Using JMP® 11 to Predict Occurrence of Seismic Activity to Prevent Hazards During Underground Mining Shifts

Zabiulla Mohammed, Graduate Student, Oklahoma State University

Sai Vijay Kishore Movva, Graduate Student, Oklahoma State University

Abstract

Mining activity ,often is associated with a high risk due to inherent ‘mining hazards’. Among these hazards, a special case of threat called a seismic hazard occurs frequently in many underground mines. Seismic hazard is hard to detect and is comparable to an earthquake. Existing techniques such as the Seismic method and the Seismo-acoustic monitoring allow a better understanding of rock mass processes but the accuracy of these methods is not precise when it comes to the prediction of seismic occurrence. The complexity of understanding the process involved and the disproportionate ratio of low energy seismic events to that of high energy seismic events make these techniques insufficient in predicting a seismic event accurately. To ensure safety at mining workplaces, it is important to predict these events accurately. Our objective to predict occurrence of seismic activity to prevent hazards during underground mining shifts

Objective

High energy seismic activities are rare events. One of the reasons that it is hard to predict these events is due their rare existence in the population dataset. Too many non events in the dataset make it tough to predict the events. Our objective is to predict these events. The target variable is modeled as a binary variable: ‘0’indicating a nonhazardous shift and ‘1’ indicating a hazardous shift.

Data Preparation

Using a dataset of 2,584 observations with input variables measuring seismic activity within a work shift, we predict the likelihood of high energy seismic bumps leading to possible hazards in the next shift. Hence, using JMP Pro 11’s subset features, we used a method in which we created a sample with 80 percent of non events and 20 percent of events. We then used this sample to build models to see if this technique l will help in identifying a high energy seismic event.

Model Building

The dichotomous response variable ‘Class’ is modeled using Binary Logistic Regression, Neural Network, Boot Strap Forest and Decision Tree. Each of the model parameters and performances are as follows:

