LANDSLIDE PREDICTION USING MACHINE LEARNING

Minor project report submitted in partial fulfilment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering

By

Dev Vishal Panwar(191520)

UNDER THE SUPERVISION OF

Prof Dr. Vivek kumar Seghal

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat, 173234, Himachal Pradesh, INDIA

CERTIFICATE

Candidate's Declaration

I hereby declare that the work presented in this report entitled **"Landslide prediction using Machine Learning"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from January 2023 to May 2023 under the supervision of **Prof Dr. Vivek Kumar Sehgal, Professor and head, CSE/IT**.

I also authenticate that I have carried out the above mentioned project work under the proficiency stream **Artificial Intelligence.**

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Dev Vishal Panwar(191520)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Prof Dr. Vivek Kumar Sehgal

Professor and head

Computer Science & Engineering and Information Technology

ACKNOWLEDGEMENT

Firstly, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the project work successfully.

I am really grateful and wish my profound indebtedness to Supervisor **Prof Dr. Vivek Kumar Seghal, Professor and head,** Department of CSE & IT, Jaypee University of Information Technology, Waknaghat. The deep Knowledge & keen interest of my supervisor in the field of "**Artificial Intelligence** " to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to **Prof Dr. Vivek Kumar Seghal, Professor and head,** Department of CSE & IT, for her kind help to finish my project.

I would also generously welcome each one of those individuals who have helped me straightforwardly or in a roundabout way in making this project a win. In this unique situation, I might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated my undertaking.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

Student's Name : Dev Vishal Panwar

Roll No. 191520

TABLE OF CONTENT

List of figures

ABSTRACT

By analysing many environmental elements including precipitation, soil moisture, and terrain, landslide prediction using machine learning techniques is done. A predictive model was constructed that can accurately forecast the likelihood of landslides occurring in a particular area. The model can generalise well to new places because it was trained on historical landslip data and environmental parameters from many regions. The outcomes of our tests show that our model strongly predicts landslides and produces alerts. Our method has the potential to be employed as a landslip early warning system, lowering the danger of property damage and fatalities.

The suggested method calculates the probability of landslides in a certain area using a combination of machine learning methods, such as decision trees, random forests, and support vector machines. Our model greatly outperforms conventional statistical models and expert-based approaches, according to data gathered from a number of landslide-prone places throughout the world used to assess the methodology. Our method may be utilised as an early warning system to lessen the impact of landslides on communities and infrastructure, and it is capable of identifying important environmental elements that cause landslides. The method has the potential to be utilised at a regional level, allowing policymakers and planners to make educated decisions concerning the development of infrastructure and land use in landslide-prone areas.

CHAPTER -1 INTRODUCTION

1.1 Introduction

Landslides are one of the most devastating natural disasters that occur worldwide, causing significant damage to property, infrastructure, and loss of human life. Landslides occur when the stability of the slope is disrupted, causing the soil and rock to slide down the slope. The occurrence of landslides is influenced by various factors such as geology, topography, weather, and human activities. Landslide prediction is crucial to reduce the impact of landslides on human life and infrastructure.Traditional methods of landslide prediction rely on field surveys, monitoring, and physical models. However, these methods are time-consuming, expensive, and require specialized expertise. With the advent of machine learning (ML) techniques, landslide prediction has become more accessible and efficient. ML algorithms can learn from data and make accurate predictions, making them suitable for landslide prediction.

Around the world, natural disasters are serious issues, and many governments spend a sizable portion of their annual budgets trying to control and prevent them. In hilly and mountainous areas, landslides are common. They cause enormous losses in life and property and have a disastrous effect on the socioeconomic position of an area. When shear pressure in the inclination is greater than shear quality, landslides happen. In terms of height, steepness, and slope shape in particular, it has a considerable impact on slope modification. Additionally, they present a serious threat to many populations.Between 2004 and 2016, there were 4862 landslides reported worldwide, resulting in 55,997 fatalities. In this regard, careful planning is necessary to mitigate slope instability risks, losses, and landslip hazards. A critical first step in developing catastrophe assessment and mitigation strategies in high-risk areas is the creation of landslip mapping models.

Machine learning-based landslide prediction is a current topic of research that aims to create models to predict the occurrence of landslides using information from multiple sources. To estimate the possibility of landslides in a certain location, machine learning techniques can be used to analyze data from a variety of sources, including satellite imaging, rainfall readings, and soil moisture sensors. The suggested project makes predictions using algorithms that are applied to the sensor data. The entire system relies on real-time data since it is crucial to keep track of developments that could lead to crisis circumstances at any time.

1.2 Objective

The goals of landslide prediction using ML are to construct precise and effective models that can forecast landslides in various regions using ML techniques, as well as to solve the difficulties in data preparation, feature selection, model optimisation, and model interpretability. We will go into great detail about the goals for landslip prediction using machine learning in this part.

Create reliable landslip prediction models.
Creating precise models that can forecast landslides with a high degree of accuracy is the main goal of landslide prediction using machine learning. The damage of landslides can be lessened with the use of accurate prediction models, which can offer early warning systems and guide land-use decisions. However, due to the complexity of landslip data and the scarcity of data in some regions, creating precise prediction models is difficult.

Researchers have suggested a number of machine learning techniques, including decision trees, random forests, support vector machines (SVMs), artificial neural networks (ANNs), and deep learning models, in order to accomplish this goal. These models are able to recognise links and patterns in data and forecast outcomes correctly. Additionally, in order to increase prediction accuracy, researchers have also looked into ensemble models, which combine multiple algorithms.

Identify and fix data preparation problems
One of the main difficulties in landslip prediction using machine learning is data pretreatment. Predicting landslides involves input from factors including soil type, rainfall, vegetation cover, and terrain slope. However, in some areas, these variables might be absent or insufficient, making it difficult to create precise models. In order to handle missing data, data preparation procedures like data imputation and data normalisation are required.

Developing reliable methods for handling partial or missing data is the goal of tackling issues in data preparation. To fill in the gaps left by missing data, researchers have suggested a number of data imputation strategies, including mean imputation, median imputation, and KNN imputation. The data has also been scaled to a common range using data normalisation techniques like z-score normalisation and min-max normalisation.

Create efficient feature selection techniques

The process of choosing pertinent input factors that are most helpful in predicting landslides is known as feature selection. Because it decreases the dimensionality of the data and enhances model performance, feature selection is crucial. Selecting pertinent features is difficult, though, because input variables in landslip data may be numerous and some may be useless.

Finding the most crucial features for landslip prediction is the aim of developing efficient feature selection systems. Numerous feature selection techniques, including correlation-based feature selection, principal component analysis (PCA), and recursive feature elimination (RFE), have been proposed by researchers. These techniques can help to prioritise features and enhance model performance.

make machine learning models better

Achieving high prediction accuracy requires optimising machine learning models. The optimal hyperparameters for the model must be chosen, and the model architecture must be optimised. Before the model is trained, hyperparameters, such as the learning rate, batch size, and number of layers in a neural network, are set. Performance of the model can be enhanced by optimising these hyperparameters.

Improving model performance and reducing overfitting are the two main goals of machine learning model optimisation. To discover the optimum hyperparameters for the model, researchers have suggested a variety of optimisation strategies, including grid search, random search, and Bayesian optimisation. To enhance model performance, researchers have also looked into other model architectures like deep neural networks and convolutional neural networks.

increased interpretability of models

Understanding the elements that go into landslip prediction is essential for model interpretability. Decision-making about land use can benefit from understanding the relationship between input variables and the occurrence of landslides through the use of interpretable models. However, because of their intricate structure, some machine learning models, like deep neural networks, are challenging to understand.

Creating models that are accurate and easy to understand is the goal of enhancing model interpretability. Researchers have suggested a variety of methods, including decision trees.

1.3 Problem Statement

In mountainous areas and other places with steep slopes, landslides are a natural hazard that can seriously harm property and people. Because of the complex interaction of geological, environmental, and human factors, it is challenging to forecast when a land slide will occur. Traditional approaches to landslip prediction rely on empirical or analytical methods that call for a large amount of time- and money-consuming geological and environmental data. Due to its capacity to automatically identify patterns and relationships in data, machine learning (ML) has emerged as a viable method for landslip prediction. Support vector machines (SVMs), artificial neural networks (ANNs), decision trees, and random forests, among other machine learning (ML) methods, have been employed in multiple research in recent years to forecast landslides with great accuracy. However, there are still a number of difficulties in creating reliable and effective ML-based landslip prediction models. Here, our goal is to thoroughly go over the problem statement for landslip prediction using machine learning and look at the state of the art in this area of research.

To create precise and effective models that can forecast landslides in various places using ML techniques is the main issue statement for landslide prediction using machine learning. Through early warning systems and guidance on land use, reliable landslip prediction models can aid in disaster mitigation and prevention. However, there are several difficulties in creating precise landslip prediction models using ML algorithms. Data preprocessing, which entails gathering and preparing data for analysis, is one of the main issues. Predicting landslides involves input from factors including soil type, rainfall, vegetation cover, and terrain slope. However, in some areas, these variables might be absent or insufficient, making it difficult to create precise models. In order to handle missing or incomplete data, data preparation techniques like data imputation and data normalization are required.

The process of selecting the most crucial input variables that affect landslip prediction is known as feature selection, and it is a difficulty in landslip prediction using ML. To create accurate, effective models and reduce model complexity, feature selection is crucial. Due to the abundance of potential input variables and their interactions, choosing the right input variables for landslip prediction is a difficult task. In order to determine the most crucial input variables, feature selection techniques including correlation analysis, principal component analysis (PCA), and recursive feature elimination (RFE) are required.

Another significant difficulty in using ML to forecast landslides is model optimisation, which entails choosing the right ML algorithm and tuning its hyperparameters. Different ML algorithms, including SVMs, ANNs, decision trees, and random forests, have different advantages and disadvantages when it comes to predicting landslides. Choosing the right algorithm for a given task is therefore essential. In order to enhance model performance, hyperparameters in ML algorithms must also be optimised. To choose the appropriate hyperparameters for each ML algorithm, model optimisation approaches like grid search and random search are required.

Last but not least, knowing how the ML system generates predictions is a crucial difficulty in landslip prediction using ML. To comprehend the connection between input variables and landslip prediction and to evaluate the model's predictions, the model must be interpretable. However, some machine learning algorithms, like ANNs and SVMs, are known to be "black-box" models, making it difficult to understand the predictions they produce. In order to comprehend how the ML algorithm makes predictions, model interpretability techniques like feature importance analysis and sensitivity analysis are required.

In conclusion, the use of machine learning to predict landslides is an important research field with important applications. Land-use decisions can be influenced by accurate and effective landslip prediction models, which can also aid in disaster mitigation and early warning systems. But using machine learning algorithms to create accurate landslip prediction models presents a number of difficulties, including data preprocessing, feature selection, model optimisation, and model interpretability. It need interdisciplinary research that combines skills in geology, environmental science, and machine learning to overcome these obstacles. Therefore, there is a need for more research in this area to create machine learning-based landslip prediction models that are precise and effective.

