Enhancing Machine Learning Algorithms to Assess Rock Burst Phenomena

Dimitrios Papadopoulos (dimtrpapadopoulos@gmail.com)

National Technical University of Athens: Ethniko Metsobio Polytechneio https://orcid.org/0000-0003-1865-2560

Andreas Benardos

National Technical University of Athens: Ethniko Metsobio Polytechneio

Research Article

Keywords: rockburst risk assessment, rockburst prediction, synthetic instances, rockburst classification

Posted Date: February 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-219424/v1

Version of Record: A version of this preprint was published at Geotechnical and Geological Engineering on June 3rd, 2021. See the published version at https://doi.org/10.1007/s10706-021-01867-z.

Enhancing machine learning algorithms to assess rock burst phenomena

Abstract: One of the main challenges that deep mining faces is the occurrence of rockburst phenomena. Rockburst risk assessment with the use of machine learning is currently gaining increased attention, due to the fact that outperforms the widely used empirical approaches. However, the limited and imbalanced instance records, combined with the multiparametric nature of the phenomenon, can lead to unstable estimations. This study focuses on the enhancement of the prediction performance of five machine learning algorithms, including Decision Trees, Naïve Bayes, K-Nearest Neighbor, Random Forest and Logistic Regression, by utilizing the oversampling technique SMOTE (Synthetic Minority Oversampling TEchnique).The initial database consists of 249 rockburst incidents, from which approximately 70% was used as the training set and the remaining 30% as the test set. Parametric analyses were conducted regarding different indicator combinations, such as the maximum tangential stress, the rock's uniaxial compressive and tensile strength, the stress coefficient, two brittleness coefficients and the elastic energy index. The models were trained with the original dataset and afterwards a gradual increase of the database with synthetic instances was made until the obtainment of a balanced dataset. Subsequently the creation of synthetic instances was continued until the real incidents used for training and the synthetic incidents were of the same amount. The results from the following analysis show that SMOTE technique has a considerable effect in the evaluation metrics of the models, even after the balancing of the dataset, and can be a valuable asset for the rockburst prediction.

Keywords: rockburst risk assessment; rockburst prediction; synthetic instances; rockburst classification

Dimitrios Papadopoulos

School of Mining and Metallurgical Engineering, National Technical University of Athens, Greece

Andreas Benardos

School of Mining and Metallurgical Engineering, National Technical University of Athens, Greece

Corresponding author: dimtrpapadopoulos@gmail.com

ORCID ID: 0000-0003-1865-2560

1. Introduction

Rockbursts are explosive failures of rock mass around an underground opening, which occur when very high stress concentrations are induced around the excavation (Hoek, 2000). Rockburst has been a serious problem in deep underground excavations and many incidents have been recorded and documented worldwide, with some of them associated with fatal results (Andrieux and Blake 2013; Shepherd 1981; Zhang 2012; Hedley 1992; Chen 1997; Zhu 2009). Brady and Brown (2004) defined rockburst as ‘‘a sudden displacement of rock that occurs in the boundary of an excavation, and cause substantial damage to the excavation’’. Cook (1963) and Salamon (1983) related rockburst and mine seismicity and characterized rockburst as part of the general term seismic event that damages mine workings. Ortlepp and Stacey (1994) distinguished rockburst from seismic events and defined rockburst as damage in a tunnel, resulting from seismic events. Muller (1991) categorized rockburst types in strain burst, pillar burst and fault slip burst. Tan (1992) reported that rockburst phenomena and rockbursting are related to the stress in the earth's crust, the rockmass characteristics, the hydrogeological conditions and the structures of the rock masses.

Two conditions are required to cause this phenomenon. Firstly, the stress that is developed in the rock or the discontinuity exceeds their strength and secondly, the energy released far exceeds the one consumed during the failure process.

