Applications in R" - Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
4. "Deep Learning" - Ian Goodfellow, Yoshua Bengio, Aaron Courville
5. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" - Trevor Hastie, Robert Tibshirani, Jerome Friedman
6. "Pattern Recognition and Machine Learning" - Christopher M. Bishop
7. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" - Aurélien Géron
8. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" - Foster Provost & Tom Fawcett