

**O‘ZBEKISTON RESPUBLIKASI  
OLIV TA‘LIM, FAN VA INNOVATSIYALAR VAZIRLIGI  
TOSHKENT DAVLAT SHARQSHUNOSLIK UNIVERSITETI**




**MA'LUMOTLAR BAZASI VA DATA MINING**

**FAN DASTURI**

<b>Bilim sohasi:</b>	200000 – San‘at va gumanitar fanlar
<b>Ta‘lim sohasi:</b>	230000 – Tillar
<b>Ta‘lim mutaxassisligi:</b>	70230401 – Kompyuter lingvistikasi

**Toshkent – 2025**

Sharq xalqlari tillari va adabiyoti  
instituti direktori:

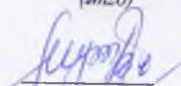
  

---

(imzo)

X.V. Mirzaxmedova

Tarjimashunoslik, tilshunoslik va  
xalqaro jumalistika oliy maktabi boshlig'i:

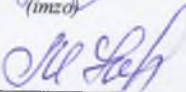
  

---

(imzo)

S.T. Mustafayeva

ARM boshlig'i:

---

(imzo)

M. Yuldasheva

<b>Fan/modul kodi</b> MBDM 11-314	<b>O'quv yili</b> 2025-2026 2026-2027	<b>Semestr</b> 1-2-3 semestr			<b>ECTS - Kreditlar</b> 1-semestr 4, 2-semestr 4, 3-semestr 6	
<b>Fan/modul turi</b> Majburiy	<b>Ta'lim tili</b> O'zbek			<b>Haftadagi dars soatlari</b> 1-semestr 4, 2-semestr 4, 3-semestr 6		
<b>1.</b>	<b>Fanning nomi</b>	<b>Auditoriya mashg'ulotlari (soat)</b>	<b>Ma'ruza</b>	<b>Amaliy</b>	<b>Mustaqil ta'lim (soat)</b>	<b>Jami yuklama (soat)</b>
	Ma'lumotlar bazasi va Data Mining	168	84	84	252	420
<b>2.</b>	<p style="text-align: center;"><b>I. Fanning mazmuni</b></p> <p>Fanni o'qitishdan <b>maqsad</b> – berilganlar bazasi haqida tushuncha berish, berilganlar bazasini tashkil qilish, berilganlar bazasini boshqarish tizimlari bilan ishlash, SQL so'rovlar, NoSQL bazalar, berilganlarning turlari, o'rgatuvchi tanlanma tushunchasi, katta hajmdagi berilganlar bilan ishlash, katta hajmdagi berilganlar ustida intellektual tahlilni amalga oshirish malakasini shakllantirish.</p> <p>Fanning <b>vazifasi</b> – berilganlarning tayanch va abstrakt turlari, berilganlar bazasini hosil qilish, berilganlar strukturalari, python dasturlash tili yordamida berilganlarni intellektual tahlil qilish, katta hajmdagi berilganlar ustida intellektual tahlil usullarini amalga oshirish orqali yashirin qonuniyatlarni aniqlash va bulutli texnologiyalardan maqsadli foydalanish bo'yicha malaka hosil qilish.</p> <p>AQSHning Oxford universiteti Data Science va Data Mining sohalarida keng qamrovli dasturlarni taklif qiladi. <i>Database Management</i> kursi, talabalar ma'lumotlar bazalari arxitekturasi, ma'lumotlarni optimallashtirish va turli ma'lumotlar bazalari tizimlari bilan ishlashni o'rganadilar. <i>Data Science and Machine Learning Bootcamp</i> kursida esa talabalarga ma'lumotlar bazalaridan ma'lumotlarni olish, ular bilan ishlash va mashina o'rganish algoritmlari yordamida ma'lumotlarni yig'ish usullari o'rgatiladi va bu tajribalar yuqori natija bergan.</p> <p>Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi top 300 talik ro'yxatga kiruvchi AQSHning Oxford universiteti (Oxford of University 6 ARWU)xorijiy tajribasini inobatga olgan holda takomillashtirildi. <a href="https://www.stats.ox.ac.uk/~teh/smlm.html">https://www.stats.ox.ac.uk/~teh/smlm.html</a></p>					

## **II. Asosiy nazariy qism (ma'ruza mashg'ulotlari)**

### **II.1. Fan tarkibiga quyidagi mavzular kiradi:**

#### **1-mavzu. Data miningda berilganlar tushunchasi.**

Berilganlar tushunchasi va uning turlari. Berilganlarning tayanch turlari. Berilganlarning abstrakt turlari.

#### **2-mavzu. Dasturlash asoslari bilan tanishish.**

Python dasturlash tili va uning imkoniyatlari. Python dasturlash tili sintaksisi. Python dasturlash tilida berilganlar strukturalari. Python dasturlash tilining asosiy operatorlari. Python dasturlash tili kutubxonalari.

#### **3-mavzu. Berilganlar ustida statistik tahlillarni amalga oshirish.**

O'rgatuvchi tanlanma tushunchasi. Tanlanmada obyekt va xossa munosabatlari. Alomatlar bo'yicha o'rtacha, mediana, medoid va standart chetlashish qiymatlarini aniqlash.

#### **4-mavzu. Chiziqli algebra bilan tanishish.**

Vektorlar ustida amallar bajarish. Vektorlarni qo'shish. Vektorlarni ko'paytirish. Vektorlarni songa ko'paytirish, bo'lish. Matritsalar ustida arifmetik amallar. Matritsalarini qo'shish. Matritsalarini ustun bo'yicha, satr bo'yicha tartiblash. Matritsani transponirlash.

#### **5-mavzu. Berilganlarga dastlabki ishlov berish texnologiyalari.**

Berilganlarni normallashtirish usullari bilan tanishish. Obyektlar orasidagi masofalarni hisoblash metrikalari bilan tanishish va ulardan foydalanish.

#### **6-mavzu. Tushirib qoldirilgan berilganlar bilan ishlash.**

Tushirib qoldirilgan qiymatlarni aniqlash. Tushirib qoldirilgan qiymatlarni KNN yordamida to'ldirish. Tushirib qoldirilgan miqdoriy alomatlar qiymatlarini regressiya masalasi yordamida to'ldirish.

#### **7-mavzu. Berilganlarni vizuallashtirish.**

Miqdoriy berilganlarni fazoda tasvirlash. PCA, t-SNE metodlari yordamida fazoning o'lchamini kichraytirish. Matplotlib kutubxonasi imkoniyatlaridan foydalangan holda berilganlarni tasvirlash.

#### **8-mavzu. Berilganlarni yig'ish va tashkillashtirish.**

Kaggle platformasi bilan tanishish. Ochiq ma'lumotlar bazalari bilan ishlash. Turli xil fayl strukturalari bilan ishlash.

#### **9-mavzu. Matnli berilganlar bilan ishlash.**

Matnli berilganlar bazalarini tashkil qilish. Tabiiy tilni qayta ishlashning asosiy texnologiyalari: tokenizatsiya, lemmatizatsiya, stemming, parsing. Matnlarni vektor ko'rinishda ifodalash.

#### **10-mavzu. Ma'lumotlar bazasini tashkillashtirish.**

Obyekt tushunchasi. Obyektlarni bog'lanishlar modeli (Entity-relationship). Obyektlar bog'lanishlar diagrammalari. Obyektlar bog'lanishlarini normallashtirish.

#### **11-mavzu. SQL asoslari.**

SQLga kirish. SQL so'rovlar va uning turlari. SQL so'rovlar yaratish va amalga oshirish.

#### **12-mavzu. Berilganlar bazasini boshqarish tizimlari.**

Berilganlar bazasini boshqarish tizimlari haqida ma'lumot. Berilganlar bazasini boshqarish tizimlarining turlari. PostgreSQL bilan tanishish.

