

**TASNIFLASH VA KLASSTERLASH*****Onarkulov Maqsadjon Karimberdiyevich****Farg'ona davlat universiteti amaliy matematika va**informatika kafedراسи dotsenti**[maxmaqsad@gmail.com](mailto:maxmaqsad@gmail.com)****Tojimamatov Israiljon Nurmamatovich****Farg'ona davlat universiteti amaliy matematika va informatika kafedراسи**katta o'qituvchisi [isik80@mail.ru](mailto:isik80@mail.ru)****Yusupov Mirsaid Abdulaziz o'g'li****Farg'ona davlat universiteti amaliy matematika va informatika kafedراسи**o'qituvchisi [mirsaidbeky@gmail.com](mailto:mirsaidbeky@gmail.com)****Jamoliddinova Diyora Umidjon qizi****Farg'ona davlat universiteti 3-kurs talabasi,**[jamoldinovadiyora07@gmail.com](mailto:jamoldinovadiyora07@gmail.com)*

***Annotatsiya:*** Tezisda ma'lumotlarni klasterlash va tasniflash haqida to'liq ma'lumot berib o'tiladi. Ularning afzalliklari, qo'llanish sohalari va ishlash prinsiplari aytib o'tiladi.

***Kalit so'zlar:*** Tasniflash, Klasterlash, Nazoratli o'rganish, Nazoratsiz o'rganish, Guruhlash, Label.

***Annotation:*** The thesis provides a detailed overview of data clustering and classification methods, discussing their advantages, application areas, and working principles.

***Keywords:*** Classification, Clustering, Supervised learning, Unsupervised learning, Grouping, Labeling.

***Аннотация:*** В тезисе подробно рассматриваются методы кластеризации и классификации данных. Описываются их преимущества, области применения и принципы работы.

***Ключевые слова:*** классификация, кластеризация, обучение с

*учителем, обучение без учителя, группировка, маркировка.*

### **Kirish**

Ma'lumotlar tahlili va sun'iy intellekt sohalarida tasniflash (classification) va klasterlash (clustering) muhim o'rin tutadi. Bu ikkala metod ma'lumotlarni tahlil qilish, o'rgatish va tuzish jarayonlarini osonlashtiradi. Tasniflash, ma'lumotlarni belgilangan kategoriyalarga ajratishga qaratilgan bo'lsa, klasterlash esa tasodifiy yoki guruhli ma'lumotlarni o'zaro o'xshashliklariga qarab guruhlashni amalga oshiradi. Ushbu maqolada, tasniflash va klasterlash usullarini ko'rib chiqamiz, ularning farqlari, qo'llanilishi va asosiy metodlarni tahlil qilamiz.

**Tasniflash** (classification) va **klasterlash** (clustering) — mashina o'rganish (machine learning) metodlarining asosiy yo'nalishlaridan bo'lib, ular ma'lumotlarni tahlil qilishda keng qo'llaniladi. Har ikkisi ham ma'lumotlardagi struktura va o'xshashliklarni aniqlashga xizmat qiladi, ammo ularning ishlash prinsiplari va qo'llanish sohaları farq qiladi.

**Tasniflash** (Classification) — bu nazoratli o'rganish (supervised learning) metodidir. Bu usulda modelni o'qitish uchun to'liq belgilangan (labellangan) ma'lumotlar to'plami ishlatiladi. Model, o'rganilgan toifalar asosida yangi, ko'rilmagan ma'lumotlarni toifalashni amalga oshiradi. Tasniflash usuli ko'pincha e-pochta spamini aniqlash, xatolarni tasniflash, bemorlarning kasalliklarini diagnostika qilish kabi vazifalarda qo'llaniladi. Tasniflashda maqsad toifalarni oldindan aniqlash va yangi ma'lumotlarni mavjud toifalarga ajratish bo'lsa, klasterlashda maqsad ma'lumotlardagi o'xshashliklarga asoslangan holda yangi guruhlar yaratishdir. Tasniflashda ma'lumotlarga belgilangan javob (label) kerak bo'lsa, klasterlashda bunday belgilash talab qilinmaydi. Shu sababli, tasniflash va klasterlash o'rtasidagi asosiy farq — tasniflash nazoratli o'rganishga, klasterlash esa nazoratsiz o'rganishga asoslanadi.

### **Misollar:**

E-pochta xabarlarini spam yoki spam emas deb tasniflash.

Xususan, tibbiyotda bemorning kasalligi bor yoki yo'qligini aniqlash.

Kredit kartasi tranzaksiyalarini firibgarlik yoki normal deb tasniflash.

