AVALANCHE FORECASTING PREDICTION

Using Gradient Boosting Classifier

Developed by: Poornima G B, Sonika Prakash, Spoorthi R S, Ramyashree G T, Anu Bai

Smart Bridge - Remote Summer Internship Program

1.INTRODUCTION

There are various areas in which Machine Learning can be used in weather forecasting sectors. One of the main benefits of introducing machine learning to weather forecasting is more accurate predictions. Machine learning can be used to process immediate comparisons between historical weather forecasts and observations. With the use of machine learning, weather models can better account for prediction inaccuracies, such as overestimated rainfall, and produce more accurate predictions.

Avalanche forecasting involves the assimilation and prediction of data and information describing weather ,snow pack and stability within a given time period and spatial extent, and assimilating this information to assess the likelihood of avalanches in the future. In conventional avalanche forecasting, this process is carried out with little or no direct use of numerical models by avalanche forecasters, who tend to apply a range of diverse and redundant data sources to the problem. Avalanche forecasts may be provided for periods in the future ranging from the next few hours through to the next day to forecasts covering periods of several days in areas with relatively low temporal variability in weather conditions. Similarly, spatial forecast scales can vary from a specific, through local forecasts for a particular region to regional forecasts for a significant part of a mountain range.

There are two most important goals for prediction and description. Prediction involves using some variables in data set to predict unknown values of other variables and description concentrates on finding patterns describing the data that can be interpreted by human. The derived knowledge must be new, not obvious, relevant and can be applied in the field where this knowledge has been obtained. It is also the process of extracting useful information from raw data.

There are huge number of phases in the prediction based on Machine Learning. Data collection is the first phase, where data should be collected. It should be huge data set according to the requirements. One should collect or create the data for the prediction. Data Preprocessing is the second phase and this contains a lot of sub-phases for the processing of the data. It includes importing libraries, data

visualization, data transformation, feature scaling, splitting and label encoding. In data splitting, the data is to be split into two as train_data and test_data for the training of the model. Then the fourth phase is model training. The last phase is model evaluation and testing and it is to develop the simplest model able to formulate a target value fast and well enough. A data scientist can achieve this goal through model tuning. That's the optimization of model parameters to achieve an algorithm's best performance.

1.1 OVERVIEW

The word Avalanche refers to snow and ice. It means a mass of snow, ice, rocks, slush falling rapidly down a mountain. Snow avalanches are among the most destructive natural hazards threatening human life, ecosystems, built structures, and landscapes in mountainous regions. Each year avalanche kills more than 150 people worldwide. The most common cause of death by avalanche is asphyxiation. If the person buried under an avalanche for more than 15 minutes, then there is no chance of surviving it. So, the life of the people in that region is difficult to live.

This project prevents the people from the avalanche by priory informing them there is a chance to the occurrence of avalanche or not. The model gets the data from the IOT based sensors. After that we want to process those data using a suitable algorithm, then our model display whether the avalanche occur or not and how strength it was. To analyze the data coming from different sensors we are applying various machine learning algorithms. If there is a chance of avalanche then the notification will be sent to people so that they can take decisions accordingly.

1.2 PURPOSE

The aim of this project is to use the IBM cloud Auto AI service to build and train the model using the Gradient Boosting technique. We then use the Node Red service for the application building. If there is a chance of occurrence of avalanche, it prints how much the danger is.

2. LITERATURE SURVEY

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process. Different data mining techniques include classification, clustering, association rule

mining, prediction and sequential patterns, neural networks, regression etc., Classification is the most commonly applied data mining technique, which employs a set of preclassified examples to develop a model that can classify the population of records at large. Avalanche detection and weather prediction are particularly well suited to classification technique. This approach frequently employs Gradient Boosting classification algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

2.1 EXISTING PROBLEM

Snow avalanches are the most destructive natural hazards threatening human life, ecosystems, built structures, and landscapes in mountainous regions. Each year avalanche kills more than 150 people worldwide due to asphyxiation. So, the life of the people in that region is difficult to live.

Many models have been built previously that predict the occurrence of avalanche. But they have high time complexity and space complexity. But this model makes use of the Gradient Boosting Classification algorithm which gives us an accuracy of more than 80%. It is also has a user friendly interface alerts the people living in those regions about the forthcoming avalanches.

2.2 PROPOSED SOLUTION

Machine Learning (Gradient Boosting Classifier):

Gradient Boosting Classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting. Gradient boosting models are becoming popular because of their effectiveness at classifying complex data sets.