The stress conditions, the geological structure, the mechanical properties of the rock mass, the human factor and their interaction are the elements responsible for triggering both seismic events and rockburst phenomena. The geological structure involves the presence of faults, shear zones, bedding planes, anticlines and synclines, stratification, bedding planes and material heterogeneity, which affect the stress distribution and can lead to high stresses. The mechanical properties of the rock mass involve the uniaxial compressive and tensile strength, the material brittleness, the heterogeneity of the rock mass, the presence of discontinuities, the friction angle and the Modulus of elasticity. The overall stiffness of the surrounding system and the deformation characteristics of the bursting material affect the intensity of rockburst. The depth of the tunnel, its support, its shape and orientation, the method of excavation and exploitation and the production rate comprise the human factor. Diederichs (2018) mentions that the evolution of a rockburst phenomenon is affected by the concentration of stresses due to cross-sectional geometry, geological parameters and creeping phenomena, the reduction of confining pressures on the shaft, the ability of the rockmass to store elastic energy and the presence of soft and stiff loading system. According to Castro (2012) strainbursts mainly take place under small confining stresses. In such conditions the failure scenarios include the creation or expansion of parallel cracks and the contribution of the spalling effect or the kinematic instability of the parts. In addition, these cracks reduce the stiffness of the loading system resulting in strainburst phenomena. In contrast, fault slip-bursts occur mostly in conditions of high confining stresses.

The complexity of rockburst and the insufficient understanding of its mechanism (Jiang 2014) hinders its prediction and mitigation. Rockburst prediction with the use of machine learning is an alternative approach adopted by many researchers that focuses on the learning by experience, while bypassing the need for knowing the cause. The major problem of this approach is the lack of sufficient amount of data, which is the key for accurate predictions. Thus, by adding synthetic instances on the initial rockburst database, this research aims in enhancing the performance of five ML classifiers regarding the rockburst prediction and classification, while concurrently investigating the effect of the SMOTE technique in the evaluation metrics of the algorithms. Instead of the common method of using SMOTE in creating synthetic instances only in the initial minority class with high oversampling rates, in this paper we add synthetic instances in the constantly changing minority classes, while keeping the oversampling at law rates in order to evaluate the process progressively.

2. Rockburst prediction methods

According to Li (2019) currently it is not possible to predict rockburst, but the areas with a rockburst tendency can be established with the use of techniques like microseismic monitoring and numerical modeling. Wang (2018) states that the accurate prediction of a seismic event is a difficult task due to the complex and multiparametric nature of the phenomenon and a fundamental step in the rockburst prediction process is the evaluation of the rockburst tendency. According to Zhang (2008) rockburst prediction can be distinguished between short term and long term. Short term prediction methods (Gu 2012; Liu 2014; Cai 2015; Cao 2015; He 2011; Cai 2015; Hosseini 2011; Cao 2016; Dou 2014; Gong 2010; Cheng 2009; Yu 2009) include drill-cutting parameters, borehole stress, backanalysis, electromagnetic emission, acoustic emission, charge method, microseismic (MS) monitoring, and active or passive seismic velocity tomography and are used during the construction stage. On the other hand, long term prediction methods are utilized mainly in the early design stage of a project and involve empirical criteria, numerical modeling, laboratory tests and currently the use of machine learning. The use of microseismic monitoring in the rockburst prediction has been a common topic by many researchers (Dou 2018; Liu 2014; Cai 2014; Cai 2015; Dou 2014). The use of numerical modeling (Vatcher 2014; Tianwei 2015; Board 2007; Vardar 2019; Khademian. 2016; Poeck 2016; Khademian 2019; He 2016; Manouchehriana 2018; Mitri 1999; Jiang 2010; Sharan 2007) in the rockburst prediction and its combination with other techniques is also a research topic that is investigated by many researchers, but it's main use focuses on the establishing of the burst prone areas and still there is not a universally accepted methodology of simulating accurately dynamic phenomena. Other research studies focus on the simulation of seismic waves generated from fault slips or from the failing rock and the associated damage that is caused in an underground excavation (Qinghua 2016; Banadaki 2012; Qiu 2019; Gao 2019; MJ Raffaldi 2017; Cho 2004; Hu 2019; He 2016). According to Kaiser (1996) numerical modeling for the rockburst prediction is based mostly on static approaches due to the complexity of the phenomenon and the difficulty to realistically simulate the dynamic procedures that are involved during a rockburst.

