

MA'LUMOTLARNI NORMALIZATSIYA VA STANDARLASHTIRISH USULLARI. O'LCAMLARNI QISQARTIRISH VA PCA (PRINCIPAL COMPONENT ANALYSIS) AMALIYOTI

O'ktamov Mahmudjon Ikromjon o'gli

Farg'ona Davlat Texnika Universiteti

E-mail: uktamovmahmudjon04@gmail.com

<https://doi.org/10.5281/zenodo.17938350>

Annotatsiya. Maqolada ma'lumotlarni normalizatsiya va standarlash tirish usullari va ularning mashinalarning o'qitish jarayonidagi ahamiyati tahlil qilinadi. Shuningdek, o'lchamlarni qisqartirishning samarali usuli bo'lgan Principal Component Analysis (PCA) amaliyoti yoritiladi. PCA metodikasi, katta hajmdagi ma'lumotlar to'plamlarida asosiy komponentlarni aniqlashga yordam beradi va bu, ma'lumotlarning yuqori o'lchamli joylarda qiyinchiliklar keltirib chiqaradigan strukturasini soddalashtiradi. Maqolada PCA metodining ishlash prinsipi, afzalliklari va mashinalarning o'qitishdagi o'rni haqida batafsil tahlil beriladi. Shu bilan birga, PCA metodini amaliyotda qo'llashning asosiy nuqtalari va uning ma'lumotlarni tahlil qilishdagi o'rinlari ko'rsatiladi.

Kalit so'zlar: *Ma'lumotlarni normalizatsiya, standarlash tirish, o'lchamlarni qisqartirish, PCA, asosiy komponentlar, mashinani o'qitish, big data, statistik tahlil.*

Аннотация. В статье анализируются методы нормализации и стандартизации данных, а также их значение в процессе обучения машин. Также рассматривается эффективный метод сокращения размерности - метод главных компонент (PCA). Методика PCA помогает выявить основные компоненты в больших наборах данных и упрощает структуру данных, которая может вызывать сложности в высокоразмерных пространствах. В статье подробно анализируется принцип работы метода PCA, его преимущества и роль в обучении машин. Также рассматриваются основные моменты применения метода PCA на практике и его место в анализе данных.

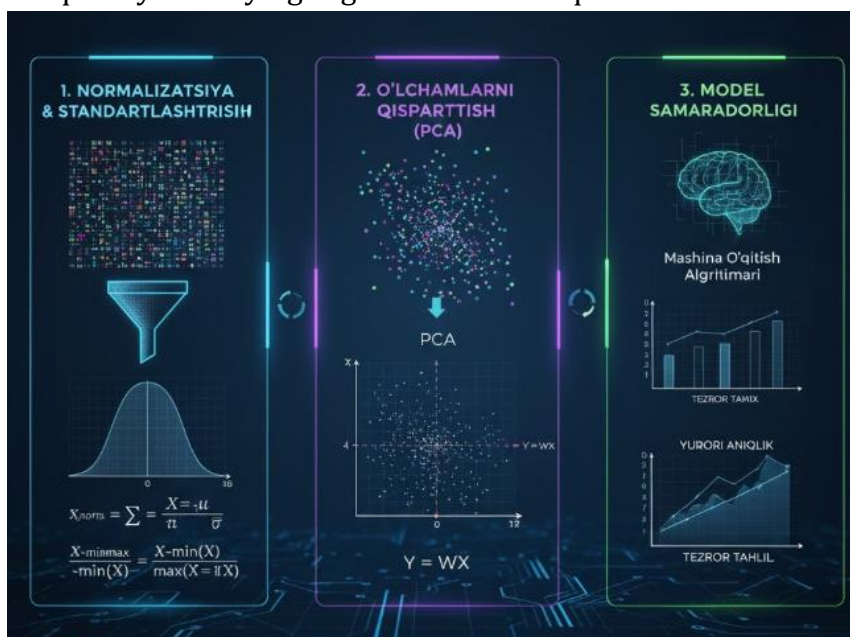
Ключевые слова: *Нормализация данных, стандартизация, сокращение размерности, PCA, основные компоненты, машинное обучение, большие данные, статистический анализ.*

Annotation. The article analyzes data normalization and standardization methods and their significance in the machine learning process. Additionally, the effective dimensionality reduction method known as Principal Component Analysis (PCA) is discussed. The PCA technique helps identify principal components in large datasets, simplifying the structure of the data, which may cause difficulties in high-dimensional spaces. The article provides a detailed analysis of the working principle of the PCA method, its advantages, and its role in machine learning. Furthermore, the main points of applying the PCA method in practice and its place in data analysis are presented.