We have also created a UI using the Node-Red service of the IBM Cloud for the avalanche prediction. This UI allows the users to predict the danger very easily and is user friendly. It predicts the occurrence of avalanche and its strength by just entering few necessary details like humidity, temperature, snow depth, forest density, etc.,

3. THEORETICAL ANALYSIS

While selecting the algorithm that gives us an accurate prediction, we went through a lot of algorithms with the help of the AutoAI experiment which gave the results abruptly accurate. We then selected the algorithm with the highest accuracy, which is the Gradient Boosting Classification algorithm which gave us an accuracy of above 80%. Classification algorithms frequently use logarithmic loss, while regression algorithms can use squared errors. Gradient boosting systems don't have to derive a new loss function every time the boosting algorithm is added, rather any differentiable loss function can be applied to the system.

The peculiarity of this problem is collecting the weather details from the user and working with the prediction at the same time. So we developed a UI for the users who want to know if there is a danger of an avalanche. Accuracy is defined as the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test data set. The formula is as follows,

$$Accuracy=TP+TN/TP+TN+FT+FN$$

Our AutoAI experiment generated 8 pipelines from the algorithms. Among these, pipeline 4 had the highest accuracy of 82.6% using the Gradient Boosting Classifier.

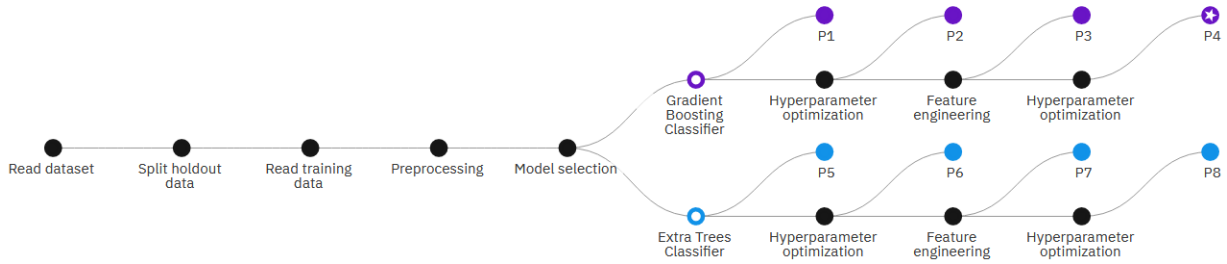
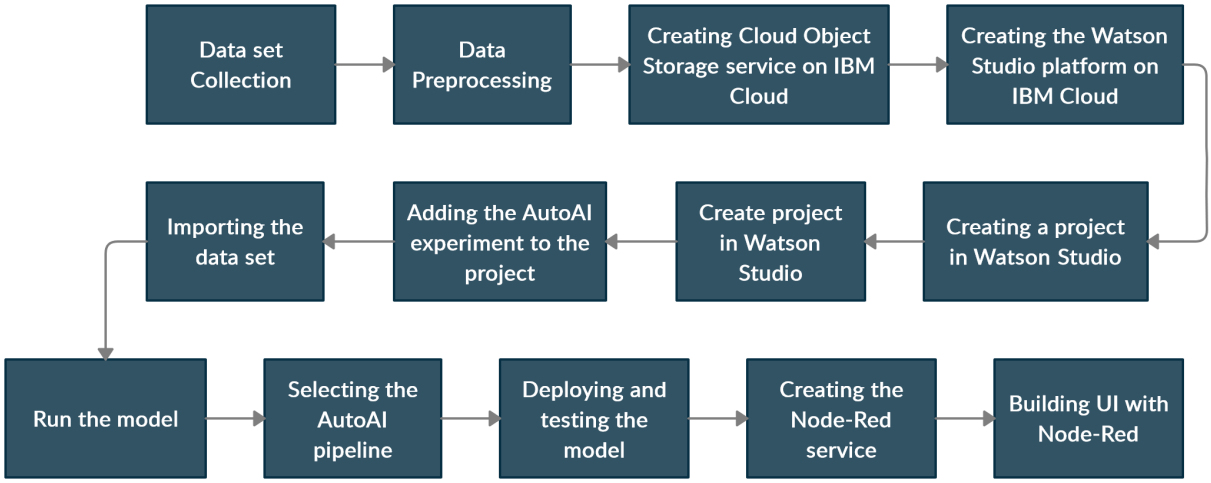


