

DATA SCIENCE UCHUN ZARUR STATISTIK KOMPETENSIYALAR: TAHLILY YONDASHUV

Bobonorova Hulkar Anvarjon qizi

Toshkent davlat iqtisodiyot universiteti

2-bosqich talabasi

Annotatsiya: Ushbu maqolada data science sohasi uchun zarur bo'lgan statistik kompetensiyalar chuqur tahlil qilinadi. Statistik tafakkur, ehtimollik nazariyasi, gipoteza sinovi, regressiya modellari va A/B testlar kabi metodologik yondashuvlarning data science doirasidagi o'rni yoritilgan. Sohaning rivoji uchun takliflar berilgan va statistik kompetensiyalarning tahliliy qaror qabul qilishdagi ahamiyati asoslab berilgan.

Kalit so'zlar: statistik kompetensiyalar, data science, ehtimollik, gipoteza sinovi, regressiya, tahliliy yondashuv, A/B test, statistik tafakkur.

Аннотация: В статье проведён углублённый анализ статистических компетенций, необходимых для специалистов в области data science. Рассматриваются ключевые методологические подходы, включая вероятностное мышление, проверку гипотез, регрессионный анализ и A/B-тестирование. Обосновано значение статистического мышления в аналитическом принятии решений и предложены направления развития данной сферы.

Ключевые слова: статистические компетенции, data science, вероятность, проверка гипотез, регрессия, аналитический подход, A/B-тест, статистическое мышление.

Abstract: This article presents an in-depth analysis of the statistical competencies essential for the field of data science. It explores the role of probabilistic reasoning, hypothesis testing, regression analysis, and A/B testing within the framework of analytical decision-making. The study emphasizes the importance of statistical thinking and proposes strategic directions for developing competencies in the field.

Keywords: statistical competencies, data science, probability, hypothesis testing, regression, analytical approach, A/B testing, statistical thinking.

Kirish

Raqamli transformatsiya jarayonlarining jadallashuvi va katta hajmdagi ma'lumotlar oqimi zamonaviy iqtisodiyotda qaror qabul qilish mexanizmlarini tubdan o'zgartirmoqda. Aynan shunday sharoitda data science – ya'ni ma'lumotlar fanining ahamiyati keskin oshib bormoqda. Ushbu fan tarmog'i nafaqat texnik ko'nikmalarni, balki statistik fikrlash, ehtimollik asosida qaror qabul qilish va model qurish kabi chuqur analitik kompetensiyalarni ham talab qiladi. Ayniqsa, data science yo'nalishida ishlayotgan mutaxassislar uchun statistik bilimlar – bu oddiy vosita emas, balki tahliliy natijalarni ishonchli, aniq va kontekstga mos talqin qilishning asosiy omilidir.

Bugungi kunda statistik kompetensiyalarning data science sohasidagi roli ikki yo'nalishda o'z aksini topmoqda: birinchidan, ular orqali katta hajmdagi ma'lumotlar tuziladi, tozalanadi va strukturaviy ko'rinishga keltiriladi; ikkinchidan esa, statistik usullar yordamida bashorat, sinov, hipotezalarni tekshirish va natijalarni optimallashtirish imkoniyati yaratiladi. Ushbu kompetensiyalar – regressiya tahlili, bayes usullari, statistik taqsimotlar, ehtimollik nazariyasi va parametrik/parametrik bo'lmagan testlar – zamonaviy data analyst va data scientist faoliyatining ajralmas qismini tashkil etadi.

Mavjud holatni tahlil qilish shuni ko'rsatmoqdaki, ko'plab sohalarda, jumladan sog'liqni saqlash, moliya, ta'lim, sanoat va marketingda statistik savodxonlik yetishmasligi natijasida ma'lumotlardan noto'g'ri xulosa chiqarish, noto'g'ri qarorlar qabul qilish holatlari kuzatilmoqda. Shu nuqtayi nazardan, statistik kompetensiyalarning chuqurroq o'rganilishi va ularni amaliy loyihalarda qo'llash bo'yicha metodik yondashuvlarni ishlab chiqish dolzarb masalaga aylanmoqda.