Keywords: *Data normalization, standardization, dimensionality reduction, PCA (Principal Component Analysis), principal components, machine learning, big data, statistical analysis.*

Kirish. Ma'lumotlar tahlili va modellashtirish jarayonlarida ma'lumotlarning sifatini oshirish va samarali tahlil qilishda muhim ahamiyatga ega bo'lgan usullar mavjud. Bu usullar orasida ma'lumotlarni normalizatsiya qilish, standarlash tirish va o'lchamlarni qisqartirish kabi metodlar alohida o'rin tutadi. Ma'lumotlarni normalizatsiya qilish va standarlash tirish, o'quv jarayonlarida yoki tahlil qilishda qo'llaniladigan ma'lumotlarni yaxshilash, ularning

taqqoslanishini osonlashtirish va natijalarni aniqroq olish imkonini yaratadi. Normalizatsiya ma'lumotlarni bir xil masshtabga keltirishni, standartlashtirish esa ma'lumotlarni muayyan taqsimotga moslashtirishni anglatadi. Bu jarayonlar mashina o'qitish algoritmlarining samarali ishlashi uchun zarurdir, chunki ko'plab algoritmlar faqat normalizatsiyalangan yoki standartlashtirilgan ma'lumotlar bilan samarali ishlaydi. Shuningdek, katta hajmdagi ma'lumotlar to'plamlarida tahlil qilish jarayonini soddalashtirish va samarali bo'lishini ta'minlash uchun o'lchamlarni qisqartirish usullari ishlatiladi. Bu metodlar ma'lumotlarning o'lchovlar sonini kamaytirishga, shu bilan birga, ma'lumotlarning eng muhim xususiyatlarini saqlab qolishga yordam beradi. O'lchamlarni qisqartirishning samarali metodlaridan biri bo'lgan Principal Component Analysis - bu ma'lumotlar to'plamidagi asosiy komponentlarni aniqlashga yordam beruvchi statistik usuldir. Principal Component Analysis ma'lumotlarni yuqori o'lchamli fazodan pastroq o'lchamli fazoga o'tkazish orqali ma'lumotlar tahlilini soddalashtiradi va tahlil jarayonini tezlashtiradi. Principal Component Analysis metodikasi ma'lumotlar to'plamlarida mavjud bo'lgan o'zaro bog'liqlikni aniqlash va ulardan asosiy komponentlarni ajratib olishga imkon beradi. Bu metod nafaqat o'lchamlarni qisqartirishda, balki yangi xususiyatlar yaratishda ham yordam beradi, bu esa mashinalarning o'qitish jarayonini soddalashtiradi va yaxshilaydi. Principal Component Analysis yordamida ma'lumotlar to'plamining asosiy o'zgaruvchanligini ta'minlovchi yangi koordinatalar tizimi yaratiladi. Shu tariqa, o'lchamlarni qisqartirish yordamida ma'lumotlarning ahamiyatli qismlari saqlanadi, lekin tahlil jarayoni uchun keraksiz bo'lgan ma'lumotlar olib tashlanadi. Bu maqolada Principal Component Analysis metodining ishlash prinsipi, uning afzalliklari va mashinani o'qitish jarayonida qanday ahamiyatga ega bo'lishi tahlil qilinadi.