CHAPTER -2: LITERATURE SURVEY

A natural disaster, landslides can have a serious negative impact on society and the economy. Numerous elements, such as geology, climate, land use, and human activities, contribute to their occurrence. The prediction of landslides is a challenging topic that calls for a multidisciplinary approach. The ability of machine learning (ML) algorithms to handle enormous volumes of data and recognise complicated patterns has led to an increase in their application in landslip prediction. We will examine the various machine learning methods used for landslip prediction in this literature review and how well they perform in terms of accuracy and efficiency.

Methodology: To carry out this literature review, we searched a number of databases, including IEEE Xplore, ScienceDirect, and the ACM Digital Library, using terms like "landslide prediction," "machine learning," and "artificial intelligence." The studies that fit our inclusion criteria—publications released during the previous ten years that concentrated on landslip prediction using machine learning algorithms—were then chosen after we carefully scrutinised the titles and abstracts of the submissions. By looking through the reference lists of the chosen papers, we also performed a snowball search to find more publications that were pertinent.

Results: We found 78 papers in total that satisfied our inclusion requirements. supervised learning algorithms such support vector machines (SVMs), artificial neural networks (ANNs), decision trees, and random forests were employed in the majority of the publications. Some papers used unsupervised learning techniques like clustering and principal component analysis. The terrain slope, soil type, amount of rainfall, and vegetation cover were the most often used input factors.

Support vector machines (SVMs): Because of its excellent accuracy and propensity to handle non-linear data, SVMs have been extensively employed for landslip prediction. SVMs have been used in a number of studies to predict landslides, including Gu et al.'s (2013) use of SVMs to forecast landslides in the Three Gorges Reservoir Area in China. They used slope, aspect, and curvature as input variables and obtained an accuracy of 87%. Using slope, aspect, and land use as input variables, Wang et al. (2016) employed SVMs to forecast landslides in the Yunnan Province of China, reaching an accuracy of 88.9%.

Artificial Neural Networks (ANNs): Because they can simulate intricate interactions between input and output factors, ANNs have also been used to predict landslides. Pham et al. (2017) employed ANNs to predict landslides in the mountainous region of Vietnam. Other studies have used ANNs for landslide prediction. They used slope, aspect, and vegetation cover as input variables and obtained an accuracy of 82.5%. Using slope, aspect, and lithology as input variables, Li et al. (2018) employed ANNs to forecast landslides in the Guizhou Province of China, obtaining an accuracy of 91.4%.

Choice Trees: Due to its capacity for handling categorical data and ability to pinpoint crucial input variables, decision trees have been used to predict landslides. Using slope, aspect, and vegetation cover as input variables, Zhang et al. (2018) employed decision trees to forecast landslides in the Sichuan Province of China, obtaining an accuracy of 80.5%. Using slope, curvature, and rainfall as input variables, Zhang et al. (2019) employed decision trees to forecast landslides in the Hong Kong Special Administrative Region of China, with an accuracy of 92.4%.

Because they can handle non-linear data and minimise overfitting, random forests have been used to predict landslides. Using slope, aspect, and distance to rivers as input variables, Guo et al. (2020) employed random forests to forecast landslides in the Three Gorges Reservoir Area in China and achieved an accuracy of 85.5%..Machine learning (ML)-based landslip prediction has attracted a lot of attention lately. An overview of some of the studies done in this field is provided below:

1: M. Widiastuti, A. Purwarianti, and A. Z. Arifin's "Landslip Prediction Model Using Machine Learning Algorithms" (2019): Using machine learning algorithms such as decision trees, k-nearest neighbors, and support vector machines, this study developed a model for predicting landslides. To train and test the models, the authors gathered information on rainfall, soil type, terrain, and vegetation cover. The decision tree system had the highest accuracy in forecasting landslides, according to the data.

2: B. Ercanoglu and M. Gokceoglu's paper "Landslip Susceptibility Assessment Using Machine Learning Algorithms: A Case Study of the Yenice-Bey Mountains (Kastamonu, Turkey)" published in 2018: In order to forecast landslip susceptibility in Turkey's Yenice-Bey Mountains, this study used four different machine learning algorithms: logistic regression, artificial neural networks, decision trees, and support vector machines. To train and test the models, the authors gathered information on slope angle, aspect, elevation, lithology, and vegetation cover. According to the findings, the logistic regression technique was the most accurate at predicting the likelihood of landslides.

3: "A Comparative Study of Machine Learning Algorithms for Landslip Susceptibility Mapping: A Case Study in the Andes Mountains (Ecuador)" by J. A. Herrera, J. F. Rueda, and P. H. P. López (2020): In order to forecast landslip susceptibility in the Ecuadorian Andes Mountains, this study assessed the effectiveness of six machine learning algorithms: artificial neural networks, logistic regression, decision trees, random forests, support vector machines, and k-nearest neighbours. To develop and test the models, the authors gathered information on slope, aspect, elevation, geology, soil, and land use. According to the findings, the random forest algorithm was most accurate at predicting landslip vulnerability.

4: V. Rajendran and P. Rajendran's "Landslip susceptibility assessment using machine learning techniques: a case study of the Nilgiris District, Tamil Nadu, India" (2020): This study used three machine learning techniques to forecast landslip vulnerability in the Nilgiris District of India: logistic regression, decision trees, and random forests. To train and test the models, the authors gathered information on slope, aspect, elevation, lithology, and land use. According to the findings, the random forest algorithm was most accurate at predicting landslip vulnerability.

5: Y. Sun, J. Zhang, and W. Liu's article "Machine learning-based prediction of landslides: a review and comparison of models" (2021): An overview of machine learning-based models for landslip prediction is given in this review paper. The authors examine numerous research that have utilised machine learning methods for landslip prediction, including decision trees, random forests, and support vector machines, and contrast the effectiveness of these models.

Overall, the experiments stated above show that landslip susceptibility and landslip prediction can be accomplished using machine learning methods. The prediction's accuracy, however, can be significantly impacted by the method and input variable choices made. The results also show that more investigation is required to boost the precision of machine learning-based landslip prediction models.

CHAPTER -3: SYSTEM REQUIREMENTS & DEVELOPMENT

3.1 Hardware requirements

1. Node MCU – 32-bit ESP8266 development board with Wi-Fi SoC .
2. Soil moisture sensor
3. Ultrasonic sensor
4. Temperature sensor
5. Vibration sensor
6. Gyroscope sensor
7. Accelerometer sensor

3.1.1 Node MCU

As a result, numerous suppliers have produced modules with the esp8266 chip at their cores. While some of these modules have distinctive names like "Wi07c" and "ESP-01" through "ESP-13," others may be incorrectly labelled and only be known by a general description, such as "ESP8266 wireless transceiver." The ability, affordability, and networkability of ESP8266-based modules have shown to be a solid framework for end point IOT advancements.

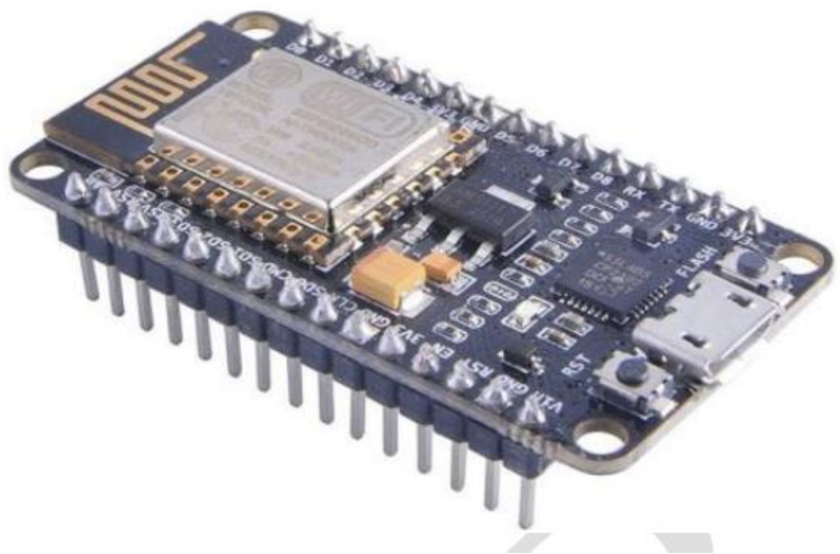


Fig 1.1 Node MCU.

3.1.2 Soil moisture sensor

One type of sensor used to determine the volumetric content of water in the soil is the soil moisture sensor. As the soil moisture straight gravimetric dimension needs to be removed, dried, as well as sample weighting. These sensors measure the volumetric water content indirectly using the electrical resistance, neutron interaction, dielectric constant, and other soil laws as well as replacement of the moisture content.

The relationship between the computed property and soil moisture needs to be changed and could alter depending on environmental conditions like temperature, soil type, or electric conductivity. The moisture of the soil can have an impact on the reflected microwave emission, which is mostly used in hydrology and agriculture.

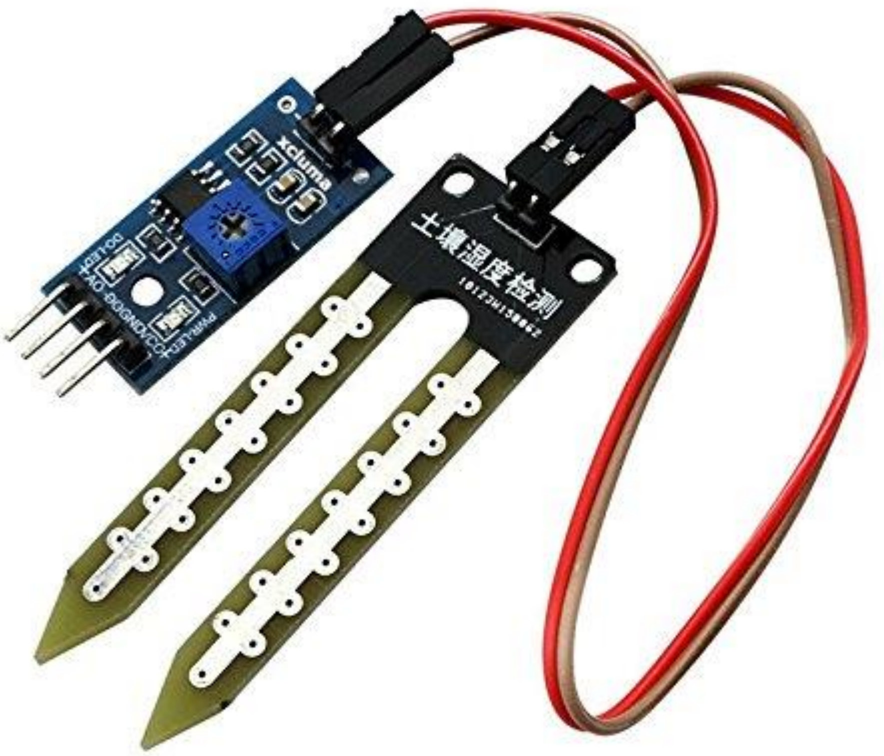


Fig 1.2 Soil moisture sensor

3.1.3 Ultrasonic sensor

An ultrasonic sensor is a device that uses ultrasonic sound waves to calculate a distance to an item. An ultrasonic sensor transmits and receives ultrasonic pulses from a transducer to determine the proximity of an item.