Mazkur maqolada data science sohasi uchun zarur bo'lgan asosiy statistik kompetensiyalar tahlil qilinadi, ularning nazariy asosi va amaliy ahamiyati yoritiladi, shuningdek, turli sohalarda qo'llanish holatlari misolida ularning amaliy samaradorligi baholanadi. Tahliliy yondashuv asosida statistik bilimlar spektri o'rganilib, data science mutaxassisi uchun zarur bo'lgan bilim, ko'nikma va kompetensiyalar majmuasi tizimli tarzda tavsiflanadi.

Mavzuga oid adabiyotlar sharhi

Zamonaviy data science sohasi statistik bilimlar majmuasining chuqur integratsiyasiga asoslanadi. Ushbu holatni ilk bor tizimli tarzda tahlil qilganlardan biri Trevor Hastie, Robert Tibshirani va Jerome Friedman bo'lib, ular o'zlarining mashhur

“The Elements of Statistical Learning” asarida statistik modellashtirish usullarini data mining jarayonlari bilan uyg‘unlashgan holda tushuntirgan. Ular klassifikatsiya, regressiya va klasterlash usullarini zamonaviy statistik kompetensiyaning asosi sifatida belgilagan.

David Hand esa statistik savodxonlik va ma’lumotlar tahlilida ehtimollik nazariyasining o’rni haqida gapirar ekan, data scientistlar uchun statistik tafakkurning muhimligini qayd etadi. Unga ko‘ra, ko‘plab xatolar noto‘g‘ri statistik talqinlardan kelib chiqadi, bu esa kompetensiyaning chuqurligini taqozo etadi. Ayniqsa, ommaviy ma’lumotlar (big data) bilan ishlaganda p-qiyamat, ishonch oralig‘i, hipoteza sinovi kabi tushunchalarning noto‘g‘ri qo‘llanilishi real qarorlar sifatiga salbiy ta’sir qiladi.

Gareth James va hammualliflar tomonidan yozilgan “An Introduction to Statistical Learning” asarida statistik metodlarning data science loyihalaridagi o’rni izchil tahlil qilinadi. Kitobda turli statistik algoritmlarni amaliy Python va R muhitida qo‘llash bo‘yicha yondashuvlar bayon etilgan, bu esa statistik kompetensiyani amaliy ko‘nikmaga aylantirishda muhim manba bo‘lib xizmat qiladi.

Shuningdek, Cathy O’Neil o‘zining “Weapons of Math Destruction” kitobida statistik modellarning noto‘g‘ri tuzilishi yoki biryoqlama yondashuvi natijasida yuzaga kelgan ijtimoiy oqibatlarni tahlil qiladi. U statistik bilim yetarli bo‘lmagan holda yaratilgan algoritmlar diskriminatsiyaga olib kelishini, ayniqsa, ta’lim, kredit baholash va jinoyatchilikni bashoratlash tizimlarida bu muammolar aniq ko‘rinishini ko‘rsatadi. Bu esa statistik kompetensiyalar faqat texnik emas, balki axloqiy va ijtimoiy nuqtayi nazardan ham muhimligini anglatadi.

Data science sohasi uchun statistik kompetensiyalarni rivojlantirishga oid nazariy yondashuvlar Daniel T. Larose tomonidan chuqur yoritilgan. U o‘zining “Discovering Knowledge in Data” nomli kitobida statistik o‘rganish jarayonida model baholash, aniqlik, sezuvchanlik kabi ko‘rsatkichlarning mohiyatini keng tahlil qiladi. Bu ko‘rsatkichlar data science sohasida qarorlar qanchalik ishonchli ekanini belgilovchi vosita sifatida xizmat qiladi.

