DETECTION OF LANDSLIDE USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

JEEVAN VIJAY

REG NO:TKM19MCA013

In partial fulfillment of the requirements for the award of the Degree of

MASTER OF COMPUTER APPLICATIONS

Thangal Kunju Musaliar College of Engineering
Kerala

DEPARTMENT OF COMPUTER APPLICATIONS

MAY 2022

DECLARATION

I undersigned hereby declare that the project report ”DETECTION OF LANDSLIDE USING MACHINE LEARNING”, submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Prof. Dr. Fousia M Shamsudeen . This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place: Kollam

Date: 12-05-22

Jeevan Vijay

Thangal Kunju Musaliar College of Engineering
Department of Computer Application

C E R T I F I C A T E

This is to certify that, the report titled "**DETECTION OF LANDSLIDE USING MACHINE LEARNING"** is a bonafide record of the project work presented by **JEEVAN VIJAY(TKM19MCA013)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree **Master of Computer Applications** in **APJ Abdul Kalam Technological University**.

Internal Supervisor | Head of the Department | External Examiner

ACKNOWLEDGEMENT

First and foremost I thank GOD almighty and my parents for the success of this project. I owe sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

I would like to thank **Dr. Fousia M Shamsudeen**,Assistant Professor and Head of the Department,MCA,TKMCE, for her constant support and encouragement throughout the project work.

With a profound sense of gratitude,I would like to express my heartfelt thanks to my guide **DR.Fousia M Shamsudeen**Assistant Professor,MCA,TKMCE, for her expert guidance, cooperation and immence encouragement. I also extend my thanks to the entire faculty and staff of the Department of Computer Application,TKMCE, who has encouraged me throughout this work.

JEEVAN VIJAY

ABSTRACT

Landslides occur when large amounts of earth, rock, sand or mud flows swiftly down hill and mountain slopes.This project proposes a novel machine-learning method to identify the reason of landslides using global landslide dataset.The five machine learning algorithm, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Boosting Method and Decision Tree are utilized and evaluated on global landslide dataset.An Ensemble Method is also used to evaluate the global landslide dataset.From the result, Ensemble method gives an accuracy of 90% . Random Forest(RF) comes in second with 89% accuracy.By using machine learning technique ,the proposed landslide reason identification shows outstanding robustness and great potential in tackling the landslide reason identification problem.

Contents

List of Figures

Chapter 1

INTRODUCTION

The movement of rock, soil, or debris down a sloped area of terrain is known as a landslide. Rain, earthquakes, volcanoes, and other elements that make the slope unstable produce landslides.Many parts of the world are affected by landslides. There are numerous different types of landslides, which can occur in a variety of materials and travel at different speeds. Earthquakes, rain, permafrost melt, deforestation, and other events can all cause landslides. Landslides often exacerbate destructive wildfires, floods, volcanic eruptions, and earthquakes. Tsunamis can be triggered by submarine landslides or surface landslides that move into water. Despite the fact that landslides can cause major human and economic losses, they also help to sustain natural diversity.Therefore landslide identification plays an important role in landslide risk assessment and management. With the advent of the remote sensing technology, landslides can be identified through visual interpretation of both remote sensing images and topographic surfaces. Although the visual interpretation has high identification accuracy, the process is time-consuming and labor-intensive. Hence, automated or semi-automated methods for landslide identification based on remote sensing techniques are highly sought after in recent years.Current studies on landslide identification are mainly based on optical images using pixel-based or object-oriented methods, and the digital terrain model (DTM) is often used as auxiliary data for such analysis.Combined optical images and digital elevation model (DEM) derivatives to identify translational landslide scars using object-oriented methods.
Machine learning technique have been proven to be a powerful and promising tool in many geotechnical applications as well as in landslide identification. So this project proposes a novel machine-learning method to identify the reason of landslides using global landslide dataset with five ma-

chine learning algorithm, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Boosting Method and Decision Tree are utilized and evaluated on global landslide dataset.An Ensemble Method is also used to evaluate the global landslide datset.

1.1 Problem Definition

A landslide's consequences can be severe, including loss of life, infrastructure ruin, land degradation, and natural resource depletion.The current system identifies the landslides but not identifies the reason for the landslide.Landslides may occur based on several reasons like earthquake, mining,rain,construction,flooding,tropical cyclone etc.So identify the reason for the landslide is very important.This helps to take precautions against the landslide.

1.2 Objective

The main goal of project is:

- To identify the reason for landslide.
- To train dataset with different machine learning algorithm and ensemble method are used.
- Compare the accuracy and find the best machine learning algorithm.
- This helps to take precautions against landslides

Chapter 2

Related Works

A literature review is a complete examination and analysis of literature on a certain topic. When research questions are identified through a literature review, one seeks to answer them by looking for and analysing relevant material. Literature reviews are important because they are a complete study and analysis of literature that relates to a specific issue. When research questions are identified through a literature review, one seeks to answer them by looking for and analysing relevant material. Re-analyzing the study's results can lead to fresh discoveries, which is why literature reviews are important. A literature review summarises and explains the whole and current state of knowledge on a topic as found in academic books and journal articles. At university, there are two types of literature reviews: one that students are asked to write as a stand-alone assignment in a course, and another that is prepared as an introduction to, or preparation for, a larger work, usually a thesis or research report. The type of review you are writing will affect the focus and perspective of your review, as well as the type of hypothesis or thesis argument you make. Reading published literature reviews or the introductory chapters of theses and dissertations in your own subject area is one approach to comprehend the differences between these two forms. Examine their arguments for structure and how they solve the challenges. By re-analyzing the study's findings, new insights can be gained. A literature review summarises and explains the whole and current state of knowledge on a topic as found in academic books and journal articles. At university, there are two types of literature reviews: one that students are asked to write as a stand-alone assignment in a course, and another that is prepared as an introduction to, or preparation for, a larger work, usually a thesis or research report. The type of review you are writing will affect the focus and perspective of

your review, as well as the type of hypothesis or thesis argument you make. Reading published literature reviews or the introductory chapters of theses and dissertations in your own subject area is one approach to comprehend the differences between these two forms. Examine their arguments for structure and how they solve the challenges.

[1]Haojie Wang a, Limin Zhang a,*, Kesheng Yin a, Hongyu Luo a, Jinhui Li b,Landslide identification using machine learning,Elseiver,Feb 2020.

Advantages:-

This paper identifies landslide with high accuracy.

Disadvantages:-

The relict landslide records of ENTLI are not completely accurate.

[2]Chen Chen a, Mo-wen Xie a, Yu-jing Jiang b, Bei-ning Jia c, Yan Dua,A new method for quantitative identification of potential landslide,Elseiver,2021.

Advantages:-

Safety factor (SF) can predict landslide Occurrence.

Disadvantages:-

Cannot identify potentially dangerous landslides.

