IQRO JURNALI , IQRO JOURNAL, ИКРО ЖУРНАЛИ
ISSN:2181-4341| wordlyknowledge.uz/index.php/iqro
Volume 22 Issue 01 IYUN 2026

JADVAL MA'LUMOTLARIDAN MA'LUMOT OLISHDA MASHINANI O'RGANISH USULLARIDAN FOYDALANISH

Raxmonov Bexzodjon Ravshanjon o'g'li
University of Business and Science 70610101
Kompyuter tizimlari va ularning dasturiy ta'minoti
mutaxassisligi magistranti

Annotatsiya: Ushbu maqolada jadval ko'rinishidagi soliq ma'lumotlaridan foydali bilimlarni ajratib olishda mashinani o'rganish usullaridan foydalanish masalasi yoritilgan. Tadqiqotda soliq faoliyatiga oid ma'lumotlarni qayta ishlash, tozalash, transformatsiya qilish, dastlabki tahlil qilish va mashinani o'rganish algoritmlari asosida soliq xavfini baholash jarayonlari ko'rib chiqilgan. Soliq tizimida shakllanadigan katta hajmdagi jadval ma'lumotlari asosida soliq to'lovchilarni xavf darajasiga ko'ra guruhlash, qarzdorlik ehtimolini aniqlash, soliq tushumlarini prognozlash va shubhali holatlarni belgilash imkoniyatlari tahlil qilingan. Tadqiqotda sintetik dataset asosida Logistic Regression, Decision Tree, Random Forest va XGBoost algoritmlarining samaradorligi taqqoslangan. Natijalar XGBoost va Random Forest algoritmlari murakkab, nochiziqli bog'liqliklarni aniqlashda yuqori aniqlikka ega ekanligini ko'rsatdi. Maqolada shuningdek, ishlab chiqilgan "Soliq Xavfi Tahlil Tizimi" dasturiy ta'minotining arxitekturasi, funksional imkoniyatlari va amaliy ahamiyati bayon etilgan.

Kalit so'zlar: jadval ma'lumotlari, mashinani o'rganish, soliq xavfi, XGBoost, Random Forest, sun'iy intellekt, dataset, prognozlash, klassifikatsiya, ma'lumotlarni tahlil qilish.

Аннотация: В данной статье рассматривается вопрос применения методов машинного обучения для извлечения полезной информации из табличных налоговых данных. В исследовании проанализированы процессы обработки, очистки, трансформации, предварительного анализа данных, относящихся к налоговой деятельности, а также оценивания налогового риска на основе алгоритмов машинного обучения. Рассмотрены возможности группировки налогоплательщиков по уровню риска, определения вероятности налоговой задолженности, прогнозирования налоговых поступлений и выявления подозрительных случаев на основе больших объемов табличных данных, формируемых в налоговой системе. В исследовании на основе синтетического набора данных проведено сравнение эффективности алгоритмов Logistic Regression, Decision Tree, Random Forest и XGBoost. Полученные результаты показали, что алгоритмы XGBoost и Random Forest обладают высокой точностью при выявлении сложных и нелинейных зависимостей. В статье также описаны архитектура, функциональные возможности и практическое значение разработанного программного обеспечения «Система анализа налогового риска».

Ключевые слова: табличные данные, машинное обучение, налоговый риск, XGBoost, Random Forest, искусственный интеллект, набор данных, прогнозирование, классификация, анализ данных.

Abstract: This article examines the use of machine learning methods for extracting useful knowledge from tabular tax data. The study analyzes the processes of processing, cleaning, transforming, and exploratory analysis of tax-related data, as well as assessing tax risk based on machine learning algorithms. The paper discusses the possibilities of grouping taxpayers by risk level, identifying the probability of tax debt, forecasting tax revenues, and detecting suspicious cases based on large volumes of tabular data generated in the tax system. The effectiveness of Logistic Regression, Decision Tree, Random Forest, and XGBoost algorithms was compared

using a synthetic dataset. The results showed that XGBoost and Random Forest algorithms provide high accuracy in identifying complex and nonlinear relationships. The article also describes the architecture, functional capabilities, and practical significance of the developed “Tax Risk Analysis System” software.