Regarding the long term rockburst prediction and its classification the empirical approaches are commonly used for the preliminary design of a deep underground construction project. Currently a geomechanical engineer can choose according to his judgement and the uniqueness of the situation between a plethora of rockburst evaluation criteria and some of those include also the prediction of the intensity of the event. Many researchers (Russenes 1974; Liu 2013; Hoek and Brown 1980; Turchaninov 1972; Martin 1999; Tajdus 1997) proposed empirical criteria based on the correlation of the stress conditions and the rock strength. Others (Cook 1966; Salamon 1984; Kaiser 1996; Mitri 1993; Mitri 1996; Brady and Brown 2004; Hedley 1992; Wang and Park 2001; Weng 2017; Kidybinski 1981; Neyman 1972; Ryder 1988) proposed rockburst energy related criteria, from which the energy release rate and excess release rate criteria are the most commonly used, especially in deep underground mines in South Africa. Other criteria that are primarily used for the pillar bursts are based in the assessment of the relative stiffness of the host rock and the failing rockmass (Wiles 2002, Gill 1993, Blake and Hedley 2003). Other empirical approaches are based on the rock brittleness (Singh 1987, Peng 1996, Feng 2000), which can be evaluated by laboratory experiments and relate the pre- and post-peak characteristics of the tested rock. Xu (2017) proposed a rockburst criterion based on the RMR classification system, while Xu (2017) and Zhou (2012) tried to correlate the rockburst potential with the RQD of the rockmass. Finally, other researches (Kaiser 1992; Durrheim 1998; Heal 2006; Qiu 2011; Zhang 2016) proposed rockburst evaluation criteria based on the combination of the above indexes and other construction factors.

3. Machine learning in Rockburst Prediction

Despite the fact that machine learning has been successfully used in a broad range of areas over the last decades, its utilization in the field of rock engineering is relatively new. Morgenroth (2019) states that machine learning can be a valuable tool to be integrated into the rock engineering practices, due to the complex nature of the geotechnical problems, the difficulty in utilizing all geotechnical data into empirical and numerical models and the rapid increase of the collected data. McGaughey (2019) stated that the application of artificial intelligence in the field of rock engineering is not a simple task, because the data required to make a prediction are sparsely scattered in space and time. However, correlations can be found between large volumes of data, the creation of statistical models through which predictions can be made, and the influence of individual factors on the overall behavior of a system can be made as well as the creation of scenarios and assumptions. Another utility of machine learning in the field of geoengineering is the addressing of issues such as the identification of terrain deformations or instability areas, with limited resources (Tsangaratos 2014).

Faradonbeh (2020) conducted 139 laboratory tests to collect data on the prediction of rockburst-induced trends, which he introduced into 2 models based on the gene expression programming (GEP) and classification regression tree algorithms (CART). He first singled out the most important and independent parameters through clustering techniques (AHC, SSE, multiple regression analysis) and then successfully trained the prediction models. Pu (2019) used the Support Vector Machine algorithm to predict rockbursts and their intensity based on 246 rockburst incidents. The data included the tangential stress, uniaxial strength, tensile strength, stress factor, brittleness index and energy index. Initially he aimed at the separation of the independent variables as well as the reduction of the data dimension by utilizing the distributed Stochastic Neighbor Embedding method (t-SNE) and then through the clustering method he grouped the remaining data. He then successfully trained a model based on the Support Vector Machine algorithm. Wu (2019) used the Least Squares Support Vector Machine algorithm to create a rockburst forecast model and by conducting sensitivity analyses reported that the ratio of tangential stress to the uniaxial compressive strength has the greatest influence on the forecast. Li (2018) used the Logistic Regression algorithm in a database consisting of rockburst and non-rockburst incidents. The input attributes included the depth, the maximum tangential stress, the elastic energy index, the uniaxial compressive and tensile strength of the rock. He reported that the depth, the uniaxial strength and the energy index have the greatest weight. In conclusion, he compared the results of the model with 6 empirical indicators and found that the algorithm performed better. Ghasemi (2019) utilized a Decision Tree algorithm to predict the occurrence and intensity of rockburst based on a dataset composed of 174 cases. Furthermore, he evaluated the importance of the input parameters and found that the energy index, the stress factor and the brittleness coefficient are the most important. Faradonbeh et al. 2018 collected a database of 134 rockburst cases and trained the algorithms Neural network, GEP and a Decision Tree. Afraei (2018) used regression models to predict rockburst and evaluated the importance of the insert attributes that contributed to the predictions. He found that the most important parameters are the maximum tangential stress, the stress factor, the elastic energy index and the uniaxial compressive strength of the rock.