[3]Haojie Cai, Tao Chen , Senior Member, IEEE, Ruiqing Niu, and Antonio Plaza, Fellow, IEEE,Landslide Detection Using Densely Connected Convolutional Networks and Environmental Conditions, IEEE Journal Of Selected Topics in Applied Earth Observations and Remote sensing ,IEEE ,May 2021.

Advantages:-

The proposed method can accurately identify landslides in a large area with fewer training samples.

Disadvantages:-

Comparing DenseNet with traditional ML algorithms, when there are scarce samples in practical applications, the performance of ML will be very poor.

[4]BISWAJEET PRADHAN 1,2, (Senior Member, IEEE), HUSAM A. H. AL-NAJJAR1, MAHER IBRAHIM SAMEEN1, MUSTAFA RIDHA MEZAAL3, AND ABDULLAH M. ALAMRI4 1Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of En-

gineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia,Landslide Detection Using a Saliency Feature Enhancement Technique From LiDAR-Derived DEM and Orthophotos,IEE, July 2020.

Advantages:-

Detect landslides in the training area with a Prediction Accuracy.

Disadvantages:-

Accuracy is less.

[5] Shuai Chen, Zelang Miao , Member, IEEE, Lixin Wu, and Yueguang He,Application of an Incomplete Landslide Inventory and One Class Classifier to Earthquake-Induced Landslide Susceptibility Mapping,2020.

Advantages:-

can be applied to determine regional landslide susceptibility after earthquakes and provide an essential reference for emergency response.

Disadvantages:-

The condition that obtaining a complete landslide inventory after an earthquake is difficult.

[6]BINGXIN SHI1,3,4, TING ZENG 2, CHUAN TANG1, LIFANG ZHANG5, ZHUOJUAN XIE5, GUOJUN LV6, AND QIHONG WU7,Landslide Risk Assessment Using Granular Fuzzy Rule-Based Modeling: A Case Study on Earthquake-Triggered Landslides,Sept 2021.

Advantages:-

To capture the main essence of landslide pattern with higher interpretability and help to reduce the computing overhead.

Disadvantages:-

Time consuming.

Chapter 3

DETECTION OF LANDSLIDE USING MACHINE LEARNING

Detection of landslide using machine learning is used to identify the reason for landslides.There are different types of landslides like rainfall,flooding,earthquake etc.The landslide is caused by the natural changes like rain,earthquake etc and human activities like construction,mining etc.There are many techniques used to identify landslides.Here the machine learning technique is used to identify the reason for the landslides.

3.1 Challenges of Detection of Landslide

Identification of landslides is essential for risk assessment and mitigation.Identify the reason for landslide is the main task.The current system identifies the landslides using machine learnng and deep learning techniques but not identify the reasons for happening landslides.So this project is to identify the reason for landslide using machine learning technique.Here five machine learning algorithms used including Random Forest(RF),Support Vector Machine(SVM), Boosting Method,Decision Tree and Logistic Regression.Ensemble method is used to increase the accuracy.

Chapter 4

METHODOLOGY

Landslides occur when large amounts of earth, rock, sand or mud flows swiftly down hill and mountain slopes.Usually triggered by natural hazards such as earthquakes, volcanic eruptions, heavy rain storms or cyclones, is increasing due to modern land-use practices, climate change and deforestation. The impact of a landslide can be extensive, including loss of life, destruction of infrastructure, damage to land and loss of natural resources.The existing system identifies the landslide.There are different types of landslides and landslides may occur based on several reasons like earthquake,flooding,mining,rain etc.

4.1 System Architecture

The main purpose of the project is to identify the reason of landslides using global landslide dataset.The five machine learning algorithm, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Boosting Method and Decision Tree are utilized and evaluated on global landslide dataset.An Ensemble Method is also utilized and evaluated on global landslide data set to increase the accuracy. The figure 3.1 below shows system architecture. The proposed system consist of four major phases:

- Data preprocessing and data augmentation
- Training the data set
- Testing the data set

- Deploying the model.

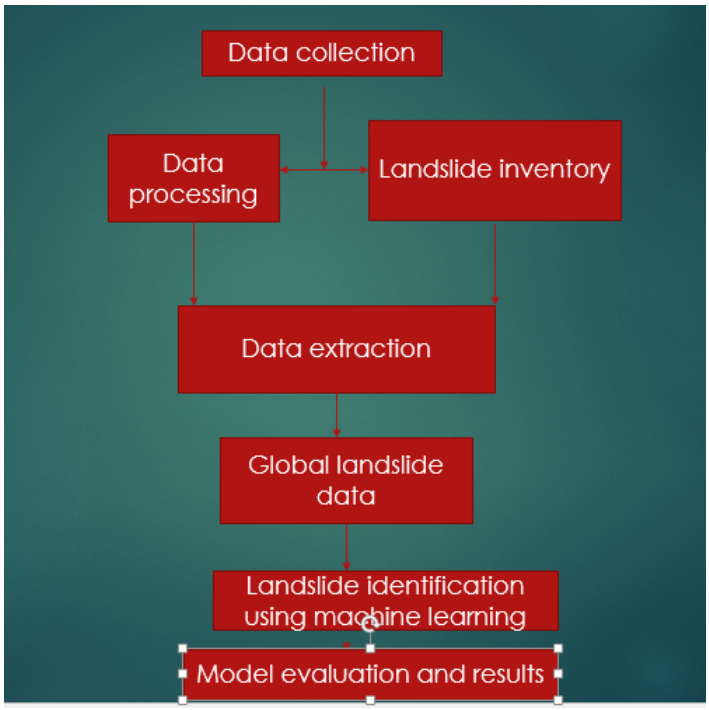


Figure 4.1: System Architecture

4.1.1 Dataset

The dataset used for landslide reason identification is based global landslide dataset which is taken from the github reposiory (link:- https://github.com/kbachler/Landslide-Trigger-Classification/blob/da040ebe2e6

Import libraries

NumPy:-

Python's NumPy package. 'Numerical Python' is what it stands for. It is a library that contains multidimensional array objects as well as a collection of array processing routines. Numeric, NumPy's forerunner, was created in the 1980s. JimHugunin's contribution Numarray, a package with some features, was also created additional capabilities Travis Oliphant wrote the NumPy package in 2005 adding Numarray's functionality into the Numeric package There are numerous this open source project's contributions .Using NumPy, a developer can perform the following operations: 1. Mathematical and logical operations on arrays. 2. Fourier transforms and routines

