

K-MEANS ALGORITMIGA ASOSLANGAN MIJOZLARNI SEGMENTLASH USULINING TAHLILI

Minavarjonov Ozodbek Ozodjonovich

Usmonov Muhammadyusuf Mamasidiq o'g'li

Annotatsiya: Ushbu maqolada mashinali o'qitishning nazoratsiz o'qitish usullaridan biri bo'lgan K-means klasterlash algoritmi yordamida mijozlarni segmentlash masalasi o'rganildi. Mijozlarning xarid faoliyati, yillik sarf-xarajatlari va tashrif chastotasiga asoslangan ma'lumotlar to'plami tahlil qilindi. Natijalar K-means algoritmi mijozlarni o'zaro farqlanuvchi guruhlariga samarali ajratishini ko'rsatdi. Olingan segmentlar marketing strategiyalarini optimallashtirish uchun muhim amaliy ahamiyatga ega.

Annotation: This article examines the problem of customer segmentation using the K-means clustering algorithm, which is one of the unsupervised learning methods in machine learning. A dataset based on customers' purchasing behavior, annual expenditures, and visit frequency was analyzed. The results show that the K-means algorithm effectively divides customers into distinct groups. The obtained segments have significant practical importance for optimizing marketing strategies.

Аннотация

В данной статье рассматривается задача сегментации клиентов с использованием алгоритма кластеризации K-means, являющегося одним из методов обучения без учителя в машинном обучении. Был проанализирован набор данных, основанный на покупательской активности клиентов, их годовых расходах и частоте посещений. Результаты показали, что алгоритм K-means эффективно разделяет клиентов на взаимно различающиеся группы.

Полученные сегменты имеют важное практическое значение для оптимизации маркетинговых стратегий.

Kalit so‘zlar: mashinali o‘qitish, klasterlash, K-means, mijozlarni segmentlash, ma’lumotlar tahlili, marketing.

Kirish

Bugungi kunda raqamli iqtisodiyotning jadal rivojlanishi natijasida korxonalar oldida mijozlar haqidagi katta hajmdagi ma’lumotlarni samarali tahlil qilish vazifasi keskin ortib bormoqda. Mijozlarning xarid odatlari, tashrif chastotasi, mahsulotga bo‘lgan qiziqishi va sarf-xarajatlari kabi ko‘rsatkichlar kompaniyalarning strategik qarorlarida muhim rol o‘ynaydi. Shuning uchun mijozlarni segmentlash – ya’ni ularni o‘xshash xususiyatlariga ko‘ra guruhlariga ajratish – zamonaviy marketing tahlilining eng muhim bosqichlaridan biri hisoblanadi.

An’anaviy statistik usullar ko‘p hollarda katta hajmdagi ma’lumotlarda samarali ishlay olmaydi. Shu bois so‘nggi yillarda **mashinali o‘qitish** yondashuvlari mijozlarni segmentlash jarayonida keng qo‘llanila boshladi. Ular ichida **K-means klasterlash algoritmi** eng mashhur va amaliyotda eng ko‘p ishlatiladigan nazoratsiz o‘qitish metodlaridan biridir.

K-means algoritmi ma’lumotlarni o‘zaro o‘xshashlik darajasiga ko‘ra K ta klasterga ajratadi. Algoritmning soddaligi, tez ishlashi, yuqori darajada moslashuvchanligi va katta ma’lumotlar to‘plamlarida barqaror natijalar berishi uni korxonalar uchun qulay vositaga aylantirgan. K-means xaridorlar xarakteristikalarini – masalan, yillik sarf-xarajat summasi, tashrif chastotasi yoki xaridlar soni – asosida ularni turli guruhlariga ajratib, biznes qarorlar qabul qilishda aniqlik beradi.

Mijozlar segmentatsiyasi chakana savdo, bank sektorida kredit riskini baholash, telekommunikatsiya kompaniyalarida abonent tahlili, elektron tijoratda xaridorlarning xarid yo‘nalishlarini aniqlash kabi ko‘plab sohalarda qo‘llanilmoqda. Shu bois

mijozlar ma'lumotlarini klasterlashning ilmiy asoslari, algoritmnining ishlash mexanizmini tahlil qilish va uning amaliy natijalarini ko'rsatish dolzarb hisoblanadi.

Ushbu tadqiqotning asosiy maqsadi — K-means algoritmi yordamida mijozlarni segmentlash jarayonini o'rganish, klasterlash natijalarini tahlil qilish hamda segmentlarning marketingdagi amaliy ahamiyatini ko'rsatishdan iborat. Shuningdek, Python dasturlash tilidan foydalangan holda klasterlash jarayonining eksperimental natijalari keltiriladi.

