

## **MACHINE LEARNING ALGORITHMS FOR AUTOMATIC DIAGNOSIS AND PREDICTION OF DISEASES BASED ON MEDICAL DATA**

**Abstract:** This article uses logical regression method and machine learning algorithms through sigmoid function for automatic diagnosis and prediction of diseases. Logical regression method is used in the article, because this method is effective in solving classification problems. Logical regression models allow predicting the presence of various diseases or the level of susceptibility to them based on the patient's data (for example, age, sex, medical indicators, laboratory results, anamnesis). This method is convenient for use in determining the risk of disease and predicting the onset or development of the disease.

**Keywords:** sigmoid function, logical regression, machine learning

### **Relevance of the topic:**

Uzbekistan aims to transition to a digital economy by 2030. In this period, digitalization in the field of IT is one of the priority areas of national development. This action strategy is aimed at wide implementation in all sectors of the country, especially in the field of medicine.

By 2030, the widespread use of IT infrastructure in the medical field will play an important role in increasing the country's competitiveness.

Early detection of diseases and their timely prediction are of great importance in modern medicine. In this process, the creation of automated diagnostic systems based on the patient's medical history and medical indicators will help to improve the quality of life of patients and increase the efficiency of the healthcare system. This project aims to create a custom web platform using machine learning algorithms for disease diagnosis and prediction.

**The logical regression method** is used in the project, because this method is effective in solving classification problems. Logical regression models allow predicting the presence of various diseases or the level of susceptibility to them based on the patient's data (for example, age, sex, medical indicators, laboratory results, anamnesis). This method is convenient for use in determining the risk of disease and predicting the onset or development of the disease.

The main functions of this web platform include:

1. Data entry: Users can enter medical data and other relevant information.

2. Diagnosis and prediction: Initial diagnosis and prediction of future disease risk using logistic regression model based on input data.

3. Visualization of results: Present results using graphs and other visual aids to explain disease risk to the patient.

During the project, machine learning algorithms are fed and optimized with medical data, which increases the accuracy of the platform. Training and testing are used to ensure the accuracy of the data and to fine-tune the models. At the same time, the security and privacy issues of the platform are also taken into account, as medical information is considered confidential and important.

This system helps health workers, saves their time and increases the possibility of early detection of diseases.

### **Setting the issue:**

There are different algorithms for machine learning. They are mainly used for data analysis and forecasting. One of the most famous of them is the following method.

**Logistic regression** is a statistical and machine learning technique that is often used in binary classification tasks. That is, with the help of this model, we can classify an object or event from two possible categories (for example, "yes" or "no", "0" or "1", "has a disease" or "does not have a disease").

Basic principles:

Logical regression mathematically uses a linear model, which means that the relationship between input features and output is linear. However, the result is only represented by values between 0 and 1. This is done by the "sigmoid" or "logistic" function.

The sigmoid function looks like this:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

In this case:  $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$  is a linear combination of the model (the features and their weights).

$e$  - is the base of the natural logarithm (about 2.718).

$\sigma(z)$ - is the result of the sigmoid function and has a value between 0 and 1.

### **The process of logistic regression:**

**Data preparation:** Based on the initial data set, the input features and output target for each observation should be clear.

**Building the model:** The model is built by training the input features and their associated weight coefficients.

**Prediction:** After training the model, classification is performed based on the new data.

### **Why Logical Regression?**

Simple and understandable: As it is based on linear models, the results are easy to understand and interpret.

Classification is very convenient: Logistic regression is very effective if the goal is binary classification.

Easy interpretation: It is very easy to understand the results of the model, that is, to determine how each feature affected the classification.

### **Problem:**

If you have a data set called "Disease presence" and you know various risk factors (such as age, physical activity level, dietary habits), you can use logistic regression to predict the presence of this disease. you can predict whether or not.

### **Formula:**

The following formula is used to calculate the logistic regression result:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

Here:

- $P(y=1|X)$  - the probability of getting a result of "1" based on certain features X.
- $\beta_0, \beta_1, \dots, \beta_n$  - are parameters of the model, we find them in the process of training.

### **Fields of application**

- **Medicine:** Predicting the presence or absence of disease.
- **Marketing:** Determining whether a customer will buy a product or not.
- **Financial sector:** Classification of loan repayment or loan borrowing.

While logistic regression is a simple and efficient model, it also works well for specific and complex data.

### **Used Literature**

- [1]. **Hastie, T., Tibshirani, R., & Friedman, J.** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- [2]. **James, G., Witten, D., Hastie, T., & Tibshirani, R.** (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- [3]. **Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer.
- [4]. **Aggarwal, C. C.** (2018). *Machine Learning for Healthcare*. Springer.