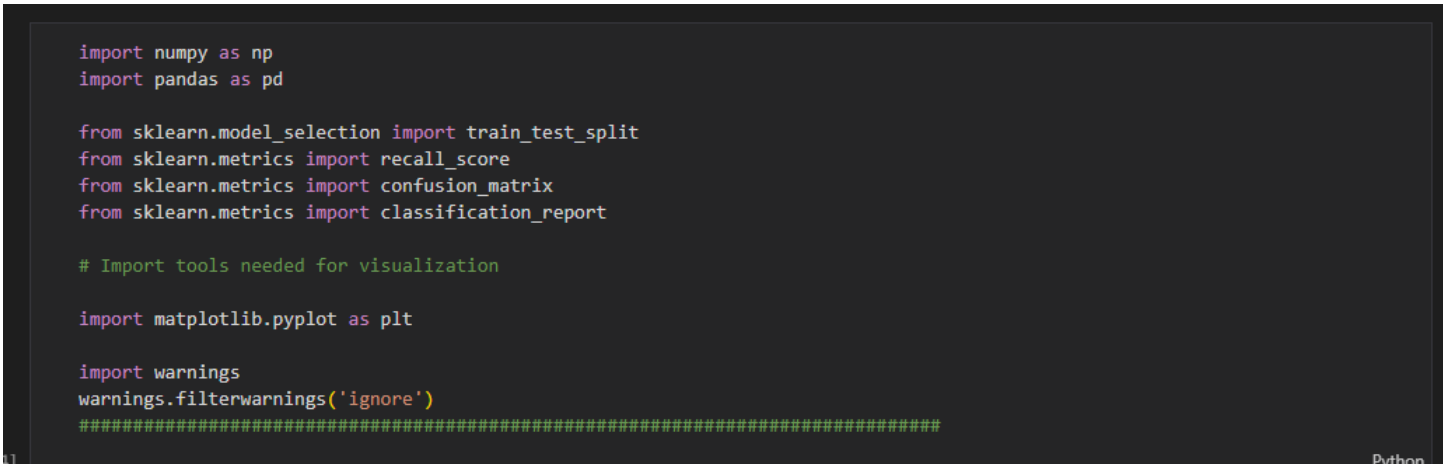
for shape manipulation 3. Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

PANDAS:-

Pandas is a BSD-licensed open-source Python library that provides high-performance, user-friendly data structures and data analysis tools for the Python programming language. Python with Pandas is utilised in many different domains, including academia and commercial fields like as finance, economics, statistics, analytics, and so on.In this lesson, we will learn about Python Pandas' numerous capabilities and how to use them in action.Pandas deals with the following three data structures : 1. Series 2. DataFrame 3. Panel

MATPLOITLIB:-

It is a set of command-style functions that make matplotlib behave similarly to MATLAB.Each pyplot function alters a figure in some way: it generates a figure, a plotting area in a figure, plots some lines in a plotting area, and decorates the plot with labels and so on Various states are retained between functions in matplotlib.pyplot calls, such that it remembers things like the current figure and plotting area, and The charting functions are aimed at the current axes (notice the word "axes" here) and in most places in the documentation refers to the axes of a figure rather than the figure itself the technical phrase for more than one axis).



```
import numpy as np
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.metrics import recall_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report

# Import tools needed for visualization

import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings('ignore')
################################################################################
```

Figure 4.2: libraries

SKLEARN:-

Scikit-learn is a Python machine learning library. SVMs, gradient boosting, kmeans, random

forests, and DBSCAN are among the regression, classification, and clustering techniques included. It is intended to be used with Python Numpy as well as Scipy.The scikit-learn project began as a Google Summer of Code (also known as GSOC) initiative. Scikits is a Google Summer of Code (GSoC) project led by David Cournapeau learn. It takes its name from ”Scikit,” a third-party extension to.

4.1.2 Data cleaning

Format, clean, and sample your selected data to organise it.The following are the common data pre-processing steps: Formatting: The data you have chosen may not be in a format suitable for to collaborate with. You may have data in a relational database and would like to access it.You would like the data in a flat file or in a proprietary file format either in a relational database or as a text file.

Cleaning: Cleaning data is the process of removing or replacing missing data. There may be data instances that are incomplete and do not include the information you assume you require to solve the problem. These instances may require removal. Additionally,Some of the properties may contain sensitive information, and these attributes may need to be completely eliminated from the data.Label encoding is also employed to sanitise the data in this case.

4.1.3 Data Preprocessing

In this module the cleaned dataset will be preprocessed where the dataset will be divided based on amount and transaction time.Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data preprocessing is a technique for converting raw data into a clean data set. Pre processing is a pre-built functionality in the scikit-learn library in Python. There are other other pre-processing alternatives that we will investigate.

4.1.4 Training the dataset

After completing the cleaning of dataset,then the dataset is used for training.In this project 80% of data is used for training.there are approximately 7304 rows and 180 features are used for training from the global landslide dataset.

```
df[ 'landslide_trigger' ] = df[ 'landslide_trigger' ].cat.codes
df.head(5)
```

[18]

	landslide_trigger	event_month	fatality_count	injury_count	population	longitude	latitude	event_time_0:00	event_time_0:30	event
0	0	7	0.0	0.0	11886.0	145.1147	-37.8558	0	0	
1	0	8	0.0	0.0	112050.0	92.7490	11.6753	0	0	
2	0	9	0.0	0.0	193750.0	28.7701	41.0863	1	0	
3	0	8	2.0	4.0	601600.0	73.0399	33.6941	0	0	
4	0	5	8.0	0.0	2628.0	44.5015	42.7204	0	0	

Figure 4.3: cleaned data

4.1.5 Testing the dataset

After dataset is cleaned 20% of the data is used for testing.There are approximately 1826 rows and 180 features are used for testing from the global landslide dataset

4.1.6 Deploying the model

In this project Five machine learning model including Random Forest(RF),Support Vector Machine(SVM),Boosting Method,Decision Tree and Logistic Regression .An Ensemble method is also build here by combining three algorithms like Random Forest,Decision Tree and Logistic Regression.

4.2 Algorithms

1.Random Forest(RF)

Random Forest is a well-known machine learning algorithm from the supervised learning technique. It can be applied to both classification and regression issues in machine learning. It is built on the notion of ensemble learning, which is a method that involves integrating several classifiers to solve a complex problem and improve the model's performance.

"Random Forest is a classifier that contains a number of decision trees on various subsets of the provided dataset and takes the average to enhance the predicted accuracy of that dataset," as the name implies. Instead than relying on a single decision tree, the random forest collects the forecasts

from each tree and predicts the final output based on the majority vote of predictions. The greater the number of trees in the forest, the higher the accuracy and the lower the risk of overfitting.

Advantages of Ranfom Forest algorithm

a. Random Forest can perform both classification and regression tasks.

b. It produces good predictions that can be easily understood.

c. It can handle huge amount of dataset efficiently.

d. It provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.