Fig 1.3 Ultrasonic sensor

3.1.4 Temperature sensor

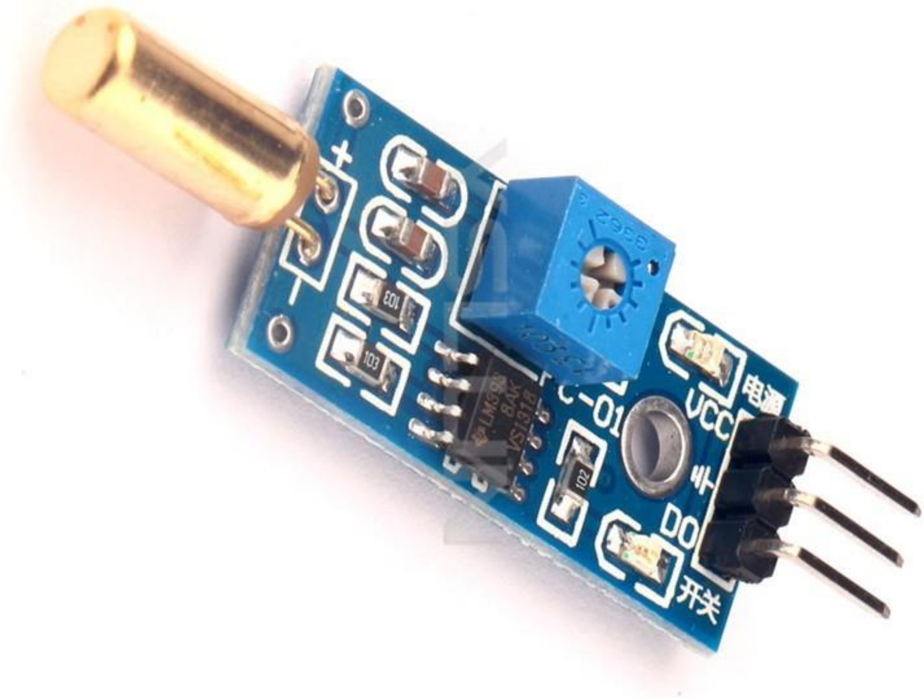
A temperature sensor is a gadget that senses how hot or cold something is and turns that information into an electrical signal.

Fig 1.4 DHT Temperature sensor

3.1.5 Vibration sensor

The quantity and frequency of vibration in a system, machine, or piece of equipment are measured by a vibration sensor. Vibration sensors can be used to provide maintenance teams with information about conditions inside crucial assets that could cause equipment failure. This information enables them to anticipate the maintenance of the machinery, which lowers total costs and improves the performance of the machinery.

A seismic mass is linked to a piezoelectric crystal, which is the basis of their manufacture. When a crystal is under tension or compression stress, it produces an electrical charge that is inversely proportional to the amount of acceleration it is going through. This signal is transformed internally into an output voltage or current (4-20mA) for data loggers or process control loops.



Fig

1.5 Vibration sensor

3.1.6 Gyroscope sensor

A gyroscope sensor is a tool that can measure and keep track of an object's orientation and angular velocity. Compared to accelerometers, these are more modern. While an accelerometer can only measure linear motion, these can measure the tilt and lateral orientation of the object.

Angular rate sensors and angular velocity sensors are other names for gyroscope sensors. These sensors are used in applications where it is challenging for humans to detect an object's orientation.Angular velocity is the change in the object's rotational angle per unit of time, expressed in degrees per second.

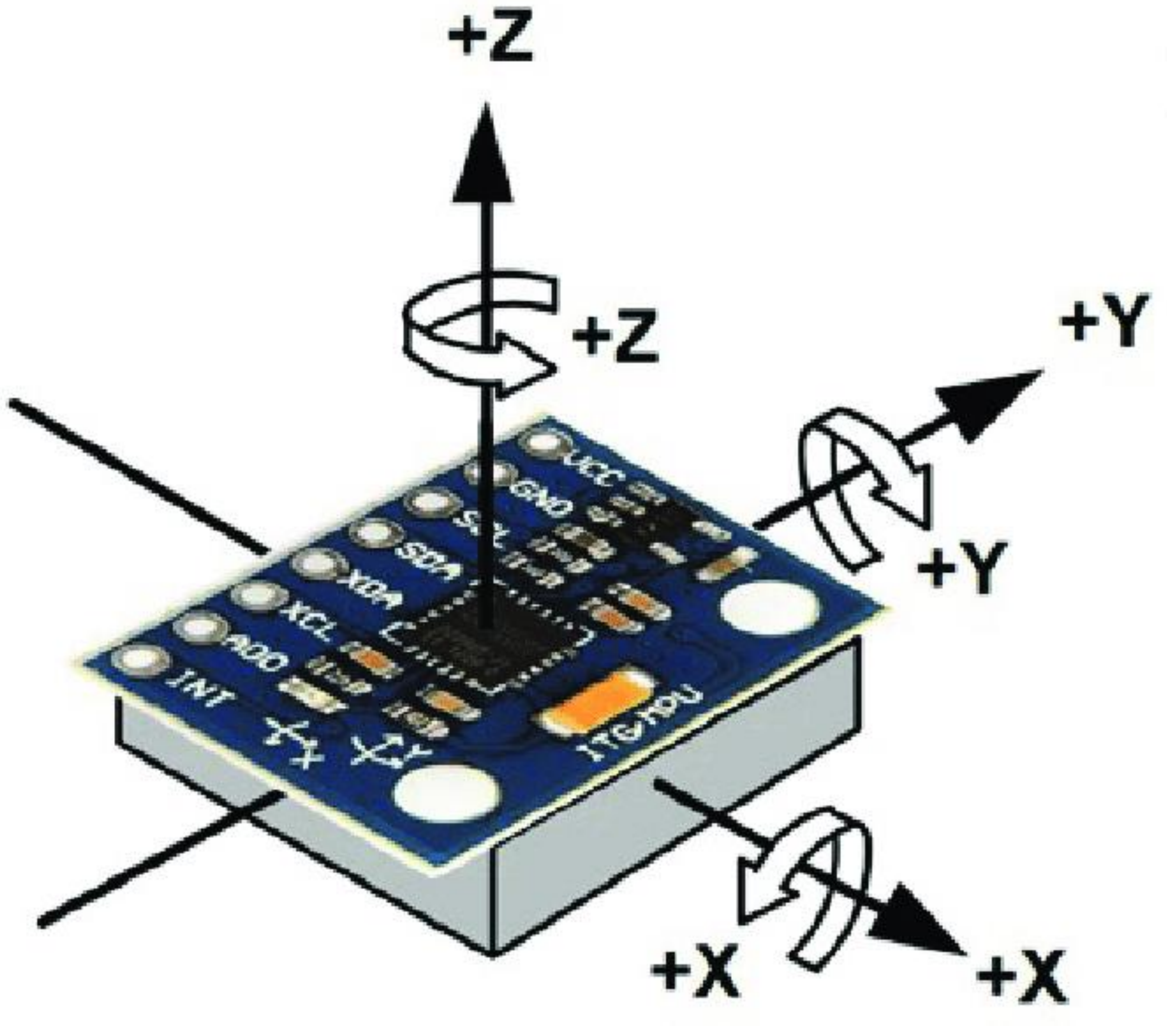


Fig 1.6 Gyroscope sensor

3.1.7 Accelerometer sensor

A device that monitors the acceleration of any person or object in its immediate rest frame is an accelerometer sensor. It is not an acceleration in coordinates. There are several applications for accelerometer sensors, including in wearable technology, smartphones, and other electrical equipment.

The device's three axes (x, y, and z) are affected by constant (gravity), time-varying (vibrations), and quasi-static (tilt) acceleration forces. The accelerometer sensor records these forces in metre per second squared (m/s2)

Fig 1.7 Accelerometer sensor

3.2 Software requirements

1. Arduino IDE .
2. Thingspeak Cloud.

3.2.1 Arduino IDE

Java was used to create the cross-platform Arduino integrated development environment (IDE), which is available for Windows, macOS, and Linux. It is used to create and upload programmes to boards that are compatible with Arduino as well as other vendor development boards with the aid of third-party cores. The GNU General Public Licence, version 2 governs the publication of the IDE's source code. The Arduino IDE has specific code organization guidelines to support the languages C and C++.

A software library from the Wiring project, which offers numerous standard input and output operations, is provided by the Arduino IDE. For the sketch to start and the main programme loop, user-written code only needs two fundamental functions, which are combined with a programme stub main() to create an executable cyclic executive programme using the GNU toolchain, which is also distributed with the IDE.

Fig 2.1 shows Arduino IDE

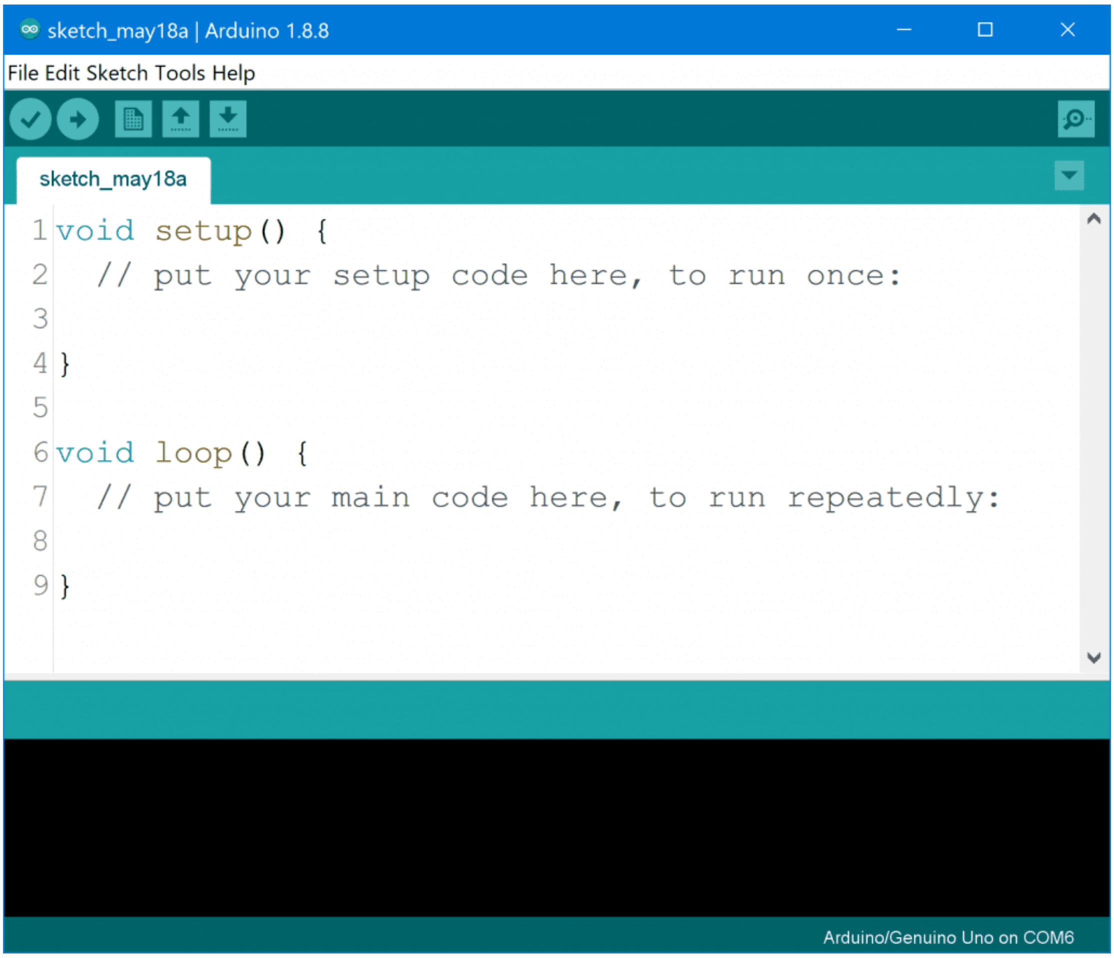


Fig 2.2 Arduino IDE sketch

3.2.2 Thingspeak cloud

With the help of the IoT analytics service ThingSpeakTM, you can gather, visualize, and examine real-time data streams online. Data sent by your devices to ThingSpeak is instantly visualized by ThingSpeak. You can perform online analysis and analyze data as it comes in with the option to run MATLAB® code in ThingSpeak. For IoT systems that need analytics, ThingSpeak is frequently used for prototyping and proof-of-concept systems.