Keywords: tabular data, machine learning, tax risk, XGBoost, Random Forest, artificial intelligence, dataset, forecasting, classification, data analysis.

Kirish

Bugungi kunda raqamli iqtisodiyotning rivojlanishi, davlat boshqaruvi tizimlarining raqamlashtirilishi hamda katta hajmdagi ma’lumotlar bilan ishlash zarurati zamonaviy axborot texnologiyalarining ahamiyatini yanada oshirmoqda. Xususan, soliq tizimida elektron hisobotlar, bank operatsiyalari, fiskal cheklar, davlat registrlari va korxonalar faoliyatiga oid boshqa manbalardan har kuni katta hajmdagi ma’lumotlar shakllanadi. Ushbu ma’lumotlarning asosiy qismi jadval koʻrinishida boʻlib, ular soliq toʻlovchilarning moliyaviy holati, toʻlov intizomi, qarzdorlik darajasi va boshqa muhim koʻrsatkichlarni oʻzida aks ettiradi.

An’anaviy statistik usullar katta hajmdagi jadval ma’lumotlarini toʻliq, tezkor va chuqur tahlil qilishda har doim ham yetarli natija bermaydi. Chunki soliq ma’lumotlari koʻp omilli, murakkab va ayrim hollarda yashirin bogʻliqliklarga ega boʻladi. Shuning uchun bunday ma’lumotlarni avtomatik tahlil qilish, ulardan foydali bilimlarni ajratib olish va qaror qabul qilish jarayonini qoʻllab-quvvatlashda mashinani oʻrganish algoritmlaridan foydalanish dolzarb hisoblanadi.

Mashinani oʻrganish algoritmlari katta hajmdagi ma’lumotlar orasidagi qonuniyatlarni aniqlash, obyektlarni sinflarga ajratish, kelajakdagi holatlarni prognozlash va xavf omillarini baholash imkonini beradi. Soliq faoliyatida bu texnologiyalar soliq tushumlarini prognozlash, soliq qarzdorligini aniqlash, shubhali moliyaviy operatsiyalarni belgilash va soliq xavfini avtomatik baholashda muhim ahamiyat kasb etadi.

Ushbu maqolaning maqsadi jadval koʻrinishidagi soliq ma’lumotlarini mashinani oʻrganish usullari asosida tahlil qilish, soliq xavfini baholashga yoʻnaltirilgan model yaratish va ishlab chiqilgan dasturiy ta’minotning amaliy imkoniyatlarini yoritishdan iborat.

Tadqiqotning dolzarbligi. Soliq tizimi har qanday davlat iqtisodiy barqarorligining muhim tarkibiy qismi hisoblanadi. Davlat budjeti daromadlarining katta qismi soliq tushumlari hisobiga shakllanadi. Shu sababli soliq tizimida mavjud ma’lumotlarni sifatli tahlil qilish, soliq xavflarini oldindan aniqlash va nazorat jarayonlarini avtomatlashtirish katta amaliy ahamiyatga ega.

Soliq organlarida shakllanadigan jadval ma’lumotlarini qoʻlda yoki oddiy avtomatlashtirilgan tizimlar orqali qayta ishlash koʻp vaqt va resurs talab qiladi. Bundan tashqari, inson omili sababli ayrim xavfli holatlar vaqtida aniqlanmasligi mumkin. Mashinani oʻrganish algoritmlaridan foydalanish esa katta hajmdagi ma’lumotlarni qisqa vaqt ichida tahlil qilish, yashirin bogʻliqliklarni aniqlash va soliq toʻlovchilarni xavf darajasiga koʻra guruhlash imkonini beradi.

Tadqiqotning dolzarbligi quyidagi omillar bilan belgilanadi:

birinchidan, soliq organlarida katta hajmdagi jadval ma’lumotlari shakllanmoqda;

ikkinchidan, bu ma’lumotlar orasida soliq xavfi, qarzdorlik, hisobot kechikishi va toʻlov intizomiga oid muhim belgilar mavjud;

uchinchidan, an’anaviy tahlil usullari murakkab va nochiziqli bogʻliqliklarni aniqlashda cheklangan imkoniyatga ega;

toʻrtinchidan, mashinani oʻrganish algoritmlari orqali soliq xavfini oldindan baholash va qaror qabul qilish jarayonini avtomatlashtirish mumkin.