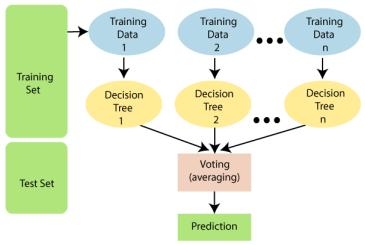


Figure 4.4: Working of Random Forest

2.Support Vector Machine(SVM) Vector Machine, or SVM, is a prominent Supervised Learning technique that is used for both classification and regression issues. However, it is mostly utilised in Machine Learning for Classification difficulties.

The SVM algorithm's purpose is to find the optimum line or decision boundary for categorising n-dimensional space so that we may easily place fresh data points in the correct category in the future. A hyperplane is the optimal choice boundary. SVM selects the extreme points/vectors that aid in the creation of the hyperplane. These extreme examples are referred to as support vectors, and the method is known as the Support Vector Machine.

Advantages of Support Vector Machine

a.When there is an understandable margin of dissociation between classes, the support vector machine performs reasonably well.

b. It is more efficient in high-dimensional spaces.

c.It is useful when the number of dimensions is greater than the number of specimens.

d.Support vector machine is memory systematic in comparison.

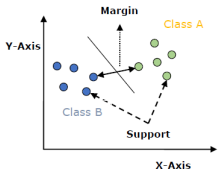


Figure 4.5: Working of SVM

3.Boosting Method

Boosting Methods gives machine learning models the ability to improve their prediction accuracy. XGBoost is a more advanced variant of the Gradient Boosting approach that literally means ”eXtreme Gradient Boosting.”

Advantages of Boosting Method

a. Implementation Ease.

b. Bias reduction.

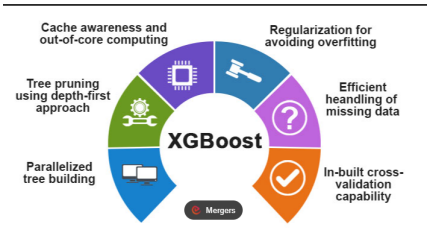


Figure 4.6: Working of Boosting Method

4.Logistic Regression(LR)

Logistic regression is a common Machine Learning method that belongs to the Supervised Learning technique. It is used to forecast the categorical dependent variable from a group of independent variables. A categorical dependent variable's output is predicted using logistic regression. As a result, the outcome must be categorical or discrete. It can be Yes or No, 0 or 1, true or False, and so on, but instead of presenting the exact values like 0 and 1, it presents the probability values that fall between 0 and 1. Except for how they are employed, Logistic Regression and Linear Regression are very similar. Logistic regression is used to solve classification difficulties, whereas linear regression is used to solve regression problems.Instead of fitting a regression line, we fit a "S" shaped logistic function that predicts two maximum values in logistic regression (0 or 1). The logistic function curve reflects the possibility of something such as whether the cells are cancerous or not, whether a mouse is obese or not based on its weight, and so on. Logistic Regression is an important machine learning approach because it can generate probabilities and classify new data using both continuous and discrete datasets.

Advantages of Logistic Regression

a. Logistic Regression is one of the most basic machine learning algorithms, and it is simple to construct while providing excellent training efficiency in some circumstances.

b. The projected parameters (trained weights) infer the significance of each characteristic. It also specifies if the relationship is positive or negative.

c. This algorithm allows models to be updated easily to reflect new data, unlike decision trees or support vector machines. The update can be done using stochastic gradient descent.

d.Logistic Regression produces calibrated probability as well as classification results. This is superior to models that merely provide the final categorization as results.

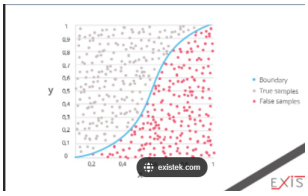


Figure 4.7: Logistic Regression

5.Decision Tree

Decision Tree is a Supervised learning technique that may be used for both classification and regression issues, but it is most commonly employed for classification. It is a tree-structured classifier in which internal nodes contain dataset attributes, branches represent decision rules, and each leaf node represents the result. A Decision tree has two nodes: the Decision Node and the Leaf Node.Decision nodes are used to make any decision and have several branches, whereas Leaf nodes are the result of those decisions and have no additional branches. The decisions or tests are based on the characteristics of the given dataset. It is a graphical representation of all possible solutions to a problem/decision given certain parameters. It is named a decision tree because, like a tree, it begins with the root node and then branches out to form a tree-like structure.

Advantages of Decision Tree

a.Comprehensive: It considers every possible outcome of a decision and follows each node to its conclusion.

b.Decision Trees assign a numerical value to each problem, decision, and outcome (s). It decreases confusion and ambiguity while simultaneously increasing clarity.

c.Decision Tree is one of the simpler and more dependable algorithms since it lacks complex equations and data structures. Calculation requires only basic statistics and algebra.

d. Decision Trees are versatile because they may be built manually using math and then used with various computer programs. Decision Trees are versatile because they may be built manually using math and then used with various computer programs.

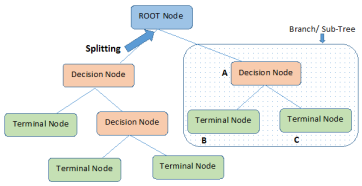


Figure 4.8: Decision Tree

6.Ensemble Method Ensemble Methods help to create multiple models and then combine them to produce improved results.In this project we can combine three algorithms like Random Forest,Logistic Regression,Decision Tree.

Advantages of Ensemble Method

a. A minimum benefit of utilising ensemble is that it reduces the spread in a prediction model's average skill. A minimum benefit of utilising ensemble is that it reduces the spread in a prediction model's average skill.

b. One significant advantage of using ensembles is that they enhance average prediction performance over every contributing member in the ensemble.

c. The reduction in the variance component of prediction mistakes made by the contributing models is frequently the mechanism for enhanced ensemble performance.

4.3 Tools Required

The software used for the project:

- Python
- Jupiter Notebook
- HTML
- CSS
- Java Script
- Flask
- SQlite

4.3.1 Python

Guido Rossum designed Python, an object-oriented programming language, in 1989. It is suitable for rapid prototyping of complex applications. It has interfaces to various OS system calls and

libraries and is difficult to program in C or C++. NASA, Google, YouTube, BitTorrent, and other large organisations employ the Python programming language. Python programming is frequently used in artificial intelligence, natural language generation, neural networks, and other sophisticated computer science domains. Python is a computer programming language.Guido van Rossum created an open supply, high-level artificial language.Python Software Foundation manages the project, which was founded in the late 1980s. It stemmed from the ABC language, which he helped create early on his professional life.Python is a strong programming language that can be used to create games, GUIs, and web applications. It's a sophisticated language. Reading and writing Python code is quite similar to reading and writing standard English sentences. As a result, they are not written in machine-readable code.Python programs must be digested before they can be executed by machines. Python is a well-known language. This means that each time a program is executed,Its interpreter goes through the code and converts it to machine-readable code.bytes of code Python is an object-oriented programming language that allows users to manage and control objects.To create and operate program, data structures or objects must be managed. Everything Python is, in reality, excellent. All objects, data types, functions, methods, and procedures Python gives equal weight to and classes. Programming languages are developed to meet the needs of programmers and users for an efficient tool for developing applications that have an impact on people's lives, lifestyles, economies, and societies. They help to improve people's life by increasing productivity, improving communication, and increasing potency. When languages fail, they die and become outdated. Languages that fail to live up to expectations are supplanted and superseded by languages that do are more potent. Python programming language is an artificial language has withstood the test of time and is still important throughout industries and enterprises, as well as among programmers and ordinary users. It's a living, vibrant thing as well as a useful language that is highly suggested as a primary Programming language for individuals who desire to learn and practise programming.

