Volume 2, Issue 11, November - 2024 **ISSN (E):** 2938-3811

CLASSIFICATION AND CLUSTERING

Axmedov Baxriddin Madaminovich
Farg‘ona politexnika institutiQurilish fakulteti “Geodeziya, kartografiya va kadastr”
kafedrasi.b.axmedov@ferpi.uz (ORCID: 0000-0002-2897-7812),

Astonaqulova Kibriyoxon Baxriddin qizi
Farg‘ona davlat universiteti Amaliy matematika yo‘nalishi 3-kurs talabasi,
kibriyoxonastonaqulova@gmail.com

Abstract
This article provides information about classification and clustering methods, classification and clustering algorithms, their workflow and application in real life, as well as examples.

Keywords: Interface, artificial intelligence, computer, classes, classification, clustering, credit, risk, big data, categorized, unclassified

Introduction

Big data technology has revolutionized the economy, science, medicine, and other industries in recent years. As data increases in size, variety and velocity, there is a need for effective methods for its analysis. Classification (*classification*) and clustering (*clustering*) are the key tools in this process.

I. Classification and Its Essence

Classification is the process of sorting given data into predefined categories (or classes). The goal is to train the model and provide the correct categorization of new data.

The process for classification includes:

1. Data preparation**:** To train a model, a data set will be needed, with each sample having an appropriate class tag.
2. Model training**:** A model is created through machine learning algorithms.
3. Test and evaluation: Validate the model on test data and evaluate its accuracy.
4. Applicability: The model is used in relation to new data.

Classification Algorithms

1. Logistic Regression: Efficient in binary classification tasks. For example, splitting *spam* or *real* email.
2. Decision Trees: Can also distinguish between more difficult categories.
3. Random Forest: A combination of multiple decision trees. Increases accuracy on large volumes of data.
4. Support Vector Machines (SVM): Good handling of large dimensional data.

Web of Scientists and Scholars: Journal of Multidisciplinary Research

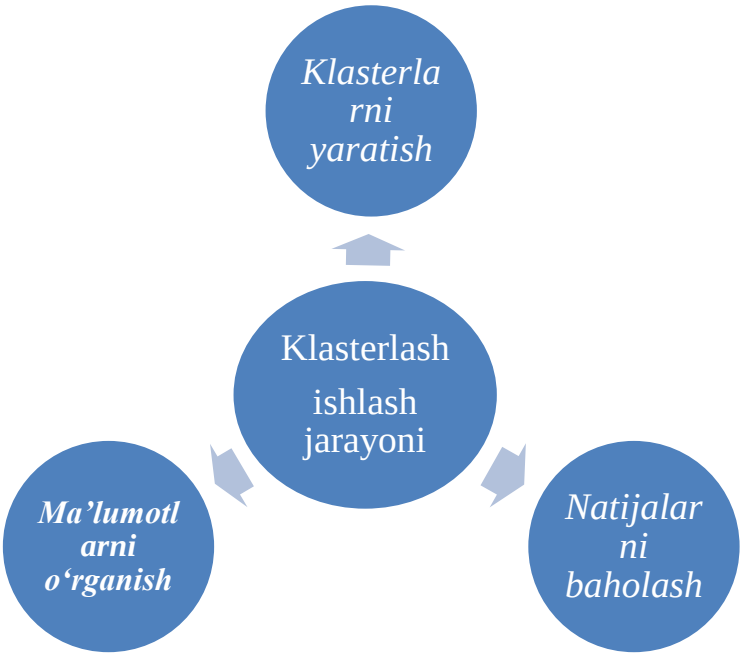
5. Neural Networks: Artificial neural networks deliver the highest results in large amounts of complex data [1-5].

Applied areas of application of classification

- Medicine: Classification of diseases (e.g., cancer detection).
- Marketing: Segmenting customers by product (e.g., new users or loyal customers).
- Cybersecurity: Classification of dangerous or unsafe internet activities.
- Financial analysis: Evaluating loan applications and determining solvency.

II. Clustering and its essence

Clustering is a method of grouping data based on mutual similarities. Unlike classification, in which the categories are not known in advance. The goal is to identify natural groups.



1. Data Exploration: Identify links between data.
2. Cluster Creation: Identify natural clusters by grouping algorithms.
3. Evaluation of results: Measurement of the quality of clusters (e.g., using the Silhouette indicator).

Clustering algorithms

1. K-Means Clustering: Divides data into K groups. It is very popular because of its simplicity.
2. Hierarchical Clustering: Places data in the form of a tree. This method is suitable for small volumes of data.
3. DBSCAN (Density-Based Spatial Clustering): Works without initial categories. The program groups the data by density.

Web of Scientists and Scholars: Journal of Multidisciplinary Research

4. Gaussian Mixture Models (GMM): Constructs clusters based on their statistical properties [5-10].

Practical application of clustering

- Customer segmentation: In marketing, grouping customers according to their shopping habits.
- Genetics: Grouping genes or DNA strands based on similarity.
- Social Networks: Clustering users based on their interests and activity.
- Geography: Grouping places on maps by a specific type (e.g., forest or urban areas).