Tadqiqot maqsadi va vazifalari. Tadqiqotning asosiy maqsadi soliq faoliyatiga oid jadval ma’lumotlarini mashinani o‘rganish usullari asosida tahlil qilish, soliq xavfini baholashga mo‘ljallangan samarali model ishlab chiqish va uning amaliy qo‘llanilishini asoslashdan iborat.

Ushbu maqsadga erishish uchun quyidagi vazifalar belgilandi:

jadval ma’lumotlarining tuzilishi, xususiyatlari va tahlil qilish imkoniyatlarini o‘rganish;

mashinani o‘rganish algoritmlarining asosiy turlarini tahlil qilish;

soliq faoliyatida ma’lumotlar tahlilining ahamiyatini asoslash;

soliq ma’lumotlari asosida dataset shakllantirish;

ma’lumotlarni tozalash, kodlash, standartlashtirish va tahlilga tayyorlash;

klassifikatsiya va regressiya algoritmlarini tanlash hamda ularning samaradorligini baholash;

soliq xavfini baholovchi model ishlab chiqish;

ishlab chiqilgan model asosida dasturiy ta’minot yaratish;

tajriba natijalarini tahlil qilish va amaliy tavsiyalar ishlab chiqish.

Jadval ma’lumotlari va ularning soliq tizimidagi ahamiyati

Jadval ma’lumotlari strukturallashgan ma’lumotlar turiga kiradi. Ular qatorlar va ustunlardan tashkil topadi. Har bir ustun ma’lum bir belgini, har bir qator esa alohida obyekt yoki yozuvni ifodalaydi. Soliq tizimida bunday obyekt sifatida yuridik shaxs, yakka tartibdagi tadbirkor yoki boshqa soliq to‘lovchi qaralishi mumkin.

Soliq ma’lumotlari jadvalida STIR, korxona nomi, viloyat, faoliyat sohasi, yillik daromad, hisoblangan soliq, to‘langan soliq, qarzdorlik, hisobot kechikishi, tekshiruvlar soni va soliq xavfi kabi ustunlar bo‘lishi mumkin. Bu ustunlarning har biri model uchun muhim xususiyat hisoblanadi.

Jadval ma’lumotlari bilan ishlashda ma’lumotlar sifati hal qiluvchi ahamiyatga ega. Agar ma’lumotlarda yetishmayotgan qiymatlar, takroriy yozuvlar, noto‘g‘ri formatlar yoki shovqinli ma’lumotlar mavjud bo‘lsa, mashinani o‘rganish modeli noto‘g‘ri natija berishi mumkin. Shu sababli tadqiqotda ma’lumotlarni dastlabki qayta ishlash bosqichiga alohida e’tibor qaratildi.

Ma’lumotlarni qayta ishlash jarayoni odatda quyidagi bosqichlardan iborat bo‘ladi: ma’lumotlarni yig‘ish, saqlash, tozalash, transformatsiya qilish, tahlil qilish va natijalarni vizualizatsiya qilish. Aynan ushbu bosqichlar mashinani o‘rganish modelining yakuniy samaradorligini belgilaydi.

Mashinani o‘rganish usullarining nazariy asoslari. Mashinani o‘rganish sun’iy intellektning muhim yo‘nalishlaridan biri bo‘lib, kompyuter tizimlariga ma’lumotlar asosida o‘rganish va qaror qabul qilish imkonini beradi. An’anaviy dasturlashda barcha qoidalar dasturchi tomonidan oldindan belgilanadi. Mashinani o‘rganishda esa algoritm ma’lumotlarni tahlil qilish orqali qonuniyatlarni mustaqil aniqlaydi.

Mashinani o‘rganishning asosiy turlari nazoratli o‘qitish, nazoratsiz o‘qitish va mustahkamlovchi o‘qitishdan iborat. Soliq xavfini baholash masalasida asosan nazoratli o‘qitish usullaridan foydalaniladi. Chunki bunda kiruvchi ma’lumotlar bilan birga ularning natijaviy sinflari ham mavjud bo‘ladi. Masalan, soliq to‘lovchining moliyaviy va ma’muriy ko‘rsatkichlari asosida uning xavf darajasi past, o‘rta yoki yuqori sinflardan biriga ajratiladi.