In the rockburst prediction topic Sousa (2017) performed such relevant research and attained a classification scheme from a dataset composed of 60 rockburst cases with the input parameters being the uniaxial compressive strength, the modulus of elasticity, the stress conditions, the excavation geometry and the equivalent cross-section of the opening. The algorithms that were utilized and compared with each other were the K-Neighbor algorithm, Decision Tree, Neural Network, Support Vector machine and Naïve Bayes. Additionally, he performed a sensitivity analysis to find the weight of each parameter in the final predictions.

Li (2017) presented the application of Bayesian networks models on rockburst prediction by using 135 rockburst cases and using as input parameters the depth, the maximum tangential stress, the uniaxial compressive and tensile strength of the rock and the elastic energy index. He reported that the Tree Augmented Naive Bayes algorithm had the best accuracy.

Zhou (2016) compared the algorithms Linear Discriminant analysis (LDA), Quadratic Discriminant Analysis (QDA), Partial Least-squares Discriminant Analysis (PLSDA), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Multilayer Perceptron Neural Network (MLPNN), Classification Tree (CT), Support Vector Machine (SVM), Random Forest (RF) and Gradient-Boosting Machine (GBM) on the prediction of rockburst intensity based on 246 incidents. The input parameters, which were examined based on their influence, included the stress factor, the depth, the uniaxial strength, the brittleness index and the elastic energy index. Random Forest showed the best performance, while the variable with the highest weight was found to be the energy index.

Dong (2013) compared the algorithms Random Forest, Artificial Neural Networks and Support Vector Machine regarding the rockburst prediction and its intensity based on 46 incidents. The Random Forest algorithm showed the best performance.

Adoko (2013) used the ANFIS method, which is a method combining neural networks with fuzzy logic, in order to predict the intensity of rockburst, based on a dataset consisting of 174 rockburst cases. Jian (2012) used the Support Vector Machines algorithm regarding rockburst prediction based on 132 rockburst incidents.

He (2012) compared the algorithms Decision Trees, K-Nearest Neighbor, Support Vector Machine and Neural Network regarding the classification of rockburst intensity based on reported rockburst cases. The input parameters included the distance of the event from the excavation, the excavation geometry, the type of support, the uniaxial strength, the modulus of elasticity, the cross-sectional area, the excavation depth, the stress factor, the existence of discontinuities and the excavation method. He reported that neural networks showed the best performance, while the decision trees showed the worst performance. Zhou (2010) introduced the Fisher Discriminant Analysis method for rockburst prediction based on 15 cases. Gong (2010) and Gong (2007) based on 21 and 15 rockburst cases respectively used the Distance Discriminant Analysis algorithm for the long-term prediction of rockburst. The input parameters consisted of the stress factor, the brittleness coefficient and the energy index.

Chen (2003) applied a Neural Network regarding the prediction of rockburst and its intensity. Zhao (2005) used the Support Vector Machine algorithm for the long-term prognosis of rockburst based on 16 rockburst cases. Ge (2008) combined Neural Networks with the AdaBoost algorithm in order to categorize and predict rockburst. Gathering data from 36 rockburst cases and using the tangential stress, the stress factor, the brittleness coefficient and the elastic energy index as input parameters, he presented a promising rockburst forecasting system. Su (2008) proposed the K-Nearest Neighbor algorithm for the rockburst prediction, which is one of the simplest and most effective algorithms in the field of machine learning.

Table 1 presents a summary of the algorithms, attributes, number of data and the classification accuracy obtained from different researchers regarding the rockburst prediction. The following results have been produced from different datasets, using various evaluation techniques and thus cannot be directly compared with each other. Nevertheless, one can get a clear idea of the main attributes used for the assessment and moreover, the estimated general accuracy level and performance attained.

4. Methodology

4.1. Proposed Methodology

The rockburst databases used have two main challenges to overcome. The first concerns the unequal distribution of cases per class and the second is the lack of sufficient amount of incidents proportional to the complexity of the phenomenon. Mainly for the qualitative and subsequently for the quantitative improvement of the database we added synthetic instances generated from the SMOTE technique. The qualitative part refers to the balancing of the database, meaning that we gradually add synthetic instances until the number of cases become equal for all classes, while the quantitative part refers to the further extension of the database with synthetic instances, that are placed uniformly in all classes after the balancing of the dataset.
Aiming at the observation and evaluation of the process in a wide range of algorithms we used five of the most common ML algorithms, while selecting a combination of attributes based on an attribute selection filter. Regarding the training and evaluation part of the procedure we applied the10-fold cross-validation technique, followed by the testing procedure. The methodology is illustrated in figure 1.