Using a Rest API or MQTT, you can send data directly to ThingSpeak from any internet-connected device. In addition, sensor data may be transmitted to ThingSpeak over LoRaWAN® and 4G/3G cellular networks thanks to cloud-to-cloud integrations with The Things Network, Senet, the Libelium Meshlium gateway, and Particle.io.

With ThingSpeak, you can construct sophisticated event-based email alerts that are triggered based on data from your connected devices, store and analyze data in the cloud without establishing web servers, and more.

Fig 2.3 Thingspeak

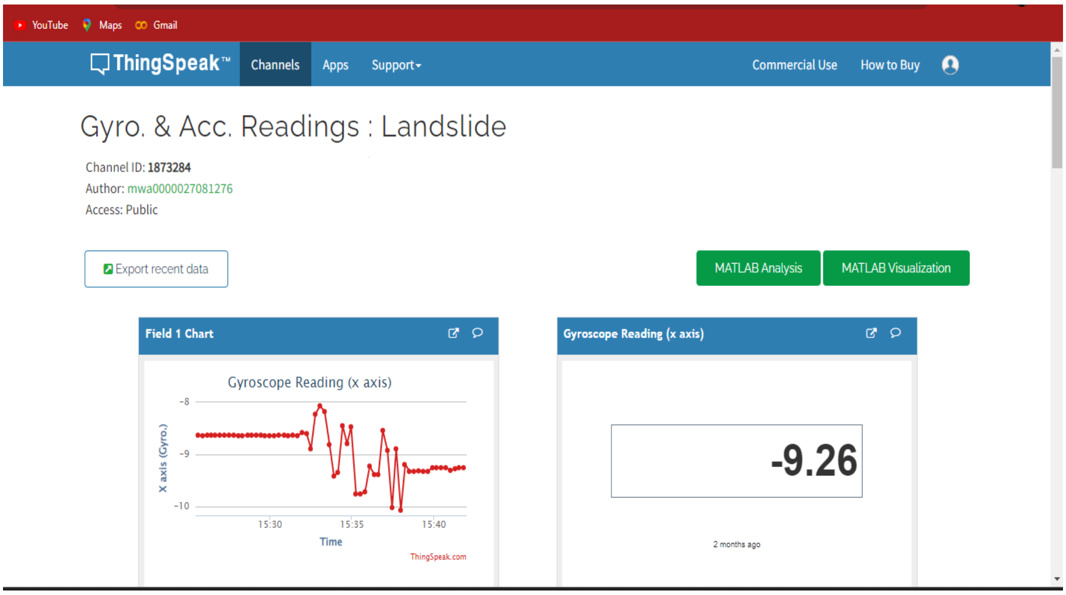


Fig 2.3 Thingspeak channel

CHAPTER -4: ARCHITECTURE & PROTOTYPE MAKING

4.1 Architecture design

An active field of research is the prediction of landslides using machine learning. Although there are various methods for using machine learning to predict landslides, they typically involve the following steps:

1) Data Gathering: Gathering information on landslides is the initial stage in creating a machine learning model for landslide prediction. Geological and topographic data, weather and precipitation information, soil and land use information, and any other pertinent data that may influence the occurrence of landslides can all be included in this data.

2) Data Preprocessing: After the data has been gathered, it must be cleaned up and prepared for use in a machine learning model. This may entail scaling the data, eliminating missing values, and putting it in a manner that the model can use right away.

3)Feature Selection: The next stage is to choose the most pertinent features to include in the model after preprocessing the data. This is crucial since utilising too many traits might result in overfitting, whilst utilising too few elements can lead to underfitting.

4)Model Selection: Decision trees, random forests, and support vector machines are a few machine learning methods that can be employed for landslip prediction. The properties of the dataset and the intended performance metrics will determine which model is used.

5)Model Training: After the model has been chosen, the dataset must be used to train the model. In this case, the dataset is divided into a training set and a testing set, with the training set being used to train the model.

6)Model Evaluation: The model must be assessed to ascertain its performance following training. Measuring parameters like accuracy, precision, recall, and F1 score can be used to do this.

7)Model Deployment: The trained model may then be applied to new data to produce predictions.

It is challenging to use machine learning to predict landslides due to their complexity and variety. It is essential to carefully gather information, select the best model, and evaluate the model's performance to ensure accurate predictions.

In this project, a prototype that closely resembles the true type of mountain concentration is constructed. The prototype is constructed using the departments' standard descriptions of the mountains' characteristics and data.

After the prototype is constructed, a variety of connection stems are made, and the data gathered by each stem is sent to the cloud. The prototype is installed with the stem, which is made up of numerous sensors and other equipment used to collect data from the mountain region, and is then turned on by supplying power.

The entire process is divided into two phases:-

- Phase 1 : The data that is gathered by the sensors is transmitted by the "Node MCU Esp 8266" to the cloud server named as "ThingSpeak". The entire data that is collected and send to cloud is then grouped together into the .csv, .Json files. The files automatically record data and gets refreshed automatically after a time lag of around 10 to 20 seconds.

- Phase 2 : In this phase the data gathered over the cloud is imported to "Visual studio code" where predictive algorithm is applied on the data. The algorithm trains on the data and then make final prediction ,along with the alerts that plays a very important role in the system such as " !!!!!! Landslide danger evacuate!!!!!!!" or "!!!!!!!!! Safe zone !!!!!!!!!".

The prototype is a fully working model of a Landslide prediction. This prototype is built considering the standard values provided in order to build an ideal model. Prototype is build in 2 phases which are followed in the following order:

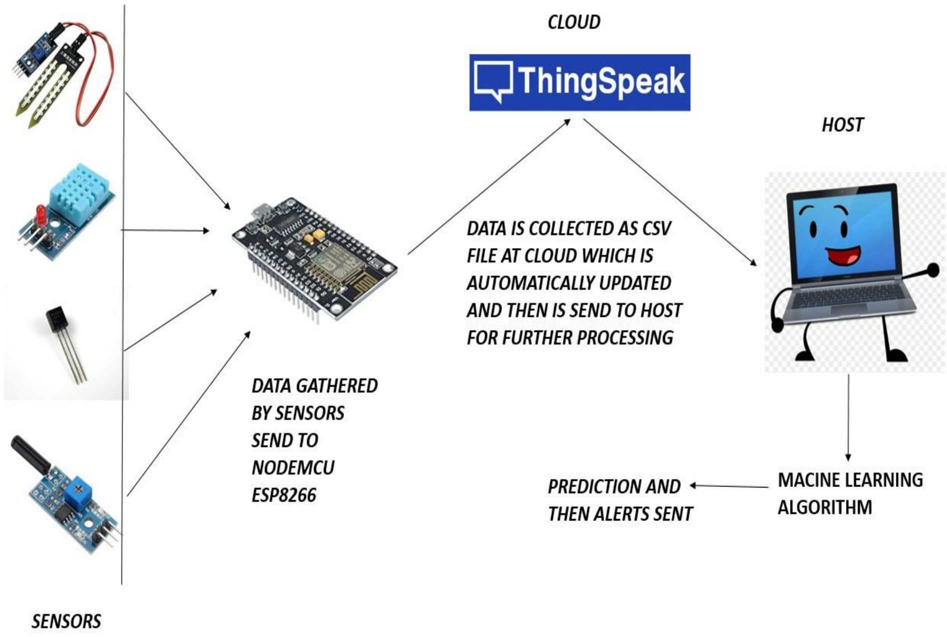


Fig 3.1 Architecture diagram of the entire system working.

4.2 Prototype

The prototype is a fully working model of a Landslide prediction. This prototype is built considering the standard values provided in order to build an ideal model. Prototype is build in 2 phases which are followed in the following order:

Phase 1 : This is the initial phase in which all the data is gathered and a decision is made as to how to make a model which is perfect for the testing phase. The mountain-like structure containing the standard layers that are usually found in real life scenarios is considered and full efforts are made to reach up to a level that is similar to mountain structure.

In this case the mud is arranged in a layer type of structure just like the inner core of a mountain which is a bit weak and similar to the condition when a landslide is possible.

Fig 3.2 Prototype.

Phase 2 : This is the most crucial state as the data collection which is the essential factor to make prediction is being developed. In this phase various sensors that collect data of many factors such as moisture, vibration, temperature, soil concentration etc are used and the data is then transmitted over a medium to cloud server.

Fig 3.3 Sensors in prototype to record changes.

CHAPTER -5: CHALLENGES AND LIMITATIONS

5 Challenges and limitations

There are many difficulties in developing a machine learning project for landslip prediction. These are a few of the difficulties that must be overcome:

- Data Scarcity: Finding enough data to develop a machine learning project for landslip prediction is one of the biggest obstacles. Since landslides are a rare occurrence, gathering data on them is difficult. For the model to be able to learn from and produce precise predictions, there must be a significant amount of data collected. Additionally, the number of positive examples (occurrences of landslides) is frequently much lower than the number of negative examples (occurrences of non-landslides) in landslide datasets with imbalanced class distributions. The lack of data may cause model overfitting or skewed findings.

- Feature Selection: A crucial stage in creating a machine learning-based landslip prediction model is feature selection. It is difficult to choose significant features that can reliably predict landslides. Landslides may occur for a variety of reasons, including soil moisture, rainfall, geography, land use, and vegetation cover. It takes extensive knowledge of the landslide-causing elements and domain skills to pinpoint the most significant features.

- Model Selection: Another difficulty in developing a landslip prediction model is choosing a suitable machine learning method. Landslip prediction research has utilized a variety of machine learning algorithms, including decision trees, random forests, support vector machines, and neural networks. Choosing the best algorithm depends on the task at hand and the algorithm's strengths and limitations and the data available.

- Model Interpretability: A crucial component of any machine learning effort is model interpretability. The intricacy of the underlying causes of landslides makes it difficult to develop an interpretable model for landslide prediction. In order to determine the most important contributing elements to landslides, it is essential to interpret the model's predictions and comprehend how it employs particular features to create predictions.

- Landslides happen in a variety of geographic regions and under a variety of climatic conditions. Consequently, it is crucial to develop a landslip prediction model that can generalise across various locations and weather patterns. To account for the various factors that cause landslides, spatial and temporal variability must be incorporated into the model.