Nazoratli o‘qitish klassifikatsiya va regressiya masalalariga bo‘linadi. Klassifikatsiya masalasida obyektlar oldindan belgilangan sinflarga ajratiladi. Regressiya masalasida esa uzluksiz sonli qiymatlar prognoz qilinadi. Soliq faoliyatida klassifikatsiya soliq xavfini baholashda, regressiya esa soliq tushumlarini prognozlashda qo‘llaniladi.

Tadqiqotda Logistic Regression, Decision Tree, Random Forest va XGBoost algoritmlari tahlil qilindi. Logistic Regression oddiy va tushunarli klassifikatsiya algoritmi hisoblanadi. Decision Tree qaror daraxti asosida ishlaydi va ma’lumotlarni shartlar bo‘yicha tarmoqlarga ajratadi. Random Forest bir nechta qaror daraxtlari kombinatsiyasidan iborat bo‘lib, aniqlik va

barqarorlikni oshiradi. XGBoost esa gradient boosting yondashuviga asoslangan bo‘lib, murakkab va katta hajmdagi ma’lumotlarda yuqori natija ko‘rsatadi.

Datasetni shakllantirish va ma’lumotlarni tayyorlash. Tadqiqotda soliq tizimining real ko‘rsatkichlariga yaqin sintetik dataset yaratildi. Haqiqiy soliq ma’lumotlari maxfiyligi sababli, sintetik ma’lumotlardan foydalanish ilmiy va amaliy jihatdan maqbul yondashuv hisoblanadi. Dataset O‘zbekiston soliq tizimiga xos STIR formati, hududlar, faoliyat sohalari, soliq stavkalari va xavf omillarini hisobga olgan holda shakllantirildi.

Dataset 12 000 ta yozuv va 15 ta xususiyatdan iborat bo‘lib, har bir yozuv bitta yuridik shaxs — soliq to‘lovchi korxonani ifodalaydi. Xususiyatlar tarkibiga identifikatsion belgilar, moliyaviy ko‘rsatkichlar va xavfga ta’sir qiluvchi omillar kiritildi. Maqsad o‘zgaruvchi sifatida “soliq_xavfi” ustuni belgilandi va u uch sinfga ajratildi: Past, O‘rta va Yuqori.

Ma’lumotlarni tayyorlash jarayonida yetishmayotgan qiymatlar aniqlanib, mediana imputation usuli yordamida to‘ldirildi. Dublikatlar olib tashlandi, outlier qiymatlar IQR metodi yordamida aniqlanib bartaraf etildi. Kategorik ustunlar modelga mos bo‘lishi uchun kodlandi, sonli qiymatlar esa standartlashtirildi. Bunday tayyorgarlik jarayoni modelning aniq, barqaror va ishonchli ishlashiga xizmat qiladi.

Exploratory Data Analysis natijalari. Ma’lumotlarni dastlabki tahlil qilish bosqichida dataset tarkibi, xususiyatlar taqsimoti va maqsad o‘zgaruvchi bilan bog‘liqlik darajalari o‘rganildi. Tahlil natijalariga ko‘ra, yillik daromad log-normal taqsimlanishga ega ekani aniqlandi. Bu holat iqtisodiy ma’lumotlar uchun tabiiy hisoblanadi, chunki korxonalar daromadi odatda bir xil taqsimlanmaydi.

Korrelyatsion tahlil natijalariga ko‘ra, soliq xavfi bilan eng kuchli bog‘liq omillar sifatida to‘lov farqi, hisobot kechikishi va tekshiruvlar soni ajralib chiqdi. To‘lov farqi hisoblangan va haqiqatda to‘langan soliq summasi o‘rtasidagi tafovutni bildiradi. Ushbu tafovut qancha katta bo‘lsa, soliq xavfi shuncha yuqori bo‘lishi mumkin. Hisobot kechikishi esa soliq to‘lovchining ma’muriy intizomini ifodalaydi. Tekshiruvlar soni avvalgi nazorat holatlari bilan bog‘liq xavf belgisi sifatida qaraladi.

Ushbu tahlillar model uchun muhim xususiyatlarni tanlashda asos bo‘lib xizmat qildi. Natijada model soliq xavfini baholashda eng ta’sirchan omillarni hisobga oladigan tarzda o‘qitildi.