Shuningdek, xalqaro tashkilotlar tomonidan ishlab chiqilgan hisobotlar ham statistik kompetensiyalar haqida muhim ma’lumotlarni taqdim etadi. Xususan, World Economic Forum tomonidan 2023-yilda e’lon qilingan “Future of Jobs” hisobotida

statistik tafakkur, raqamli savodxonlik va ma'lumotlar asosida tahlil yuritish ko'nikmalari eng talab yuqori kompetensiyalar qatorida qayd etilgan.

Shu asosda umumlashtirish mumkinki, statistik kompetensiyalar nafaqat data science sohasiga kirish uchun zarur boshlang'ich to'siq, balki chuqur tahliliy yondashuvlar, bashorat modellari, qarorlarni asoslash, xatoliklarni kamaytirish va natijalarni ishonchli taqdim etish uchun muhim poydevor hisoblanadi. Ilmiy manbalar ushbu kompetensiyalarni zamonaviy iqtisodiyot, tibbiyot, ta'lim va ijtimoiy sohalardagi raqamli islohotlarning muhim tarkibiy qismi sifatida baholaydi.

Tadqiqot metodologiyasi

Ushbu tadqiqotda statistik kompetensiyalarga oid ma'lumotlar ilmiy maqolalar, darsliklar va xalqaro hisobotlardan kontent-tahlil usuli orqali to'plandi. Tahlil jarayonida mavzuga oid asosiy tushunchalar ajratilib, ularning data science doirasidagi amaliy qo'llanilishi sifat jihatdan tahlil qilindi hamda ekspert yondashuvi asosida baholandi.

Tahlil va natijalar

Zamonaviy iqtisodiyot, raqamli transformatsiya va avtomatlashtirilgan qaror qabul qilish tizimlarining markazida turgan data science sohasi turli ko'nikma va bilimlar majmuasini talab qiladi. Bu soha faqat dasturlash, vizualizatsiya va sun'iy intellektga asoslanmasdan, o'zining ilmiy poydevorini statistik tafakkur va metodologiyadan oladi. Shu nuqtayi nazardan, statistik kompetensiyalar data science mutaxassislarining eng muhim malaka toifasiga aylanmoqda. Bunday kompetensiyalar data science'ni klassik informatika fanlaridan farqlovchi asosiy mezon bo'lib, bu fanni tahliliy ishonchlilik, empirik dalillar bilan ishlash va bashorat qilish salohiyati bilan boyitadi.

Statistik kompetensiyalarni tushunishda birinchi navbatda ularning tuzilmasini ko'rib chiqish lozim. Statistik kompetensiyalar deganda, faqatgina formulalarni yodlash yoki usullarni bilish emas, balki muammoni statistik nuqtayi nazardan ko'ra olish, ma'lumotlardagi yashirin tendensiyalarni aniqlash, noto'g'ri xulosalardan saqlanish, ehtimollik mezonlaridan foydalangan holda qarorlar qabul qilish kabi qator kognitiv, amaliy va tanqidiy ko'nikmalar tushuniladi. Data science kontekstida esa bu kompetensiyalar quyidagicha shakllanadi: ma'lumotlarni to'plash, tozalash (data cleaning), transformatsiya qilish (feature engineering), statistik modellashtirish (regressiya, klassifikatsiya, klasterlash), baholash (model evaluation) va talqin (interpretation).

Birinchi darajali statistik kompetensiyalar orasida ehtimollik nazariyasi asoslari, statistik taqsimotlar va ularning xususiyatlari, namuna olish usullari, parametrik va parametrik bo'lmagan testlar, hipoteza sinovi va ishonch oralig'ini qurish bo'yicha bilimlar ajralib turadi. Ayniqsa, data scientistlar uchun statistik inference, ya'ni tanlanmadan asoslangan holda umumiy populyatsiya haqida xulosa chiqarish zaruriy funksional kompetensiyadir. Chunki real dunyoda mavjud barcha ma'lumotlar bilan ishlash imkoni mavjud emas, qarorlar esa doimo cheklangan va buzilgan (noisy) ma'lumotlar asosida qabul qilinadi.

