International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Predictive Analysis of Snow-slide based on Temperature Monitoring

Dinesh PM, Pavithran S, Priyanka S

Abstract*: Avalanche prediction is a complex process which involves the study and understanding to predict certain parameters that are varying continuously. We predict the output using the decision tree algorithm. We apply this algorithm to the dataset. Dataset is formed by studying the factors affecting avalanche and we fix five parameters which mostly affects the occurrence of avalanche. By applying the decision tree algorithm we will be able to predict the output. Using this we can save the life of people who live in the snow regions. In this, the user should enter the values of the parameters. In future, we can use sensors to determine the value of all the five parameters.

***Keywords* : Avalanche Prediction using machine learning, Snow slide prediction.**

I. INTRODUCTION

The word Avalanche refers to snow and ice. It means a mass of snow, ice, rocks, slush falling rapidly down a mountain. Avalanche has two types; one is snow slide avalanche and another one is rockslide avalanche. Slides of snow and ice from the mountains is known as snow slide avalanche. Slides of rocks and debris from the mountains is known as rockslide avalanche. The remainder of the paper refers to snow slide avalanche.

The avalanche occurs mainly due to the gravity, because gravity always tries to get the object or material close to the earth. This is applicable in the case of avalanche also.

Avalanche in past were caused by earthquakes, temperature, wind, rain, slope and sometimes by the human activities also. In recent times, avalanche is caused due to the human activities. The use of snow mobiles, terrain vehicles, skiers creates vibrations. These vibrations combined with the gravitational pull is the quickest way to cause an avalanche. This paper predict that the avalanche will occur due to some natural parameters and not by the above-mentioned human activities.

Many type of snow avalanches are available. They are

- Wet avalanches
- Powder avalanches
- Dry slab avalanches
- Loose snow avalanches
- Ice and cornice fall avalanchese.

A. Wet avalanches

Wet avalanches are due to the increase in temperature which makes the snow to melt and also due to the rainfall causing the water entering the layers of snow which reduce the strength of the snow pack and cause an avalanche. This type of avalanche travels 10 to 20 miles per hour.

B. Powder avalanches

Powder avalanche is also called as dry avalanche. This has powder cloud which lies above the slab snow. It can be formed by any type of snow and initiation mechanism. But it mainly occurs due to the fresh snow fallen recently. This powdered cloud consists of 70 percentage of the air and 30 percentage of the dust snow. This type of snow reaches the speed of 190 miles per hour.

C. Dry slab avalanches

Slab means large thick piece of rock or anything. In our case, it is a snow slab. It is formed due to the old snow and decrease in the temperature which makes the snow like slab. Slab is about half the size of the football ground and two or three feet deep. This slab slides easily from the mountain. It reaches the speed of 220 miles per hour.

D. Loose snow avalanches

Loose snow slides from the mountain is called as loose snow avalanche. Small loose snow avalanche is also called as sluffs. This type of snow avalanche occurs in very steep terrain. The climbers and the skiers are mostly killed by this type of snow avalanche.

E. Ice and cornice fall avalanches

It occurs due to the breaking off over hanging ice or snow. This can trigger another type of avalanches also.

II. BLOCK DIAGRAM

In this paper, we follow some steps to build and test our model as shown in figure 1. First step is the collection of data. Data can be taken from websites or taken the survey or by with the domain knowledge. By using those data, we want to pre-process it. Pre-Processing means remove or alter the unwanted data. Then by using that data we want to choose the suitable algorithm and by using that algorithm we want to train our model. After this we want to test our model and deploy the model for the real time use.

Manuscript published on 30 September 2019
* Correspondence Author
Dinesh P M*, Assistant professor, Electronics and communication Engineering, Sona College of Technology, Salem, India. Email: pmdineshece@live.com
Pavithran S, Electronics and communication Engineering, Sona College of Technology, Salem, India. Email: 619pavithran@gmail.com
Priyanka S, Electronics and communication Engineering, Sona College of Technology, Salem, India. Email: priyasenkamal29@gmail.com