4.2. Data Sources and Description

The database is composed of 249 published rockburst cases over the period 1991–2013, as collected and compiled by various researchers. This database is given as a supplementary data to this paper and is available for use from other researchers.

The database is consisted of a number of parameters including the maximum tangential stress ($\sigma\theta$), the uniaxial compressive strength (σc), the tensile strength (σt), the stress coefficient (SCF= $\sigma\theta/\sigma c$) as given by Martin (1999), the brittleness coefficient (B1=$\sigma c/\sigma t$) as proposed by Peng (1996), the brittleness coefficient (B2=$(\sigma c-\sigma t)/(\sigma c+\sigma t)$ as proposed by Singh (1987) and, finally, the elastic energy index (Wet). These attributes, which represent the basic conditions needed for the initiation and propagation of the rockburst phenomenon, consist the inputs of the analysis. They are used by the majority of the researchers for the long term rockburst prediction and are part of the most empirical indexes for rockburst assessment.

The output of the database represent the rockburst's intensity, which corresponds to the input set. This is given and discerned into four categories: none, light, moderate, and strong. This intensity based classification is presented in Table 2, as proposed by Zhou (2012).

In figure 2 an overview of the distribution of all attributes in the dataset is given, both in terms of values, and in terms of the rockburst intensity class occurrence (None, Low, Moderate and Heavy). It can be easily seen that the rockburst intensity is actually spreading throughout the value range of all parameters, without having a clearly defined trend or pattern. In addition in Table 3 some basic statistical information regarding the input attributes are presented, covering the minimum and maximum values, the mean values and the standard deviation of all parameters.

What is most important however is the imbalanced nature of the database, meaning that the classes are composed of unequally quantities of instances. This is a common issue especially when dealing with phenomena like rockbursts, where the occurrence of certain intensity class is more scarce than some other. Thus, the "None" class participates at a rate of 19% (47 cases), the "Low" class has 29% of the total (73 cases), the "Moderate" class 33% (83 cases) and finally the "Heavy" class consists 18% of the dataset (46 cases).

The dataset is divided in two parts, the training and the testing subset. The division has been made using the 70-30 rule, with 71% of the data consisting the training set (178 cases) of the ML model and the rest 29% (71 cases) forming the testing hold-out set, which is to be introduced to the finally trained model for assessing its performance. The division of the dataset was made randomly, while finally the distribution per class in both training and testing subsets are approximately the same as in the total database.

4.3. Synthetic Minority Oversampling Technique - SMOTE

An imbalanced database can create poor performance results or overfitting problems, as often the databases' minority class or classes can be overlooked by the machine learning algorithms. Sun (2009), stated that database imbalances is a key issue and an obvious problem in employing machine learning algorithms for classification applications, accompanied by other factors such as small databases, class separability issues, etc. Chawla (2004) outlined the importance of the class imbalance problem along with the data distribution within each class in the classifier's performance.

One method for dealing with imbalanced datasets is the adoption of the Synthetic Minority Oversampling Technique (SMOTE) technique (Chawla 2000), which increases the quantity of the minority class with new instances synthesized from existing instances of the minority class. According to Fernandez (2018) the utilization of SMOTE preprocessing algorithm is considered "de facto" standard in the framework of learning from imbalanced data.

This technique, which is illustrated in figure 3, actually injects new synthetic data into the database so to increase the available number of instances in the databases' minority class and hence strengthen its presence. It is an oversampling method and it generates new instances with the help of interpolation between the positive instances that lie together. The procedure involves the following steps. Firstly the minority class is set where A = {x1, x2, …xt}. For each x Є A the k-nearest neighbors are obtained based on the calculation of the Euclidean distance between x and every other minority points in set A. Next, for each x belongs to A, n minority points from its k-nearest neighbors are chosen and form the set A1.Lastly for every sample x_k Є A1 new synthetic instances are interpolated based on the following formula:

$x' = x + rand(0, 1)*|x - x_k|$, where rand(0,1) represents the random number between 0 and 1.