4.3.2 Jupyter Notebook

Jupyter Notebook is a free and open source web tool that allows you to create and share documents with live code, equations, visualisations, and text. People at Project Jupyter manage Jupyter Note-

book. Jupyter Notebooks are a fork of the IPython project, which used to have its own IPython Notebook project. Jupyter gets its name from the core programming languages it supports: Julia, Python, and R. Jupyter comes with the IPython kernel, which allows you to develop Python programmes, although there are presently over 100 alternative kernels available. Jupyter notebooks are designed to give a more user-friendly interface for code used in digitally aided research or instruction. . Jupyter notebooks, for example, are less relevant to study or teach about in isolation because they do little to immediately further research or instruction. Jupyter has gained popularity as an open-source environment that is compatible with a variety of programming languages. Jupyter is named after the three basic languages supported by the project (Julia, Python, and R), however kernels are available that make Jupyter compatible with dozens of others, including Ruby, PHP, Javascript, SQL, and Node.js.It may not make sense to use Jupyter notebooks to implement projects in all of these languages (for example, Omeka will not let you install a plugin written as a Jupyter notebook), but the Jupyter environment can still be useful for documenting code, teaching programming languages, and providing students with a space where they can easily experiment with provided examples. As of late 2019, there are two major environments for running Jupyter Notebooks: Jupyter Notebook (not to be confused with Jupyter Notebook files, which have an.ipynb extension) and the newer Jupyter Lab. Jupyter Notebook is a popular and well-documented programming environment that includes a simple file browser as well as an environment for creating, editing, and running notebooks. Jupyter Lab is more advanced, featuring a user interface that resembles an Integrated Development Environment (discussed in previous Programming Historian tutorials for Windows, Mac, and Linux).

The Jupyter Notebook is great for not just studying and teaching a programming language like Python, but also for sharing your data. You may make your Notebook into a slideshow or upload it to GitHub. Binder is a great way to share a Notebook without having your users to install anything. Google Colaboratory and Microsoft Azure Notebooks both feature their own versions of the Notebook that you may use to build and share your Notebooks. You can also look at some extremely intriguing Notebooks there. JupyterLab, the latest product from Project Jupyter, was recently released. JupyterLab integrates Jupyter Notebook into an Integrated Development Environment (IDE) that runs in your browser. JupyterLab can be thought of as a more complex version of Jupyter Notebook. In addition to Notebooks, JupyterLab allows you to run terminals,

text editors, and code consoles in your browser.

4.3.3 HTML

HTML stands for Hyper Text Markup Language, and it is used to create web pages and web applications. Let's define Hypertext Markup Language (HTML) and Web page.

HyperText is an abbreviation for "Text inside Text." A hypertext is a text that has a link. When you click on a link that takes you to a new webpage, you have used hypertext. HyperText is a method of connecting two or more web pages (HTML documents).

A markup language is a computer language used to apply style and formatting principles to text documents. Markup language enhances the interactive and dynamic nature of text. It can convert text into graphics, tables, links, and so forth.
A web page is a document that is typically authored in HTML and translated by a web browser. An URL can be used to locate a web page. A static or dynamic Web page can exist. We can design static web pages entirely with HTML. As a result, HTML is a markup language that is used to create attractive web pages with the help of styling that appear in a nice format on a web browser. An HTML document is made up of many HTML tags, each of which includes unique content.

4.3.4 CSS

CSS is an abbreviation for Cascading Style Sheets. It is a style sheet language that is used to describe the look and formatting of a markup language document. It adds a new functionality to HTML. It is commonly used in conjunction with HTML to alter the appearance of web pages and user interfaces. It can also be used with any type of XML document, such as plain XML, SVG, or XUL.

CSS is used in most websites, together with HTML and JavaScript, to construct user interfaces for web applications and many mobile applications.

4.3.5 JavaScript

is a lightweight programming language used by web developers to generate more dynamic interactions while creating web pages, applications, servers, and even games.

JavaScript is commonly used by developers in conjunction with HTML and CSS. In formatting HTML components, the scripting language works well with CSS. However, JavaScript retains user interaction, which CSS cannot achieve on its own.

JavaScript's implementations in web, mobile application, and game development make it worthwhile to understand the scripting language. You can do so by exploring free JavaScript templates and applications on code hosting platforms like GitHub or via learning platforms like BitDegree.

4.3.6 Flask

Flask is a Python-based web application framework. It is created by Armin Ronacher, who leads the Pocco international network of Python aficionados. Flask is built with the Werkzeug WSGI toolkit and the Jinja2 template engine. Both are Pocco initiatives. Python 2.6 or higher is typically used.Flask installation is required. Despite the fact that Flask and its dependencies operate well with Python 3,Many Flask extensions do not properly support it (Python 3.3 and later). As a result, it is advised.Python 2.7 should have Flask installed. Flask is a Python API that allows us to create web applications.Armin Ronacher came up with the idea. Flask's framework is more explicit than Django's framework, and it is also easier to learn because it requires less basic code to construct a simple web application. Flask is built on the WSGI (Web Server Gateway Interface) toolkit and the Jinja2 template engine. Flask was created to be simple to use and extend. The goal of Flask is to create a strong foundation for web applications of varying complexity You are then free to connect in any extensions you believe you may require You can also create your own modules. The flask is ideal for a variety of projects.It's very useful for prototyping. Flask is dependent on two third-party libraries: the Jinja2 template engine and the Werkzeug WSGI framework.
Flask has a lightweight and modular design, making it simple to adapt to the web framework of your choice required with a few extensions without clogging it up The documentation for Flask is extensive and complete a lot of examples and well-structured You may also test out some sample applications to get a feel for them from Flask Flask is extremely simple to deploy in production ("Flask is 100 percent WSGI 1.0 compatible"),High Flexibility in HTTP request handling functionality. The setup is now significantly more adaptable than Django, providing plenty of solutions for any production need.

