2nd International Conference on Problems of Logistics, Management and Operation in The East-West Transport Corridor (PLMO 2023) Baku, Azerbaijan
DOI:10.54381/plmo2023.15

Data mining on base of artificial intelligence

Shahnaz SHAHBAZOVA, Nigar ALIYEVA
Azerbaijan State Economic University, UNEC, Department of Digital Technologies and Applied Informatics, 6 Istiqlaliyat str., Baku, Azerbaijan
E-mail: shhabazova@gmail.com, shahnaz_shahbazova@unec.edu.az

Abstract. Data mining on the base of the artificial intelligence is using to create solutions; help regular databases to perform faster and extract hidden, useful part and information from data. Using data mining based on artificial intelligence in nowadays technology helps businesses to predict future trends and analysis behaviors, allowing them to make proactive, knowledge driven decisions and contain huge amount of data. Due to the requirements of the modern world data mining on the base of artificial intelligence can automate the most of manual processes, increase sales and boost business. Intelligent data risk analysis organizational decision making through deep data analysis. The data mining techniques underlying this analysis can be divided into two main objectives, they can either describe a data set or predict results using machine learning algorithms. These methods are used to identify and filter data, revealing the most interesting information, from fraud detection to user usage and even individual security incidents. My research has moved from the old point of view of manual data processing to an automated data management approach using the Big Data Analysis (BDA) paradigm [1]. BDA merges Big Data, Data Analytics, Data mining, and Artificial Intelligence in order to achieve data.

Keywords**:** *artificial intelligence; data mining; prediction; machine learning; automation; big data; BDA; analytics; solutions*

1. INTRODUCTION

The data mining process involves a series of steps from data collection to visualization to extract valuable insights from large datasets. Data mining techniques are used to create descriptions and predictions of a target data set.

Data scientists describe data by observing patterns, associations and correlations, classify and group data using classification and regression methods and identify outliers for use cases such as spam detection.

Data mining techniques are used in business intelligence and in data analysis sphere to extract knowledge for organization and industry. Some use cases for data mining include such spheres as banking sectors, sales and marketing, education, operational optimization, fraud detection and etc. Over the past years the number of the projects based on big data analytics into the business sector is huge. As we mentioned above, data mining plays a big role in banking sector for extracting more information from existing data and reveal useful information to better decision making. So in banking sector we can trace such data mining techniques as clustering, association, Data Cleaning, Data Visualization, classification, Machine Learning, prediction, Neural Networks Outlier Detection and Data Warehousing [2].

At the last few years we can see increasing the automation in financial sector. Intelligently analyzed data is a valuable resource. It can lead to new insights, and in commercial settings, to competitive advantages [3]

This use case is the best Al use case in this sector, as in this sector AI provide huge potentials. Al software helps banks streamline and automate every job currently done by humans, making the process simple and virtual. As a result, Al technologies can help to minimize workload as well as improve the quality of the work [4].

At the end, this work aims to advance our understanding of current challenges and the prospects of knowledge extracting from data mining. By using this techniques companies can make faster and cleaner decisions to provide their customers best solutions, keeping benefits.

2. DATA MINING AND AI PROCESS REVIEW

Knowledge Discovery in Databases (KDD): KDD is organized procedure of recognizing valid, useful, clear patterns from large data sets. Data Mining is the root of the KDD. Data mining is positioned as a step in the overall KDD process where the user applied selected data mining algorithms to identify interesting patterns [5].

The process of the data mining including such steps as collecting data, visualization and extracting valuable information from the huge data sets.

Data mining techniques are important for generating descriptions and predictions about targeted data set. In the case of our observation I also want to touching steps of data mining: business understanding, data preparation, data understanding, model building, evaluating and deployment steps. This steps also known as Cross Industry Standard Process for Data Mining (CRISP-DM) Model:

1. Business understanding - it seems to be harder part of the data mining process where data scientists with business stakeholders must work together to define the business problem and resolving cases, predict parameters for a given project.
2. Data preparation - process of preparing raw data so that it is suitable for further processing and analysis is named data preparation. After detecting problem data scientists can identify which set of data will help answer the pertinent questions to the business easily. Collecting datas must be cleaned, removing any noise, such as duplicates, missing values, labeling raw data into a form suitable for machine learning (ML) algorithms and then exploring and visualizing the data. Data scientists trying to keep the most important predictors to ensure optimal accuracy within any models.
3. Data Understanding - for understanding data we must dig deeper, defined quality of the data and
differences between other subsets.
4. Model building - a mining model is created by applying an algorithm to data and becoming into a set of data, statistics, and patterns that can be applied to new data to generate predictions and make inferences about relationships. Model building focuses on desired algorithms and use symbolic regression techniques nearby the other ones.
5. Evaluation - in this phase discovering additional constraints for projects final statement, redirected to the quality assurance (QA) and at last make decisions to proceed with deployment or accept changes for improvement.
6. Deployment - as a final stage model must deploy in the real world setting by monitoring results.

Also I want to mention about pattern mining that based on identifying rules which describe specific patterns within the data. Depending on the type of analysis, data scientists can investigate interesting data relationships such as sequential patterns, association rules and correlations AI work is based on the fast combining huge amounts of data with iterative processing, intelligent algorithms which allow the software automatically learn from patterns and features in the data. In generally Al project includes tree main stages.

- Project planning and data collection.
- Design and training of the Machine Learning (ML) model.

- Deployment and maintenance

These technologies are connected but not the same. Creating an artificial intelligence program requires a combination of data analysis, data mining, and machine learning and after creating an AI model new data can be entered to create predictions, without requiring manual intervention. As AI learns from experience incorrect data leads to inaccurate models if not programmed efficiently. For resolving this kind of problem we use Machine Learning (ML) [6].