#### **13-mavzu. PostgreSQL bilan ishlash.**

PostgreSQLni o'rnatish. PostgreSQLda ishlash muhitlari. PostgreSQL bilan dasturlash tillarini bog'lash.

#### **14-mavzu. NoSQL berilganlar bazasi.**

NoSQL berilganlar bazasining asosiy xususiyatlari. NoSQL berilganlar bazasining avzalliklari va kamchiliklari. Keng tarqalgan NoSQL berilganlar bazalari.

#### **15-mavzu. BigData tushunchasi.**

BigData tushunchasi. BigData xususiyatlari. BigData hosil qilish texnologiyalari. Katta hajmdagi berilganlar muammolari. Ma'lumotlarni boshqarish evolyutsiyasi. Katta hajmdagi ma'lumotlar uchun asosiy texnologiyalar. Katta ma'lumotlarning ta'siri.

#### **16-mavzu. Bulutli texnologiyalar.**

Bulutli hisoblashning rivojlanishi. Bulutli hisoblashning asosiy tamoyillari. Bulutli xizmat modellari. Bulutli hisoblashning afzalliklari va muammolari.

#### **17-mavzu. DataMining tushunchasi.**

Data Mining texnologiyalari va dasturiy vositalari. Data Mining tadbirlari.

#### **18-mavzu. Klassifikatsiya algoritmlari.**

Qaror daraxtlari. ID3, C4.5 va CART algoritmlari. Qaror daraxtlarini qurish, kesish va talqin qilish. Naiv Bayes tasniflagichlari. Ehtimollarga asoslangan tasniflash. Matnlarni tasniflash va spamni aniqlashdagi ilovalar. Tayanch Vektorlar Usuli (SVM). Ma'lumotlarni ajratish uchun optimal gipertekislikni topish. Yadro funksiyalari va ularning tasnifga ta'siri. K eng yaqin qo'shnilar usuli (KNN). K ning optimal qiymatini tanlash.

### **19-mavzu. Berilganlar ustida klaster tahlilni amalga oshirish.**

K-means algoritmi. Klasterlarning optimal sonini tanlash. Ierarxiy klasterlash. O'xshashlik asosida klasterlar ierarxiyasini qurish. Aglomerativ va bo'linuvchi klasterlash usullari. Zichlikka asoslangan klasterlash. Turli xil shakl va zichlikdagi klasterlarni topish uchun DBSCAN algoritmi. Turli xil zichlik darajalariga ega bo'lgan klasterlarni aniqlash.

### **20-mavzu. Assotsiatsiya qoidalarini qazib olish.**

Apriori algoritmi. Tez-tez topiladigan narsalar to'plami va birlashma qoidalari. Qoidalarni baholash uchun qo'llab-quvvatlash, ishonch va ko'tarilish ko'rsatkichlari. FP-o'sish algoritmi. Tez-tez naqshlarni qazib olish uchun samarali algoritim. Bozor savatchasini tahlil qilish va tavsiya qilish tizimlarida qo'llanilishi.

### **21-mavzu. Regressiya algoritmlari.**

Chiziqli regressiya. O'zgaruvchilar orasidagi eng yaxshi chiziqli munosabatni topish. R-squared va MSE kabi ko'rsatkichlar yordamida regressiya modellarini baholash. Logistik regressiya. Binar klassifikatsiya. Firitbgarlikni aniqlash va kredit reytingidagi ilovalar. Regressiya masalasi uchun qarorlar daraxti. Uzluksiz o'zgaruvchilarni bashorat qilish uchun qarorlar daraxtidan foydalanish.

### **22-mavzu. Modelni baholash va tanlash.**

Modelni baholash uchun ko'rsatkichlar. Aniqlik, aniqlik, eslab qolish, tasniflash vazifalari uchun F1 ball. Regressiya vazifalari uchun o'rtacha kvadrat xatolik (MSE), ildiz o'rtacha kvadrat xatosi (RMSE). O'zaro tekshirish usullari. Kuchli modelni baholash uchun K-katlama o'zaro tekshirish.

### **23-mavzu. Berilganlar ombori (Data Warehouse) tushunchasi**

Ushbu mavzuda berilganlar omborining (Data Warehouse) asosiy tushunchalari, maqsadi va vazifalari bilan tanishiladi. Talabalar konsolidatsiyalangan (birlashtirilgan) ma'lumotlarni uzoq muddatli saqlash, ularni tahlil qilish va qaror qabul qilishda qanday foydalanilishini o'rganadilar. Mavzu davomida OLAP kublari (ko'p o'lchovli tahlil modellari) va ular orqali tezkor analitik so'rovlarni bajarish texnologiyasi ko'rib chiqiladi. Shuningdek, saqlash arxitekturasi, yulduz (star), qor yog'ishi (snowflake) va galaktika (galaxy) modellari bilan tanishiladi.

### **24-mavzu. ETL jarayoni: Extract, Transform, Load amaliyoti**

Ushbu mavzuda ETL (Extract, Transform, Load) — ma'lumotlarni olish, o'zgartirish va yuklash jarayoni bilan tanishiladi. Talabalar turli manbalardan (ma'lumotlar bazalari, fayllar, API) ma'lumotlarni qanday olish (extract), ularni kerakli formatga mos ravishda tozalash, normallashtirish va konvertatsiya qilish (transform), so'ngra maqsadli omborga (Data Warehouse yoki boshqaruv tizimiga) qanday yuklash (load) jarayonlari bilan amaliy tarzda shug'ullanishadi.

Shuningdek, ETL jarayonini avtomatlashtirishda ishlatiladigan vositalar (masalan, Apache Airflow, Talend, SSIS) haqida ham tushuncha beriladi.

### **25-mavzu. OLAP va OLTP tizimlari farqi**

Ushbu mavzuda OLTP (Online Transaction Processing) va OLAP (Online Analytical Processing) tizimlarining mohiyati, vazifalari va o'zaro farqlari o'rganiladi. OLTP tizimlari odatda real vaqтли tranzaksiyalarni (masalan, bank operatsiyalari, buyurtmalar) qayd etish uchun ishlatiladi, OLAP esa katta hajmdagi tarixiy ma'lumotlarni ko'p o'lchovli tahlil qilish va strategik qaror qabul qilish uchun mo'ljallangan. Mavzuda ularning arxitekturasi, so'rov turlari, ishlash samaradorligi va qo'llanilish sohalari o'rtasidagi farqlar amaliy misollar bilan tushuntiriladi. OLAP analitik tizimlari va OLTP tranzaksion tizimlari farqi, har birining vazifalari.

### **26-mavzu. Data Mining algoritmlariga kirish**

Ushbu mavzuda data mining (ma'lumotlardan bilim ajratish) jarayonining asosiy maqsadlari va algoritmlari bilan tanishiladi. Talabalar ma'lumotlarni guruhlash (klasterlash), tasniflash (klassifikatsiya), bog'lanish qoidalarini aniqlash (association rules) kabi asosiy usullarni o'rganadilar. Mavzuda har bir algoritmnining umumiy ishlash prinsipi, real hayotdagi qo'llanilishi va foydaliligi ko'rib chiqiladi. Bu kirish bosqichi bo'lib, keyingi mavzularda har bir algoritmnin chuqurroq yoritiladi.

### **27-mavzu. K-means klasterlash algoritmi**

Ushbu mavzuda mashhur va keng qo'llaniladigan klasterlash algoritmlaridan biri bo'lgan **K-means** algoritmi bilan tanishiladi. K-means — nazoratsiz o'rganishga asoslangan algoritmn bo'lib, berilganlarni o'zaro o'xshashliklariga qarab guruhlarga (klasterlarga) ajratadi. Mavzuda algoritmnining ishlash bosqichlari, klaster markazlarini aniqlash, Euclid masofasi bilan hisoblash, "elbow method" yordamida optimal **K** qiymatini topish usullari o'rganiladi. Talabalar algoritmnin Python dasturlash tilida amaliy misollar bilan bajarishni ham o'rganadilar.