4.3.7 SQlite

SQLite is a C-language library that provides a SQL database engine that is tiny, fast, self-contained, high-reliability, and full-featured. SQLite is the most widely used database engine on the planet. SQLite is incorporated into all mobile phones and most PCs, as well as innumerable other apps that users use on a daily basis. SQLite file format is stable, cross-platform, and backwards compatible, and the developers promise to maintain it that way until 2050. SQLite database files are commonly used as containers for rich content transmission across computers and as a long-term data storage format. There are currently over 1 trillion (1e12) SQLite databases in use. SQLite source code is in the public domain and can be used for any purpose.

Chapter 5

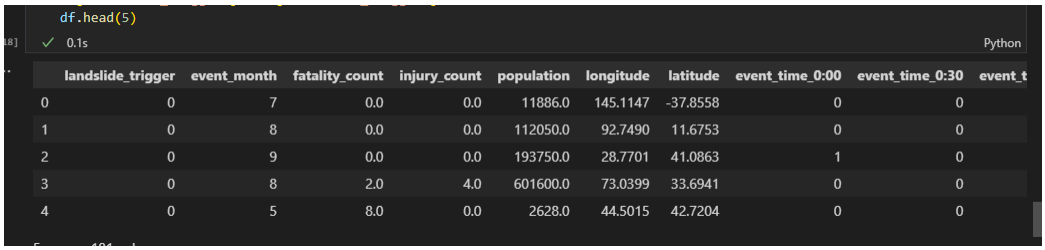
EXPERIMENTAL SETUPS AND RESULTS

This project identify the reason for the landslide with machine learning.The five machine learning algorithm, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Boosting Method and Decision Tree are utilized and evaluated on global landslide dataset.An Ensemble Method is also used to evaluate on global landslide dataset.From the result, Ensemble method gives an accuracy of 90%.Random Forest comes in second with 89% accuracy followed by Boosting Method gives accuracy of 87%,SVM gives accuracy of 86% ,Decision Tree gives accuracy of 85% and Logistic Regression gives accuracy of 75%.

Data preprocessing

	landslide_trigger	event_id	formatted_date	event_date	event_month	event_day	event_time	event_title	event_description	location_
0	Construction	7544	7/14/2015 23:00	7/14/2015	7	14	23:00	Highbury Rd and Huntingdale Rd	WITNESSES have detailed the dramatic moment th...	Hight Hun
1	Construction	7616	8/6/2015 0:00	8/6/2015	8	6	9:00	Atlanta Point	The PBMC Chairperson, Mr. Armugam has made an ...	A
2	Construction	9558	9/27/2016 0:00	9/27/2016	9	27	0:00	Basaksehir district	People had been working both at the top and at...	Basaks
3	Construction	6480	8/12/2014 0:00	8/12/2014	8	12	2:30	the metro bus flyover project in Sector G9	The labourers were busy in work on the metro b...	the metro project i
4	Construction	6030	5/17/2014 0:00	5/17/2014	5	17	9:00	Dariali Gorge	A landslide in the Dariali Gorge in Georgia ha...	D

Figure 5.1: Dataset before preprocessing



```
df.head(5)
```

0.1s Python

	landslide_trigger	event_month	fatality_count	injury_count	population	longitude	latitude	event_time_0:00	event_time_0:30	event_t
0	0	7	0.0	0.0	11886.0	145.1147	-37.8558	0	0	
1	0	8	0.0	0.0	112050.0	92.7490	11.6753	0	0	
2	0	9	0.0	0.0	193750.0	28.7701	41.0863	1	0	
3	0	8	2.0	4.0	601600.0	73.0399	33.6941	0	0	
4	0	5	8.0	0.0	2628.0	44.5015	42.7204	0	0	

Figure 5.2: Dataset after preprocessing

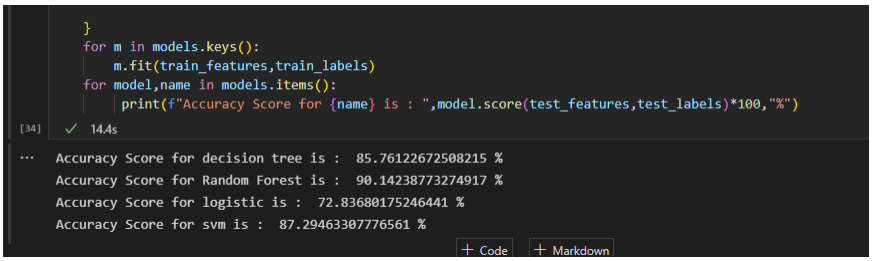


Figure 5.3: Accuracy of 4 algorithms

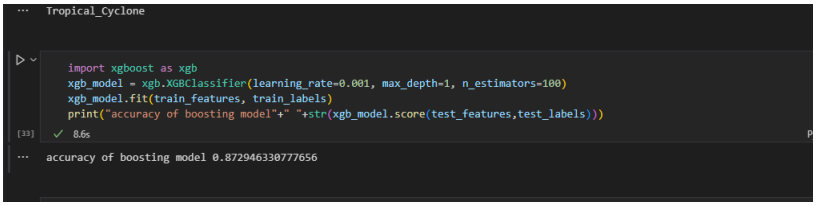


Figure 5.4: Accuracy of boosting method

5.1 Performance metrics for validation phase

The below figure indicates the performance metrics to calculate accuracy based on the classification results. The most important performance matrix is precision.Precision is defined as the proportion of relevant examples found among the recovered instances of categorization procedures.

$$Precision = TP/(TP+FP) \text{ -(1)}$$

Recall is also one of the most commonly used performance indicators for estimating classification model performance. The recall is the percentage of the total number of relevant instances that were actually recovered by the categorization method.

$$Recall = TP/(TP+TN) \text{ -(2)}$$

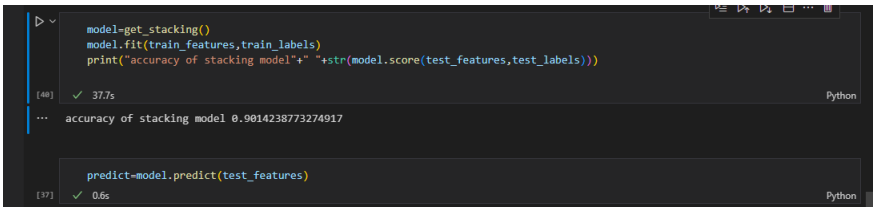


Figure 5.5: Accuracy of ensemble method

The f1 score is a measure of testing accuracy in binary classification statistical analysis. To compute the f1 score, it takes into account both the precision and recall of the classification system.