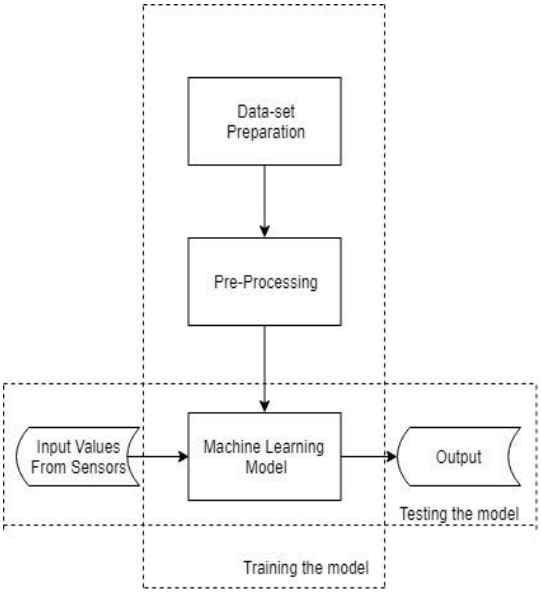

Fig. 1.Bock Diagram

III. DATA PREPARATION

We cannot take any existing dataset. So, we decide to create our own dataset based on some parameters. We studied and found five parameters which causes the occurrence of avalanche mostly and the parameters are

- Temperature
- Slope
- Wind
- Snow Density
- Forest Density

A. Temperature

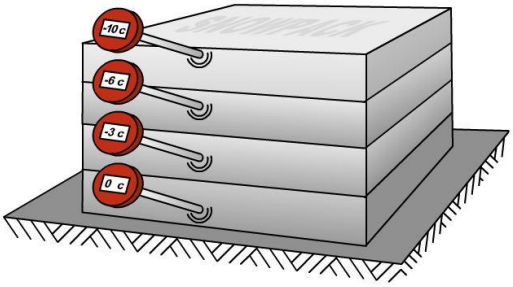

Fig. 2.Temperature of different layers

The temperature greatly affects the behavior of the avalanche. From figure 2, If increase in temperature leads to melting of the snow, then the melted water enters the layers of the snow and it reduce the strength of the layers. If decrease in temperature leads to the formation of the slab in the upper layer of the snow. Under the slab layer the temperature is reduced to 1 degree per 10 centimeters. As a result, the last layer is in high temperature and upper layer is in the form of slab and which causes the slab avalanche. When the temperature is less than -10 degree Celsius and more than 8 degree Celsius then the chance of occurrence is very high. When the temperature is between -10 degree Celsius and 8 degree Celsius then the chance of occurrence is low.

B. Slope

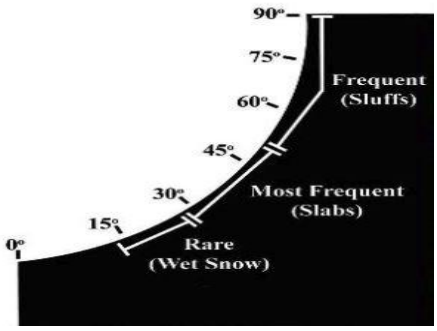

Fig. 3.Gives the idea of slop

Slope is one of the significant parameters which affects the behavior of the avalanche. In this parameter, gravity places a vital role. Slab avalanche most frequently occur in steep regions. From the figure 3 when the slope is in the range of 0 degree to 30 degree then the chance of avalanche occurrence is low. When the slope is between 30 degree and 50 degree then the chance of avalanche occurrence is medium. When the slope is more than 50 degree then the chance of avalanche occurrence is high.

C. Wind

Wind erodes the snow from one side of the slop of the mountain to another side. So, it makes larger layer of snow at one side of the mountain. This condition leads to the occurrence of an avalanche. When the wind speed is more than 30 miles per hour then that condition is critical. When the wind speed is less than 10 miles per hour then the condition is normal. When the wind speed is between 11 miles per hour to 22 miles per hour then that condition is moderate occurrence of avalanche.

D. Snow Density

If the snow density is more, then the chance of occurrence of the avalanche is higher. Because due to the gravity the upper layer of the snow is having more chance to slide than the lower layer snows. When the snow density is less than 20 meters then there is no chance of occurrence of avalanche. When the snow density is between 20 meters and 50 meters then medium possibility of occurrence of avalanche. When the snow density is more than 50 meters then there is a high possibility of occurrence of avalanche.

E. Forest Density

Forest protect against avalanche because the trees in the accident zone which reduce the impact of the descending snow from the mountain. If there are no trees in that region then the side of the mountain is plain, then descending snow slides without any obstacles and it destroys more. So, the forest density decides the size of avalanche.

Slope	Forest De	Snow Den	Air Tempe	Wind	Prediction
59	H	52	-17	19	2
24	L	23	-8	19	0
15	H	30	7	19	0
27	L	61	-13	25	1
34	H	4	-20	7	0
34	L	76	-7	40	1
34	M	73	-32	27	2
34	L	4	3	9	1

Fig. 4.Sample dataset

IV. ALGORITHM

We use decision tree algorithm to predict the avalanche occurrence. This algorithm comes under the category of supervised learning. It solves both regression and classification problems. Here our problem is classification. Regression means predicting the continuous data. Classification means predicting the categorical data. Our categories are high, medium and low. High means occurrence of avalanche is high. Medium means may be the avalanche will occur or not. Low means there is no possibility of occurrence of avalanche.