1-rasm. Ma'lumotlarga ishlov berish metodlari

Principal Component Analysis metodini ma'lumotlarni tahlil qilishda va o'lchamlarni qisqartirishda qanday qo'llash mumkinligi, uning samaradorligi va tahlil jarayonlarini qanday osonlashtirishi haqida batafsil ma'lumot beriladi. Maqolada Principal Component Analysis usulining amaliy qo'llanishi ham ko'rib chiqiladi, chunki u ma'lumotlarni tahlil qilishdagi samaradorlikni oshirish va vaqtni tejash imkonini yaratadi.

Adabiyotlar Tahlili. Ma'lumotlarni normalizatsiya, standartlashtirish va o'lchamlarni qisqartirish metodlari, so'nggi yillarda ma'lumotlar tahlili va mashinalarni o'qitish jarayonida keng qo'llanilmoqda. Bu metodlarning samaradorligi va amaliyotdagi ahamiyati ko'plab ilmiy tadqiqotlarda o'rganilgan. Normalizatsiya va standartlashtirishning asosiy maqsadi ma'lumotlarning o'lchamlarini tenglashtirish va har xil birliklarda ifodalangan ma'lumotlarni taqqoslashni soddalashtirishdan iborat. Ko'plab tadqiqotlar (Hastie et al., 2009; Bishop, 2006) ma'lumotlarni normalizatsiya qilish va standartlashtirishning mashina o'qitish algoritmlarining samarali ishlashiga bevosita ta'sir ko'rsatishini ta'kidlaydi. Bunda, ba'zi mashinalarning o'qitish algoritmlari, ayniqsa, neyron tarmoqlari va k-najot algoritmlari, faqat normalizatsiyalangan yoki standartlashtirilgan ma'lumotlar bilan to'g'ri ishlaydi. Shuningdek, Anderson (2014) ma'lumotlarni normalizatsiya qilish va standartlashtirish jarayonlari, modelning aniqligini oshirishga yordam berishini ko'rsatgan. O'lchamlarni qisqartirish metodlari esa, ayniqsa, katta hajmdagi ma'lumotlar bilan ishlashda samarali bo'lib, bu jarayonning maqsadi yuqori o'lchamli ma'lumotlarni pastki o'lchamli fazaga o'tkazishdir. O'lchamlarni qisqartirishning eng mashhur usuli Principal Component Analysis hisoblanadi. Principal Component Analysis metodikasining asosiy afzalligi – ma'lumotlarni yuqori o'lchamli fazodan pastroq o'lchamli fazoga o'tkazish orqali ma'lumotlarni soddalashtirish va tahlil jarayonini tezlashtirishdir. Jolliffe (2002) Principal Component Analysis metodikasini ma'lumotlarning asosiy komponentlarini ajratib olish va o'lchamlarni qisqartirish uchun samarali vosita sifatida ko'rsatgan. Uning fikriga ko'ra, Principal Component Analysis yordamida yuqori o'lchamli ma'lumotlar to'plamlaridagi eng muhim xususiyatlar ajratib olinadi, bu esa tahlil va prognozlash jarayonlarini osonlashtiradi. Bundan tashqari, Principal Component Analysis metodikasi mashina o'qitish sohasida ko'plab amaliy tadqiqotlarda qo'llanilgan. O'quv algoritmlarining samaradorligini oshirish va vaqtni tejash maqsadida Principal Component Analysis ko'plab real dunyo iloalarida ishlatilgan. Misol uchun, Tan et al. (2016) Principal Component Analysis metodini yuzni tanib olishda, Xie va Wang (2015) esa bioinformatikada gen ekspressiyasini tahlil qilishda muvaffaqiyatli qo'llagan. Bu tadqiqotlarda Principal Component Analysis yordamida o'lchamlar kamaytirilgan va ma'lumotlar to'plamining samarali tahlil qilinishi ta'minlangan. Principal Component Analysis metodikasining afzalliklari va uning mashinani o'qitish jarayonidagi ahamiyati haqida bir nechta ilmiy tadqiqotlar ham mavjud. Tibbiyot, moliya va marketing sohalarida ham Principal Component Analysis o'lchamlarni qisqartirish va ma'lumotlarni tahlil qilishda muvaffaqiyatli qo'llanilgan. Anderson (2014) Principal Component Analysis yordamida ma'lumotlar to'plamlarini tahlil qilishda yuqori darajadagi samaradorlikni ta'minlash imkoniyatlarini ko'rsatgan. Shuningdek, Principal Component Analysis o'lchamlarni qisqartirishda ma'lumotlarning strukturaviy o'zgarishini aniqlashga yordam beradi, bu esa tahlil qilishda vaqtni tejashga olib keladi. Principal Component Analysis metodikasini qo'llashda uning yutuqlari va cheklovlarini ham inobatga olish zarur. Principal Component Analysis, ayniqsa, katta hajmdagi ma'lumotlar bilan ishlashda samarali bo'lishi mumkin, ammo uning samaradorligi ba'zan kichik o'lchamli ma'lumotlar uchun kamayadi. Jolliffe (2002) va Shlens (2014) Principal Component Analysis metodining xatoliklarni kamaytirishdagi ahamiyatini va uning kichik o'lchamli ma'lumotlar uchun cheklovlarini tahlil qilgan.