Masalan, korxona foydalanuvchilari tomonidan bildirilgan fikrlar asosida xizmat sifati baholanayotgan bo'lsa, bu yerda to'liq populyatsiya emas, balki tasodifiy tanlangan fikrlar majmuasi mavjud bo'ladi. Ana shunday vaziyatda statistik metodlar, ayniqsa ishonch oralig'i, hipoteza sinovi va ehtimollik funksiyalari orqali ishonchli xulosa chiqarish mumkin bo'ladi. Bu esa statistik kompetensiyaning faqat texnik emas, balki konseptual ahamiyatga ham ega ekanini ko'rsatadi.

Statistik kompetensiyaning yana bir muhim elementi – deskriptiv statistika va vizualizatsiyadir. Data science jarayonida ko'pincha analitik ishlov berilayotgan ma'lumotlar juda katta hajmda, strukturasiz va murakkab tuzilmali bo'ladi. Bunday sharoitda oddiy o'rtacha qiymat, median, dispersiya, kvartillar, qiyshoqlik va cho'zinchalik kabi statistik o'lchovlar orqali ma'lumotlarga dastlabki "diagnostika" o'tkaziladi. Ushbu kompetensiyalar data analystlar uchun asosiy vosita bo'lib, real biznes qarorlarida foydali insightlar (ichki xulosalar) chiqarish uchun dastlabki bosqichdir.

Shu bilan birga, regressiya tahlili va klassifikatsiya modellarini qurish statistika va machine learning o'rtasidagi murakkab o'zaro aloqani aks ettiradi. Klassik statistik metodlar (masalan, oddiy chiziqli regressiya, logistik regressiya, ANOVA) statistik asoslarda ishlaydi, natijalarni talqin qilish imkonini beradi va ularning xulosalari nazariy asosga ega. Machine learning modellar esa ba'zida interpretatsiyadan ko'ra aniqlikni afzal ko'radi. Biroq, ularning ishlashi ham statistik g'oyalarga — namuna olish, overfitting va bias-variance tradeoff kabi tushunchalarga bog'liq. Statistik kompetensiyasiz bu tushunchalarni tushunish, modelni to'g'ri sozlash (tuning) va baholash amri mahol.

Bugungi kunda data science'da muhim o'rin tutadigan A/B testlar ham statistik kompetensiyalarning muhim namunasidir. Mahsulot dizaynini o'zgartirishdan tortib foydalanuvchi interfeysidagi tugmalar rangini tanlashgacha bo'lgan jarayonlarda

kompaniyalar turli variantlarni sinovdan o'tkazadilar. Ushbu testlarning asosida statistik gipotezalarni sinash, p-value ni baholash, test kuchini hisoblash (power analysis) kabi murakkab kompetensiyalar yotadi. Statistik bilimlarsiz bu testlar noto'g'ri natijalar berishi, shuning uchun kompaniya noto'g'ri strategik qaror qabul qilishi mumkin.

