

«М.А. Гендельманнның 110 жылдығына арналған «Сейфуллин оқулары – 19» халықаралық ғылыми-практикалық конференциясының материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 19», посвященной 110 - летию М.А. Гендельмана» - 2023.- Т. II, Ч.1.- Б. 339-340.

**ОӘЖ 004.042**

## **PYOD КІТАПХАНАСЫН ҚОЛДАНА ОТЫРЫП, ДЕРЕКТЕРДЕГІ АУЫТҚУЛАРДЫ АНЫҚТАУ СЕРВИСІН ӘЗІРЛЕУ ШАРТТАРЫ**

*Такеров Ш.О., 4 курс студенті  
«С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті», КеАҚ Астана қ.*

*Кіріспе.* Кәсіпорындар жасайтын деректер көлемінің артуы деректердегі ауытқулар санының артуына әкелді. Деректердегі ауытқуларды анықтау деректер сапасын қамтамасыз ету, бизнес шешімдерін қабылдау және алаяқтықтың алдын алу үшін өте маңызды. Бұл зерттеу жұмысы Python тіліндегі PyOD (Python Outlier Detection) кітапханасын, көп айнымалы деректердегі шектен тыс мәндерді анықтауға арналған Python жан-жақты кітапханасын пайдаланып деректердегі ауытқуларды анықтау қызметін құруға бағытталған.

PyOD кітапханасы ауытқуларды анықтауға арналған алгоритмдердің кең ауқымын қамтамасыз етеді, соның ішінде кластерлеуге негізделген, қашықтыққа негізделген, тығыздыққа негізделген және ықтималдық модельдер [1]. Кітапхананы орнату оңай және оны Scikit-learn, NumPy және Pandas сияқты басқа деректерді талдау және машиналық оқыту кітапханаларымен бірге пайдалануға болады.

*Деректердегі ауытқуларды анықтау қызметін құру.* Деректердегі ауытқуларды анықтауға арналған ұсынылатын қызмет Python тіліндегі PyOD кітапханасы арқылы жасалады. Бұл қызмет әртүрлі өлшемдер мен типтегі деректер жиынын өңдеуге арналған және пайдаланушыларға деректермен өзара әрекеттесу үшін ыңғайлы интерфейсті қамтамасыз етеді. Бұл қызмет пайдаланушыларға өз деректерін жүктеп салуға және ауытқуды анықтауға мүмкіндік беретін веб-интерфейс арқылы қол жетімді болады. Қызметті құру кезінде келесі қадамдар орындалады:

*1-қадам: Деректерді тазалау және дайындау.*

Ауытқуды анықтау алдында деректер жинағы тазартылады және дайындалады. Деректерді тазалау жетіспейтін мәндерді, көшірмелерді және маңызды емес деректер нүктелерін жоюды қамтиды. Деректерді дайындау категориялық деректерді сандық деректерге түрлендіру сияқты ауытқуларды анықтау үшін деректерді сәйкес пішімге түрлендіруді қамтиды.

*2-қадам: Ауытқуды анықтау алгоритмін таңдау.*

Келесі қадам ауытқуды анықтау үшін сәйкес алгоритмді таңдау болып табылады. Таңдалған алгоритм деректер түріне және нақты пайдалану жағдайына байланысты болады.

*3-қадам: Үлгіні оқыту.*

Алгоритм таңдалғаннан кейін модель деректер жиынында оқытылады. Оқу барысында модель күтілетін нормадан айтарлықтай ауытқыған деректер нүктелерін анықтауды үйренеді.

*4-қадам: Ауытқуды анықтау.*

Модель оқытылғаннан кейін ауытқуларды анықтау үшін деректер жиынына қолданылады. PyOD кітапханасы болжау, шешім функциясы және ауытқу ұпайларын қоса, ауытқуды анықтауға арналған функциялар ауқымын қамтамасыз етеді. Бұл функциялар ауытқу дәрежесін көрсететін әрбір деректер нүктесі үшін ұпай немесе белгіні қайтарады.

*5-қадам: Ауытқулар туралы хабарлау.*

Соңғы қадам - анықталған ауытқулар туралы хабарлау. Ауытқулар туралы веб-негізделген интерфейс, электрондық пошта немесе басқа хабарландыру әдістері арқылы хабарлауға болады. Есеп деректер нүктесі, ұпай немесе белгі және кез келген басқа сәйкес ақпарат сияқты ауытқу туралы мәліметтерді қамтиды.

PyOD кітапханасын пайдаланып деректердегі ауытқуларды анықтау қызметін жасау бизнеске деректерді басқару стратегияларын жақсарту үшін пайдалы құралды ұсына алады. Қызмет өте ауқымды және теңшелетін болады, бұл бизнеске ауытқуларды анықтау процесін олардың нақты қажеттіліктеріне бейімдеуге мүмкіндік береді. Деректердегі ауытқуларды анықтау арқылы бизнес шешімдер қабылдауды жақсартып алады, деректер сапасын қамтамасыз етеді және алаяқтықтың алдын алады. Дегенмен, деректер жиынтығын мұқият бағалау және ең жақсы тәсілді анықтау үшін әртүрлі алгоритмдермен тәжірибе жасау маңызды [2].

Қорытындылай келе, PyOD кітапханасының көмегімен деректердегі ауытқуларды анықтау қызметін құру бизнес үшін өз деректерін тиімді басқарудың тиімді шешімі болып табылады. PyOD кітапханасының кең ауқымды алгоритмдерін пайдалана отырып, бизнес шешімдер қабылдауды жақсарту, деректер сапасын қамтамасыз ету және алаяқтықтың алдын алу үшін ауытқуларды анықтаудың тиімді жүйелерін жасай алады [3]. Дегенмен, нақты деректер жиынтығы үшін ең жақсы тәсілді анықтау үшін мұқият бағалау және әртүрлі алгоритмдермен тәжірибе қажет

### **Пайдаланылған әдебиеттер тізімі**

1. Kannan R. et al. Outlier Detection for Text Data: An Extended Version [Text]. // arXiv e-prints. – 2017. – С. arXiv: 1701.01325

2. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL. Feasibility of Obtaining Measures of Lifestyle from a Smartphone App: The My Heart Counts Cardiovascular Health Study [Text]. // JAMA Cardiol. – 2017. Vol. 2(1). – P.67–76

3. Zhao, Y., Nasrullah, Z., Li, J., & Liu, Y. PyOD: A python toolbox for scalable outlier detection [Text]. // Journal of Machine Learning Research. – 2019. Vol. 20(96). P. 1-7

*Ғылыми жетекшісі, PhD, қауымдастырылған профессор Исмаилова А.А.*