### **28-mavzu. Beshrik (Association Rule Mining) – Apriori algoritmi**

Ushbu mavzuda **beshrik** (assotsiativ qoidalar asosida ma'lumotlar qazib olish) tushunchasi va **Apriori algoritmi** bilan tanishiladi. Bu usul asosan xaridlar savati tahlili (market basket analysis) yoki biror hodisani bashorat qilishda ishlatiladi. Talabalar **support** (qo'llab-quvvatlash), **confidence** (ishonchlilik) va **lift** kabi asosiy statistik mezonlar bilan tanishadilar. Apriori algoritmnining ishlash prinsipi, qadam-baqadam qoidalarni aniqlash usullari va ularni Python kutubxonalari yordamida qanday amalga oshirish amaliyotda ko'rib chiqiladi.

### **29-mavzu. K-nafari (KNN) algoritmi orqali tasniflash**

Ushbu mavzuda **KNN (K-Nearest Neighbors)** algoritmi orqali ma'lumotlarni

tasniflash usuli o'rganiladi. Bu nazoratli o'rganishga (supervised learning) asoslangan oddiy, ammo samarali klassifikatsiya usullaridan biridir. Har bir yangi kuzatuv qiymati unga eng yaqin bo'lgan  $K$  ta qo'shning sinfiga qarab belgilab beriladi. Mavzuda Euclid masofasi,  $K$  qiymatini tanlash, KNNning afzalliklari va cheklovlari hamda Python orqali amaliy qo'llanilishi yoritiladi.

### 30-mavzu. Beshlik qaror daraxtlari (Decision Trees)

Ushbu mavzuda **Decision Tree (qaror daraxti)** algoritmining ishlash prinsipi, tahlil jarayonidagi o'rni va amaliy qo'llanilishi o'rganiladi. Qaror daraxtlari ma'lumotlarni tahlil qilish va tasniflash uchun ishlatiladigan intuitiv va tushunarli model bo'lib, u har bir savolga "ha/yo'q" shaklida javob berish orqali qarorga olib keladi. Mavzuda **Gini impurtasi, axborot darajasi (information gain), entropy** kabi tushunchalar, daraxtni o'stirish va ortiqcha moslashuv (overfitting) muammolarining oldini olish usullari (pruning) yoritiladi. Talabalar ushbu algoritmni Python yordamida amaliyotda sinab ko'radilar.

### 31-mavzu. Naive Bayes tasniflovchi

Ushbu mavzuda **Naive Bayes** algoritmi — ehtimolliklarga asoslangan tasniflash (klassifikatsiya) usuli o'rganiladi. Ushbu algoritm **Bayes teoremasi** asosida ishlaydi va har bir xususiyat (feature) mustaqil deb faraz qilinadi ("naive" – soddaligi shundan). Mavzuda Bayes formulasi, sinflar ehtimolini hisoblash, shuningdek, bu usulning **matnli berilganlar (spam filter, sentiment tahlil)** bilan ishlashdagi afzalliklari tushuntiriladi. Amaliyotda Python kutubxonasi yordamida model yaratish, o'rgatish va test qilish ko'nikmalari shakllantiriladi.

### 32-mavzu. PCA – Asosiy komponentlarga ajratish

Ushbu mavzuda **PCA (Principal Component Analysis)** — o'lchamlarni kamaytirish (dimensionality reduction) usuli bilan tanishiladi. PCA yordamida ko'p xususiyatli (feature) ma'lumotlar ichidan eng asosiylarini aniqlab, ularni kamroq o'lchovli makonga proyeksiya qilish mumkin. Bu usul vizualizatsiya, klasterlash va model samaradorligini oshirishda keng qo'llaniladi. Mavzuda kovariatsiya, eigenvalues/eigenvectors, dispersiyani saqlab qolish kabi tushunchalar, shuningdek, PCA'ni Python yordamida amalga oshirish bosqichlari o'rganiladi.

### 33-mavzu. Outlier detection – noodatiy qiymatlarni aniqlash

Ushbu mavzuda **noodatiy qiymatlar (outlier)** tushunchasi va ularni aniqlash usullari o'rganiladi. Noodatiy qiymatlar — bu asosiy tendensiyalardan keskin farq qiladigan, tahlilga salbiy ta'sir ko'rsatishi mumkin bo'lgan qiymatlardir. Mavzuda statistik yondashuvlar (Z-score, IQR), grafik metodlar (boxplot, scatter plot) va mashinaviy yondashuvlar (Isolation Forest, DBSCAN) orqali noodatiy qiymatlarni aniqlash va ularni boshqarish usullari ko'rib chiqiladi. Talabalar Python vositalari yordamida real ma'lumotlarda outlierni tahlil qilishni o'rganadilar.

### **34-mavzu. Data miningda etik muammolar va maxfiylik**

Ushbu mavzuda data mining jarayonida yuzaga keladigan etik muammolar va foydalanuvchi ma'lumotlarining maxfiyligini ta'minlash masalalari muhokama qilinadi. Talabalar shaxsiy ma'lumotlarni to'plash, saqlash va tahlil qilishda yuzaga keladigan xavf-xatarlar, foydalanuvchi roziligi, diskriminatsiya xavfi, noto'g'ri qarorlar va algoritmik noxolislik kabi muammolar bilan tanishadilar. Shuningdek, axborot maxfiyligini himoya qilish usullari, anonimlashtirish, ma'lumotlarni shifrlash va tegishli qonunchilik (masalan, GDPR) haqida umumiy tushuncha beriladi.

### **35-mavzu. Data Mining va Machine Learning: o'xshashlik va farqlar**

Ushbu mavzuda **Data Mining (ma'lumotlardan bilim ajratish)** va **Machine Learning (mashinali o'rganish)** tushunchalari, ularning maqsadlari, metodlari va qo'llanilish sohalari o'rganiladi. Ikkala soha bir-biriga yaqin bo'lsa-da, ularning yondashuvi va qo'llanilishi jihatidan farqli tomonlari mavjud. Mavzuda talabalarga nazoratli/nazoratsiz o'rganish, bashoratli modellar, statistik tahlil, va avtomatik o'rganishga asoslangan tizimlar orasidagi farqlar haqida tushuncha beriladi. Real hayotdagi misollar orqali o'xshashlik va farqlar aniq ko'rsatib beriladi.

### **36-mavzu. Big Data tushunchasi va texnologiyalari**

Ushbu mavzuda **Big Data (katta ma'lumotlar)** tushunchasi, uning xususiyatlari va zamonaviy texnologiyalardagi ahamiyati o'rganiladi. Talabalar katta hajmli, yuqori tezlikda va xilma-xil shakldagi ma'lumotlar oqimini qanday boshqarish va tahlil qilish haqida asosiy bilimlarga ega bo'ladilar. Shuningdek, **5V modeli** (Volume, Velocity, Variety, Veracity, Value), an'anaviy tizimlar bilan farqlari, va Big Data uchun mo'ljallangan texnologiyalar — **Hadoop, Spark, Kafka, NoSQL** bazalari bilan tanishadilar.

