



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

STROKE PREDICTION USING MACHINE LEARNING

Dr. Harish B.G* and Noorul Huda Khanum

Department of Master of Computer Applications, University BDT College of Engineering,
Visvesvaraya Technological University, Davanagere. Karnataka, INDIA

Abstract:

A Stroke is a health condition that causes damage by tearing the blood vessels in the brain. It can also occur when there is a halt in the blood flow and other nutrients to the brain. According to the World Health Organization (WHO), stroke is the leading cause of death and disability globally. Most of the work has been carried out on the prediction of heart stroke but very few works show the risk of a brain stroke. With this thought, various machine learning models are built to predict the possibility of stroke in the brain.

Stroke is a destructive illness that typically influences individuals over the age of 65 years age. Prediction of stroke is a time consuming and tedious for doctors. Therefore, the project mainly aims at predicting the chances of occurrence of stroke using the emerging Machine Learning techniques. Aim is to create an application with a user-friendly interface which is easy to navigate and enter inputs.

Introduction:

“The prime objective of this project is to construct a prediction model for predicting stroke using machine learning algorithms. The dataset was obtained from "Healthcare dataset stroke data".

With the advancement of technology in the medical field, predicting the occurrence of a stroke can be made using Machine Learning. The algorithms present in Machine Learning are constructive in making an accurate prediction and give correct analysis. The works previously performed on stroke mostly include the ones on Heart stroke prediction. Very less works have been performed on Brain stroke. This paper is based on predicting the occurrence of a brain stroke using Machine Learning. The key components of the approaches used and results obtained are that among the five different classification algorithms used Naïve Bayes has best performed obtaining a higher accuracy metric. The limitation with this model is that it is

being trained on textual data and not on real time brain images. The paper shows the implementation of six Machine Learning classification algorithms. This paper can be further extended to implementing all the current machine learning algorithms. A dataset is chosen from Kaggle with various physiological traits as its attributes to proceed with this task.

These traits are later analyzed and used for the final prediction. The dataset is initially cleaned and made ready for the machine learning model to understand. This step is called Data Preprocessing. For this, the dataset is checked for null values and fill them. Then Label encoding is performed to convert string values into integers followed by one-hot encoding, if necessary. After Data Preprocessing, the dataset is split into train and test data. A model is then built using this new data using various Classification Algorithms. Accuracy is calculated for all these algorithms and compared to get the best-trained model for prediction. After training the model and calculating the accuracy, an HTML page and a Flask application are developed. The web application is for the user to enter the values for prediction. The flask application is a framework that connects the trained model and the web application. After proper analysis, the paper concludes which algorithm is most appropriate for the prediction of stroke.

Here in this application I am going to use three modules namely, admin, users, patients. Admin is the module in this application and have the facility to view the list of users. Users are the main module of this application they can register to the application. Mainly users can predict stroke can occur or not, by filing some parameters. Patients can login and book the appointment.

Literature Survey:

Geethanjali et al, In their paper, stroke attack can be predicted accurately. They have used three classifiers such as logistics regression, Decision tree classifier and support vector classifier for the prediction of stroke. The classification model is based on 5110 records. And they have provided the result with 95.49% of accuracy. **Veena potdar et al**, In their paper they have described different methods of prediction of stroke and concluded that each technique for prediction of stroke has its own advantages and disadvantages. One must perform statistical analysis and initialization to decide on the specific technique to use. However, they suggested random forest technique is considered as one of the most precise technique for analyzing which shows promising result. **ShraddhaMainali et al**, In their paper they have described the stroke diagnosis and outcome prediction using supervised and unsupervised machine learning types. **Ankitha et al**, In their paper they have used machine learning or AI intelligence algorithm to predict the type of stroke a patient is suffering through an application where admin, user and interaction with particular doctor come into play. And finally gives the result for the user which type of stroke he is suffering. **Vida Abedi et al**, In their paper they have used artificial intelligence for improving stroke diagnosis in emergency department. They have mentioned the key step for stroke ML-enabled decision support system

for EDs. **Harshitha et al**, In this paper, they have constructed a model for predicting stroke using machine learning algorithms and given the accuracy for each model.

Proposed System:

The proposed system acts as a prediction support machine and will prove as an aid for the user with diagnosis. The algorithms used to predict the output have potential in obtaining a much better accuracy than the existing system. In proposed system, the practical use of various collected data has turned out to be less time consuming.