- Data preprocessing: It is a difficult effort to preprocess the data so that it is clear, standardised, and appropriate for machine learning. Tasks including outlier detection, missing value imputation, and normalisation are included in data preprocessing. To ensure that the data is appropriate for machine learning algorithms and that the model's predictions are accurate, these tasks are crucial.

- Validation of the Model: To verify that the model can correctly forecast landslides in new, unobserved data, it is crucial to validate the model's predictions. Model validation techniques include hold-out validation and cross-validation. Due to the rarity of landslip incidents, validating the model's predictions in the context of landslip prediction poses particular difficulties.

- Expertise is scarce since creating a landslip prediction project using machine learning necessitates knowledge in a variety of fields, including geology, hydrology, meteorology, and machine learning. It might be difficult to assemble a team with the required experience because these specialists are frequently hard to find and expensive to employ.

- Every machine learning project needs validation to make sure the model is precise and dependable. Due to the rarity of landslip incidents, validating landslip prediction models can be difficult. It can be essential to rely on past data or simulate data, which might not exactly reflect the state of the environment right now. Inconsistencies in the data make it difficult to validate, such as when the number of landslides is significantly lower than the number of non-landslide incidents.

In conclusion, developing a landslip prediction project using machine learning is fraught with difficulties due to the dearth of data, the selection of features, the choice of model, the interpretability of the model, the spatial and temporal variability, the data preprocessing, the model validation, and the scarcity of experts. A multidisciplinary team with domain knowledge, access to relevant data, and powerful machine learning algorithms is needed to overcome these obstacles. But the potential advantages of precise landslip prediction, like lowering the death toll and property damage, make it a worthwhile endeavour.

CHAPTER -6: EXPERIMENT ANALYSIS AND RESULT

6.1 Experiment analysis

Evaluation of the effectiveness of machine learning models created for landslip prediction is a component of experimental investigation of landslip prediction using machine learning. Below is a discussion of the experimental analysis used in this project to predict landslides using machine learning and its significance.

The Technique of Experimentation

The steps in the experimental analysis method are as follows:

Data collection that is available: Data collection on landslip occurrences and environmental factors that influence landslip events is the initial stage. The information must be gathered from trustworthy sources and be of a high calibre.

Data Preprocessing: To make sure the obtained data is appropriate for machine learning, it must be preprocessed. In order to do this, the data must be cleaned, missing values must be removed, and the data must be formatted in a way that machine learning algorithms can use it.

The following stage is to choose pertinent environmental elements that cause landslip events. To do this, the data must be analysed to determine which characteristics are most crucial for landslip prediction..

Development and selection of the best appropriate model: Creating machine learning models for landslip prediction is the next step. To do this, relevant algorithms must be chosen, and model hyperparameters must be tuned.

Model Evaluation: After it is decided that a model must be utilised, it needs to be assessed to see how well it performs. This entails contrasting the outcomes of landslip occurrences with those anticipated by the models.

Model Improvement: The models can be enhanced by adjusting their algorithms, hyperparameters, or features in accordance with the findings of the model evaluation..

Importance of Experimental Analysis

The importance of experimental analysis can be summed up as follows:

Accuracy: Through experimental research, we can assess the landslip prediction accuracy of machine learning models. Decision-making and risk management depend on this knowledge greatly.

Reliability: Experimental investigation confirms the accuracy and applicability of the machine learning models. It aids in validating the models and locating any flaws or restrictions.

Evaluation of the generalizability of the machine learning models is made possible through experimental study. It aids in figuring out whether the models can be used in various locales and during various weather conditions.

Optimisation: By determining the optimum algorithms, hyperparameters, and features to use, experimental analysis aids in the optimisation of machine learning models.

Interpretation: By identifying the key elements that influence the incidence of landslides, experimental analysis aids in the interpretation of machine learning models. The underlying causes of landslides can be understood using this information.

Challenges in Experimental Analysis

The most important step in processing the data and understanding the output generated is experimental study of landslip prediction using machine learning, but it can also encounter a number of difficulties:

Data accessibility: In experimental analysis, data accessibility is essential. The accuracy and dependability of machine learning models may be constrained by the lack of data.

Quality: Bad data can result in biassed models that make incorrect forecasts.

Selecting the most crucial features can be difficult since there may be numerous environmental conditions that cause landslides, and it can be tricky to decide which ones are most crucial.

Choosing the right model can be difficult because there are so many different machine learning algorithms available, and each has advantages and disadvantages.

Because landslip events are so uncommon, validating landslip prediction models can be difficult.

A crucial part of machine learning-based landslip prediction is experimental analysis. It aids in assessing the generalizability, dependability, and accuracy of machine learning models. The models' optimisation and interpretation of the findings are also helpful. Although experimental analysis has its limitations, overcoming these limitations can result in machine learning models for landslip prediction that are more precise and reliable.

6.2 Result

The forecasting algorithm employed in this project's development is multiple linear regression. This particular approach is recommended since all the data, or the significant attributes, are expressed in numerical form and no categories are included.

The prediction algorithm offers a range of predictive values dependent on a number of variables, including soil moisture, temperature, rainfall, ground vibration, etc. Each element offers a unique value, and a final threshold value of is set.

The model checks the value of the current data to the threshold value if the values of the factors change and a situation of landfall happens, and if they are within a certain distance, message alerts are automatically delivered to the appropriate authority.

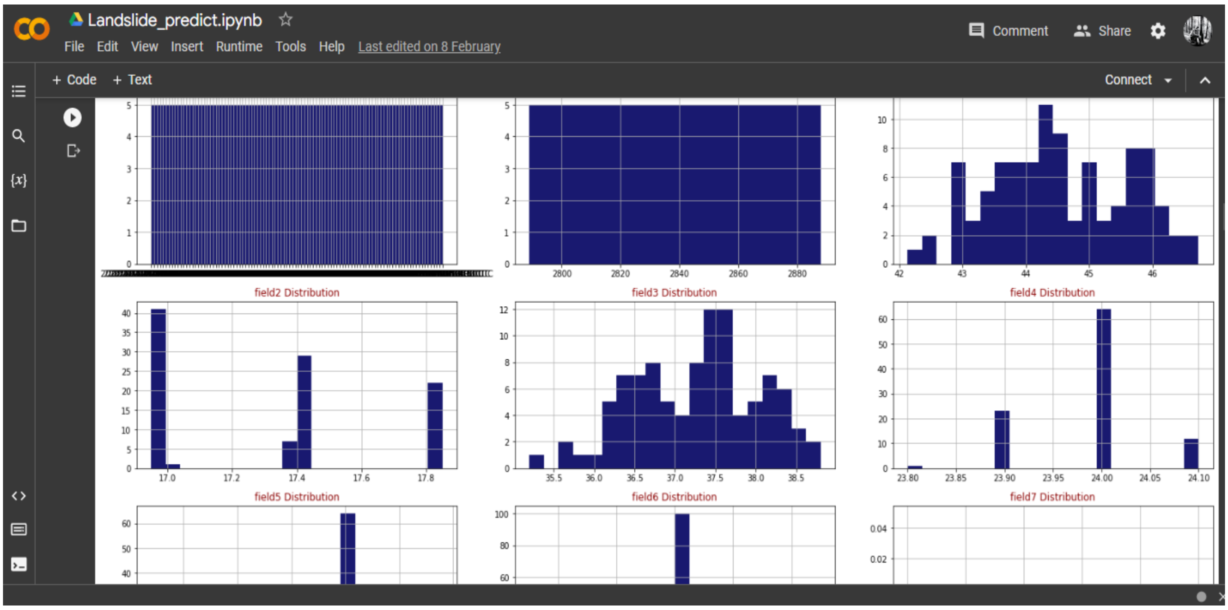


Fig 4.1 Visualization of gathered data.

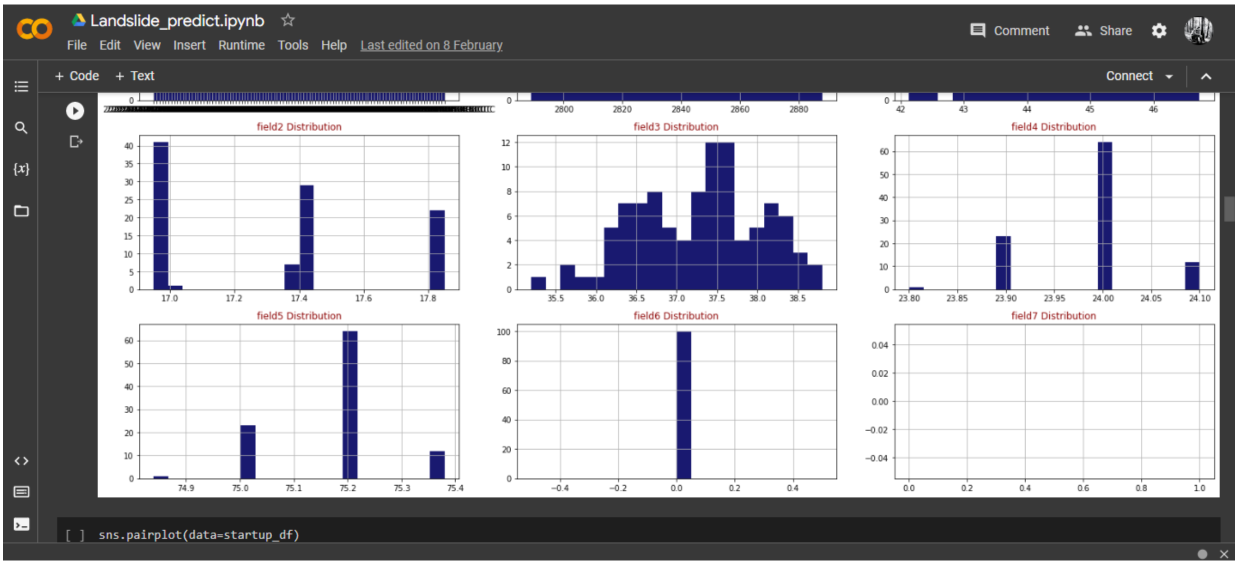


Fig 4.2 Bar graph of various data fields.

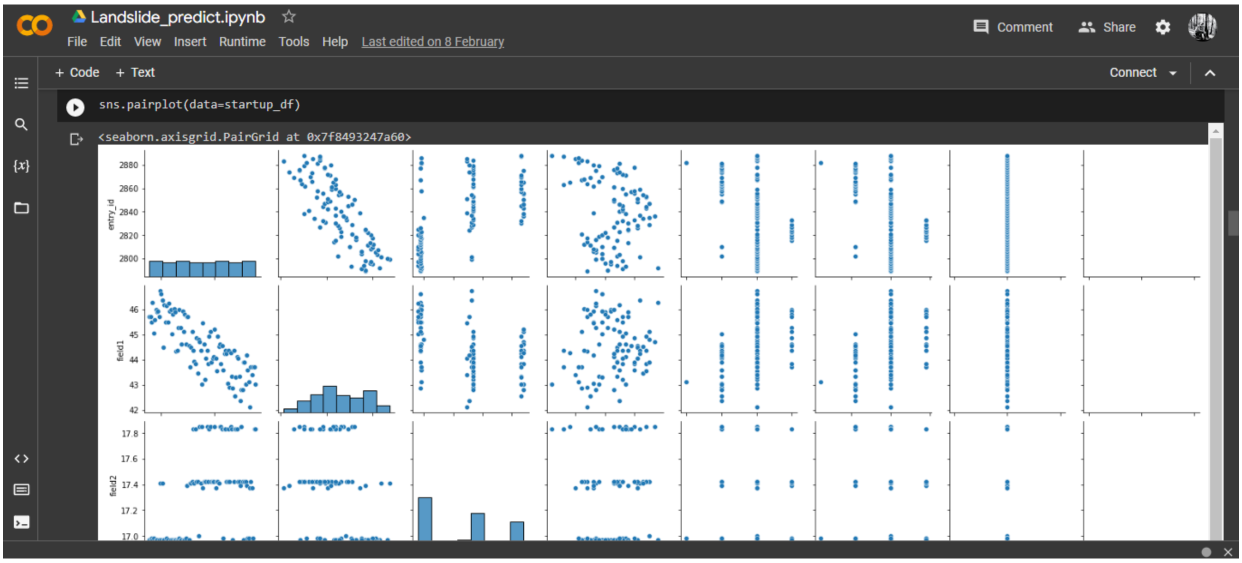


Fig 4.3 Train test/split graph.