Decision tree algorithm uses tree representation to solve the problems. A decision tree is a one of the ensemble techniques where each node represents a feature/attribute, each link/Branch represents a decision/Rule and each leaf represents an outcome. Decision tree has terminology like root node, Parent/child node, leaf node, Splitting and pruning. Root node represents the starting point of the tree and this node gets divided into two or more homogenous sets. Root node is the parent node and all the other nodes branched from it is known as child node. Leaf node represents the end of the tree or a branch and this node cannot be further classified. Splitting divides, the root node or sub node into different parts on the basis of some condition. Pruning is the opposite of splitting and it removes the unwanted branches from the tree.

We can implement the decision tree algorithm in two methods, one is CART (Classification and Regression Trees) which has the concept of Gini Index and ID3 algorithm which has the concept of entropy function and information gain. Here we use ID3 algorithm to decide which is the root node and what are all the sub nodes.

First step in the algorithm is to set the root node from the five-independent variable. Entropy and Information gain is used to decide the root node.

A. Entropy

Entropy which measures the amount of uncertainty in the dataset. It means that it gives the result based on the number of positive and the number of negative samples. For example, if we have equal number of positive and negative samples then the entropy is one. If we have only positive samples or only negative samples then the entropy is zero.

$$Entropy = -\frac{p}{p+n}log2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}log2\left(\frac{n}{p+n}\right)$$

Where p represents the number of positive samples and n represents the number of negative samples.

B. Information Gain

Information gain is the difference of entropy for all the datasets with the entropy for individual attributes.

$$Gain = Entropy(s) - I(Attribute)$$

Where Entropy(s) represents the entropy for the dataset S and I (Attribute) represents the entropy for the individual independent parameters.

IV. PSEUDOCODE

First, we want to find the entropy of whole the data set. Then find the entropy for each attribute they are temperature, slope, wind, snow density, forest density. Next find the information gain which means differentiate total entropy with the individual entropy. Then find which information gain is maximum. Then that attribute is selected for the root node. This procedure is repeated until all the branches are end with leaves or labels.

V. PREPROCESSING

For our data set we have Forest density with the values L, M, H. L means minimum trees in the slope of the mountain or a terrain. M means moderate amount of trees are present in the slope of the mountain or a terrain. As same as H which means more number of trees are in the side of the mountain or a terrain. Machine learning which uses mathematical formulas to predict the output. But here we have character as an input. So, we need to convert those characters into numeric values. For that we use LabelEncoder and OneHotEncoder to convert character data into numeric data. For our case LabelEncoder converts L as 0, M as 1 and H as 2. Then OneHotEncoder coverts those numeric data into binary values with separate columns.

VI. TESTING

Our dataset has 204 rows, so, we split those data for training and testing with the ratio of 8: 2. Then 163 rows for training and 41 rows for testing. It predicts 0, 1 or 2. 0 for the avalanche occurrence is minimum, 1 for moderate possibilities of occurrence of avalanche and 2 for the maximum possibilities of occurrence of avalanche. After training and testing we get the accuracy score of 0.8292 which means our model predicts the 80 percentage correctly.

VII. VALIDATION

Below is the confusion matrix for our result. What confusion matrix does? The Confusion matrix gives the summary of the predicted result. It is mainly used for the classification problem. It gives the number of correctly predicted and incorrectly predicted counts. Why it called as Confusion matrix? It shows the way in which how our model is confused with the predicted value.Confusion matrix of two dependent variable classification model is shown below

TABLE 1: CONFUSION MATRIX OF TWO CLASSES

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	True Positive	False Negative
Class 2 Actual	False Positive	True Negative

True Positive means number of times actual positive value is predicted as positive. True Negative means number of times actual negative value is predicted as negative. False Positive means number of times actual negative value is predicted as positive. False negative means number of times actual positive value is predicted as negative.

But our model has three classes, so our confusion matrix looks like below table

TABLE 2 : CONFUSION MATRIX OF THREE CLASSES

	Class 1 Predicted	Class 2 Predicted	Class 3 Predicted
Class 1 Actual	True Class 1	False Class 2	False Class 3
Class 2 Actual	False Class 1	True Class 2	False Class 3
Class 3 Actual	False Class 1	False Class 2	True Class 3

Here True Class 1, 2 and 3 are the output values of actual class 1, 2 and 3 are predicted correctly as Class 1, 2 and 3. False class 1 means number of times wrongly predicted as class 1. False class 2 means number of times wrongly predicted as class 2. False class 3 means number of times wrongly predicted as class 3.