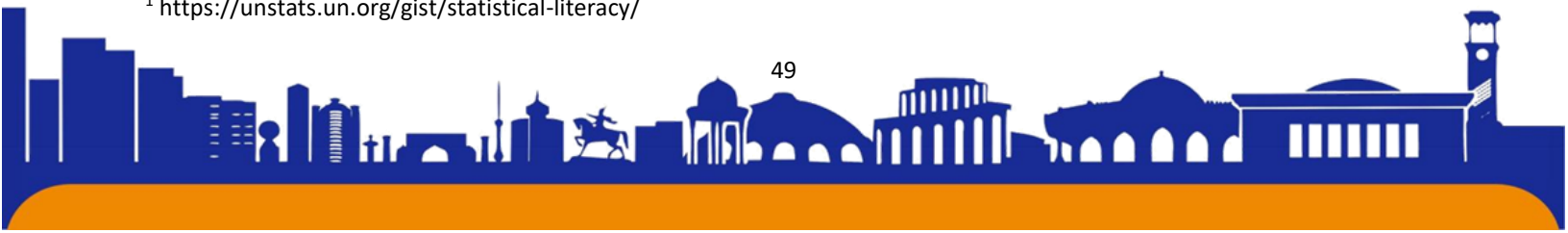
Yana bir muhim masala — statistik tafakkur va tanqidiy yondashuv. Data scientist nafaqat model tuzuvchi dasturchi, balki ma'lumotga asoslangan xulosa chiqaruvchi tadqiqotchidir. Shu sababli, statistik kompetensiya doirasida har qanday ma'lumotga tanqidiy yondashish, noto'g'ri to'plangan ma'lumotni aniqlash, ofatlar va ekstremal qiymatlarni (outlier) ajratish, missing values bilan ishlash, normal taqsimotga bog'liq bo'lmagan metodlarni tanlay olish kompetensiyalari shakllanishi zarur.

Statistik kompetensiyalarning shakllanishida ta'lim va o'rganish manbalarining sifati muhim ahamiyatga ega. Bugungi kunda open-source muhitdagi statistika bo'yicha qo'llanmalar, interaktiv kurslar (masalan, Coursera'dagi "Statistics with R" yoki edX'dagi "Data Science: Probability"), shuningdek, R va Python kutubxonalaridagi (scikit-learn, statsmodels) tayyor modellar orqali statistik bilimlarni amaliyotda qo'llash imkoniyati kengaygan. Biroq bular bilan ishlash uchun dastlab nazariy bilimlar, ya'ni statistik kompetensiyalar puxta egallangan bo'lishi zarur (1-jadval).

1-jadval. Turli mamlakatlarda statistik kompetensiyalar va data literacy darajasining taqqoslamasi¹

№	Mamlakat	Statistik savodxonlik (%)	Data literacy darajasi (%)	Statistika asosidagi qaror qabul qilish kompetensiyasi (%)
1	Finlyandiya	84	79	72
2	Kanada	81	76	68
3	Yaponiya	79	74	70
4	Germaniya	77	72	67
5	AQSh	74	70	62

¹ <https://unstats.un.org/gist/statistical-literacy/>



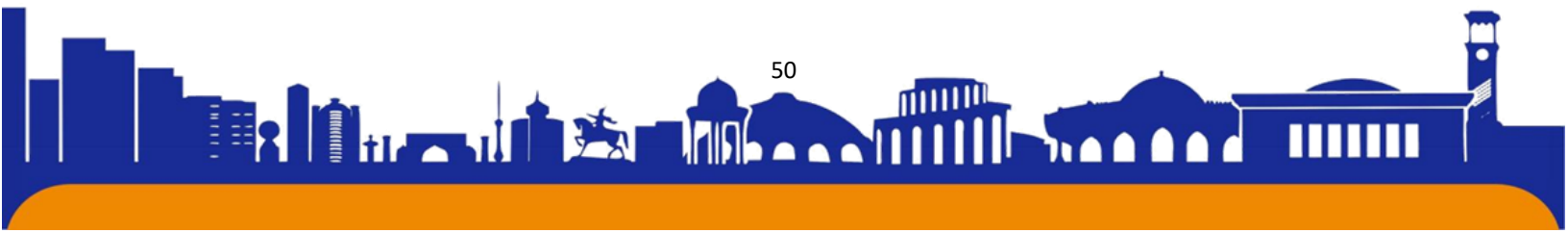
6	Janubiy Koreya	71	68	60
7	Turkiya	56	52	44

Yuqoridagi jadval tahlili shuni ko'rsatadiki, statistik kompetensiyalar darajasi mamlakatning raqamli tayyorgarlik holati, ta'lim tizimi va innovatsion salohiyati bilan bevosita bog'liq. Finlyandiya, Kanada va Yaponiya yuqori statistik savodxonlik va data literacy darajasi bilan ajralib turadi, bu esa ularning ta'lim tizimida statistik tafakkurga alohida e'tibor berilayotganini ko'rsatadi. Ayniqsa, bu davlatlarda qaror qabul qilishda statistik yondashuvning ustuvorligi dolzarb hisoblanadi.