ML is a subset of AI that focuses on algorithms that learn from data without being explicitly programmed [7]. Machine leaming can be used to automatically discover patterns in data and then use that knowledge to make predictions or recommendations [8].

Artificial Intelligence helps data mining:

- Identifying relevant datasets that may contain valuable information. This kind tasksloften a time consuming for humans but for AI can fast scan large amounts of data and find the right datasets.
- Extracting useful information from these datasets is a task that will take a very long time for humans and quickly but efficiently performing by AI.
- Analyze the extracted information and identify any patterns or trends - this analysis would be extremely difficult to do manually but can be easily done using Al algorithms.

Summary, the goal of the data mining based on Al is to provide software that can reason on input and explain on output. Al help to improve the accuracy of predictions that made from data mining models by providing more accurate results based on huge amounts of training data. Al will provide human-like interactions with software and offer decision support for specific tasks and predictions from huge and unstructured data sets [9].

3. APPROACHES TO DATA MINING ON BASE OF ARTIFICIAL INTELLIGENCE

Data mining works on base of different algorithms and techniques turning huge volumes of data into the useful information. Below we can track some of the most common ones:

Neural networks: neural networks process training data faster than human and mimicking how the human brain operates. Neural networks consist of three main layers such as input, hidden and output layers. Data at the first stage entered through the input layer, then processed in the hidden layer and resolved in the output layer and providing relevant actions on the base of taken data. The hidden layer can include many of processing layers that depend on the amount of data.

- Clustering: This technique used to represent data visually. Grouping various data points based on characteristics is clustering. Below we can introduce with methods for Data Clustering:
- Partitioning method: Involves dividing an input data set into a group of specific clusters, so data points belong to one group or cluster.
- Hierarchical method: Data points are a single cluster that grouped based on similarities. These clusters can be separately analyzed
- Density-based method: A machine learning method where for further analyzing used plotted together or vacant data points.
- Grid-based method: Involves dividing data into cells on a grid and clustered data by individual cells.
- Model-based method: Statistical approach to data clustering. In this method for each data cluster are created models that located the best data to fit that particular model.

Clustering can facilitate businesses manage their data. As we mentioned before, for understanding and predicting customer requirements banks and financial institutions must use clustering [10].

Association rules: this is method for finding relationships between differ in a given dataset.

Decision tree: for classifying or predicting potential outcomes based on a set of decisions of decision tree technique must use classification or regression methods.

K- nearest neighbor (KNN): KNN algorithm is a non-parametric algorithm. Using KNN algorithm data points based on their proximity are classified and associated to other available data. This algorithm suppose that similar data points can be found next each other.

Also the powerful and fast sorting of data to save time is the Fuzzy data mining [11]. The principle of Fuzzy logic and how it works in data mining, the system first grouping the data into one class or into the Fuzzy set of elements. After grouping, the centroid of the data is determined. This centroid helps to divide the data into cluster. Two or more clusters can be formed and it is depended on how many data in the database [12]. Fuzzy data is uncertain, inconsistent and incomplete data. Because of data from existing sources can be fuzzy data so it is important to develop methods from others. There are algorithms for clustering, classifying or visualizing fuzzy data [13].

4. CONCLUSION

At the end as we mentioned before data mining based on artificial intelligence. Data Mining was appeal as convenient approach that allowed the deployment of a broad range of suitable for transmission and distribution networks applications. The most supported areas are where uses plenty data and high system uncertainties cases. Data Mining Platform to Big Data Analytics. In this regard, using new generated data, data-driven solutions have been used to offer new modeling alternatives that treat data as practical knowledge rather than just isolated repositories of information. To provide solutions to the modernization of the distribution sector needs, the role of the different AI approaches (including Machine Learning) become perceptible. Nowadays AI can implement a large amount of data and exploited by companies for extracting huge data without limitations. Due these cases data mining based on artificial intelligence might provide quick and pure data to the large amount of companies without loss.

References

1. IEEE Smart Grid Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid Working Group, Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid: Introduction, Benefits, Challenges and Issues, White Paper, 2017.
2. B. D. V. Association, "European big data value strategic research and innovation agenda," Big Data Value Association, Tech. Rep., October 2017.
3. Aba.D.Andrew S and Bay S.D., 2014.
4. Bazarbash. M. 2019. "FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk." IMF Working Paper 19/109, International Monetary Fund, Washington, DC.
5. Reinastz 2002; Han and Kamber 2006; Kurgan and Musilek 2006.
6. SalitiME, and ShawlaSt 2014. "Data Mining and the Review of Database", IEEE Trans as ComRuter Technology and Processing. pp. 45 - 61.
7. Quinlan, JR.64.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
8. Quinlan, JR.. Induction on decision trees. Machine Learning, 1(1): 81-106, 1986.
9. Fayyad, UM.. Diatestku-Shapiro, G. &mSmsth. P. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence. 17:37-54, 1996.
10. Letb. B. Aboaxi, J. & Szeifert. F. Model order selection of nonlinear input-output models - a clustering based approach. Journal of Process Control, 14(6):593-602, 2004.
11. Bandemes. 2006.
12. Pal, N.R. Bezdak. J.C. & Isal E.C.K Generalized clustering networks and Kehensa's self-organization scheme. IEEE Transactions on Neural Networks, 4(4):549-557, 1993
13. Kuok, C.M., Fu, A. & Wong. M.I. Mining fuzzy association rules in databases. A CM SIGMOD Record, 27(1):41-46, 1998.