Below is the confusion matrix of our model from that our model predicts 35 prediction correctly out of 41 that means our model predicts 80 percentage of the data correctly.

TABLE 3: CONFUSION MATRIX OF OUR MODEL

	Class 1 Predicted	Class 2 Predicted	Class 3 Predicted
Class 1 Actual	09	00	01
Class 2 Actual	04	16	00
Class 3 Actual	00	01	10

VIII. CONCLUSION

By using this method, we achieved 80 percentage of accuracy in the prediction of avalanche. In this paper we are using the values of the five parameters from the user. In future, we can use sensors like gyroscope sensors which measure the slope of the mountain. The remaining parameters can also be found using their respective sensors.

REFERENCES

1. Avalanche.org » Connecting the public to avalanche information and education [Internet]. Avalanche.org. 2019 [cited 13 August 2019]. Available from: https://avalanche.org/
2. Pozdnoukhov A, Purves R, Kanevski M. Applying machine learning methods to avalanche forecasting. Annals of Glaciology. 2008; 49:107-113.
3. Choubin B, Borji M, Mosavi A, Sajedi-Hosseini F, Singh V, Shamshirband S. Snow avalanche hazard prediction using machine learning methods. 2019.
4. Northwest Avalanche Center [Internet]. Northwest Avalanche Center. 2019 [cited 13 August 2019]. Available from: https://www.nwac.us/
5. Avalanche [Internet]. En.wikipedia.org. 2019 [cited 13 August 2019]. Available from: https://en.wikipedia.org/wiki/Avalanche
6. How to predict an avalanche with radar and GPS [Internet]. Science | AAAS. 2019 [cited 9 September 2019]. Available from: https://www.sciencemag.org/news/2015/04/how-predict-avalanche-radar-and-gps
7. Howley E. Snow avalanche prediction in Grasdalen, Norway. Saarbrücken: VDM Verlag Dr. Müller; 2008.
8. Mitchell T. Machine learning. New York: McGraw Hill; 2017.
9. Machine Learning - GeeksforGeeks [Internet]. GeeksforGeeks. 2019 [cited 9 September 2019]. Available from: https://www.geeksforgeeks.org/machine-learning/
10. National Geographic: Stories of Animals, Nature, and Culture [Internet]. Nationalgeographic.com. 2019 [cited 9 September 2019]. Available from: https://www.nationalgeographic.com/
11. Predicting the 2020 Stanley cup using machine learning text prediction, part 1 [Internet]. Reddit. 2019 [cited 9 September 2019]. Available from: https://www.reddit.com/r/hockey/comments/c55o25/predicting_the_2020_stanley_cup_using_machine/
12. How (not) to use Machine Learning for time series forecasting: Avoiding the pitfalls [Internet]. Medium. 2019 [cited 9 September 2019]. Available from: https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424
13. Brownlee J. How To Implement The Decision Tree Algorithm From Scratch In Python [Internet]. Machine Learning Mastery. 2019 [cited 9 September 2019]. Available from: https://machinelearningmastery.com/implement-decision-tree-algorithm-scratch-python/
14. Himalayas | History, Map, & Facts [Internet]. Encyclopedia Britannica. 2019 [cited 9 September 2019]. Available from: https://www.britannica.com/place/Himalayas
15. Yahoo is now part of Oath [Internet]. News.yahoo.com. 2019 [cited 9 September 2019]. Available from: https://news.yahoo.com/indian-team-treks-missing-climbers-himalayan-slope-132329988.html

AUTHORS PROFILE

P.M.Dinesh is currently working as Assistant Professor of ECE Department, at Sona College of Technology, Salem. He is also a Team member of Sona SIPRO, Sona Signal and Image PROcessing Research Centre. He completed his Bachelors in Engineering in Electronics and Communication Engineering in the year 2009 and Masters in Engineering in VLSI Design in the year 2011. He is currently pursuing his Doctoral Research in the area of Image Processing. His areas of interest include pattern recognition, pattern classification, text recognition and texture analysis. He has published more than twenty papers in leading conferences and journals in the country.

Pavithran S is currently a final year student of Electronics and communication Engineering, Sona College of Technology, Salem, India. His area of interest are Machine learning, text processing and digital image processing.

Priyanka S is currently a final year student of Electronics and communication Engineering, Sona College of Technology, Salem, India. Her area of interest are Embedded systems, machine learning and digital electronics.