### **37-mavzu. Vaqtli qatorlar (Time series) tahlili**

Ushbu mavzuda **vaqtli qatorlar (time series)** — vaqtga bog'liq bo'lgan ma'lumotlar ketma-ketligini tahlil qilish usullari o'rganiladi. Vaqtli qatorlar moliyaviy bozorlar, ob-havo, sensor ma'lumotlari, sotuvlar statistikasi kabi ko'plab sohalarda uchraydi. Mavzuda asosiy komponentlar — **trend, sezonsizlik va shovqin (noise)** tushunchalari, shuningdek, **harakatlanuvchi o'rtacha (moving average), dekompozitsiya**, va **ARIMA** modellarning nazariyasi va Python yordamida amaliy qo'llanilishi yoritiladi.

### **38-mavzu. Web scraping va API orqali Internetdan ma'lumotlarni avtomatik yig'ish usullari**

Ushbu mavzuda **web scraping** — internetdagi ochiq veb-sahifalardan avtomatik tarzda ma'lumot yig'ish usullari o'rganiladi. Talabalar HTML tuzilmasi, elementlar atributlari, URL'lar va sahifa navigatsiyasi bilan tanishadilar hamda **Python** dasturlash tilida requests, BeautifulSoup, Selenium kabi kutubxonalar yordamida **amaliy scraping** bajaradilar. Shuningdek, **veb-saytlarning**

foydalanish shartlari, **robots.txt**, va etik muammolar haqida tushunchalar ham beriladi.

### **39-mavzu. Data Miningda Python kutubxonalari va Mashinaviy o'rganish orqali ma'lumotlardan bilim olish usullari**

Ushbu mavzuda **Data Mining** (ma'lumotlardan bilim ajratish) jarayonlarida keng qo'llaniladigan **Python kutubxonalari** bilan tanishiladi. Talabalar ma'lumotlarni yuklash, tozalash, tahlil qilish, modellashtirish va vizualizatsiya qilish jarayonida foydalaniladigan kutubxonalar — **pandas, NumPy, scikit-learn, matplotlib, seaborn** va **mlxtend** kabi vositalarning asosiy funksiyalari bilan amaliyotda ishlashni o'rganadilar. **Data Mining** jarayonining asosiy maqsadi — **katta hajmdagi ma'lumotlardan foydali bilim va naqshlarni ajratib olish** (knowledge discovery) tushunchasi yoritiladi. Talabalar **Data Mining** jarayonining bosqichlari, tahliliy usullari, ma'lumotdan to bilimgacha bo'lgan yo'lni o'rganadilar. Shuningdek, **Knowledge Discovery in Databases (KDD)** modeli asosida bilim chiqarishning nazariy asoslari va amaliy uslublari ko'rib chiqiladi.

### **40-mavzu. SQL'da murakkab so'rovlar va optimallashtirish. MongoDB va JSON formatda ma'lumotlar bilan ishlash**

Ushbu mavzuda **SQL tilida murakkab so'rovlar yozish** va ularning samaradorligini **optimallashtirish** usullari o'rganiladi. Talabalar bir nechta jadvallar bilan ishlash, **INNER JOIN, LEFT JOIN, UNION, SUBQUERY, GROUP BY, HAVING**, hamda **CTE (Common Table Expressions)** kabi murakkab operatorlar yordamida tahliliy so'rovlar tuzishni o'rganadilar. Shuningdek, indekslardan foydalanish, **EXPLAIN** buyrug'i yordamida so'rov rejalarini tahlil qilish va **so'rovlarni tezlashtirish texnikalari** bilan tanishadilar. **MongoDB** — NoSQL tipidagi hujjatga asoslangan ma'lumotlar bazasi bilan tanishiladi. Talabalar **JSON (JavaScript Object Notation)** formatidagi ma'lumotlar bilan qanday ishlash, ma'lumotlar qo'shish, yangilash, qidirish va o'chirish amallarini bajarishni o'rganadilar. **MongoDB**'ning an'anaviy (relatsion) bazalardan farqlari, moslashuvchan tuzilmasi va katta hajmdagi yarimstrukturalangan ma'lumotlar bilan ishlashdagi ustunliklari yoritiladi.

### **41-mavzu. Berilganlar normallashtirilishi va transformatsiyasi**

Ushbu mavzuda ma'lumotlar (berilganlar) ustida oldindan ishlov berish bosqichlaridan biri bo'lgan normallashtirish (normalization) va transformatsiya (o'zgartirish) jarayonlari o'rganiladi. Ushbu jarayonlar model sifatini oshirish, o'rganish tezligini yaxshilash va aniq natijaga erishish uchun muhimdir. Talabalar min-max scaling, z-score, log transformatsiya, categorical encoding (one-hot, label encoding) kabi usullar bilan tanishadilar va ularni Python vositalari bilan amalda bajaradilar.

#### **42-mavzu. Model baholash metrikalari**

Ushbu mavzuda mashinaviy o'rganish (Machine Learning) modellari natijalarini baholash uchun ishlatiladigan asosiy metrikalar bilan tanishiladi. To'g'ri baholash modelning sifatini tushunish, uni takomillashtirish va real sohalarida ishonchli ishlatish uchun muhimdir. Talabalar klassifikatsiya va regressiya modellari uchun turli metrikalarni, shuningdek confusion matrix, precision, recall, F1-score, ROC-AUC, va MSE, MAE,  $R^2$  kabi baholovchilarni o'rganadilar.

### **III. Amaliy mashg'ulotlari bo'yicha ko'rsatma va tavsiyalar**

*"Ma'lumotlar bazasi va Data Mining" fani bo'yicha amaliy mashg'ulotlar uchun quyidagi mavzular tavsiya etiladi:*

#### **1-mavzu. Data miningda berilganlar tushunchasi.**

Berilganlar tushunchasi va uning turlari. Berilganlarning tayanch turlari. Berilganlarning abstrakt turlari.

#### **2-mavzu. Dasturlash asoslari bilan tanishish.**

Python dasturlash tili va uning imkoniyatlari. Python dasturlash tili sintaksisi. Python dasturlash tilida berilganlar strukturallari. Python dasturlash tilining asosiy operatorlari. Python dasturlash tili kutubxonalari.

#### **3-mavzu. Berilganlar ustida statistik tahlillarni amalga oshirish.**

O'rgatuvchi tanlanma tushunchasi. Tanlanmada obyekt va xossa munosabatlari. Alomatlar bo'yicha o'rtacha, mediana, medoid va standart chetlashish qiymatlarini aniqlash.

#### **4-mavzu. Chiziqli algebra bilan tanishish.**

Vektorlar ustida amallar bajarish. Vektorlarni qo'shish. Vektorlarni ko'paytirish. Vektorlarni songa ko'paytirish, bo'lish. Matritsalar ustida arifmetik amallar. Matritsalarini qo'shish. Matritsalarini ustun bo'yicha, satr bo'yicha tartiblash. Matritsani transponirlash.

#### **5-mavzu. Berilganlarga dastlabki ishlov berish texnologiyalari.**

Berilganlarni normallashtirish usullari bilan tanishish. Obyektlar orasidagi masofalarni hisoblash metrikalari bilan tanishish va ulardan foydalanish.

#### **6-mavzu. Tushirib qoldirilgan berilganlar bilan ishlash.**

Tushirib qoldirilgan qiymatlarni aniqlash. Tushirib qoldirilgan qiymatlarni KNN yordamida to'ldirish. Tushirib qoldirilgan miqdoriy alomatlar qiymatlarini regressiya masalasi yordamida to'ldirish.

#### **7-mavzu. Berilganlarni vizuallashtirish.**

Miqdoriy berilganlarni fazoda tasvirlash. PCA, t-SNE metodlari yordamida fazoning o'lchamini kichraytirish. Matplotlib kutubxonasi imkoniyatlaridan foydalangan holda berilganlarni tasvirlash.

**8-mavzu. Berilganlarni yig'ish va tashkillashtirish.**

Kaggle platformasi bilan tanishish. Ochiq ma'lumotlar bazalari bilan ishlash. Turli xil fayl strukturalari bilan ishlash.