Fig-3.1 Progress map

3.1 BLOCK DIAGRAM



3.2 SOFTWARE DESIGNING

- Watson Studio
- Machine Learning service
- AutoAI experiment
- Node-Red

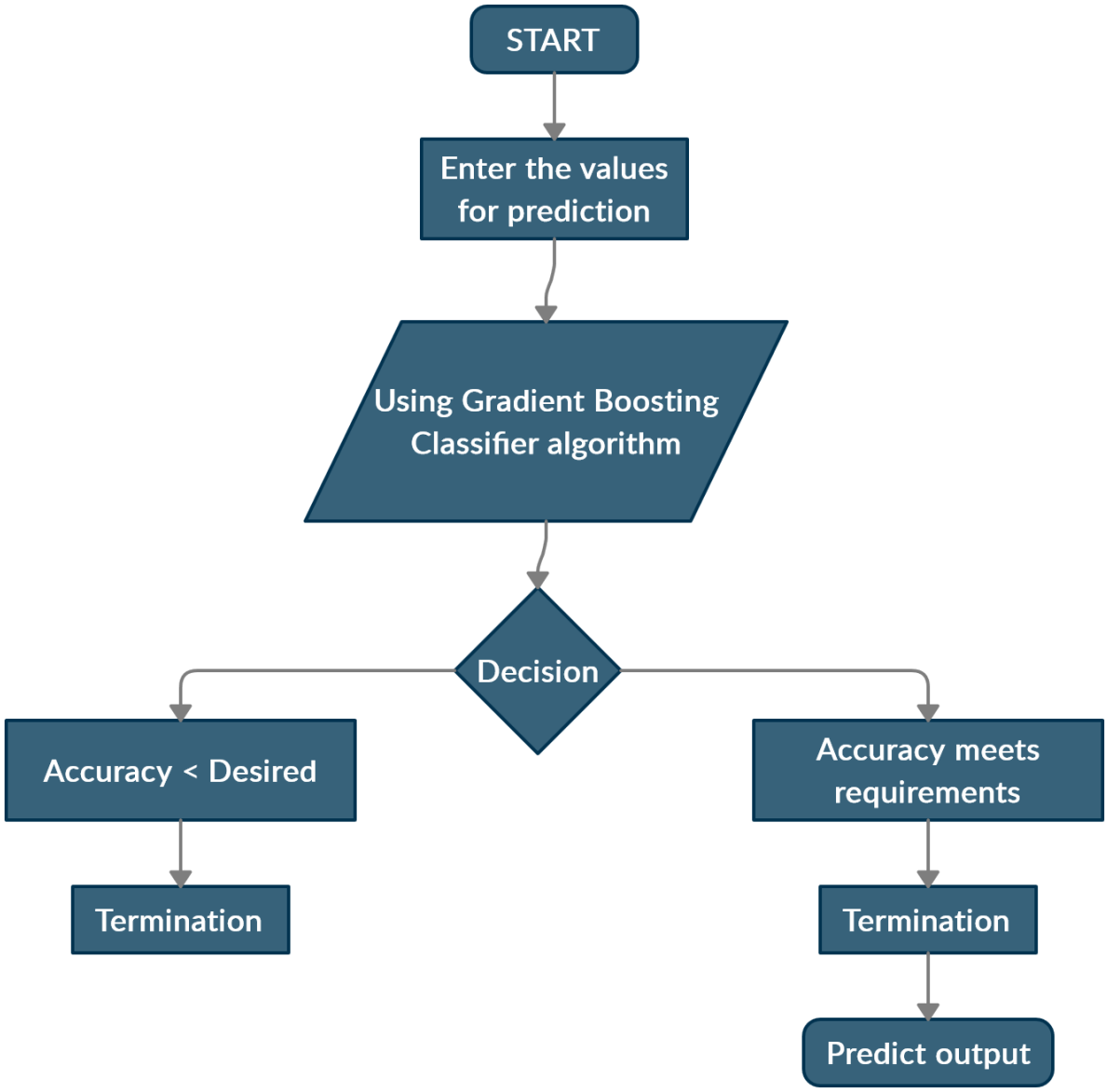
We developed this project in Watson Studio. We used the AutoAI experiment and used Machine Learning as the associated service to build the model. We then deployed the model in Watson Studio. For the application building, we used the Node-Red service.

4. EXPERIMENTAL INVESTIGATION

The data set we used contains 199 rows and 8 columns. It had no missing values. There were no outliers found. The below figure shows the first 5 rows of the data set.

	A	B	C	D	E	F	G	H
1	Date/Time (PST)	Relative Humidity (%)	Slope	Forest Density	Snow Density	Air Temperature	Wind	Prediction
2	25-06-2019 15:00	89.9	59	High	52	-17	19	High
3	25-06-2019 14:00	69.82	24	Low	23	-8	19	Less
4	25-06-2019 13:00	65.32	15	High	30	7	19	Less
5	25-06-2019 12:00	54.77	27	Low	61	-13	25	Moderate
6	25-06-2019 11:00	58.87	34	High	4	-20	7	Less

5. FLOWCHART



6. RESULT

Here, the Gradient Boosting Classifier is used for the prediction and compared with the other algorithm, Extra Trees Classifier.