**Klasterlash** (Clustering) — bu nazoratsiz o'rganish (unsupervised learning) usulidir. Klasterlashda ma'lumotlar to'plamidagi obyektlar oldindan toifalanmagan bo'lib, model o'xshashliklar asosida ma'lumotlarni guruhlashga harakat qiladi. Bu usul, odatda, mijozlarni segmentlash, tasvirlarni o'xshashliklarga qarab guruhlash yoki yangi ma'lumotlarning tuzilishini o'rganishda qo'llaniladi.

**Farqlar:**

1. Tasniflash uchun ma'lumotlar toifalangan va model bu etiketlarga asoslanib o'qitiladi, klasterlashda esa ma'lumotlar toifalanmagan va model o'z-o'zidan guruhlashni amalga oshiradi.

2. Tasniflash ko'proq oldindan belgilangan javobga (label) ega bo'lishini talab qiladi, klasterlash esa bu kabi javoblarga muhtoj emas. Shunday qilib, tasniflash aniq va belgilangan toifalarga ajratish uchun ishlatiladi, klasterlash esa o'xshashliklarga asoslangan holda guruhlash uchun ishlatiladi.

Sun'iy intellekt (SI) sohasida o'rganish jarayonlari turli metodlarga bo'linadi. Ularning eng keng tarqalganlaridan biri nazoratli o'rganish (supervised learning) metodidir. Nazoratli o'rganish, tizimga ma'lumotlar va ularga tegishli javoblar (etiketlar) berilib, bu orqali modelni o'rgatish jarayonini anglatadi. Bu metod ko'plab muammolarni hal qilishda, jumladan, tasniflash (classification) va regressiya (regression) masalalarida keng qo'llaniladi.

**Nazoratli o'rganish** – bu o'rganish usulidan biri bo'lib, ma'lumotlar to'plami bilan birga har bir namunaga tegishli to'g'ri javob (etiket) beriladi. Model, ushbu ma'lumotlar to'plamidan foydalangan holda, yangi, ilgari ko'rilmagan ma'lumotlar uchun prognozlar qilishni o'rganadi.

Nazoratli o'rganishning asosiy komponentlari:

1. Kiruvchi ma'lumotlar (features): Bu modelga o'rganish uchun berilgan xususiyatlar, ya'ni ma'lumotning atributlari. Masalan, uy sotish narxini prognoz qilishda xususiyatlar sifatida uyning maydoni, yoshi, joylashuvi va boshqalar bo'lishi mumkin.

2. Chiqish yoki etiket (labels): Har bir kiruvchi ma'lumotga tegishli bo'lgan natija yoki to'g'ri javob. Tasniflashda bu sinflar (mavjud kategoriyalar), regressiyada esa sonli qiymatlar bo'lishi mumkin.

3. Model: Kiruvchi ma'lumotlar va chiqish etiketlari asosida o'rganiladigan matematik yoki statistik model. Modelning maqsadi yangi, ilgari ko'rilmagan ma'lumotlar uchun prognoz qilishdir.

4. Yo'qotish funksiyasi (Loss function): Modelning noto'g'ri prognozlari necha foizni tashkil etayotganini o'lchovchi funksiya. Modelni o'rgatishda yo'qotish funksiyasini minimallashtirishga harakat qilinadi.

### **Nazoratli o'rganish ikkita asosiy yo'nalishda ishlaydi:**

Tasniflash va regressiya:

1. Tasniflash (Classification): Tasniflash - bu kiruvchi ma'lumotlarni oldindan belgilangan kategoriyalarga (sinflarga) ajratish jarayonidir. Tasniflashda chiqish etiketlari diskret (butun sonlar yoki kategoriyalar) bo'ladi

2. Regressiya (Regression): Regressiya - bu o'zgaruvchilar orasidagi uzviy bog'lanishni o'rganish va prognoz qilish jarayonidir. Regressiya masalalarida chiqish etiketlari uzluksiz qiymatlar bo'ladi.

**Nazoratsiz o'rganishning** asosiy maqsadi - ma'lumotlar orasidagi o'xshashliklarni yoki tuzilmalarni aniqlash, ulardan yangi va foydali bilimlarni chiqarishdir. Bu usulda ma'lumotlarga hech qanday etiketlar (klassifikatsiya) berilmaydi, lekin tizim ularni o'zining ichki xususiyatlariga ko'ra guruhlashga yoki modellashga harakat qiladi.