Algoritmlarni tanlash va modellarni o‘qitish. Tadqiqotda bir nechta mashinani o‘rganish algoritmlari o‘zaro taqqoslandi. Dataset o‘quv va test qismlariga ajratildi. O‘quv qismida model ma’lumotlardagi qonuniyatlarni o‘rgandi, test qismida esa uning yangi ma’lumotlarda qanday ishlashi baholandi.

Logistic Regression oddiy model sifatida bazaviy natijani aniqlash uchun qo‘llanildi. Decision Tree modeli esa qaror qabul qilish jarayonining tushunarli bo‘lishi uchun tanlandi. Random Forest va XGBoost algoritmlari murakkab bog‘liqliklarni aniqlash, yuqori aniqlik va barqaror natija olish maqsadida qo‘llanildi.

Algoritmlarni taqqoslash natijalariga ko‘ra, XGBoost modeli eng yuqori samaradorlikni ko‘rsatdi. Random Forest algoritmi ham yuqori aniqlik va barqarorlikka ega bo‘ldi. Logistic Regression va Decision Tree algoritmlari esa sodda bo‘lgani sababli murakkab bog‘liqliklarni aniqlashda nisbatan pastroq natija berdi.

XGBoost modelining ustunligi shundaki, u xatoliklarni bosqichma-bosqich kamaytirish orqali o‘rganadi, murakkab nochiziqli bog‘lanishlarni aniqlaydi va katta hajmdagi ma’lumotlarda yuqori aniqlik beradi. Random Forest esa bir nechta qaror daraxtlari natijalarini birlashtirib, modelning haddan tashqari moslashib ketish xavfini kamaytiradi.

Model samaradorligini baholash. Klassifikatsiya modellari Accuracy, Precision, Recall, F1-score va ROC-AUC ko‘rsatkichlari asosida baholandi. Accuracy umumiy to‘g‘ri bashoratlar ulushini bildiradi. Precision model tomonidan “yuqori xavf” deb belgilangan obyektlarning

qanchasi haqiqatda yuqori xavfli ekanini ko‘rsatadi. Recall haqiqiy yuqori xavfli obyektlarning qanchasi model tomonidan aniqlanganini bildiradi. F1-score esa Precision va Recall o‘rtasidagi muvozanatni ifodalaydi. ROC-AUC modelning sinflarni ajratish qobiliyatini baholashda qo‘llaniladi.

Natijalarga ko‘ra, Random Forest modeli yuqori aniqlik ko‘rsatdi, XGBoost modeli esa undan ham yuqori natija qayd etdi. XGBoost modelining Accuracy va ROC-AUC ko‘rsatkichlari soliq xavfini baholashda ushbu algoritmning samarali ekanligini tasdiqladi. Shuningdek, Soft Voting ensambli yordamida Random Forest va XGBoost modellari birlashtirildi va umumiy samaradorlik yanada oshirildi.

Regressiya modeli bo‘yicha soliq tushumlarini prognozlashda RMSE va MAPE ko‘rsatkichlari qo‘llanildi. MAPE ko‘rsatkichi past bo‘lgani modelning soliq tushumlarini prognozlashda ishonchli natija berishini ko‘rsatadi. Bu esa soliq organlari uchun budjet tushumlarini rejalashtirishda muhim amaliy ahamiyatga ega.

Xususiyatlar muhimligi tahlili. Model natijalarini tahlil qilishda xususiyatlar muhimligi alohida o‘rganildi. Bu tahlil qaysi omillar soliq xavfiga ko‘proq ta’sir qilishini aniqlash imkonini beradi. Natijalarga ko‘ra, to‘lov farqi, hisobot kechikishi va yillik daromad eng muhim omillar sifatida ajralib chiqdi.

To‘lov farqi soliq xavfini baholashda asosiy indikatorlardan biridir. Agar hisoblangan soliq summasi va haqiqatda to‘langan soliq o‘rtasida katta tafovut bo‘lsa, bu soliq xavfi yuqori bo‘lishi mumkinligini anglatadi. Hisobot kechikishi soliq to‘lovchining intizomini ko‘rsatadi. Yillik daromad esa korxona faoliyati hajmini bildiradi va risklarni baholashda muhim omil hisoblanadi.