$$F1score = precision * Recall/(precision + Recall) \text{ -(3)}$$

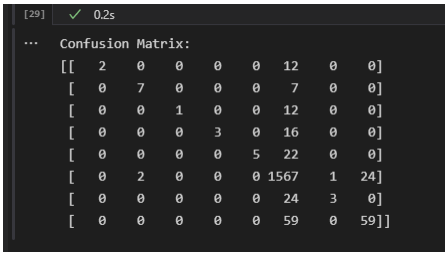


Figure 5.6: confusion matrix of random forest

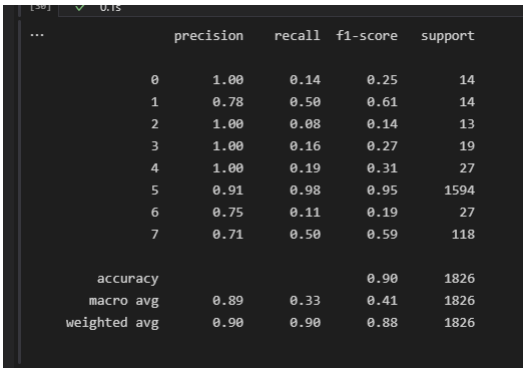


Figure 5.7: classification report of random forest

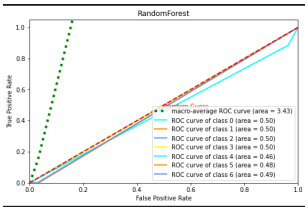


Figure 5.8: Graph of Random Forest

...	precision	recall	f1-score	support
0	0.67	0.14	0.24	14
1	0.78	0.50	0.61	14
2	0.00	0.00	0.00	13
3	1.00	0.16	0.27	19
4	0.33	0.04	0.07	27
5	0.91	0.98	0.95	1594
6	0.50	0.15	0.23	27
7	0.70	0.58	0.63	118
accuracy			0.90	1826
macro avg	0.61	0.32	0.37	1826
weighted avg	0.88	0.90	0.88	1826

Figure 5.9: classification report of ensemble method

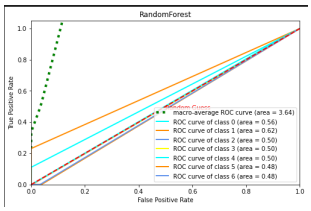


Figure 5.10: Graph of Ensemble Method

5.2 Graphical User Interface

In this project a Graphical User Interface(GUI) is created.The front end of this project is created by using html,css and javascript.Also Flask is the web framework is used.SQlite database is used for creating front end.In this GUI system he/she can put values of the attributes and the system will predict the reason for the landslide.