### **Nazoratsiz o'rganish ikkita asosiy yo'nalishda ishlaydi:**

Klasterlash va dimensiyaning kamaytirish:

**Guruhlash**, nazoratsiz o'rganish metodining bir qismi bo'lib, uning maqsadi - ma'lumotlarni xususiyatlari asosida o'xshashliklariga qarab guruhlariga ajratishdir. Guruhlar ichida bo'lgan ma'lumotlar o'zaro juda o'xshash bo'ladi, lekin turli guruhlardagi ma'lumotlar orasida katta farqlar mavjud bo'ladi.

Sun'iy intellekt (SI) va mashina o'rganish (ML) sohalarida **label** (etiket) — bu ma'lumotlarga berilgan tavsif yoki klassifikatsiyadir. Etiketlar, odatda

ma'lumotlar nuqtalari bilan bog'langan va tizimga ma'lum bir kategoriyani yoki natijani bildiradi. Etiketlar, asosan, nazoratli o'rganish (supervised learning) algoritmlarida ishlatiladi, chunki tizim etiketlangan ma'lumotlardan o'rganish orqali to'g'ri qarorlar qabul qilishni o'rganadi. Etiketlar yordamida model ma'lumotlar va javoblar o'rtasidagi bog'lanishni o'rganadi va shu asosda kelajakda yangi, ko'rsatilmagan ma'lumotlarga asoslangan prediktsiyalar qiladi.

**Etiketlar** (Label) Nima? Etiket (label) — bu ma'lumotning ma'lum bir xususiyatini yoki klassifikatsiyasini ifodalovchi qiymat yoki tavsifdir.

### **Xulosa:**

Tasniflash va klasterlash mashina o'rganishning asosiy metodlaridan bo'lib, ular ma'lumotlarni tahlil qilishda keng qo'llaniladi. Tasniflashda ma'lumotlar oldindan belgilangan toifalarga ajratiladi, klasterlashda esa ma'lumotlar tabiiy guruhlariga ajratiladi. Har bir metodning o'ziga xos afzalliklari bor va ular turli vazifalarni hal qilishda samarali ishlaydi. Ikkala metod ham turli tahlil vazifalarini amalga oshirishda muhim ahamiyatga ega bo'lib, ular so'nggi yillarda katta ma'lumotlar va sun'iy intellekt texnologiyalari rivojlanishi bilan yanada kengroq qo'llanilmoqda.

### **FOYDALANILGAN ADABIYOTLAR**

- 1.Pétrowski, Alain; Ben-Hamida, Sana (2017). Evolutionary algorithms. John Wiley & Sons. p. 30. ISBN 978-1-119-13638-5.
- 2.Mitchell 1996, p. 2.
- 3.Gerges, Firas; Zouein, Germain; Azar, Danielle (12 March 2018). "Genetic Algorithms with Local Optima Handling to Solve Sudoku Puzzles". Proceedings of the 2018 International Conference on Computing and Artificial Intelligence. ICCAI 2018. New York, NY, USA: Association for Computing Machinery. pp. 19–22. doi:10.1145/3194452.3194463. ISBN 978-1-4503-6419-5. S2CID 44152535.
- 4.Burkhardt, Michael C.; Ruiz, Gabriel (2023). "Neuroevolutionary representations for learning heterogeneous treatment effects". Journal of

- Computational Science. 71: 102054. doi:10.1016/j.jocs.2023.102054. S2CID 258752823.
5. Whitley 1994, p. 66.
6. Luque-Rodriguez, Maria; Molina-Baena, Jose; Jimenez-Vilchez, Alfonso; Arauzo-Azofra, Antonio (2022). "Initialization of Feature Selection Search for Classification (sec. 3)". *Journal of Artificial Intelligence Research*. 75: 953–983. doi:10.1613/jair.1.14015.
7. Eiben, A. E. et al (1994). "Genetic algorithms with multi-parent recombination". *PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature*: 78–87. ISBN 3-540-58484-6.
8. Ting, Chuan-Kang (2005). "On the Mean Convergence Time of Multi-parent Genetic Algorithms Without Selection". *Advances in Artificial Life*: 403–412. ISBN 978-3-540-28848-0.
9. Deb, Kalyanmoy; Spears, William M. (1997). "C6.2: Speciation methods". *Handbook of Evolutionary Computation*. Institute of Physics Publishing. S2CID 3547258.
10. Shir, Ofer M. (2012). "Niching in Evolutionary Algorithms". In Rozenberg, Grzegorz; Bäck, Thomas; Kok, Joost N. (eds.). *Handbook of Natural Computing*. Springer Berlin Heidelberg. pp. 1035–1069. doi:10.1007/978-3-540-92910-9\_32. ISBN 9783540929093.