Logistic Regression

RSquare (U)	0.8217
AICc	-22564
BIC	5443.1
Observations (or Sum Wgts)	750

Fig1.1:Logistic Performance

Confusion Matrix

Actual / Training	Predicted 0	Predicted 1
0	597	3
1	21	129

Fig1.2:Confusion Matrix

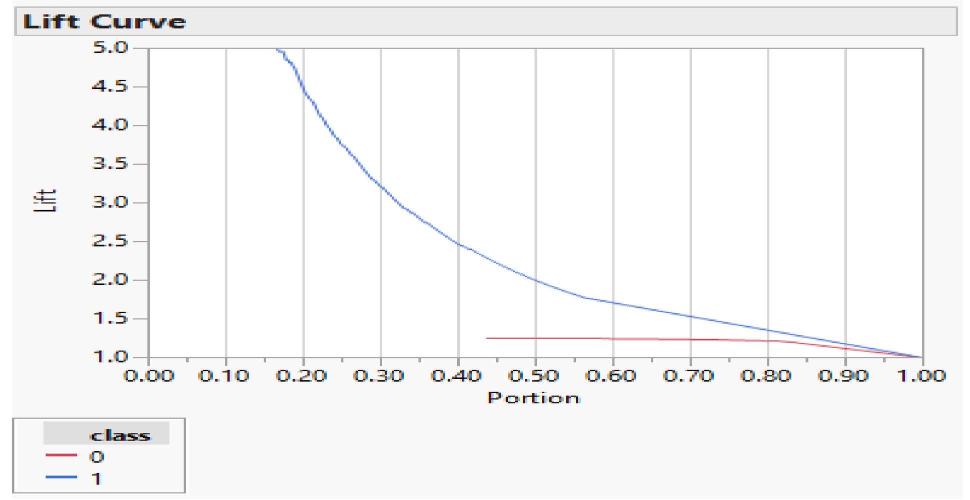

Fig1.3:Lift Curve

Neural Network

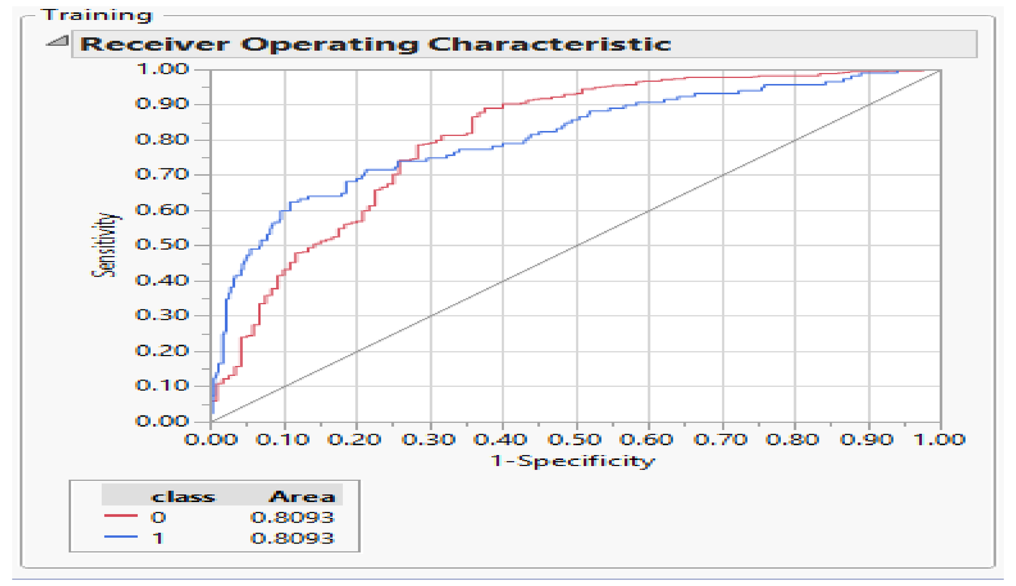

Fig2.1:Training ROC Curve

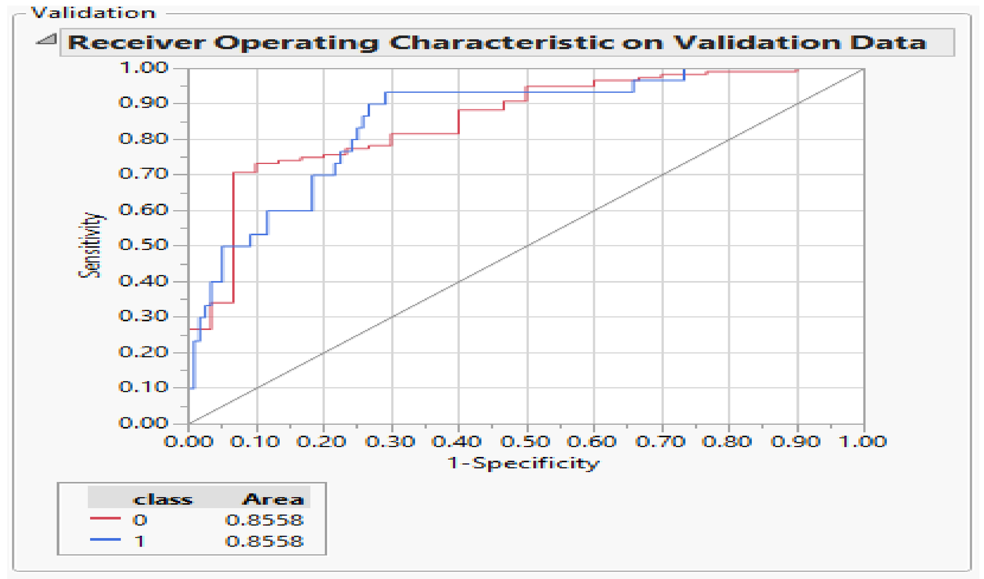

Fig2.2:Validation ROC Curve

Boot Strap Forest

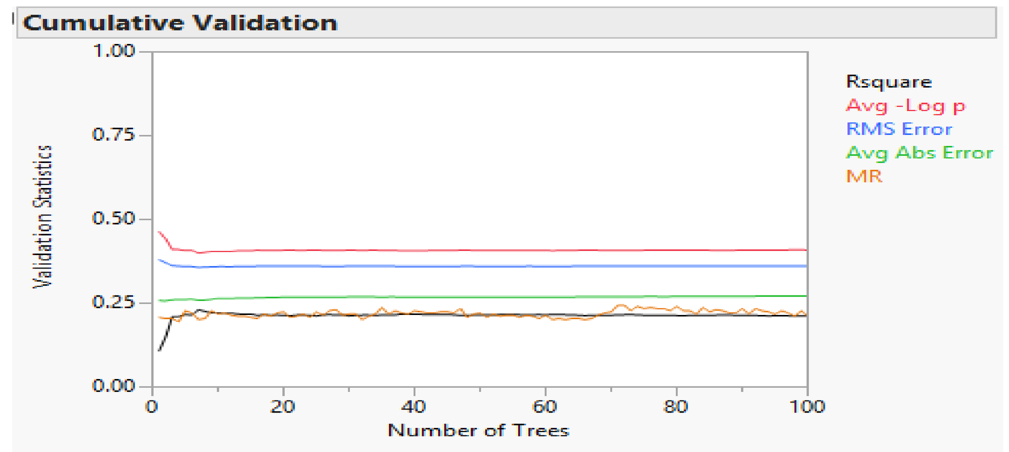

Fig3.1:Cumulative Validation

Confusion Matrix

Actual / Training	Predicted 0	Predicted 1
0	362	0
1	77	8

Actual / Validation	Predicted 0	Predicted 1
0	236	2
1	59	6

Fig3.2:Confusion Matrix

Decision Tree

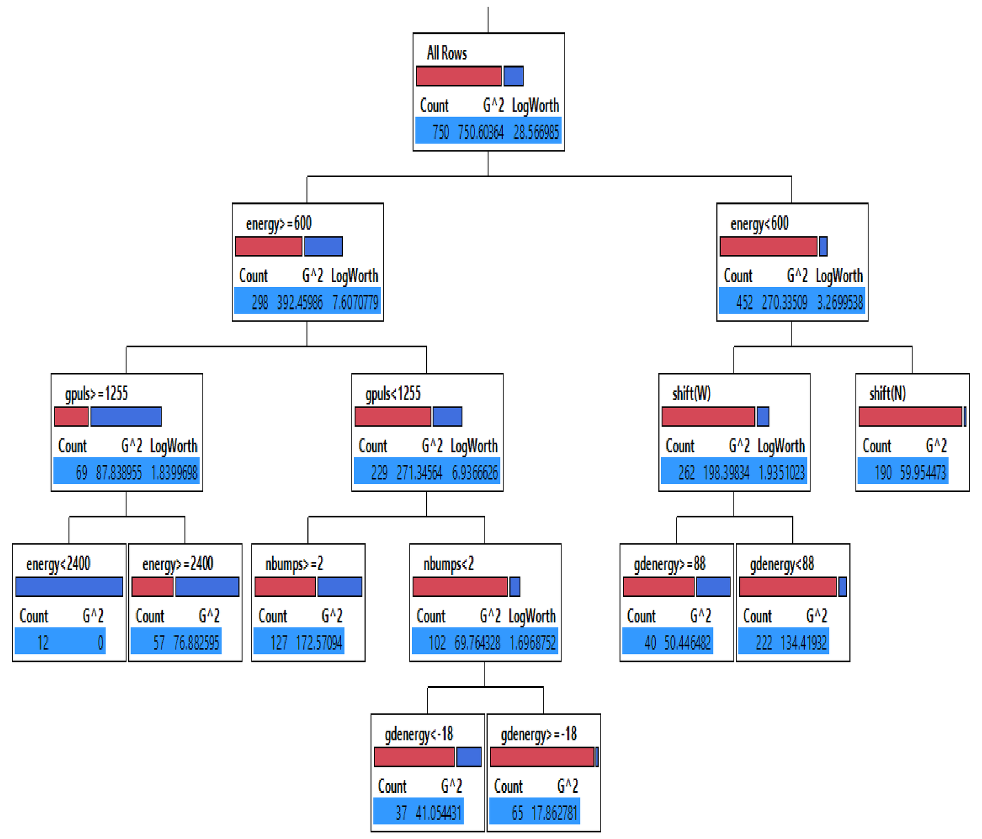

Fig 4.1: Decision Tree

Fit Details

Measure	Training	Definition
Entropy RSquare	0.2611	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.3636	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3698	$\sum -\text{Log}(\rho[j])/n$
RMSE	0.3408	$\sqrt{\sum(y[j]-\rho[j])^2/n}$
Mean Abs Dev	0.2336	$\sum \lvert y[j]-\rho[j]\rvert/n$
Misclassification Rate	0.1693	$\sum (\rho[j]\neq\rho\text{Max})/n$
N	750	n

Confusion Matrix

Actual / Training	Predicted 0	Predicted 1
0	577	23
1	104	46

Fig 4.2: Fit Statistics

Using JMP® 11 to Predict Occurrence of Seismic Activity to Prevent Hazards During Underground Mining Shifts

Zabiulla Mohammed, Graduate Student, Oklahoma State University

Sai Vijay Kishore Movva, Graduate Student, Oklahoma State University

Model Comparison

Using JMP Pro 11's Model Comparison feature the nominal logistic regression model, the neural network model, the boot strap forest and the decision tree models are compared with each other.