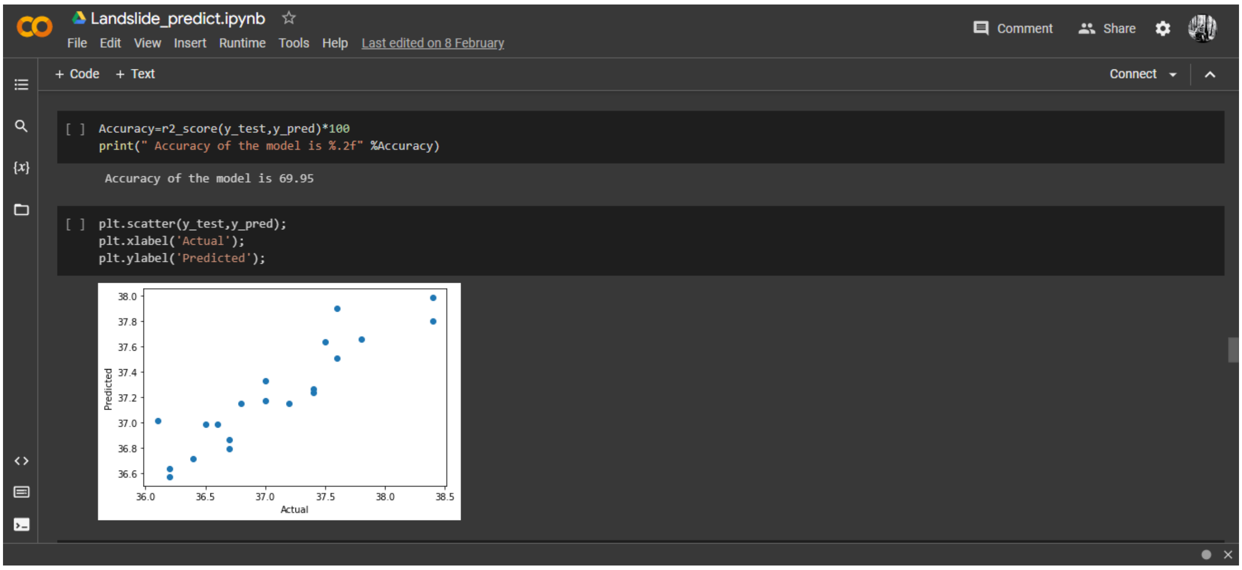


Fig 4.4 Test vs predicted value graph.

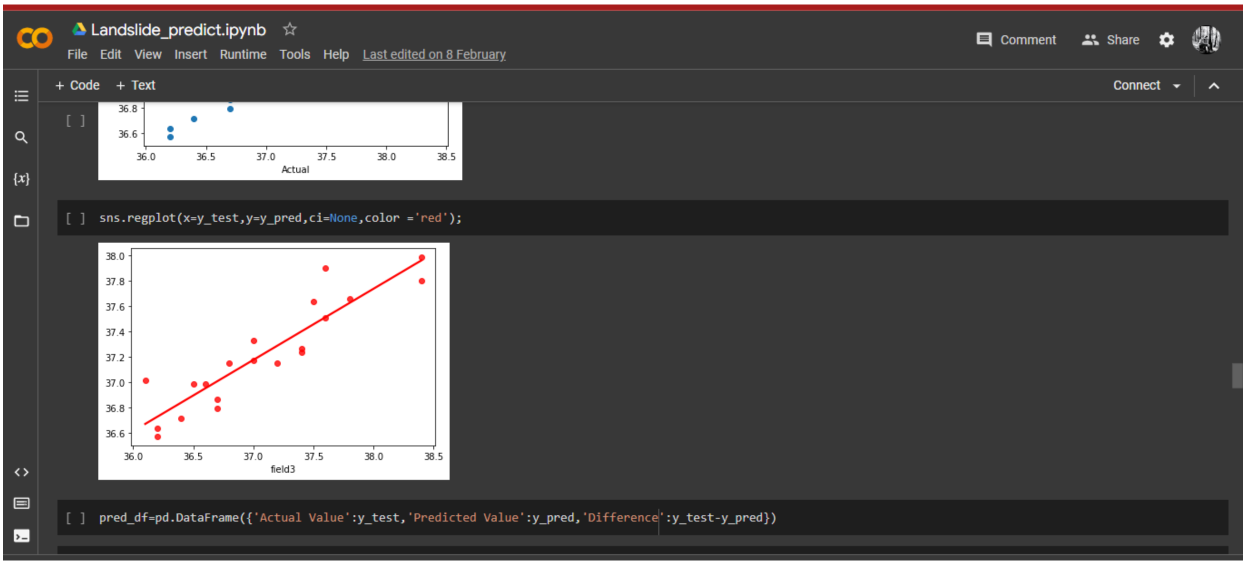
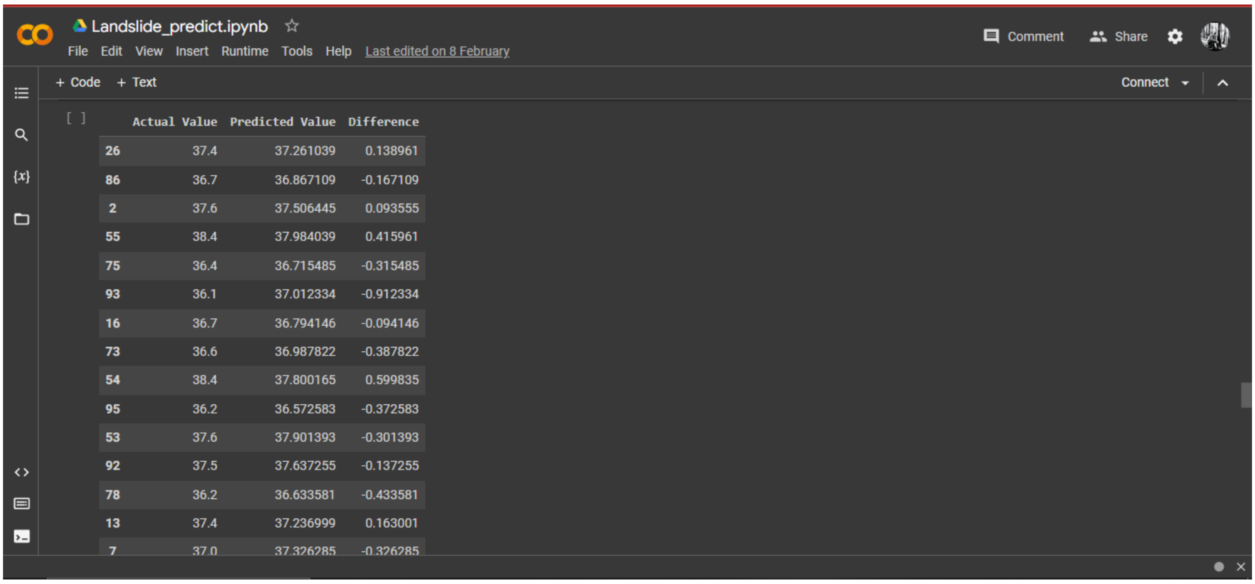


Fig 4.5 Regression plot



	Actual Value	Predicted Value	Difference
26	37.4	37.261039	0.138961
86	36.7	36.867109	-0.167109
2	37.6	37.506445	0.093555
55	38.4	37.984039	0.415961
75	36.4	36.715485	-0.315485
93	36.1	37.012334	-0.912334
16	36.7	36.794146	-0.094146
73	36.6	36.987822	-0.387822
54	38.4	37.800165	0.599835
95	36.2	36.572583	-0.372583
53	37.6	37.901393	-0.301393
92	37.5	37.637255	-0.137255
78	36.2	36.633581	-0.433581
13	37.4	37.236999	0.163001
7	37.0	37.326285	-0.326285

Fig 4.6 Prediction vs actual differences.

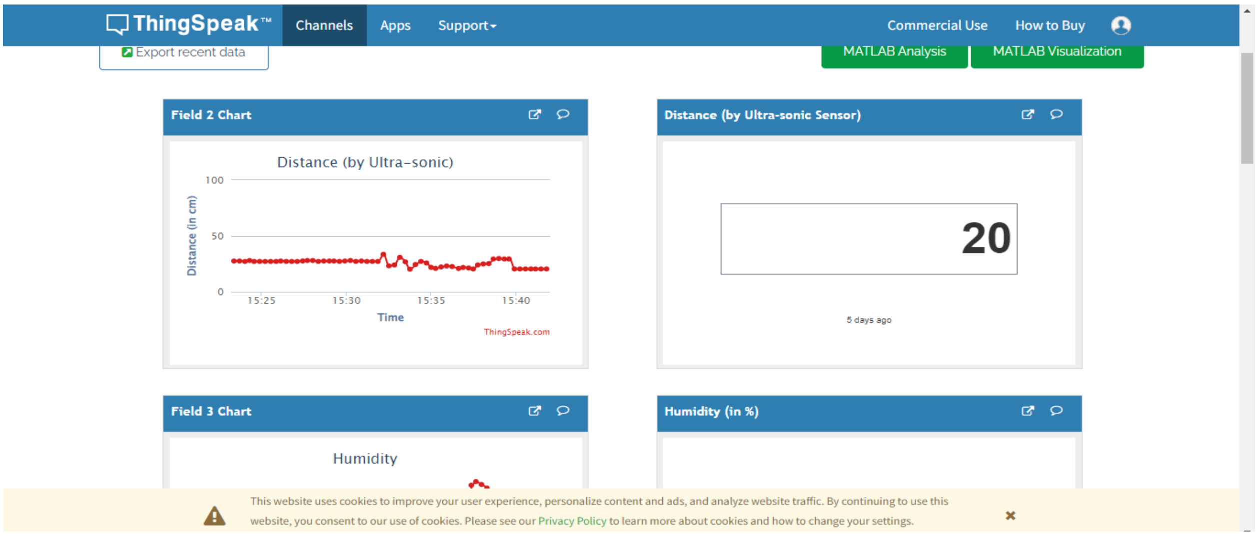


Fig 4.7 Motion detection during landslide phase.