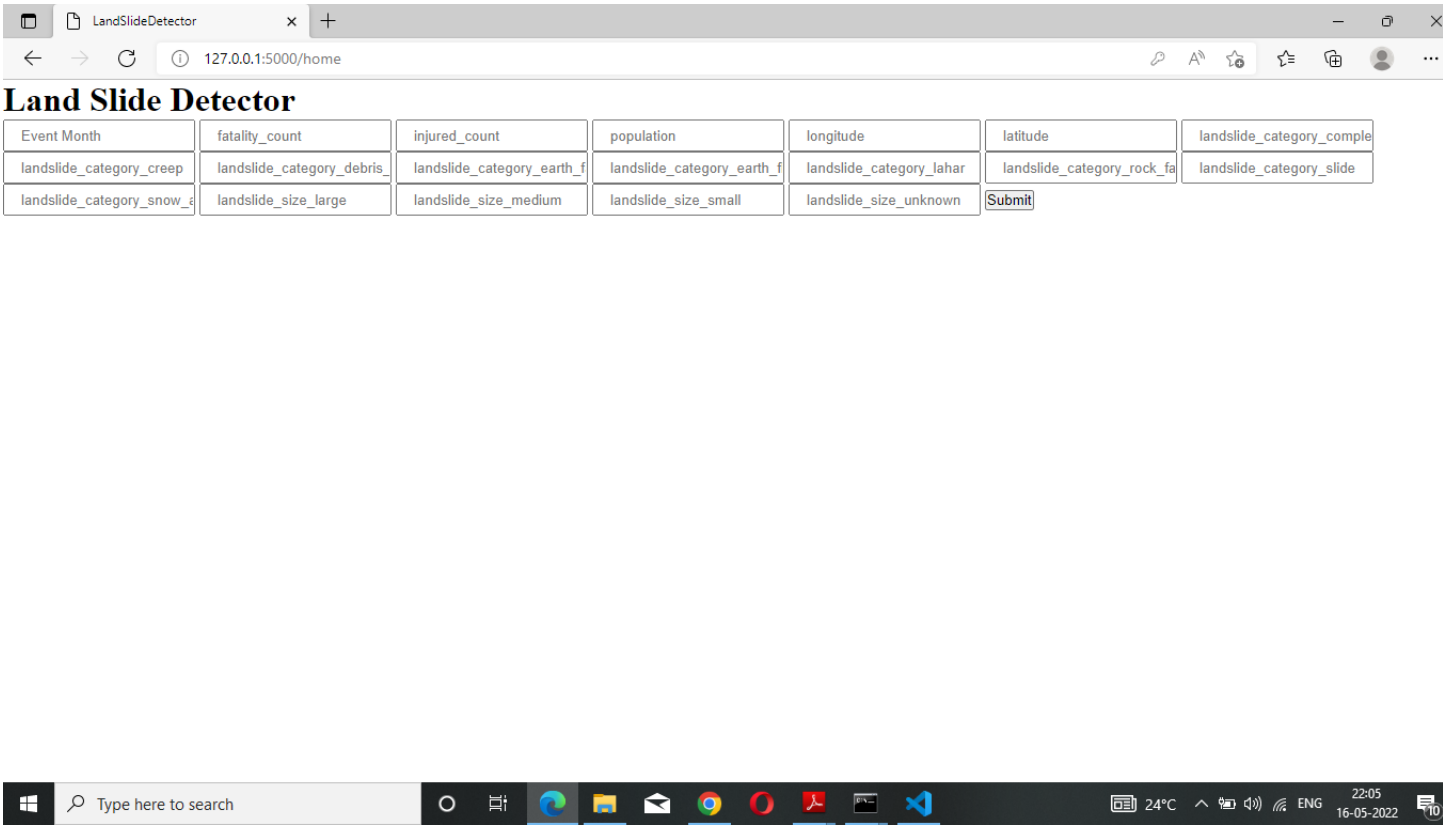


Figure 5.11: Input Values

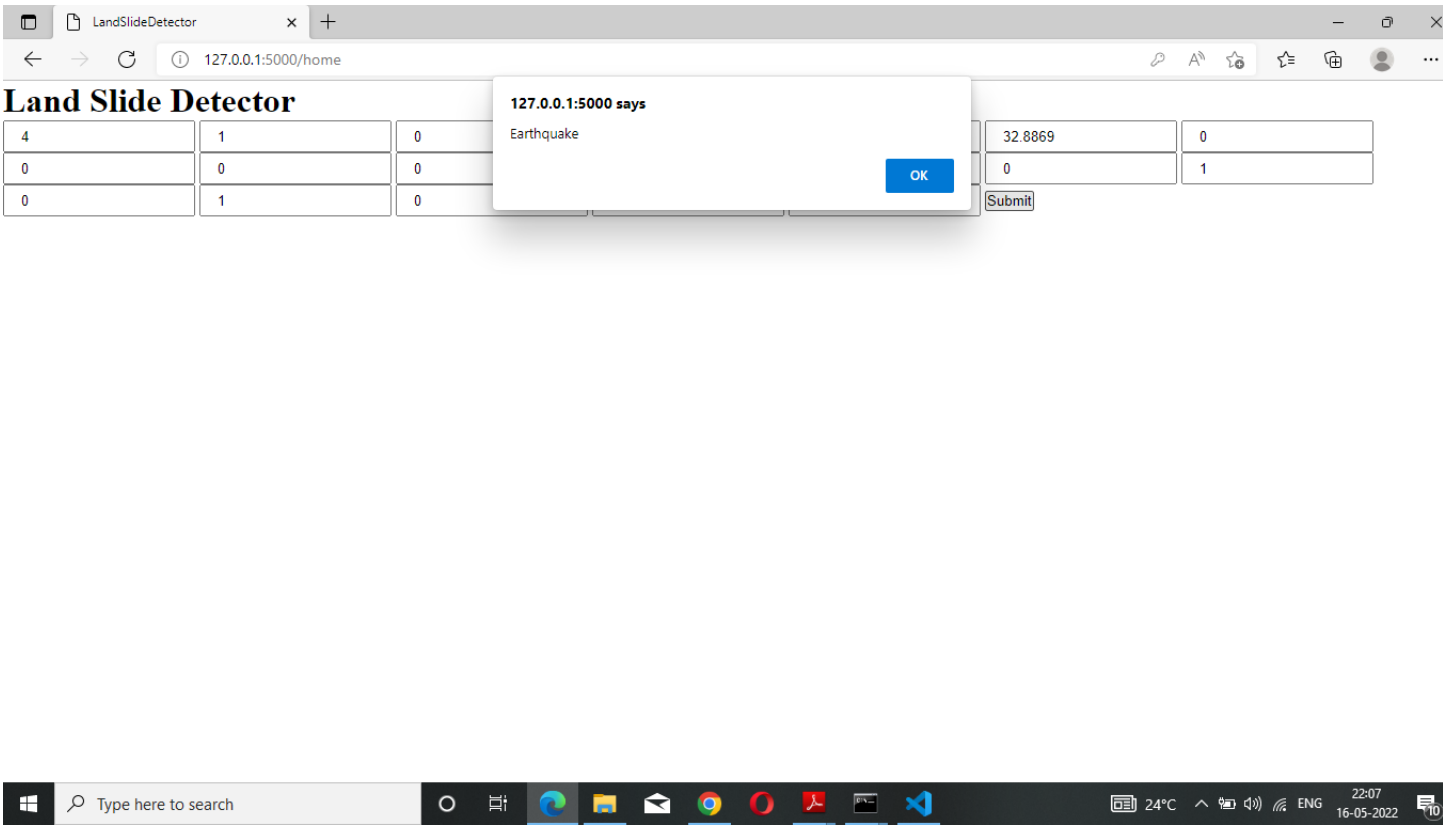


Figure 5.12: Output

Chapter 6

CONCLUSION

This project identify the reason for the landslide with machine learning.The five machine learning algorithm, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Boosting Method and Decision Tree are utilized and evaluated on global landslide dataset.An Ensemble Method is also used to evaluate on global landslide dataset..From the result, .Random Forest gives an accuracy of 89% followed by Boosting Method gives accuracy of 87%,SVM gives accuracy of 86% ,Decision Tree gives accuracy of 85% and Logistic Regression gives accuracy of 75%.By using Ensemble Method to increase the accuracy.It gives the accuracy of 90%.The reasons for landslide identification is useful to take precautions against landslide.

6.1 Advantages

The main merits of proposed model are:

- It helps to identify the reason for landslide.
- It is very useful to take precautions against landslide.
- The system provide more accuracy and efficiency for identfy the reason for landslide.

6.2 Future Enhancement

This project has an immense scope in the field of remote sensing and study of natural hazards.This is very helpful in prevention and take precautions against landslide.

REFERENCES

[1] Haojie Wang a, Limin Zhang a,*, Kesheng Yin a, Hongyu Luo a, Jinhui Li b,Landslide identification using machine learning,Elseiver,Feb 2020.

[2]]Chen Chen a, Mo-wen Xie a, Yu-jing Jiang b, Bei-ning Jia c, Yan Dua,A new method for quantitative identification of potential landslide,Elseiver,2021.

[3] Haojie Cai, Tao Chen , Senior Member, IEEE, Ruiqing Niu, and Antonio Plaza, Fellow, IEEE,Landslide Detection Using Densely Connected Convolutional Networks and Environmental Conditions, IEEE Journal Of Selected Topics in Applied Earth Observations and Remote sensing ,IEEE ,May 2021.

[4] BISWAJEET PRADHAN 1,2, (Senior Member, IEEE), HUSAM A. H. AL-NAJJAR1, MAHER IBRAHIM SAMEEN1, MUSTAFA RIDHA MEZAAL3, AND ABDULLAH M. ALAMRI4 1Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia,Landslide Detection Using a Saliency Feature Enhancement Technique From LiDAR-Derived DEM and Orthophotos,IEE, July 2020.

[5] Shuai Chen, Zelang Miao , Member, IEEE, Lixin Wu, and Yueguang He,Application of an Incomplete Landslide Inventory and One Class Classifier to Earthquake-Induced Landslide Susceptibility Mapping,2020.

[6] BINGXIN SHI1,3,4, TING ZENG 2, CHUAN TANG1, LIFANG ZHANG5, ZHUOJUAN XIE5, GUOJUN LV6, AND QIHONG WU7,Landslide Risk Assessment Using Granular Fuzzy Rule-Based Modeling: A Case Study on Earthquake-Triggered Landslides,Sept 2021.

[7] Gao, L., Zhang, L.M., Lu, M., 2017. Characterizing the spatial variations and correlations of large rainstorms for landslide study. Hydrol. Earth Syst. Sci. 21 (9), 4573.

[8] Gao, L., Zhang, L.M., Cheung, R.W.M., 2018. Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. Landslides 15, 727–740.

[9] Geo, 1996. Compilation of a Database on Landslide Consequence. Submitted by Mitchell, McFarlane, Brentnall Partners Int. Ltd., Agreement No. GEO 8/95. Geotechnical Engineering Office, Hong Kong Special Administration Region.

[10] Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Rem. Sens. 11 (2), 196.

[11] Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. Griffiths, D.V., Huang, J., Fenton, G.A., 2011. Probabilistic infinite slope analysis. Comput. Geotech. 38 (4), 577–584.

[12] Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. Earth Sci. Rev. 112 (1–2), 42–66.