# AN ALGORITHM FOR EXTRACTING INFORMATICS FEATURES FROM DATA USING ARTIFICIAL INTELLIGENCE METHODS

***Begmatov Komiljon Uktam uglu***,

*Qarshi State Technical University,*

*Student of the Department of Telecommunication Technologies*

***Annotation.*** *The article reviews algorithms for extracting informatics features from data using artificial intelligence methods. Extracting informatics features is the process of identifying important and useful information from a data set. This process is carried out, in particular, using natural language processing (NLP), machine learning (ML), and deep learning (DL) technologies. The article covers in detail the stages of data collection, preprocessing, feature extraction, and selection, as well as algorithms for extracting informatics features using machine learning and deep learning methods (e.g., SVM, KNN, CNN, RNN).*

***Keywords:*** *Artificial intelligence, data, computer science, algorithm, natural language processing (NLP), machine learning (ML), deep learning (DL), technology, machine learning, deep learning methods, SVM, KNN, CNN, RNN, text, image and audio, format.*

***Аннотация.*** *Аннотация. В статье рассматриваются алгоритмы извлечения признаков информатики из данных с использованием методов искусственного интеллекта. Извлечение признаков информатики — это процесс выявления важной и полезной информации из набора данных. Этот процесс осуществляется, в частности, с использованием технологий обработки естественного языка (NLP), машинного обучения (ML) и глубокого обучения (DL). В статье подробно рассматриваются этапы сбора данных, предварительной обработки, извлечения и отбора признаков, а также алгоритмы извлечения признаков информатики с использованием методов машинного обучения и глубокого обучения (например, SVM, KNN, CNN, RNN).*

Algorithms are used to identify important features from data in various formats, such as text, images, and audio, and are of great importance for effective data analysis and decision-making. The article provides a detailed understanding of the practical application of informatics features extraction technologies using artificial intelligence and their role in data analysis.

In data analysis, extracting important information from data sets or texts is one of the main tasks of artificial intelligence (AI) technologies. Extracting information from data sets involves identifying important and useful information. This process is carried out, in particular, using natural language processing (NLP), machine learning (ML), and deep learning methods.

Basic Steps in Extracting Information from Data. The basic steps in extracting information from data are as follows.

Data Collection and Aggregation. The first step in extracting information from data is to collect data. Data can be in different formats: text, images, video, or audio. Specific algorithms and AI methods are used for each format. For example, NLP is used for text data, convolutional neural networks (CNN) for images, and recurrent neural networks (RNN) for audio data.

Data Preprocessing. Before data extraction, the data must be cleaned, normalized, and encoded. This process includes the following.

Text cleaning. Cleaning texts from stop words, special characters, and non-standard words.

Lexical normalization: Lemmatization or stemming of words.

Tokenization. Dividing text into small pieces (tokens).

Feature Selection and Extraction.

In feature selection, the main and computationally important parts of the data are separated using a specific algorithm.

TF-IDF (Term Frequency-Inverse Document Frequency): Used to identify important words for textual data. This method takes into account how often a word occurs in a document and how widespread it is in the entire data set.

Word Embeddings (Word2Vec, GloVe): Represents words using vectors, taking into account the semantic meaning of words, which provides understanding of words in relation to each other.

Feature Selection. When selecting features, not all features are equally important. Some features may be more important than others in identifying information. The following methods are used in feature selection.

Mutual Information: Measures the relationship between features.

PCA (Principal Component Analysis): Combines similar features and reduces the dimensionality of the data.

Algorithms and Methods. Machine Learning (ML) Algorithms. Machine learning algorithms are used to extract informational features from data. The following methods are widely used.

KNN (K-Nearest Neighbors): Used to group objects with similar features.

SVM (Support Vector Machine): Effective in separating categories and selecting features.

Random Forest. Determines the relationship between features through many decision trees.

Deep Learning Methods. Deep learning can extract informational features from complex data. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are especially effective in analyzing text, image, and audio data.

CNN (Convolutional Neural Networks). Used to extract features from images and video data. In each layer, important elements of the image are extracted.

RNN (Recurrent Neural Networks). Used to extract features from sequential data, such as text or time series data. Advanced models such as LSTM (Long Short-

Term Memory) or GRU (Gated Recurrent Unit) help extract informatics features by remembering past states.

Emerging Feature Analysis. The extracted features are then analyzed and based on them, model outputs or decisions are made. Machine learning or deep learning models are often used in feature analysis. Classification, regression, and other forecasting methods are performed using the analyzed features.

Practical Applications. The practical applications of informatics feature extraction are very wide. Here are some examples. Text data analysis: Data classification, sentiment analysis, extracting important information.