Germaniya va AQShda ham ko'rsatkichlar nisbatan yuqori bo'lsa-da, ularning amaliy qarorlar darajasida biroz pastroq ekanligi nazariy bilimlarning amaliyot bilan yetarlicha integratsiyalanmaganligini ko'rsatadi. Janubiy Koreya va Turkiya misolida esa ma'lumotlar savodxonligi bo'yicha o'sish borligi kuzatiladi, ammo hali raqamli qarorlar qabul qilish darajasi zaifroq. O'zbekiston misolida esa statistik kompetensiyalar darajasi past bo'lib, bu sohaga e'tibor kuchaytirilishi zarurligini anglatadi. Bu holat ta'lim tizimida statistik tafakkurni rivojlantirish, data science yo'nalishidagi o'quv dasturlarni joriy qilish va xalqaro metodikalarga asoslangan mashg'ulotlarni kengaytirish zaruratini ko'rsatadi.

Xalqaro darajadagi tajribalar statistik kompetensiyalarning nafaqat texnik, balki ijtimoiy va boshqaruv jihatidan ham ahamiyatga ega ekanini ko'rsatmoqda. Masalan, sog'liqni saqlash sohasida COVID-19 pandemiyasi davrida statistik kompetensiyaga ega bo'lmagan tahlilchilar tomonidan noto'g'ri interpretatsiya qilingan epidemiologik ko'rsatkichlar ijtimoiy vahima keltirib chiqargan. Aksincha, statistik kompetensiyaga ega mutaxassislar tomonidan ishlab chiqilgan grafiklar, bashoratlar va tavsiyalar siyosiy qarorlar uchun asos bo'ldi.

Bundan tashqari, raqamli marketing, moliyaviy modellashtirish, iqlim bashorati, sanoat diagnostikasi va boshqa ko'plab sohalarda statistik kompetensiyalarsiz data science samarali bo'la olmaydi. Har bir model, har bir taxmin, har bir algoritm orqasida statistik asos borligini anglash — bu statistik tafakkurga asoslangan yondashuvdir. Ushbu yondashuv esa aynan statistik kompetensiyalarni chuqur egallagan mutaxassislarda shakllanadi.