**9-mavzu. Matnli berilganlar bilan ishlash.**

Matnli berilganlar bazalarini tashkil qilish. Tabiiy tilni qayta ishlashning asosiy texnologiyalari: tokenizatsiya, lemmatizatsiya, stemming, parsing. Matnlarni vektor ko'rinishda ifodalash.

**10-mavzu. Ma'lumotlar bazasini tashkillashtirish.**

Obyekt tushunchasi. Obyektlarni bog'lanishlar modeli (Entity-relationship). Obyektlar bog'lanishlar diagrammalari. Obyektlar bog'lanishlarini normallashtirish.

**11-mavzu. SQL asoslari.**

SQLga kirish. SQL so'rovlar va uning turlari. SQL so'rovlar yaratish va amalga oshirish.

**12-mavzu. Berilganlar bazasini boshqarish tizimlari.**

Berilganlar bazasini boshqarish tizimlari haqida ma'lumot. Berilganlar bazasini boshqarish tizimlarining turlari. PostgreSQL bilan tanishish.

**13-mavzu. PostgreSQL bilan ishlash.**

PostgreSQLni o'rnatish. PostgreSQLda ishlash muhitlari. PostgreSQL bilan dasturlash tillarini bog'lash.

**14-mavzu. NoSQL berilganlar bazasi.**

NoSQL berilganlar bazasining asosiy xususiyatlari. NoSQL berilganlar bazasining avzalliklari va kamchiliklari. Keng tarqalgan NoSQL berilganlar bazalari.

**15-mavzu. BigData tushunchasi.**

BigData tushunchasi. BigData xususiyatlari. BigData hosil qilish texnologiyalari. Katta hajmdagi berilganlar muammolari. Ma'lumotlarni boshqarish evolyutsiyasi. Katta hajmdagi ma'lumotlar uchun asosiy texnologiyalar. Katta ma'lumotlarning ta'siri.

**16-mavzu. Bulutli texnologiyalar.**

Bulutli hisoblashning rivojlanishi. Bulutli hisoblashning asosiy tamoyillari. Bulutli xizmat modellari. Bulutli hisoblashning afzalliklari va muammolari.

**17-mavzu. DataMining tushunchasi.**

Data Mining texnologiyalari va dasturiy vositalari. Data Mining tadbirlari.

### **18-mavzu. Klassifikatsiya algoritmlari.**

Qaror daraxtlari. ID3, C4.5 va CART algoritmlari. Qaror daraxtlarini qurish, kesish va talqin qilish. Naiv Bayes tasniflagichlari. Ehtimollarga asoslangan tasniflash. Matnlarni tasniflash va spamni aniqlashdagi ilovalar. Tayanch Vektorlar Usuli (SVM). Ma'lumotlarni ajratish uchun optimal gipertekislikni topish. Yadro funksiyalari va ularning tasnifga ta'siri. K eng yaqin qo'shnilar usuli (KNN). K ning optimal qiymatini tanlash.

### **19-mavzu. Berilganlar ustida klaster tahlilni amalga oshirish.**

K-means algoritmi. Klasterlarning optimal sonini tanlash. Ierarxik klasterlash. O'xshashlik asosida klasterlar ierarxiyasini qurish. Aglomerativ va bo'linuvchi klasterlash usullari. Zichlikka asoslangan klasterlash. Turli xil shakl va zichlikdagi klasterlarni topish uchun DBSCAN algoritmi. Turli xil zichlik darajalariga ega bo'lgan klasterlarni aniqlash.

### **20-mavzu. Assotsiatsiya qoidalarini qazib olish.**

Apriori algoritmi. Tez-tez topiladigan narsalar to'plami va birlashma qoidalari. Qoidalarni baholash uchun qo'llab-quvvatlash, ishonch va ko'tarilish ko'rsatkichlari. FP-o'sish algoritmi. Tez-tez naqshlarni qazib olish uchun samarali algoritim. Bozor savatchasini tahlil qilish va tavsiya qilish tizimlarida qo'llanilishi.

### **21-mavzu. Regressiya algoritmlari.**

Chiziqli regressiya. O'zgaruvchilar orasidagi eng yaxshi chiziqli munosabatni topish. R-squared va MSE kabi ko'rsatkichlar yordamida regressiya modellarini baholash. Logistik regressiya. Binar klassifikatsiya. Firibgarlikni aniqlash va kredit reytingidagi ilovalar. Regressiya masalasi uchun qarorlar daraxti. Uzlüksiz o'zgaruvchilarni bashorat qilish uchun qarorlar daraxtidan foydalanish.

### **22-mavzu. Modelni baholash va taulash.**

Modelni baholash uchun ko'rsatkichlar. Aniqlik, aniqlik, eslab qolish, tasniflash vazifalari uchun F1 ball. Regressiya vazifalari uchun o'rtacha kvadrat xatolik (MSE), ildiz o'rtacha kvadrat xatosi (RMSE). O'zaro tekshirish usullari. Kuchli modelni baholash uchun K-katlama o'zaro tekshirish.

### **23-mavzu. Berilganlar ombori (Data Warehouse) tushunchasi**

Ushbu mavzuda berilganlar omborining (Data Warehouse) asosiy tushunchalari, maqsadi va vazifalari bilan tanishiladi. Talabalar konsolidatsiyalangan (birlashtirilgan) ma'lumotlarni uzoq muddatli saqlash, ularni tahlil qilish va qaror qabul qilishda qanday foydalanilishini o'rganadilar. Mavzu davomida OLAP kublari (ko'p o'lchovli tahlil modellari) va ular orqali tezkor analitik so'rovlarni

bajarish texnologiyasi ko'rib chiqiladi. Shuningdek, saqlash arxitekturasi, yulduz (star), qor yog'ishi (snowflake) va galaktika (galaxy) modellari bilan tanishiladi.

#### **24-mavzu. ETL jarayoni: Extract, Transform, Load amaliyoti**

Ushbu mavzuda ETL (Extract, Transform, Load) — ma'lumotlarni olish, o'zgartirish va yuklash jarayoni bilan tanishiladi. Talabalar turli manbalardan (ma'lumotlar bazalari, fayllar, API) ma'lumotlarni qanday olish (extract), ularni kerakli formatga mos ravishda tozalash, normallashtirish va konvertatsiya qilish (transform), so'ngra maqsadli omborga (Data Warehouse yoki boshqaruv tizimiga) qanday yuklash (load) jarayonlari bilan amaliy tarzda shug'ullanishadi. Shuningdek, ETL jarayonini avtomatlashtirishda ishlatiladigan vositalar (masalan, Apache Airflow, Talend, SSIS) haqida ham tushuncha beriladi.

#### **25-mavzu. OLAP va OLTP tizimlari farqi**

Ushbu mavzuda OLTP (Online Transaction Processing) va OLAP (Online Analytical Processing) tizimlarining mohiyati, vazifalari va o'zaro farqlari o'rganiladi. OLTP tizimlari odatda real vaqtli tranzaksiyalarni (masalan, bank operatsiyalari, buyurtmalar) qayd etish uchun ishlatiladi, OLAP esa katta hajmdagi tarixiy ma'lumotlarni ko'p o'lchovli tahlil qilish va strategik qaror qabul qilish uchun mo'ljallangan. Mavzuda ularning arxitekturasi, so'rov turlari, ishlash samaradorligi va qo'llanilish sohalari o'rtasidagi farqlar amaliy misollar bilan tushuntiriladi. OLAP analitik tizimlari va OLTP tranzaksion tizimlari farqi, har birining vazifalari.