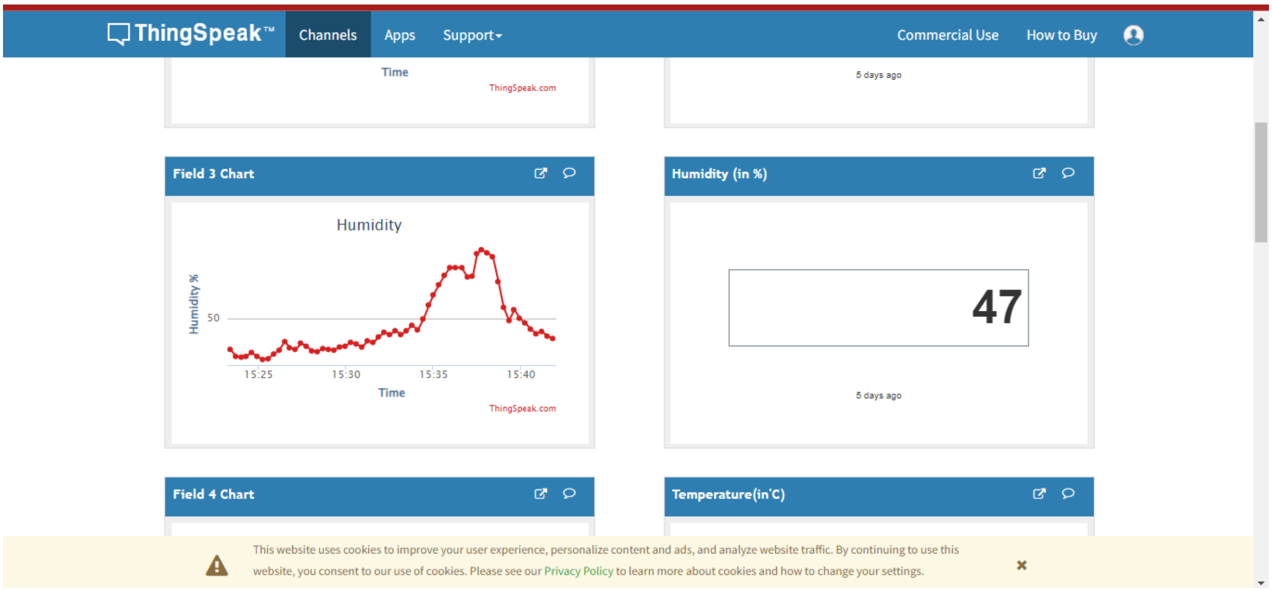


Fig 4.8 Humidity changes observed during landslide phase.

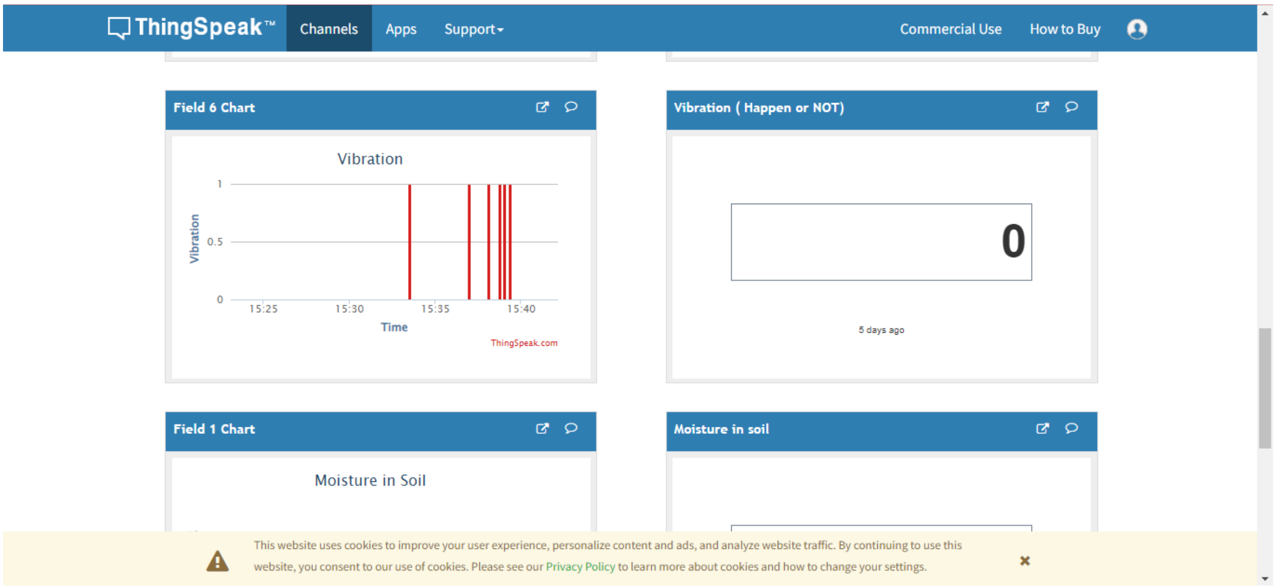


Fig 4.9 Vibration in structure observed during landslide phase.

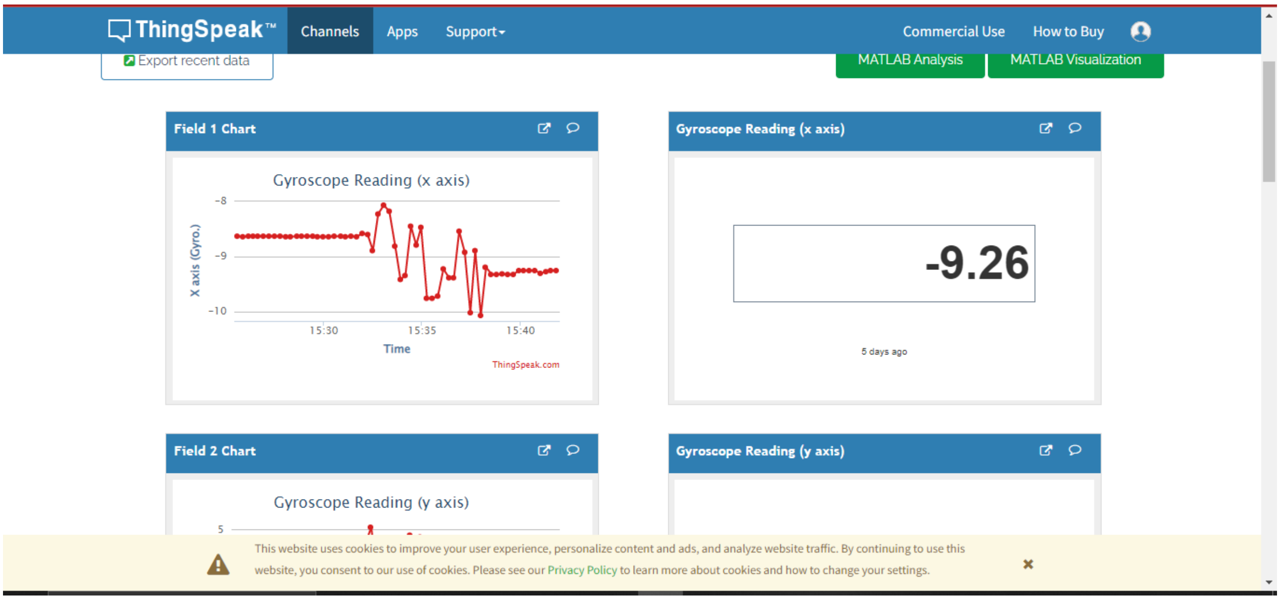


Fig 4.10 Changes in the core of the mountain at x-axis observed during landslide phase.

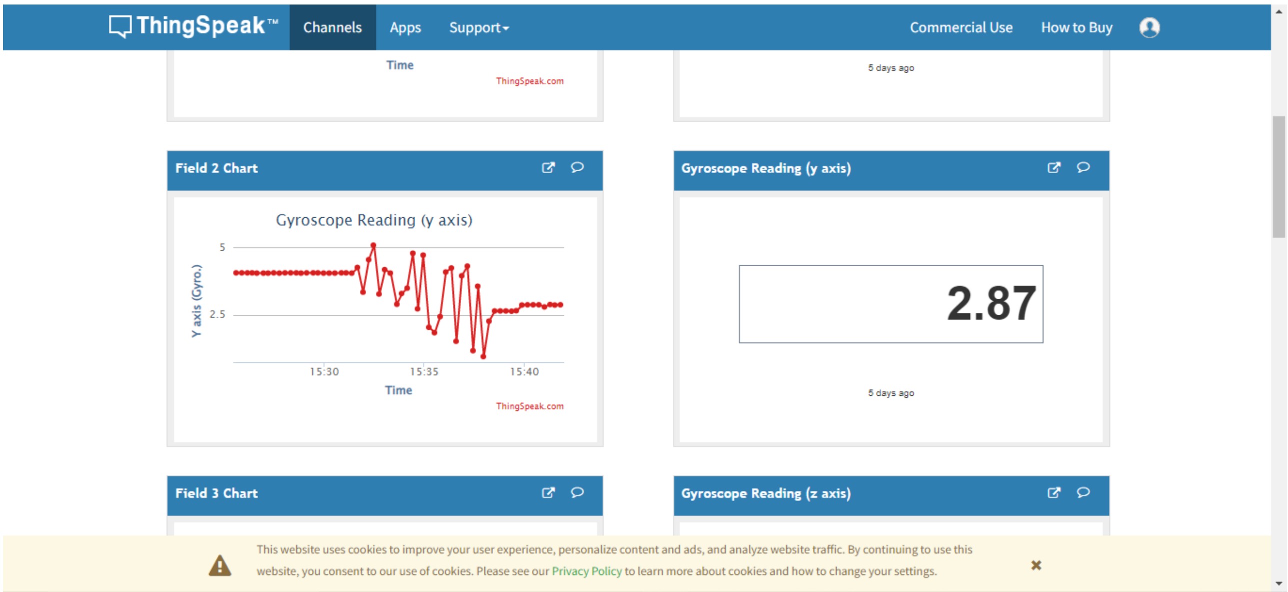


Fig 4.11 Changes in the core of the mountain at y-axis observed during landslide phase.

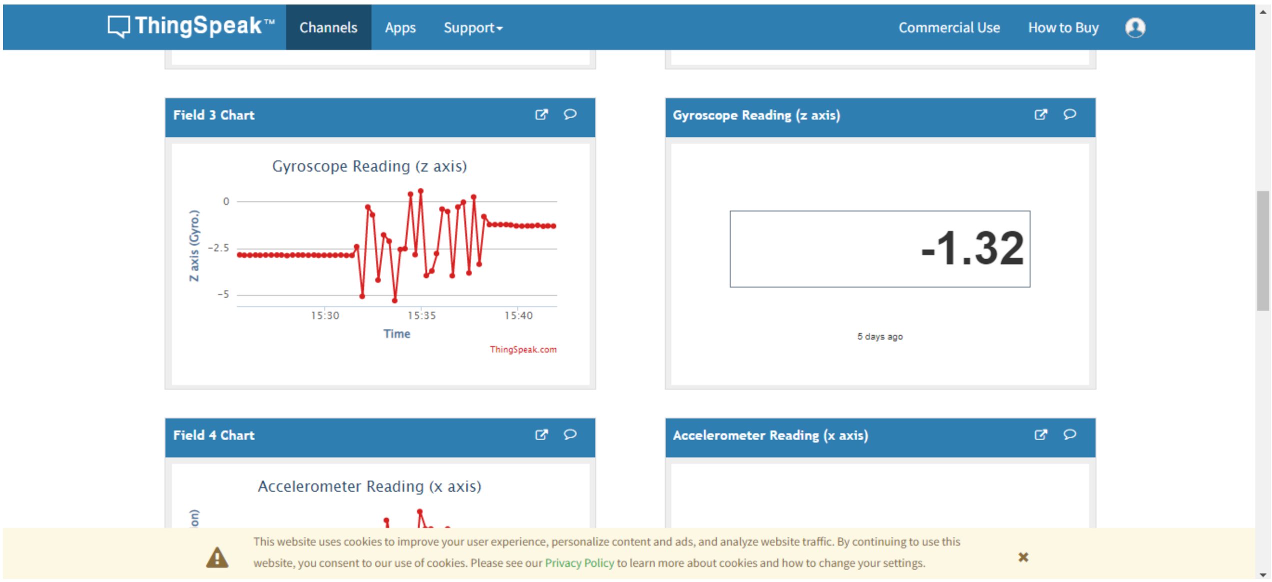


Fig 4.12 Changes in the core of the mountain at z-axis observed during landslide phase.