SMOTE is defined by the k and n indices where, k = nearest neighbors and n = no. of samples to be generated.

The SMOTE process was utilized trough the WEKA sofware and during the procedure five nearest neighbors were used for the creation of the instances, while the oversampling was kept at low rates (5 – 10%), meaning that the synthesized data created 3 or 4 instances per step. The new synthetic data were inserted to the rockburst classes "None", "Low" and "Heavy", until the balancing of the dataset was succeeded. After that point new synthetic instances were placed successively to all classes. In total 182 synthetic instances were added in the starting training set in 48 steps, from which 32%, 19%, 16% and 33% correspond to the classes "None", "Low", "Moderate" and "Heavy" respectively.

4.4. ML model building

The development of the ML model is made though the WEKA open source software. Weka is a robust platform for data mining experiments containing four application environments (Explorer, Experimenter, KnowledgeFlow and Simple CLI). For this study's experiments the Explorer application was used, due to its user friendly environment, the simplicity in visualizing the data and the easy access to plenty of tools and data analytic processes. The use of this software provides a great degree of automation and flexibility in the design model, as well as consistency and confidence in the overall results obtained. Through the next paragraphs the steps to develop and build the ML model are given.

4.4.1. Attribute Selection

Aiming at the optimization of a classifier's performance the Correlation Attribute Evaluation filter combined with the Ranker search method, provided by the WEKA software (Waikato Environment for Knowledge Analysis), was adopted for the purpose of a targeted reduction of the amount of attribute combinations for our analysis. This filter weights and ranks features based on Pearson's product moment correlation (Hall 1999). The results of the filter in the rockburst database are presented in the following figure 4.

From the above figure it is observed that the maximum tangential stress has the biggest weighting factor, followed by the energy index, the stress factor, the brittleness coefficient B1, the brittleness coefficient B2, the tensile strength and the uniaxial compressive strength. Hence in order to gradually decrease the number of inputs and witness the effect on the prediction capability of the ML models, based on the most important parameters, we designed our analysis on the following five attribute combinations leading in twenty – five basic classifiers:

I. 7 attributes: $\sigma\theta$, Wet, SCF, B1, B2, σt, σc
II. 6 attributes: $\sigma\theta$, Wet, SCF, B1, B2, σt
III. 5 attributes: $\sigma\theta$, Wet, SCF, B1, B2
IV. 4 attributes: $\sigma\theta$, Wet, SCF, B1
V. 3 attributes: $\sigma\theta$, Wet, SCF

4.4.2. Stratified Cross Validation

Stratified cross-validation is a resampling technique for performance evaluation purposes, in which a systematic way of running repeated percentage splits is done, in an effort to minimize bias from the training and testing subset selection procedure. Cross validation offers two main advantages. Firstly a model is trained with every instance of a dataset and secondly overfitting problems can be reduced. According to Ian Witten (2005) cross-validation—is gaining ascendance and is probably the evaluation method of choice in most practical limited-data situations.
Our models were trained and evaluated with the 10-fold cross-validation method in the whole training subset. The process involves the division of a dataset into 10 equally proportional folds with class values, from which 9 folds are used for training and the remaining fold is used for testing. Thus 10 evaluation results are originated and averaged. Having done this 10-fold cross-validation and computed the evaluation results, Weka invokes the learning algorithm a final (11th) time on the entire dataset so as to have a final working model that can be used for the case selected. In this 11-th ML model the testing subset, that is consisted of completely new data, is introduced so as to attain the final performance in the classification accuracy of the rockburst classes.

4.4.3. Building Classifiers

A total of five ML algorithms have been selected to perform the classification of the rockburst, namely J48, Naïve Bayes, Logistic Regression, Random Forest and K-Nearest Neighbor. The analysis is made by taking into account the attribute combinations as given in section 4.4.1 by using the open source software WEKA.

Though the result obtained, the overall ML performance evaluation was made, for all the 25 classifier configurations, with and without the use of the synthetic data (SMOTE on/off). For the assessment of the predictions' accuracy classifications of the models used, a set of 4 major performance evaluation indices have been employed, namely the Accuracy (percentage of correctly classified instances), K-statistic, F-Measure and Area Under the Curve (AUC).