#### **26-mavzu. Data Mining algoritmlariga kirish**

Ushbu mavzuda data mining (ma'lumotlardan bilim ajratish) jarayonining asosiy maqsadlari va algoritmlari bilan tanishiladi. Talabalar ma'lumotlarni guruhlash (klasterlash), tasniflash (klassifikatsiya), bog'lanish qoidalarini aniqlash (association rules) kabi asosiy usullarni o'rganadilar. Mavzuda har bir algoritmning umumiy ishlash prinsipi, real hayotdagi qo'llanilishi va foydaliligi ko'rib chiqiladi. Bu kirish bosqichi bo'lib, keyingi mavzularda har bir algoritm chuqurroq yoritiladi.

#### **27-mavzu. K-means klasterlash algoritmi**

Ushbu mavzuda mashhur va keng qo'llaniladigan klasterlash algoritmlaridan biri bo'lgan **K-means** algoritmi bilan tanishiladi. K-means — nazoratsiz o'rganishga asoslangan algoritim bo'lib, berilganlarni o'zaro o'xshashliklariga qarab guruhlariga (klasterlarga) ajratadi. Mavzuda algoritimning ishlash bosqichlari, klaster markazlarini aniqlash, Euclid masofasi bilan hisoblash, "elbow method" yordamida optimal **K** qiymatini topish usullari o'rganiladi. Talabalar algoritimni Python dasturlash tilida amaliy misollar bilan bajarishni ham o'rganadilar.

#### **28-mavzu. Beshrik (Association Rule Mining) – Apriori algoritmi**

Ushbu mavzuda **beshrik** (assotsiativ qoidalar asosida ma'lumotlar qazib

**olish**) tushunchasi va **Apriori algoritmi** bilan tanishiladi. Bu usul asosan xaridlar savati tahlili (market basket analysis) yoki biror hodisani bashorat qilishda ishlatiladi. Talabalar **support (qo'llab-quvvatlash)**, **confidence (ishonchlilik)** va **lift** kabi asosiy statistik mezonlar bilan tanishadilar. Apriori algoritmining ishlash prinsipi, qadam-baqadam qoidalarni aniqlash usullari va ularni Python kutubxonalari yordamida qanday amalga oshirish amaliyotda ko'rib chiqiladi.

### **29-mavzu. K-nafari (KNN) algoritmi orqali tasniflash**

Ushbu mavzuda **KNN (K-Nearest Neighbors)** algoritmi orqali ma'lumotlarni tasniflash usuli o'rganiladi. Bu nazoratli o'rganishga (supervised learning) asoslangan oddiy, ammo samarali klassifikatsiya usullaridan biridir. Har bir yangi kuzatuv qiymati unga eng yaqin bo'lgan **K** ta qo'shning sinfiga qarab belgilab beriladi. Mavzuda Euclid masofasi, **K** qiymatini tanlash, **KNN**ning afzalliklari va cheklovlari hamda Python orqali amaliy qo'llanilishi yoritiladi.

### **30-mavzu. Beshlik qaror daraxtlari (Decision Trees)**

Ushbu mavzuda **Decision Tree (qaror daraxti)** algoritmining ishlash prinsipi, tahlil jarayonidagi o'rni va amaliy qo'llanilishi o'rganiladi. Qaror daraxtlari ma'lumotlarni tahlil qilish va tasniflash uchun ishlatiladigan intuitiv va tushunarli model bo'lib, u har bir savolga "ha/yo'q" shaklida javob berish orqali qarorga olib keladi. Mavzuda **Gini impurtasi**, **axborot darajasi (information gain)**, **entropy** kabi tushunchalar, daraxtni o'stirish va ortiqcha moslashuv (overfitting) muammolarining oldini olish usullari (pruning) yoritiladi. Talabalar ushbu algoritmi Python yordamida amaliyotda sinab ko'radilar.

### **31-mavzu. Naive Bayes tasniflovchi**

Ushbu mavzuda **Naive Bayes** algoritmi — ehtimolliklarga asoslangan tasniflash (klassifikatsiya) usuli o'rganiladi. Ushbu algoritm **Bayes teoremasi** asosida ishlaydi va har bir xususiyat (feature) mustaqil deb faraz qilinadi ("naive" – soddaligi shundan). Mavzuda Bayes formulasi, sinflar ehtimolini hisoblash, shuningdek, bu usulning **matnli berilganlar (spam filter, sentiment tahlil)** bilan ishlashdagi afzalliklari tushuntiriladi. Amaliyotda Python kutubxonasi yordamida model yaratish, o'rgatish va test qilish ko'nikmalari shakllantiriladi.

### **32-mavzu. PCA – Asosiy komponentlarga ajratish**

Ushbu mavzuda **PCA (Principal Component Analysis)** — o'lchamlarni kamaytirish (dimensionality reduction) usuli bilan tanishiladi. PCA yordamida ko'p xususiyatli (feature) ma'lumotlar ichidan eng asosiyalarini aniqlab, ularni kamroq o'lchovli makonga proyeksiya qilish mumkin. Bu usul vizualizatsiya, klasterlash va model samaradorligini oshirishda keng qo'llaniladi. Mavzuda kovariatsiya, eigenvalues/eigenvectors, dispersiyani saqlab qolish kabi tushunchalar, shuningdek, PCA'ni Python yordamida amalga oshirish bosqichlari o'rganiladi.

### **33-mavzu. Outlier detection – noodatiy qiymatlarni aniqlash**

Ushbu mavzuda **noodatiy qiymatlar (outlier)** tushunchasi va ularni aniqlash usullari o'rganiladi. Noodatiy qiymatlar — bu asosiy tendensiyalardan keskin farq qiladigan, tahlilga salbiy ta'sir ko'rsatishi mumkin bo'lgan qiymatlardir. Mavzuda statistik yondashuvlar (Z-score, IQR), grafik metodlar (boxplot, scatter plot) va mashinaviy yondashuvlar (Isolation Forest, DBSCAN) orqali noodatiy qiymatlarni aniqlash va ularni boshqarish usullari ko'rib chiqiladi. Talabalar Python vositalari yordamida real ma'lumotlarda outlierlarni tahlil qilishni o'rganadilar.

### **34-mavzu. Data miningda etik muammolar va maxfiylik**

Ushbu mavzuda data mining jarayonida yuzaga keladigan etik muammolar va foydalanuvchi ma'lumotlarining maxfiyligini ta'minlash masalalari muhokama qilinadi. Talabalar shaxsiy ma'lumotlarni to'plash, saqlash va tahlil qilishda yuzaga keladigan xavf-xatarlar, foydalanuvchi roziligi, diskriminatsiya xavfi, noto'g'ri qarorlar va algoritmik noxolislik kabi muammolar bilan tanishadilar. Shuningdek, axborot maxfiyligini himoya qilish usullari, anonimlashtirish, ma'lumotlarni shifrlash va tegishli qonunchilik (masalan, GDPR) haqida umumiy tushuncha beriladi.

### **35-mavzu. Data Mining va Machine Learning: o'xshashlik va farqlar**

Ushbu mavzuda **Data Mining (ma'lumotlardan bilim ajratish)** va **Machine Learning (mashinali o'rganish)** tushunchalari, ularning maqsadlari, metodlari va qo'llanilish sohalari o'rganiladi. Ikkala soha bir-biriga yaqin bo'lsa-da, ularning yondashuvi va qo'llanilishi jihatidan farqli tomonlari mavjud. Mavzuda talabalarga nazoratli/nazoratsiz o'rganish, bashoratli modellar, statistik tahlil, va avtomatik o'rganishga asoslangan tizimlar orasidagi farqlar haqida tushuncha beriladi. Real hayotdagi misollar orqali o'xshashlik va farqlar aniq ko'rsatib beriladi.