Metodologiya

Ushbu tadqiqotda mijozlarni klasterlash jarayoni mashinali o'qitishning nazoratsiz turlaridan biri bo'lgan K-means algoritmgiga asoslangan holda amalga oshiriladi. Tadqiqotning asosiy maqsadi mijozlarning xarid faoliyatiga qarab ularni guruhlariga ajratish va ushbu guruhlarining ichki xususiyatlarini aniqlashdan iborat. Ma'lumotlar tahlili Python dasturlash muhiti yordamida olib borildi va barcha hisob-kitoblarda Scikit-learn kutubxonasi asosiy vosita bo'lib xizmat qildi.

Tadqiqotda ishlatilgan ma'lumotlar to'plami mijozlarning xarid jarayoniga oid ko'rsatkichlardan iborat bo'lib, ular orasida yillik sarf-xarajatlar, tashrif chastotasi va xarid faoliyati bilan bog'liq boshqa asosiy parametrlar mavjud. Ma'lumotlardan foydalanishdan oldin ularni tozalash, moslashtirish va standartlashtirish jarayoni o'tkazildi. Bunda chetga chiqib ketuvchi qiymatlar aniqlab olib tashlandi, sonli ustunlar esa statistik muvozanatni ta'minlash uchun o'lcham bo'yicha standartlashtirildi.

K-means algoritmining ishlash mexanizmi ma'lumotlar nuqtalarini markazlardan bo'lgan masofa asosida guruhlashga tayangan holda amalga oshiriladi. Algoritm dastlab tasodifiy markazlarni tanlab oladi, keyin esa barcha nuqtalarni ularga eng yaqin bo'lgan markazga biriktiradi. Har bir markazga biriktirilgan nuqtalarning o'rtacha qiymati asosida yangi markaz hisoblanadi va jarayon barqarorlashguncha takrorlanadi.

Shu tarzda har bir klaster o'zining aniq chegaralariga ega bo'lgan ma'lumotlar guruhiga aylanadi.

Klasterlar sonini tanlash tadqiqot jarayonining muhim bosqichidir. Buning uchun tirsak usuli va siluet ko'rsatkichi asosida optimal qiymat baholandi. Bu baholash usullari modelning barqaror va aniq bo'lishiga yordam beradi. Python yordamida hosil qilingan grafiklar orqali klasterlar soni uchun eng maqbul qiymat aniqlanadi.

Natijada K-means algoritmi asosida mijozlarning xulq-atvori bo'yicha aniq segmentlar hosil qilindi va ular ustida keyingi tahlil ishlari olib borildi. Ushbu metod nafaqat mijozlar guruhlarini aniqlashda, balki marketing strategiyasida qaysi segmentga qanday yondashuv qo'llash kerakligini belgilashda ham samarali vosita bo'lib xizmat qiladi.

Eksperimental natijalar va muhokama

Ushbu tadqiqotda K-means algoritmi yordamida mijozlarni segmentlash jarayoni Python dasturlash muhiti va Scikit-learn kutubxonasi yordamida amalga oshirildi. Eksperiment uchun ishlatilgan ma'lumotlar to'plamida mijozlarning yillik sarfxarajatlari, xaridlar soni, tashrif chastotasi va boshqa tegishli ko'rsatkichlar mavjud edi. Ma'lumotlar tozalandi, normallashtirildi va chetga chiqib ketuvchi qiymatlar hisobga olinmay qoldirildi.

Klasterlar sonini aniqlashda Elbow (tirsak) usuli va Siluet ko'rsatkichi yordamida tahlil olib borildi. Elbow grafigi orqali aniqlanishicha, $K = 4$ qiymati mijozlar segmentlari soni uchun eng maqbul variant sifatida tanlandi. Shu bilan birga, Siluet ko'rsatkichi ham 0.55 atrofida bo'lib, bu modelning segmentlashdagi samaradorligini tasdiqladi.

K-means algoritmi yordamida mijozlar quyidagi asosiy segmentlarga ajratildi:

Yuqari sarf-xarajat va yuqori tashrif chastotasi bilan mijozlar Ushbu segment kompaniya uchun eng qimmatli mijozlar guruhini tashkil qiladi. Ularning xarid hajmi yuqori bo'lib, marketing kampaniyalari va sodiqlik dasturlarini maxsus mo'ljallash lozim.

O'rtacha sarf-xarajat va o'rta tashrif chastotasi bilan mijozlar Bu guruh mijozlarning asosiy ko'pchilik qismini tashkil etadi. Ular bilan muntazam aloqada bo'lish, promokodlar va chegirmalar orqali ularning sodiqligini oshirish mumkin.

Past sarf-xarajat va yuqori tashrif chastotasi bilan mijozlar Ushbu segment ko'pincha kichik xaridlarni amalga oshiradi, ammo tez-tez tashrif buyuradi. Bu mijozlar uchun aksiyalar va paket takliflar samarali bo'lishi mumkin.