SHAP texnologiyasidan foydalanish model qarorlarini izohlash imkonini berdi. Bu ayniqsa soliq tizimi uchun muhimdir, chunki modelning nima uchun aynan shunday qaror chiqargani tushunarli bo‘lishi kerak. SHAP tahlili har bir omilning yakuniy bashoratga ta’sirini ko‘rsatadi va auditorlar hamda soliq xodimlari uchun qarorlarni asoslash imkonini beradi.

Dasturiy ta’minotni ishlab chiqish. Tadqiqotning amaliy natijasi sifatida “Soliq Xavfi Tahlil Tizimi” dasturiy ta’minoti ishlab chiqildi. Modelning o‘zi ilmiy jihatdan muhim bo‘lsa-da, uni amaliyotda qo‘llash uchun qulay, tez va ishonchli dasturiy tizimga aylantirish zarur. Shu sababli ishlab chiqilgan model dasturiy muhitga integratsiya qilindi.

Dasturiy ta’minot uch qatlamli arxitektura asosida qurildi. Birinchi qatlam — ma’lumotlar qatlami bo‘lib, unda soliq ma’lumotlari saqlanadi va qayta ishlashga tayyorlanadi. Ikkinchi qatlam — mashinani o‘rganish qatlami bo‘lib, unda model bashorat qiladi. Uchinchi qatlam — foydalanuvchi interfeysi bo‘lib, foydalanuvchi ma’lumot kiritadi va natijalarni ko‘radi.

Tizimda PostgreSQL va Pandas ma’lumotlarni saqlash hamda qayta ishlash uchun, Python, Scikit-learn va XGBoost model yaratish va bashorat qilish uchun, FastAPI va Uvicorn REST API yaratish uchun, Joblib o‘qitilgan modelni saqlash uchun, Streamlit interaktiv dashboard yaratish uchun, SHAP esa model qarorlarini izohlash uchun qo‘llanildi.

Foydalanuvchi interfeysida soliq to‘lovchining STIR raqami, viloyati, faoliyat turi, yillik daromadi, hisoblangan soliq summasi, to‘langan soliq summasi, qarzdorligi, hisobot kechikishi va tekshiruvlar soni kiritiladi. Tizim ushbu ma’lumotlarni qayta ishlaydi va soliq to‘lovchining xavf darajasini past, o‘rta yoki yuqori sinflardan biriga ajratadi. Natijalar grafiklar, diagrammalar va ehtimollik ko‘rsatkichlari orqali foydalanuvchiga taqdim etiladi.

Amaliy ahamiyati. Ishlab chiqilgan tizim soliq organlari faoliyatida bir nechta muhim vazifani bajarishi mumkin. Birinchidan, soliq to‘lovchilarni xavf darajasiga ko‘ra avtomatik tasniflaydi. Ikkinchidan, qarzdorlik va to‘lov intizomi bilan bog‘liq muammolarni oldindan aniqlashga yordam beradi. Uchinchidan, soliq tushumlarini prognozlash orqali davlat budjeti rejalashtirish jarayonlarini qo‘llab-quvvatlaydi. To‘rtinchidan, yuqori xavfli subyektlarni aniqlab, soliq nazorati resurslarini samarali taqsimlash imkonini beradi.

Bunday tizim inson omilini kamaytiradi, nazorat jarayonlarini tezlashtiradi va qaror qabul qilishni ma’lumotlarga asoslangan holda tashkil etishga yordam beradi. Ayniqsa, katta hajmdagi soliq ma’lumotlari bilan ishlashda avtomatlashtirilgan tahlil vositalaridan foydalanish soliq organlari samaradorligini oshiradi.

Natijalar muhokamasi. Tadqiqot natijalari mashinani o‘rganish algoritmlaridan soliq ma’lumotlarini tahlil qilishda samarali foydalanish mumkinligini ko‘rsatdi. XGBoost va Random Forest algoritmlari soliq xavfini baholashda yuqori aniqlik qayd etdi. Bu algoritmlar murakkab va nochiziqli bog‘liqliklarni aniqlay olgani sababli an’anaviy modellarga nisbatan ustunlikka ega bo‘ldi.