### **36-mavzu. Big Data tushunchasi va texnologiyalari**

Ushbu mavzuda **Big Data (katta ma'lumotlar)** tushunchasi, uning xususiyatlari va zamonaviy texnologiyalardagi ahamiyati o'rganiladi. Talabalar katta hajmli, yuqori tezlikda va xilma-xil shakldagi ma'lumotlar oqimini qanday boshqarish va tahlil qilish haqida asosiy bilimlarga ega bo'ladilar. Shuningdek, **5V modeli** (Volume, Velocity, Variety, Veracity, Value), an'anaviy tizimlar bilan farqlari, va Big Data uchun mo'ljallangan texnologiyalar — **Hadoop, Spark, Kafka, NoSQL** bazalari bilan tanishadilar.

### **37-mavzu. Vaqtli qatorlar (Time series) tahlili**

Ushbu mavzuda **vaqtli qatorlar (time series)** — vaqtga bog'liq bo'lgan ma'lumotlar ketma-ketligini tahlil qilish usullari o'rganiladi. Vaqtli qatorlar moliyaviy bozorlar, ob-havo, sensor ma'lumotlari, sotuvlar statistikasi kabi ko'plab sohalarda uchraydi. Mavzuda asosiy komponentlar — **trend, sezonsizlik va shovqin (noise)** tushunchalari, shuningdek, **harakatlanuvchi**

**o'rtacha (moving average), dekompozitsiya, va ARIMA** modellarning nazariyasi va Python yordamida amaliy qo'llanilishi yoritiladi.

### **38-mavzu. Web scraping va API orqali Internetdan ma'lumotlarni avtomatik yig'ish usullari**

Ushbu mavzuda **web scraping** — internetdagi ochiq veb-sahifalardan avtomatik tarzda ma'lumot yig'ish usullari o'rganiladi. Talabalar HTML tuzilmasi, elementlar atributlari, URL'lar va sahifa navigatsiyasi bilan tanishadilar hamda **Python** dasturlash tilida requests, BeautifulSoup, Selenium kabi kutubxonalar yordamida amaliy scraping bajaradilar. Shuningdek, veb-saytlarning foydalanish shartlari, robots.txt, va etik muammolar haqida tushunchalar ham beriladi.

### **39-mavzu. Data Miningda Python kutubxonalari va Mashinaviy o'rganish orqali ma'lumotlardan bilim olish usullari**

Ushbu mavzuda **Data Mining** (ma'lumotlardan bilim ajratish) jarayonlarida keng qo'llaniladigan **Python kutubxonalari** bilan tanishiladi. Talabalar ma'lumotlarni yuklash, tozalash, tahlil qilish, modellashtirish va vizualizatsiya qilish jarayonida foydalaniladigan kutubxonalar — **pandas, NumPy, scikit-learn, matplotlib, seaborn** va **mlxtend** kabi vositalarning asosiy funksiyalari bilan amaliyotda ishlashni o'rganadilar. **Data Mining** jarayonining asosiy maqsadi — **katta hajmdagi ma'lumotlardan foydali bilim va naqshlarni ajratib olish** (knowledge discovery) tushunchasi yoritiladi. Talabalar Data Mining jarayonining bosqichlari, tahliliy usullari, ma'lumotdan to bilimgacha bo'lgan yo'lni o'rganadilar. Shuningdek, **Knowledge Discovery in Databases (KDD)** modeli asosida bilim chiqarishning nazariy asoslari va amaliy uslublari ko'rib chiqiladi.

### **40-mavzu. SQL'da murakkab so'rovlar va optimallashtirish. MongoDB va JSON formatda ma'lumotlar bilan ishlash**

Ushbu mavzuda **SQL** tilida murakkab so'rovlar yozish va ularning samaradorligini optimallashtirish usullari o'rganiladi. Talabalar bir nechta jadvallar bilan ishlash, **INNER JOIN, LEFT JOIN, UNION, SUBQUERY, GROUP BY, HAVING**, hamda **CTE (Common Table Expressions)** kabi murakkab operatorlar yordamida tahliliy so'rovlar tuzishni o'rganadilar. Shuningdek, indekslardan foydalanish, **EXPLAIN** buyrug'i yordamida so'rov rejalarini tahlil qilish va so'rovlarni **tezlashtirish texnikalari** bilan tanishadilar. **MongoDB** — NoSQL tipidagi hujjatga asoslangan ma'lumotlar bazasi bilan tanishiladi. Talabalar **JSON (JavaScript Object Notation)** formatidagi ma'lumotlar bilan qanday ishlash, ma'lumotlar qo'shish, yangilash, qidirish va o'chirish amallarini bajarishni o'rganadilar. **MongoDB**'ning an'anaviy (relatsion) bazalardan farqlari, moslashuvchan tuzilmasi va katta hajmdagi yarimstrukturalangan ma'lumotlar bilan ishlashdagi ustunliklari yoritiladi.

#### **41-mavzu. Berilganlar normallashtirilishi va transformatsiyasi**

Ushbu mavzuda ma'lumotlar (berilganlar) ustida oldindan ishlov berish bosqichlaridan biri bo'lgan normallashtirish (normalization) va transformatsiya (o'zgartirish) jarayonlari o'rganiladi. Ushbu jarayonlar model sifatini oshirish, o'rganish tezligini yaxshilash va aniq natijaga erishish uchun muhimdir. Talabalar min-max scaling, z-score, log transformatsiya, categorical encoding (one-hot, label encoding) kabi usullar bilan tanishadilar va ularni Python vositalari bilan amalda bajaradilar.

#### **42-mavzu. Model baholash metrikalari**

Ushbu mavzuda mashinaviy o'rganish (Machine Learning) modellari natijalarini baholash uchun ishlatiladigan asosiy metrikalar bilan tanishiladi. To'g'ri baholash modelning sifatini tushunish, uni takomillashtirish va real sohalarda ishonchli ishlatish uchun muhimdir. Talabalar klassifikatsiya va regressiya modellari uchun turli metrikalarni, shuningdek confusion matrix, precision, recall, F1-score, ROC-AUC, va MSE, MAE,  $R^2$  kabi baholovchilarni o'rganadilar.

### **IV. Mustaqil ta'lim\***

#### **IV.1. Mustaqil ta'lim va mustaqil ishlar**

Talabalarni mustaqil ta'lim shaklini tashkil etishga qo'yilgan talablar O'zbekiston Respublikasi oliy ta'lim, fan va innovatsiyalar vazirligining 2024-yil 29-apreldagi 136-sonli "Oliy ta'lim muassasalari talabalari mustaqil ta'limini tashkil etish bo'yicha namunaviy tartibni tasdiqlash to'g'risida"gi buyrug'i asosida ishlab chiqilgan.

Mustaqil ta'limni baholash semestr davomida berilgan topshiriq asosida bajarilgan ishlarni HEMISda ilova qilish, shuningdek, oraliq va yakuniy test va savollarga javob berish asosida oshiriladi.

1. Obyektlar o'xshashligini baholash. Cosinus masofasi.
2. Sifatli berilganlar bazasiga qo'yiladigan talablar.
3. Inforamtiv alomatlarni ajratish algoritmlari.
4. Mashinali o'rgatish va uning turlari.
5. Keras kutubxonasi va uning imkoniyatlari.
6. Tensorflow kutubxonasi va uning imkoniyatlari.
7. Scikit-learn kutubxonasi va uning imkoniyatlari.
8. Google Colab muhiti va uning imkoniyatlari.
9. Alomatlar fazosining o'lchamini kamaytirish usullari.
10. Chuqur o'rgatish. Asosiy tamoyillari.
11. Neyron to'ri va uning modellari.
12. Egiluvchan neyron to'rlar.
13. Sun'iy neyron to'rlarining LSTM modeli.