5. Results of the ML methodology

All M.L models performed relatively well in classifying the rockburst classes of the unknown testing subset that was introduced to them. The results below are focusing on the attained performance prediction capability in terms of accuracy with and without the use of synthetic data (SMOTE methodology) with respect to the selected number of the input attributes/parameters used (from 3 to 7 attributes).

The results for the ML models used are given in the following Figures 5 to 9. In each diagram, the y-axis represent the accuracy level, while the x-axis denotes the total instances that were used for the training of the ML models. Their start position is the value of the 178 instances (initial training dataset), from which new synthetic data are added in increment steps until the final value of 360 instances is reached, attaining the doubling of their initial data. The vertical line at the point of 248 instances represents the threshold where the balancing of the dataset is reached, meaning that all the rockburst classes of the training dataset contain the exact number of data (instances). Thus, it can be seen that the diagrams can be discerned in two parts, the first until the balancing is reached and the second part where new synthetic instances are continuously added, until the training dataset doubles in size. Furthermore each line represents each one of the 5 attribute combinations.

The Random Forest algorithm has the best accuracy when the initial training dataset is used. This is shown for all attributes combination, that yield consistently high accuracy levels ranging from 71.8% (3 attributes) to 74.6% (4 and 6 attributes). At the early stage of the SMOTE process, before the balancing of the dataset is achieved, the classifiers showed an improvement in their performance. The maximum attained accuracy scores during this stage ranged from 74.6% (3 and 4 attributes) to 76% (5, 6 and 7 attributes). After the balancing of the dataset the accuracy scores of the classifiers dropped in general, except the one of the 5-attribute classifier, whose performance was steadily increased after the point of 288 instances. The 5-attribute classifier achieved the highest accuracy (77.5%) at the points 340 and 356, which is the best score in this study.

As for the KNN algorithm the starting accuracy varies between 60.6% (7 attributes) and 69% (4 attributes). During the balancing of the dataset, the addition of synthetic instances improved the performance of the classifiers with 5, 7 and 3 attributes, but the highest accuracy score is maintained by the 4 attribute starting classifier. SMOTE enhanced further the predictive ability of the classifiers after the balancing of the dataset. The classifiers with 3 and 7 attributes achieved their highest scores (67.6%) at the point of 256 instances, while the classifiers with 5 and 6 attributes outperformed the highest starting score (73.2% and 71.8%) at the points 356 and 344 respectively.

Regarding the J48 diagrams the starting accuracy scores range between 60.6% (7 attribute) and 69% (4 attribute). The classifier with 6 attributes starts with the second lowest accuracy (63.38%), but before the balancing of the dataset (at 233 instances) achieves the highest accuracy of all the J48 classifiers (71.8%). Similarly, during the balancing stage the performance of the classifiers with 3 and 7 attributes reached their peak scores (70.4% and 69%), while the 5-attribute classifier increased its accuracy at the point 214. After the balancing stage, the addition of synthetic instances enhanced the performance of the 5-attribute classifier, which attained its highest accuracy (70.4%) at the point 320.

The starting accuracy scores, that were obtained by the Naïve Bayes Algorithm, range between 57.7% (7 attributes) and 66.2% (4 and 3 attributes). The highest scores attained by the classifiers are achieved at the first stage of the SMOTE procedure, before the balancing of the dataset, between the points at 187 and 200 instances. The 5-attribute classifier obtained the highest accuracy (70.4%) in comparison with the rest Naïve Bayes classifiers, followed by the classifiers with 3 and 4 attributes(69%).

Finally the Logistic Regression algorithm presented the worst starting scores, which vary between the values 54.9% (5 attributes) and 57.7% (4, 6 and 7 attributes). Similarly to the Naïve Bayes algorithm the performance enhancement of the Logistic Regression classifiers occurs at the early stage of the procedure. The classifiers with 3 and 5 attributes attain improved scores (57.4% and 56.3%), while the 7 attribute classifier obtains the highest accuracy (59.15) regarding the Logistic Regression algorithm.