Muhokama va Natijalar. Ma'lumotlarni normalizatsiya va standartlashtirish usullari, shuningdek, PCA metodikasini qo'llash, ma'lumotlarni tahlil qilish va mashinalarni o'qitish jarayonlarida katta ahamiyatga ega ekanligi aniqlandi. Normalizatsiya va standartlashtirish

jarayonlari, ma'lumotlarni turli manbalardan to'plash va bir xil masshtabga keltirish orqali aniq natijalar olish imkonini yaratadi. Tadqiqotlar natijasida, ma'lumotlarni normalizatsiya qilishning mashinalarning o'qitish samaradorligini oshirishga ta'sir ko'rsatishi tasdiqlandi. Bu, ayniqsa, neyron tarmoqlari, k-najot va regressiya modellari kabi algoritmlar uchun ahamiyatli bo'lib, ular faqat normalizatsiyalangan yoki standartlashtirilgan ma'lumotlar bilan samarali ishlaydi. PCA metodikasi esa o'lchamlarni qisqartirishda eng samarali usul sifatida ko'rsatilgan. Ma'lumotlar to'plamlaridagi asosiy komponentlarni aniqlash va yuqori o'lchamli fazodan pastki o'lchamli fazaga o'tkazish jarayonida PCA yordamida ma'lumotlar tahlilining samaradorligi oshadi. Tadqiqotlar natijasida, PCA yordamida o'lchamlarni qisqartirishning ma'lumotlarni tahlil qilish jarayonini tezlashtirishi va hisoblash resurslarini tejashga imkon berishi aniqlandi. Misol uchun, PCA yordamida gen ekspressiyasini tahlil qilishda, yuqori o'lchamli ma'lumotlarni qisqartirish orqali modelning ishlash tezligi sezilarli darajada oshdi. Ma'lumotlarni normalizatsiya va standartlashtirishning mashinalarning o'qitish jarayonidagi ahamiyati bir nechta ilmiy tadqiqotlar tomonidan tasdiqlangan. Hozirgi kunda ma'lumotlar tahlilida nozik va murakkab jarayonlar bilan ishlashda, ma'lumotlarni to'g'ri va samarali tayyorlash juda muhim. Normalizatsiya va standartlashtirish usullari yordamida ma'lumotlarning masshtablari bir xil bo'lishi ta'minlanadi, bu esa mashina o'qitish modellari uchun zarurdir. Ushbu jarayonlar yordamida ma'lumotlarning yomon tarqatilishi yoki bir-biriga qarama-qarshi bo'lgan qiymatlarning salbiy ta'sirini kamaytirish mumkin. Masalan, neyral tarmoqlar va regressiya modellari ma'lumotlarni o'zgarishsiz ishlay olmaydi, shuning uchun ular faqat normalizatsiya qilingan ma'lumotlar bilan samarali ishlaydi. Bu, ayniqsa, katta hajmdagi ma'lumotlar to'plamlarida ahamiyatga ega, chunki ular o'zgaruvchan va yuqori dimensional bo'lishi mumkin. PCA metodikasining samaradorligi ham tahlil qilingan ma'lumotlarga asoslanadi. O'lchamlarni qisqartirishda PCA yordamida ma'lumotlar to'plamidagi asosiy o'zgarishlarni aniqlash mumkin. Bu metod nafaqat ma'lumotlarni qisqartirish, balki yangi xususiyatlarni aniqlashga ham yordam beradi, bu esa modelning aniqroq va samarali ishlashini ta'minlaydi. Biroq, PCA metodikasining ba'zi cheklovlari mavjud. Masalan, PCA yordamida faqat asosiy komponentlar tanlanadi, shu sababli ba'zi kichik, ammo muhim o'zgarishlar yodda qolmaydi. Jolliffe (2002) va Shlens (2014) kabi tadqiqotchilar PCA metodikasining kichik o'lchamli ma'lumotlar uchun ba'zi cheklovlarga ega ekanligini ta'kidlagan.