Classification and Clustering Comparison Table

PROPERTY	CLASSIFICATION	KLASTERLASH
Ma'lumot turi	Categorized information	Unlabeled information
Purpose	Do not categorize	Definition of natural groups
Control Type	Supervised Learning	Unsupervised teaching
Basic algorithms	Logistic Regression, Random Forest, SVM	K-Means, DBSCAN, GMM
Applications	Credit ratings, medicine	Customer segmentation, genomics

Examples of Using Classification and Clustering

Classification Example: Credit Risk Assessment

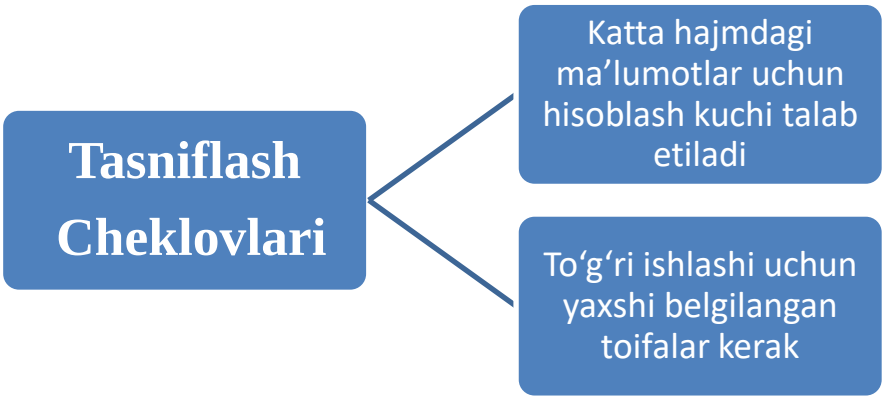
The bank will assess the financial condition of the customer before approving the loan. Using a model classification algorithm, it classifies a client into a *solvent* or *incapable* class.

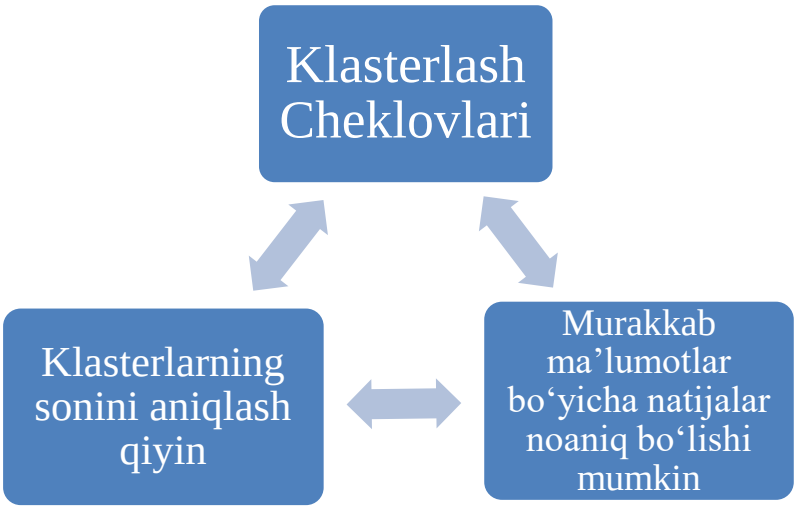
Clustering example: Customer segmentation

An e-commerce company can segment customers based on purchase data into the following clusters:

- Customers who shop frequently.
- New users.
- Customers who buy expensive products [10-15].

III. Limitations of Classification and Clustering





Conclusion

Classification and clustering methods are important in big data analysis. Whereas classification works based on predefined categories, clustering focuses on identifying natural groups. Both methods complement each other and help to get maximum benefit from data in different areas. In the future, these methods will come to a more perfect level with the development of machine learning and artificial intelligence.

References

1. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
3. Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.
4. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer
5. Kumar, V., Steinbach, M., & Tan, P.-N. (2005). *Introduction to Data Mining*. Addison Wesley
6. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). *Data Clustering: A Review*. ACM Computing Surveys.
- A review on the basic clustering algorithms and their application.
7. Akhmedov B. M. and Astonakulova K. B. Neurochips and their characteristics
8. Tojimamatov I. N. Madaminova Kibriikhon Bakhriddinovna FORMAL GRAMMAR AND SEMANTIC GRID Online Conference.
9. Madaminovich A. B. MEASUREMENT OF DEFORMATION OF BUILDINGS AND STRUCTURES BY GEODETIC METHODS //Western European Journal of Modern Experiments and Scientific Methods. – 2024. – Т. 2. – №. 6. – С. 170-179.

10. Akhmedov B. M. Methods of Calculating Function Range Calculations in Accuracy Assessment. Evaluation of Parametric Determination of Equation //Texas Journal of Engineering and Technology. – 2023. – T. 21. – C. 57-62.
11. Akhmedov B. Using the fundamentals of the theory of measurement errors in performing geodesic measurement and calculation works //E3S Web of Conferences. – EDP Sciences, 2023. – T. 452. – C. 03012.
12. Axmedov B. M., Qosimov L. M. Knauf Insulation is Effective Isolation //Central Asian Journal of Theoretical and Applied Science. – 2022. – T. 3. – №. 6. – C. 298-302.
13. [https://scikit-learn.org] (https://scikit-learn.org)
14. https://www.kdnuggets.com
15. https://scholar.google.com

Web of Scientists and Scholars: Journal of Multidisciplinary Research