The results show that Gradient Boosting Classifier algorithm has the highest accuracy

of 82.6%. The Extra Trees Classifier algorithm has an accuracy of 77.6%. So we chose the Pipeline 4 which is ranked the highest with the highest accuracy.
The below table shows the pipelines generated by AutoAI.

Rank ↑	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1	Pipeline 4	Gradient Boosting Classifier	**0.826**	HPO-1 FE HPO-2	00:00:37
2	Pipeline 3	Gradient Boosting Classifier	**0.815**	HPO-1 FE	00:01:04
3	Pipeline 2	Gradient Boosting Classifier	**0.776**	HPO-1	00:00:14
4	Pipeline 7	Extra Trees Classifier	**0.776**	HPO-1 FE	00:01:12
5	Pipeline 8	Extra Trees Classifier	**0.776**	HPO-1 FE HPO-2	00:00:30
6	Pipeline 1	Gradient Boosting Classifier	**0.771**	*None*	00:00:01
7	Pipeline 5	Extra Trees Classifier	**0.737**	*None*	00:00:04
8	Pipeline 6	Extra Trees Classifier	**0.737**	HPO-1	00:00:12

Fig-6.1 Pipeline leader board

7. ADVANTAGES AND DISADVANTAGES

Advantages:

- Easy and user friendly interface for the users who want to check the occurrence of an avalanche.
- Gradient Boosting Classifier gave us an accurate prediction up to 82.6%.
- It is widely used for managing risks of snow slide in the mountain regions.
- It can work in real time and can make predictions as soon as the required details are given to the model.

Disadvantages:

- Gives only 82.6% of accuracy.
- Needs more than a single value for prediction.
- It cannot be used from anywhere like a web application and can be used by only one user at a time.

8. APPLICATIONS

- It can be widely used to manage the risks in the mountain regions by notifying the people about the avalanche.
- To have an idea about the weather and avalanche conditions of that particular area.
- It is one of the most widely used areas of data mining in the forecasting predictions.
- Due to tremendous growth in the data the weather forecasting uses, the analysis and transformation of the huge data into useful knowledge has become a task beyond human ability.

9. CONCLUSION

In this project, we have used the Gradient Boosting Classifier to predict the occurrence of avalanche. The AutoAI experiment is used which generated 8 pipelines of Extra Trees Classifier and Gradient Boosting Classifier algorithms. The experiment showed that Gradient Boosting Classifier gave the highest accuracy. The Node-Red service is used to build the UI. There is no definite guide of which algorithm to use at any given situation. What may work on some data sets may not necessarily work on the other data sets. Therefore, always evaluate the methods using cross validation to get the reliable estimate.

10. FUTURE SCOPE

In future, more data sets about the weather conditions of different years of that region can be collected and merged into one single data set which can improve the accuracy. In further study, we can try to tune the model so as to achieve the state - of - art performance of the model. The UI can be improved and it can be made into a complete web application model which can be accessed from anyone and from anywhere.

11. BIBLIOGRAPHY

- Challa, M.L., Malepati, V. & Kolusu , S.N.R. Forecasting risk using auto regressive integrated moving average approach: an evidence from S&P BSE Sensex. *Financ Innov***4,** 24 (2018). https://doi.org/10.1186/s40854-018-0107-z
- McClung, D. M.1994. Computer assisted avalanche forecasting. *In*Hipel, K.W. and Liping, Fang, *eds.*Statistical and stochastic methods in hydrology and environmental engineering. Vol. 4. Dordrecht, etc., Kluwer Academic Publishers, 347–358. https://link.springer.com/chapter/10.1007%2F978-94-017-3081-5_26
- Schweizer, J. and Fohn, P. M. B.. 1996. Avalanche forecasting — an expert system approach. J. Glaciol., 42(141), 318–332. http://dx.doi.org/10.1017/S0022143000004172
- Durand, Y, Giraud, G. and Mérindol, L.. 1998. Short-term numerical avalanche forecast used operationally at Météo-France over the Alps and Pyrenees. Ann. Glaciol., 26, 357–366. http://dx.doi.org/10.1017/S0260305500015093
- Cagnati, A, Valt, M., Soratroi, G., Gavalda, J. and Selles, C. G.. 1998. A field method for avalanche danger-level verification. Ann. Glacial., 26, 343–346. http://dx.doi.org/10.1017/S026030550001507X