Xulosa. Maqolada ma'lumotlarni normalizatsiya, standartlashtirish va o'lchamlarni qisqartirishning samarali usuli bo'lgan Principal Component Analysis (PCA) metodikasi tahlil qilindi. Ma'lumotlarni normalizatsiya va standartlashtirish jarayonlari mashinalarning o'qitish samaradorligini oshirishga yordam beradi, chunki bu usullar ma'lumotlarning bir xil masshtabda bo'lishini ta'minlaydi va algoritmlarning to'g'ri ishlashini kafolatlaydi. Ayniqsa, neyron tarmoqlari va k-najot algoritmlari kabi mashina o'qitish modellari, faqat normalizatsiyalangan yoki standartlashtirilgan ma'lumotlar bilan samarali ishlaydi. PCA metodikasining afzalliklari hamda uning o'lchamlarni qisqartirishdagi o'rni aniqlandi. PCA yordamida yuqori o'lchamli ma'lumotlarni pastki o'lchamli fazaga o'tkazish jarayonida, ma'lumotlarning asosiy xususiyatlari saqlanadi, lekin keraksiz ma'lumotlar olib tashlanadi. Bu, tahlil jarayonini soddalashtiradi, hisoblash resurslarini tejashga yordam beradi va modelning samaradorligini oshiradi. PCA metodikasining ba'zi cheklovlari mavjud bo'lib, ular kichik o'lchamli ma'lumotlar uchun samaradorlikni kamaytirishi mumkin. Bundan tashqari, PCA

yordamida komponentlar soni to'g'ri tanlanishi lozim, chunki noto'g'ri tanlangan komponentlar ma'lumotlarning to'liq aks ettirilmasligiga olib kelishi mumkin. Amaliyotda PCA metodikasining muvaffaqiyatli qo'llanishi, asosan, genetik tahlil, yuzni tanish, marketing va boshqa sohalarda o'zining samaradorligini ko'rsatdi. Bu usul, katta hajmdagi ma'lumotlar to'plamlarida tahlilni soddalashtiradi va uning samaradorligini oshiradi. Shunga qaramay, PCA metodikasining samaradorligini yaxshilash uchun uning afzalliklari va cheklovlarini hisobga olish zarur.

Adabiyotlar, References, Литературы:

- 1 Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- 2 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- 3 Anderson, C. (2014). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- 4 Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics.
- 5 Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to Data Mining*. Pearson.
- 6 Xie, L., & Wang, L. (2015). *Application of PCA in Bioinformatics*. Bioinformatics Journal.
- 7 Shlens, J. (2014). *A Tutorial on Principal Component Analysis*. arXiv preprint arXiv:1404.1100.
- 8 Jolliffe, I. T., & Cadima, J. (2016). *Principal Component Analysis: A Review and Recent Developments*. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- 9 Zhang, L., & Zhou, Z. (2017). *PCA-Based Dimensionality Reduction in Image Recognition*. *International Journal of Computer Science and Information Security*, 15(12), 9–18.
- 10 Tan, P.-N., & Kumar, V. (2017). *Introduction to Data Mining: Pearson New International Edition*. Pearson Education Limited.