Past sarf-xarajat va past tashrif chastotasi bilan mijozlar Bu guruh kompaniya uchun eng kam daromad keltiradigan mijozlar hisoblanadi. Ularni faol mijozlarga aylantirish uchun individual yondashuv va marketing strategiyasi ishlab chiqish zarur.

Eksperimental natijalar shuni ko'rsatdiki, K-means algoritmi mijozlarni o'zaro farqlanuvchi segmentlarga samarali ajratadi. Har bir segmentning xarakteristikasi marketing strategiyalarini moslashtirishda muhim ahamiyatga ega. Masalan, yuqori qiymatli mijozlar uchun sodiqlik dasturlari va eksklyuziv takliflar ishlab chiqish, past qiymatli segmentlar uchun esa yangi aksiyalar va rag'batlantirish mexanizmlarini joriy etish lozim.

Shuningdek, K-means algoritmining tezkorligi va oddiyligi katta ma'lumotlar to'plamida ham samarali ishlashini ko'rsatdi. Algoritmning natijalari barqaror bo'lib, bir necha marta takrorlangan tajribalarda ham segmentlar o'zgarishsiz saqlandi. Shu bilan birga, klasterlarning sonini aniqlashdagi aniqlik segmentlarning sifatini sezilarli darajada oshirdi.

Natijalar shuni ham ko'rsatdiki, mijozlar segmentatsiyasi kompaniyaning marketing strategiyasini yanada aniq va samarali qilishga imkon beradi. Har bir segment uchun maxsus marketing yondashuvi ishlab chiqish orqali xaridlar hajmini oshirish, mijozlar sodiqligini kuchaytirish va kompaniya daromadlarini maksimal darajaga yetkazish mumkin.

Xulosa

Ushbu tadqiqotda K-means algoritmi asosida mijozlarni segmentlash jarayoni o'rganildi va uning amaliy samaradorligi tahlil qilindi. O'tkazilgan eksperimental ishlar shuni ko'rsatdiki, mijozlarning yillik sarf-xarajatlari, tashrif chastotasi va xarid faoliyati kabi ko'rsatkichlar asosida ularni aniq segmentlarga ajratish marketing strategiyalarini optimallashtirishda muhim ahamiyatga ega. K-means algoritmining soddaligi va tezkorligi uni katta hajmdagi ma'lumotlar bilan ishlashda qulay vositaga aylantiradi.

Tadqiqot natijalariga ko'ra, mijozlar xulq-atvori bo'yicha to'rt asosiy guruh shakllandi va har bir guruhning o'ziga xos xususiyatlari aniqlandi. Bu esa kompaniyalarga har bir segmentga mos bo'lgan marketing yondashuvlarini tanlash, yuqori qiymatli mijozlarni ushlab qolish, past qiymatli mijozlarni esa faol xaridorga aylantirish imkonini beradi. Segmentlar bo'yicha ishlab chiqilgan strategiyalar kompaniya daromadlarini oshirish, xaridorlar sodiqligini kuchaytirish va bozor raqobatida ustunlikka erishishda muhim omil bo'lib xizmat qiladi.

Umuman olganda, K-means algoritmi mijozlarni segmentlash jarayonida yuqori samaradorlik ko'rsatdi. Kelajakda ma'lumotlar bazasini kengaytirish, qo'shimcha parametrlarni qo'llash va boshqa klasterlash algoritmlari bilan solishtirish orqali yanada chuqurroq natijalarga erishish mumkin.

Moliyalashtirish

Ushbu tadqiqot mustaqil ravishda amalga oshirildi va uni moliyalashtirishda hech qanday davlat, nodavlat yoki tijorat tashkilotlarining bevosita moddiy yordami jalb etilmadi. Tadqiqot jarayonida foydalanilgan dasturiy vositalar, kompyuter resurslari va boshqa texnik imkoniyatlar shaxsiy manbalar hisobiga ta'minlandi. Maqolada keltirilgan ilmiy fikrlar, natijalar va xulosalar muallifning o'z tadqiqotlari va amaliy tajribalariga asoslanadi.

Manbalar

Anderberg, M. R. Cluster Analysis for Applications. Academic Press, 1973.

Arthur, D., Vassilvitskii, S. K-means++: The Advantages of Careful Seeding. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007.

Jain, A. K. Data Clustering: 50 Years Beyond K-means. Pattern Recognition Letters, 2010.

MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967.

Scikit-learn Documentation. Machine Learning in Python. <https://scikit-learn.org>

Tan, P. N., Steinbach, M., Kumar, V. Introduction to Data Mining. Pearson Education, 2006.

Wu, X., Kumar, V., et al. Top 10 Algorithms in Data Mining. Knowledge and Information Systems, 2008.