	<p>Talabalarni mustaqil ta'lim shaklini tashkil etishga qo'yilgan talablar O'zbekiston Respublikasi oliy ta'lim, fan va innovatsiyalar vazirligining 2024-yil 29-apreldagi 136-sonli "Oliy ta'lim muassasalari talabalari mustaqil ta'limini tashkil etish bo'yicha namunaviy tartibni tasdiqlash to'g'risida"gi buyrug'i asosida ishlab chiqilgan.</p> <p>Mustaqil ta'limni baholash semestr davomida berilgan topshiriq asosida bajarilgan ishlarni HEMISda ilova qilish, shuningdek, oraliq va yakuniy test va savollarga javob berish asosida oshiriladi.</p> <p><i>* Izoh: Mustaqil ta'lim ishi mavzulari fan doirasida o'zgarishi mumkin.</i></p>
3.	<p><b>V. Fan o'qitilishining natijalari (shakllanadigan kompetensiyalar)</b></p> <p>Fanni o'zlashtirish natijasida talaba:</p> <ul style="list-style-type: none"> <li>• magistrlar ma'lumotlar bazasi va data mining obyekti, predmeti, vazifalari, berilganlarning intellektual tahliliga doir tadqiqotlar yuzasidan chuqur <i>bilimlarga ega bo'lishi</i>, berilganlarning intellektual tahlili ahamiyati borasida <i>tasavvurga ega bo'lishi</i>;</li> <li>• berilganlarni yig'ish va ularga dastlabki ishlov berish texnologiyalari, regerssiya va obrazlarni anglashning asosiy masalasini qo'yish va yechish usullaridan foydalanish, berilganlar bazasidan yashirin qonuniyatlarni topish algoritmlarini qo'llash <i>ko'nikmalariga ega bo'lishi</i>;</li> <li>• berilganlarning intellektual tahlil qilish usullarini qo'llash, obrazlarni anglash muammolari bo'yicha yechimlar qabul qilish <i>malakalariga ega bo'lishi kerak.</i></li> </ul>
4.	<p><b>VI. Ta'lim texnologiyalari va metodlari:</b></p> <ul style="list-style-type: none"> <li>• ma'ruzalar;</li> <li>• interfaol keys-stadilar;</li> <li>• seminarlar (mantiqiy fiklash, tezkor savol-javoblar);</li> <li>• guruhlarda ishlash;</li> <li>• taqdimotlar (PPT)ni qilish;</li> <li>• individual loyihalar;</li> <li>• jamoa bo'lib ishlash va himoya qilish uchun loyihalar.</li> </ul>
5.	<p><b>VII. Kreditlarni olish uchun talablar:</b></p> <p>Fanga oid nazariy va uslubiy tushunchalarni to'liq o'zlashtirish, tahlil natijalarini to'g'ri aks ettira olish, o'rganilayotgan jarayonlar haqida mustaqil mushohada yuritish va joriy, oraliq nazorat shakllarida berilgan vazifa va topshiriqlarni bajarish, yakuniy nazorat bo'yicha yozma ishni topshirishi lozim.</p>
6.	<p><b>VIII. ASOSIY VA QO'SHIMCHA O'QUV ADABIYOTLAR HAMDA AXBOROT MANBALARI</b></p> <p><b>Asosiy adabiyotlar</b></p> <p>1. Iawei H., Jian P., Hanghang T. Data Mining: Concepts and Techniques 4th Edition, Morgan Kaufmann, 2023, 786 p. ISBN 978-0128117606</p>

2. Васильев Юлий Python для data science. — СПб.: Питер, 2023. — 272 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-2392-6

#### Qo‘shimcha adabiyotlar

3. Потапов А.С. Технологии искусственного интеллекта - СПб: СПбГУ ИТМО, 2010.-218 с.
4. Морозова В. А., Паутов В. И. Представление знаний в экспертных системах: учебное пособие - Екатеринбург : Изд-во Урал. ун-та, 2017. - 120 с.
5. Бурков А. Машинное обучение без лишних слов: Изд-тво Питер, 2020 г. - ISBN: 978-5-4461-1560-0. -192 с.
6. Günther S., Gunther R., Jan Philipp Prote a, Frederick S, Julia H., Florian O., Dino K. Data Mining Definitions and Applications for the Management of Production Complexity.// ScienceDirect Available online at [www.sciencedirect.com](http://www.sciencedirect.com) [www.elsevier.com/locate/procedia](http://www.elsevier.com/locate/procedia) Procedia CIRP 81 (2019) 874–879. <https://doi.org/10.1016/j.procir.2019.03.217>
7. Carlo G.H., Jordan C.L. Cybersecurity Scientometric Analysis: Mapping of Scientific Articles using Scopus API for Data Mining and Webscraping//See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/361602937>
8. Представление знаний и операции над ними : учебное пособие / В. М. Куклии. – Харьков : ХНУ имени В. Н. Каразина, 2019. – 180 с.

#### Scopusdan maqolalar

1. Retrieval-style In-context Learning for Few-shot Hierarchical Text Classification Open Access Huiyao Chen, Yu Zhao, Zulong Chen, Mengjia Wang, Liangyue Li, Meishan ZhangCorresponding Author, Min Zhang Transactions of the Association for Computational Linguistics (2024) 12: 1214–1231. [https://doi.org/10.1162/tacl\\_a\\_00697](https://doi.org/10.1162/tacl_a_00697) [https://direct.mit.edu/tacl/article/doi/10.1162/tacl\\_a\\_00697/124630/Retrieval-style-In-context-Learning-for-Few-shot](https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00697/124630/Retrieval-style-In-context-Learning-for-Few-shot)
2. Unifying Structured Data as Graph for Data-to-Text Pre-Training Open Access Shujie Li, Liang Li, Ruiying Geng, Min YangCorresponding Author, Binhua Li, Guanghu Yuan, Wanwei He, Shao Yuan, Can Ma, Fei Huang, Yongbin LiCorresponding Author Transactions of the Association for Computational Linguistics (2024) 12: 210–228. [https://doi.org/10.1162/tacl\\_a\\_00641](https://doi.org/10.1162/tacl_a_00641) [https://direct.mit.edu/tacl/article/doi/10.1162/tacl\\_a\\_00641/119991/Unifying-Structured-Data-as-Graph-for-Data-to-Text](https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00641/119991/Unifying-Structured-Data-as-Graph-for-Data-to-Text)

<b>Axborot manbalari</b>	
	<p>9. <a href="http://lib.nuu.uz/">http://lib.nuu.uz/</a> – Ўзбекистон Миллий университети электрон кутубхонаси</p> <p>10. <a href="http://www.raai.org/library">http://www.raai.org/library</a> – Российская ассоциация искусственного интеллекта</p> <p>11. <a href="http://www.intuit.ru">http://www.intuit.ru</a> – Национальиом Открытом Университет (Россия)</p>
7.	<p>Fanning o'quv dasturi Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabida ishlab chiqilgan va 202 <u>5</u> yil "<u>18</u>" <u>06</u> dagi <u>24</u>-sonli bayonnomasi bilan ma'qullangan.</p> <p>O'quv dasturi Toshkent davlat sharqshunoslik universiteti Kengashining 202 <u>5</u> yil "<u>28</u>" <u>06</u> <u>11</u> sonli bayoni bilan tasdiqlangan.</p>
8.	<p>S.T.Mustafayeva - Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi f. f. d., dotsent</p> <p>R.B.Allanyazov- Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi o'qituvchisi</p>
9.	<p><b>Taqrizchilar:</b></p> <p>N. Abduraxmonova- Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi professori</p> <p>Madraximov Sh.F. – O'zMU, "Sun'iy intellekt" kafedrası professori, f.-m.f.d.</p>

11/13