Image analysis. Object detection, face recognition, image classification.

Medical data. Identify diseases by extracting important features from medical images and laboratory results.

Audio data. Speech recognition, music analysis, voice command recognition.

Data mining algorithms using artificial intelligence techniques are used to identify important and useful features from various types of data. Machine learning and deep learning algorithms play an important role in automating and making this process efficient. Proper analysis and extraction of data provides greater accuracy and efficiency.

**REFERECEN:**

1.      Daminova B. E., Bozorova I. J., Jumayeva N. X. FORMATION OF TEXT DATA PROCESSING SKILLS //Экономика и социум. – 2024. – №. 4-2 (119). – С. 110-119.

2.      Daminova B. E. et al. USE OF ONLINE ELECTRONIC DICTIONARIES IN ENGLISH LANGUAGE LESSONS //Экономика и социум. – 2024. – №. 5-1 (120). – С. 193-196.

3.      Daminova B. E. et al. ADVANTAGES OF USING MULTIMEDIA RESOURCES IN ENGLISH LANGUAGE LESSONS //Экономика и социум. – 2024. – №. 5-1 (120). – С. 207-210.

4.      Daminova B. E. et al. SCIENTIFIC AND METHODOLOGICAL SUPPORT OF EDUCATIONAL INFORMATION INTERACTION IN THE EDUCATIONAL

PROCESS BASED ON INTERACTIVE ELECTRONIC EDUCATIONAL RESOURCES: USING THE EXAMPLE OF TEACHING ENGLISH //Экономика и социум. – 2024. – №. 5-1 (120). – С. 233-236.

5.      Daminova B. E. et al. THE ROLE AND FEATURES OF THE USE OF INFORMATION TECHNOLOGY IN TEACHING A FOREIGN LANGUAGE //Экономика и социум. – 2024. – №. 5-1 (120). – С. 184-188.

6.      Daminova B. E. et al. USING THE GOOGLE CLASSROOM WEB SERVICE AND PREPARING INTERACTIVE PRESENTATIONS //Экономика и социум. – 2024. – №. 5-1 (120). – С. 216-225.

7.      Daminova B. E., Bozorova I. J., Jumayeva N. X. CREATION OF ELECTRONIC LEARNING MATERIALS USING MICROSOFT WORD PROGRAM //Экономика и социум. – 2024. – №. 4-2 (119). – С. 104-109. 1. – S. 1169-1172.

8.      Daminova B. E. et al. APPLICATION OF MODERN INFORMATION AND COMMUNICATION TECHNOLOGIES IN TEACHING ENGLISH //Экономика и социум. – 2024. – №. 5-1 (120). – С. 197-201.

9.      Daminova B. E. et al. SOFTWARE TOOLS FOR CREATING MULTIMEDIA RESOURCES IN TEACHING ENGLISH //Экономика и социум. – 2024. – №. 5-1 (120). – С. 202-206.

10.      Daminova B. E. et al. THE MAIN ADVANTAGES, PROBLEMS AND DISADVANTAGES OF USING MULTIMEDIA IN TEACHING FOREIGN LANGUAGES //Экономика и социум. – 2024. – №. 5-1 (120). – С. 189-192.

11.      Даминова Б. Э. и др. ОБРАБОТКА ВИДЕОМАТЕРИАЛОВ ПРИ РАЗРАБОТКЕ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ //Экономика и социум. – 2024. – №. 2-2 (117). – С. 435-443.

12.      Daminova B. E. GAUSS AND ITERATION METHODS FOR SOLVING A SYSTEM OF LINEAR ALGEBRAIC EQUATIONS //Экономика и социум. – 2024. – №. 2 (117)-1. – С. 235-239.

13.      Daminova B. E., Oripova M. O. METHODS OF USING MODERN METHODS BY TEACHERS OF MATHEMATICS AND INFORMATION

TECHNOLOGIES IN THE CLASSROOM //Экономика и социум. – 2024. – №. 2 (117)-1. – С. 256-261.

14.     Daminova B. E. et al. USE OF ELECTRONIC EDUCATIONAL RESOURCES IN THE PROCESS OF TEACHING A FOREIGN LANGUAGE //Экономика и социум. – 2024. – №. 5-1 (120). – С. 230-232.

15.     Daminova B. E. et al. USING COMPUTER PRESENTATIONS IN TEACHING FOREIGN LANGUAGES //Экономика и социум. – 2024. – №. 5-1 (120). – С. 211-215.

16.     Daminova B. E. et al. USING DIGITAL TECHNOLOGIES IN FOREIGN LANGUAGE LESSONS //Экономика и социум. – 2024. – №. 5-1 (120). – С. 226-229.