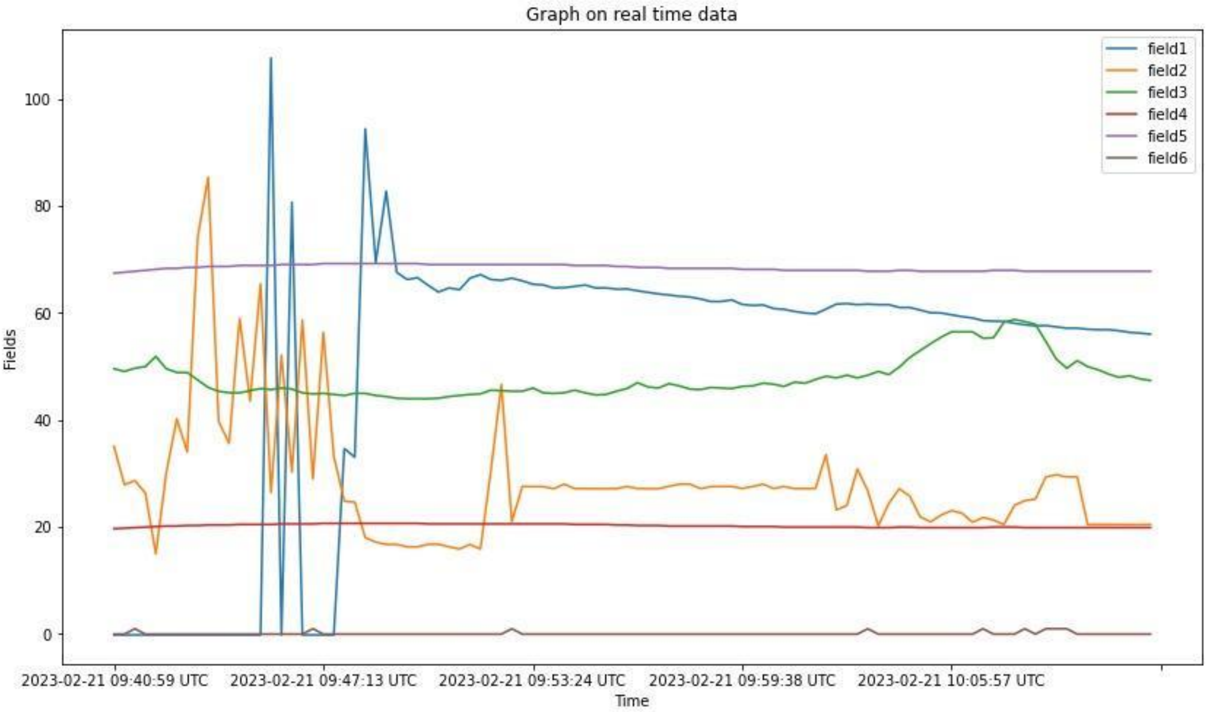


Fig 4.12 Changes in all the factors observed during the landslide phase.

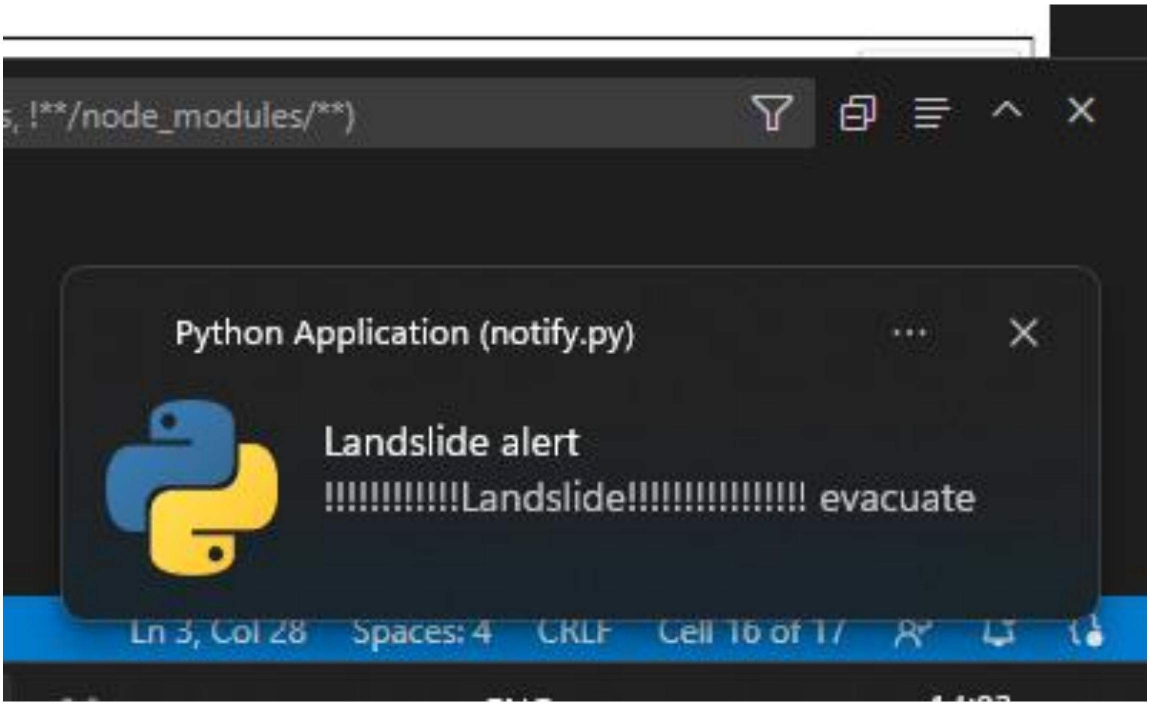


Fig 4.10 After analyzing changes in the data, the alert system is activated and messages are transmitted to the users.

Conclusion

In the recent years, landslides have grown to be a very serious issue in the Uttarkashi and Himachal region. Due to the area's very fresh, young, and delicate geological formations, land slides have been a nightmare for the locals. An key step in doing this kind of research is to thoroughly evaluate the likelihood of landslides in the future. In this situation, a landslip prediction modelling is a useful tool. In the past several years, machine learning techniques have gained popularity since they have already delivered results that are incredibly accurate. Various machine learning techniques were selected to examine slope instability.

This project's primary objective is to work with the real-time data that is being gathered and produce forecasts based on various parameters. The methodology and predictive variables used in this study were selected based on the data that was gathered, the project's objectives, and the local environmental conditions. When dealing with the landslip prediction model, visualisation has shown to be a reliable tool for quickly examining prediction outputs. By adding dynamic and detailed map layers, ML techniques can enhance model performance and offer strong visualisation tools that are helpful for managers and decision-makers in the prevention of landslip disasters.

References

- https://www.google.com/search?q=moisture+sensor&rlz=1C1ONGR_enIN1032IN1032&sxsrf=APwXEdfoQr6cdbX4VAH0rwujay_7e1DBmg:1682238450762&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjy2qPKyr_-AhUZ82EKHdspAkEQ_AUoAnoECAEQBA&biw=1366&bih=617&dpr=1#imgrc=IEXIszSMMR_sZM
- https://www.google.com/search?q=ultrasonic+sensor&tbm=isch&ved=2ahUKEwjGpsSBy7_-AhWwwTgGHUSeAuAQ2-cCegQIABAA&oq=ultrasonic+sensor&gs_lcp=CgNpbWcQAzIKCAAQigUQsQMQQzIHCAAQigUQQzIFCAAQgAQyBQgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEOgQIIxAnOgcIIxDqAhAnOggIABCxAxCDAToICAAQgAQQsQM6BAgAEAM6CwgAEIAEELEDEIMBUPEFWJY5YNo6aAFwAHgFgAG0BYgBzyqSAQwwLjEwLjkuMC4yLjGYAQCgAQGqAQtnd3Mtd2l6LWltZ7ABCsABAQ&sclient=img&ei=ZuxEZMbbJrCD4-EPxLyKgA4&bih=617&biw=1366&rlz=1C1ONGR_enIN1032IN1032#imgrc=WQ5JeSFJfT5NAM
- https://www.google.com/search?q=temperature+sensor&rlz=1C1ONGR_enIN1032IN1032&hl=en&sxsrf=APwXEdeu8kKL-am9V5iS1mjcsXB3_92vSQ:1682239003842&source=lnms&tbm=isch&sa=X&ved=2ahUKEwin8oDSzL_-AhXNqlYBHYFkCWkQ_AUoAnoECAEQBA&biw=1366&bih=560&dpr=1#imgrc=NW_KVGPuUinP_M
- https://www.google.com/search?q=Vibration+sensor&rlz=1C1ONGR_enIN1032IN1032&hl=en&sxsrf=APwXEdfe0Z89J0VBvEGgqF6Hj6lKWcN_6Q:1682239244321&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj50tbEzb_-AhVbzzgGHZCwD5UQ_AUoAXoECAEQAw&biw=1366&bih=560&dpr=1#imgrc=VU3lxMCV6gKbHM
- https://www.google.com/search?q=gyroscope+sensor&rlz=1C1ONGR_enIN1032IN1032&hl=en&sxsrf=APwXEdfG_631PNggKc3Ckc-TwVrMKmHqVQ:1682239465716&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjbzZ-uzr_-AhVYMd4KHUUGC64Q_AUoAnoECAEQBA&biw=1366&bih=560&dpr=1#imgrc=X7KCIxiBmBdcBM
- https://www.google.com/search?q=accelerometer+sensor&rlz=1C1ONGR_enIN1032IN1032&hl=en&sxsrf=APwXEddUJyj2FM-qrkHysRqGQudEbZ5ZLQ:1682239683219&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjK5fqVz7_-AhUtsFYBHTTGC1QQ_AUoAnoECAEQBA&biw=1366&bih=560&dpr=1#imgrc=cEuXZnJ-R63T_M
- https://www.researchgate.net/publication/339840571_Landslide_identification_using_machine_learning
- Gariano, S. L. & Guzzetti, F. Landslides in a changing climate. Earth Sci. Rev. 162, 227–252 (2016)
- Poonam Kainthura & Neelan Sharma, Hybrid machine learning approach for landslide prediction, Uttarakhand, India(2022)