Advantages:

1. High performance and accuracy rate.
2. Data and information collected for prediction is easily available to the users.

Tools and Technologies used:

Visual Studio:

A built-in improvement surroundings (IDE) is a feature-rich software that helps many components of software program development. The Visual Studio IDE is a innovative launching pad that you can use to edit, debug, and construct code, and then put up an app. Over and above the widespread editor and debugger that most IDEs provide, Visual Studio consists of compilers, code completion tools, graphical designers, and many extra facets to decorate the software program improvement process.

A built-in enhancement environment (IDE) is a feature-rich utility that helps many elements of software program software development. The Visual Studio IDE is a modern launching pad that you can use to edit, debug, and assemble code, and then put up an app. Over and above the today's editor and debugger that most IDEs provide, Visual Studio consists of compilers, code completion tools, graphical designers, and many higher elements to enhance the software program application enhancement process.

The preceding photo shows Visual Studio with an open undertaking that suggests key domestic home windows and their functionality: In Solution Explorer, at greater right, you can view, navigate, and manipulate your code files. Solution Explorer can aid put together your code by way of grouping the archives into options and projects. The central editor window, the vicinity you will in all possibility spend most of your time, suggests file contents. In the editor window, you can edit code or diagram a customer interface such as a window with buttons and textual content material boxes. In Git Changes at limit right, you can tune work objects and share code with others thru the utilization of mannequin manipulate utilized sciences like Git.

The previous photo indicates Visual Studio with an open challenge that suggests key home windows and their functionality: In Solution Explorer, at top right, you can view, navigate, and control your code files. Solution Explorer can assist prepare your code by way of grouping the archives into options and projects. The central editor window, the place you may likely spend most of your time, shows file contents. In the editor window, you can edit code or graph a person interface such as a window with buttons and textual content boxes. In Git Changes at decrease right, you can song work gadgets and share code with others with the aid of the usage of model manipulate applied sciences like Git.

HTML

Stands for Hyper Text Markup Language, which is the most broadly used language on Web to enhance net pages. HTML used to be created through Berners-Lee in late 1991 however "HTML 2.0" was once the first widespread HTML specification which used to be posted in 1995. HTML 4.01 was once a predominant model of HTML and it was once posted in late 1999. Though HTML 4.01 model is broadly used however presently we are having HTML-5 model which is an extension to HTML 4.01, and this model used to be posted in 2012. Tags outline factors and inform the browser how to show them (for example, they can inform the browser whether or not a thing is an photograph or a paragraph of text).

CSS:

Like HTML, CSS is now not a programming language. It's now not a markup language either. CSS is a fashion sheet language. CSS is what you use to selectively fashion HTML elements. CSS tutorial or CSS three tutorial offers fundamental and superior ideas of CSS technology. Our CSS tutorial is developed for novices and professionals. The main factors of CSS are given below: CSS stands for Cascading Style Sheet. CSS is used to diagram HTML tags. CSS is a extensively used language on the web.

SQLite3:

SQLite is a self-contained, high-reliability, embedded, full-featured, public-domain, SQL database engine. It is the most used database engine in the world. It is an in-process library and its code is publicly available. It is free for use for any purpose, industrial or private. It is essentially an embedded SQL database engine. Ordinary disk archives can be without problems study and write by way of SQLite due to the fact it does now not have any separate server like SQL.

The SQLite database file structure is cross-platform so that all and sundry can effortlessly reproduction a database between 32-bit and 64-bit systems. Due to all these features, it is a famous

preference as an Application File Format. History: It was once designed by way of D. Richard Hipp for the cause of no administration required for working a program. in August 2000.

As it is very light-weight in contrast to others like MySQL and Oracle, it is known as SQLite. Different variations of SQLite are launched considering 2000. Installation on Windows: 1. Visit the reliable internet site of SQLite for downloading the zip file.

2. Download that zip file.
3. Create a folder in C or D (at any place you prefer) for storing SQLite with the aid of increasing zip file.
4. Open the command instant and set the direction for the place of SQLite folder given in the preceding step. After that write “sqlite3” and press enter.