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RMSE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.8217	0.8865	0.0892	0.1594	0.0508	0.0320	750
Bootstrap Forest		0.2801	0.3866	0.3602	0.3328	0.2458	0.1453	750
Neural		0.2681	0.3721	0.3662	0.3383	0.2329	0.1613	750
Partition		0.2611	0.3636	0.3698	0.3408	0.2336	0.1693	750

Fig 5.1: Model Comparison

Conclusion

The best model chosen among the four is the nominal logistic regression model which is able to predict a high energy seismic event at a misclassification rate of 3.2% as compared to other models whose misclassification prediction rates are high. The logistic regression model also has better Entropy R Square, RMSE and Mean absolute deviation values. The significant interactions in this model are between the results of seismic method used , seismic energy recordings and number of energy bumps recorded.

References

https://archive.ics.uci.edu/ml/datasets/seismic-bumps

Sikora M., Wrobel L.: Application of rule induction algorithms for analysis of data collected by seismic hazard monitoring systems in coal mines. Archives of Mining Sciences, 55(1), 2010, 91-114.

Acknowledgement

We thank Dr. Goutam Chakraborty, Professor (Marketing), Spears School of Business and the founder of SAS and OSU Data Mining Certificate Program – Oklahoma State University for his encouragement and support.

Disclaimer : We have analyzed this topic using standard data mining and statistical techniques and we do not claim to have any kind of high level expertise in understanding hard rock mass processes.