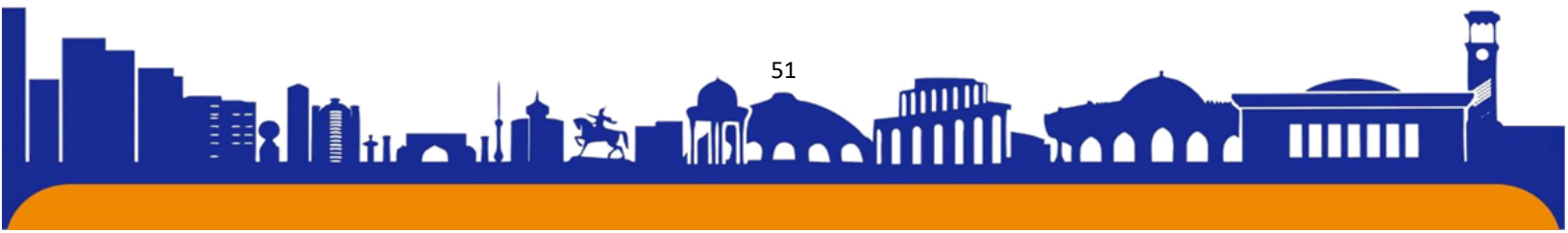
Xulosa va takliflar

Data science sohasida statistik kompetensiyalar nafaqat texnik vosita, balki chuqur tahliliy tafakkur, aniqlik, ishonchlilik va ilmiy yondashuvning asosiy omili sifatida namoyon bo'lmoqda. Tadqiqotlar shuni ko'rsatmoqdaki, statistik bilimlarni yetarli darajada egallamagan mutaxassislar tomonidan yaratilgan modellar ko'pincha noto'g'ri talqin qilinadi, natijada xatolik ehtimoli ortadi va qaror qabul qilish jarayoni zaiflashadi. Bu esa, zamonaviy raqamli iqtisodiyotda data science mutaxassislarning statistik tafakkur, ehtimollik asosidagi qarorlar va gipotezalarni sinovdan o'tkazish malakalarini chuqurlashtirish zarurligini taqozo etadi. Statistik kompetensiyalar, ayniqsa, real ma'lumotlardagi noaniqlik, ekstremal qiymatlar, bias-variance muvozanati, gipoteza sinovlari va regressiya modellari bilan ishlashda hal qiluvchi rol o'ynaydi.

Sohani rivojlantirish uchun, avvalo, statistik kompetensiyalarni bosqichma-bosqich egallashga qaratilgan tizimli o'quv modullari ishlab chiqilishi lozim. Bu jarayonda nafaqat nazariy bilim, balki real ma'lumotlar bazasi asosida mustahkamlangan amaliy mashg'ulotlar, analitik loyihalar va A/B testlar asosida o'qitish zarur. Ta'lim muassasalarida data science kurslari statistik tafakkurga asoslangan holatda qayta tuzilishi, ish beruvchilar esa mutaxassislarni tanlashda statistik savodxonlikni asosiy mezon sifatida belgilashi lozim. Bundan tashqari, ilmiy-tadqiqot institutlari va tarmoq markazlarida statistik metodlar va ularning sun'iy intellekt, iqtisodiyot, tibbiyot, ekologiya sohalaridagi amaliyotini integratsiyalovchi tahliliy markazlar tashkil etilishi sohaning amaliy ahamiyatini yanada oshiradi.

Foydalanilgan adabiyotlar ro'yxati:

1. T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning. – New York: Springer, 2009. – 745 p.
2. G. James, D. Witten, T. Hastie, R. Tibshirani. An Introduction to Statistical Learning. – New York: Springer, 2013. – 426 p.
3. D.J. Hand. Statistics: A Very Short Introduction. – Oxford: Oxford University Press, 2008. – 144 p.
4. C. O'Neil. Weapons of Math Destruction. – New York: Crown Publishing, 2016. – 259 p.



5. D.T. Larose. Discovering Knowledge in Data. – New York: Wiley, 2014. – 336 p.
6. F. Provost, T. Fawcett. Data Science for Business. – Sebastopol: O'Reilly Media, 2013. – 414 p.
7. D.S. Moore, W.I. Notz, M.A. Fligner. The Basic Practice of Statistics. – New York: W.H. Freeman, 2017. – 720 p.
8. D.C. Montgomery, G.C. Runger. Applied Statistics and Probability for Engineers. – Hoboken: Wiley, 2018. – 768 p.
9. H. Wickham, G. Grolemund. R for Data Science. – Sebastopol: O'Reilly Media, 2017. – 522 p.
10. A. Gelman, J. Hill. Data Analysis Using Regression and Multilevel/Hierarchical Models. – Cambridge: Cambridge University Press, 2007. – 625 p.
11. P. Bruce, A. Bruce. Practical Statistics for Data Scientists. – Sebastopol: O'Reilly Media, 2017. – 318 p.
12. <https://unstats.un.org/gist/statistical-literacy/>