Dataset tahlili soliq xavfiga eng kuchli ta’sir qiluvchi omillar to‘lov farqi, hisobot kechikishi va tekshiruvlar soni ekanini ko‘rsatdi. Bu natijalar soliq nazorati amaliyoti uchun ham muhim hisoblanadi. Chunki soliq organlari yuqori xavfli subyektlarni aniqlashda aynan shu omillarga alohida e’tibor qaratishi mumkin.

Dasturiy ta’minot natijalarni foydalanuvchiga tushunarli ko‘rinishda taqdim etgani sababli, undan amaliy faoliyatda foydalanish imkoniyati yuqori. Dashboard orqali soliq xavfi taqsimoti, yuqori xavfli subyektlar ro‘yxati, hududlar bo‘yicha xavf darajasi va model izohlari ko‘rsatiladi. Bu esa tahliliy qarorlar qabul qilishni soddalashtiradi.

Xulosa

Jadval ko‘rinishidagi soliq ma’lumotlarini tahlil qilishda mashinani o‘rganish usullaridan foydalanish zamonaviy raqamli boshqaruv uchun muhim ahamiyatga ega. Tadqiqotda soliq faoliyatiga oid jadval ma’lumotlarini yig‘ish, tozalash, transformatsiya qilish, tahlil qilish va modelga tayyorlash jarayonlari ko‘rib chiqildi. Sintetik dataset asosida bir nechta mashinani o‘rganish algoritmlari sinovdan o‘tkazildi va ularning samaradorligi baholandi.

Natijalar XGBoost va Random Forest algoritmlari soliq xavfini baholashda yuqori samaradorlikka ega ekanini ko‘rsatdi. Xususiyatlar muhimligi tahlili to‘lov farqi, hisobot kechikishi va yillik daromad soliq xavfiga sezilarli ta’sir ko‘rsatishini aniqladi. SHAP texnologiyasi yordamida model qarorlarini izohlash imkoniyati yaratildi.

Tadqiqotning amaliy natijasi sifatida “Soliq Xavfi Tahlil Tizimi” dasturiy ta’minoti ishlab chiqildi. Ushbu tizim soliq ma’lumotlarini yuklash, qayta ishlash, xavf darajasini aniqlash, natijalarni vizualizatsiya qilish va hisobot shaklida taqdim etish imkonini beradi. Mazkur tizim soliq organlarida monitoring, prognozlash, risklarni baholash va nazorat jarayonlarini avtomatlashtirishda foydalanilishi mumkin.

Umuman olganda, mashinani o‘rganish algoritmlaridan foydalanish soliq tizimida raqamli transformatsiyani jadallashtirish, ma’lumotlarga asoslangan qaror qabul qilishni kuchaytirish va soliq nazorati samaradorligini oshirishga xizmat qiladi.

Foydalanilgan adabiyotlar

1. O‘zbekiston Respublikasi Prezidentining 2020-yil 5-oktabrdagi PF–6079-son Farmoni “Raqamli O‘zbekiston – 2030” strategiyasini tasdiqlash to‘g‘risida.
2. O‘zbekiston Respublikasi Prezidentining 2021-yil 17-fevraldagi PQ–4996-son qarori “Sun’iy intellekt texnologiyalarini yaratish va ularni keng joriy etish chora-tadbirlari to‘g‘risida”.
3. O‘zbekiston Respublikasining Soliq kodeksi. 2019-yil 30-dekabrdagi O‘RQ–599-son Qonuni.
4. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2022.
5. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O’Reilly Media, 2022.
6. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2009.

7. Mitchell T. Machine Learning. McGraw-Hill, 1997.
8. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference, 2016.
9. Breiman L. Random Forests. Machine Learning, 2001.
10. Lundberg S.M., Lee S.I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 2017.
11. OECD. Tax Administration 2023: Comparative Information on OECD and other Advanced and Emerging Economies. OECD Publishing, 2023.
12. Jordon J., Yoon J., van der Schaar M. PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees. ICLR, 2019.
13. Pedregosa F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2011.
14. McKinney W. Python for Data Analysis. O'Reilly Media, 2018.
15. Fowler M. Patterns of Enterprise Application Architecture. Addison-Wesley, 2002.