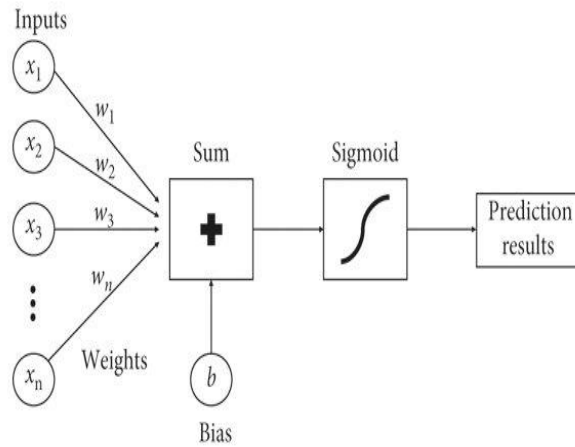
Python:

Python has a clean syntax corresponding to the English language. Python has syntax that approves builders to write down programs with fewer strains than a few awesome programming languages. Python runs on an interpreter system, which means that code may also be carried out as fast as it is miles written. Python is a high-degree object-orientated programming language that modified into as quickly as created with the aid of Guido van Rossum. It is in addition referred to as general-reason programming language as it is miles utilized in nearly each and every location we will assume of as stated below: Web Development Software Development Game Development AI & ML Data Analytics.

Method used in this Project:

Logistic Regression:

This algorithm is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. This algorithm is based on Supervised Learning technique.



This method is used to calculate or predict the probability of binary (yes/no) event occurring. An example of this method could be applying machine learning if a person is likely to be infected with stroke or not.

This Logistic Regression make prediction only within the Range of Data. In other words, we don't know whether the shape of the curve changes. If it does, our prediction will be invalid. The graph shows that the observe **BMI (Body Mass Index)** values range. We should not make prediction outside of the given range.

Result:

As mentioned above, about the method logistic regression used in this project which basically works on the given BMI value range. In this project BMI value range is given to predict the type of stroke. If the BMI value is >30 and <35 then it is a "Ischemic Stroke", if BMI value is >35 and <40 then it is a "Hemorrhagic Stroke" and if the BMI value is >40 then it is "Transient Ischemic Stroke". And array known as disease is created to which if the patient has heart disease and hypertension will append to the array disease along with the type of stroke the person is suffering. This result is fetched by inserting attribute information to gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, smoke_status and bmi. After predicting the type of stroke the user can enter the phase of the stroke to know what activities or exercise should be done for the type of stroke for that particular phase. This project gives 95% accuracy for imbalance dataset.

Conclusion:

The importance of knowing and understanding the risks of brain stroke is very much in these trying times. The model predicts the probability of brain stroke on the basis of very trivial day-to-day and known to all parameters. This makes this project highly relevant and of need to society. The objective of implementing the project on a web platform was to reach as many individuals as possible. The early warning can save someone's life who might have a probability of a stroke.

References:

1. Veena Potdar, LavanyaSanthosg, Yashu Raj Gowda CY, A survey on stroke disease classification and prediction using machine learning algorithms,IJERT,Vol. 10, 2021.
2. T.M, Geethanjali, Divyashree M.D, Monisha S.K, Sahana M.K, Stroke prediction using machine learning, JETIR,2349-5162,2014.
3. Ankitha S,Jhanvi V, Harshavardhan N, Deepthi M, Remanth M, An artificial approach for predicting different types of stroke, IJert,ISSN:2278-0181,Vol. 8 Issue 12, 2019.
4. ShraddhaMainali, Marin E. Darsie and Keaton S. Smetana, frontiers in Neurology (fneur), 2021.
5. Vida Abedi, Ayesha khan, Using AI for improving stroke diagnosis in emergency departments: a practical framework, The AdvNeuroIDisord. Vol. 13, 2020.
6. Harshitha K V, Harshitha P, Stroke prediction using machine learning algorithms, IJIREM, ISSN:2350-0557,Volume-8, Issue-4, 2021.
7. <https://machinelearningmastery.com/machine-learning-in-python-step-by-step/>
8. <https://www.coursera.org/learn/machine-learning-with-python>
9. https://www.tutorialspoint.com/machine_learning_with_python/index.htm
10. https://www.w3schools.com/python/python_ml_getting_started.asp
11. <https://www.edx.org/course/machine-learning-with-python-a-practical-introduct>