In table 4 the maximum increase (Max) and decrease (Min) of the evaluation metrics (Accuracy-ACC, k-Statistic-k, F-Measure-F-M and AUC) that were achieved compared with the starting scores (Start) of the classifiers during the SMOTE process is presented. The table is split based on the number of attributes (No) and the machine learning algorithm. The blue boxes represent the maximum increase (%) per algorithm per evaluation metric, while the yellow boxes represent the maximum scores of the evaluation metrics before and after the use of SMOTE. Taking into account Accuracy, K-statistic and F-Measure, the highest starting score was obtained by the 4-attribute Random Forest classifier (74.6%, 0.66, 0.7434), while, the best overall results were achieved by the 5-attribute Random Forest classifier (77.5%, 0.7, 0.771) meaning that SMOTE managed an increase of 4%, 6% and 4% in the selected metrics, respectively.

In general, 20 out of the 25 starting ML classifiers performed better scores with the use of SMOTE, indicating the positive effect of the technique in the rockburst classification and prediction. The maximum increase due to SMOTE took place in the J48 algorithm that used the 7 input attributes. The increase in its Accuracy, K-statistic, AUC and F-Measure was 14%, 25.5%, 10.2% and 13.9%, respectively.

Furthermore it is observed, that all the starting ML algorithms (before SMOTE) consisting of 4 attributes achieved the best starting scores. In these classifiers SMOTE had negative effect in their evaluation metrics. On the contrary the highest scores, due to the utilization of SMOTE, were obtained by the algorithms with 5, 6 and 7 attributes that attained lower starting scores. The same trend is observed also in the maximum increase percentages, in which the low-attribute algorithms (3 and 4 attributes) have the smallest increase rates, indicating that SMOTE performs better when dealing with an increased number of evaluation attributes.

In table 5 a comparison is given between the starting classifiers (Before SMOTE) and the best classifiers (after SMOTE) focusing not on the overall classification performance but rather taking into account the classification attained within any of the individual rockburst intensity classes (within class classification metrics). The evaluation metrics are composed of the True Positive Rate (TP Rate), and F-Measure.

Overall the values of the metrics after SMOTE were greatly improved between 3% and 33.5%. SMOTE affected positively the capability of the ML algorithms in distinguishing the classes “None”, “Low” and “Moderate” and more specifically J48 and Random Forest algorithms were benefited the most. These algorithms achieved 100% accuracy in distinguishing the existence of rockburst. An issue exists in the classification accuracy of the “Heavy” rockburst cases, which is due to the fact that the rockburst database consists of both strainbursts and fault-slip bursts, that lead to within-class sub concepts.

It is clear though that after utilizing SMOTE, the differences between the metrics per class are smoothed and the overall results are more homogeneous in all ML algorithms. For example regarding the J48 algorithm and the True Positive evaluation metric the classifier scores 0.476 at the class Low, which is a value significantly lower than those of the other classes. This metric improved (0.619) after SMOTE and the relative metrics in all classes were both improved and uniformed. This type of improvement occurs also in the F-Measure index, indicating that SMOTE enhanced both the overall performance of the algorithms as well their classification performance between the classes.

6. Summary and Conclusions

This study examined the effect of SMOTE in five machine learning algorithms (J48, Naïve Bayes, K-Nearest Neighbor, Random Forest and Logistic Regression) regarding rockburst long-term prediction with respect to its expected intensity. The initial database is composed of 249 rockburst cases, from which 178 instances were used for the training and evaluation of the models with the 10-fold cross validation technique, while also an additional hold-out set of the remaining 71 instances were used for the final testing of the algorithms. Five different attribute combinations were obtained, based on the Correlation Attribute Evaluation filter, resulting in 25 basic classifiers. These classifiers were then trained in a constantly increasing dataset with synthetic instances generated from the SMOTE algorithm. The experiments stopped after the generation of 182 synthetic instances. The evaluation metrics used in this study involved the accuracy, the K-statistic, the ROC Area (AUC) and the F-measure.
Based on the test results the following conclusions can be made:

- The maximum classification accuracy scores obtained by the algorithms (J48 - 71.83%, Random Forest - 77.46%, K-Nearest Neighbors - 73.24% and Naïve Bayes - 70.42%) are among the highest in the current literature, taking into account that the test set is approximately 30% of the database.
- SMOTE managed to increase the evaluation metrics of twenty out of twenty five basic classifiers, thus proving its value as a tool for enhancing the capability of ML algorithms when dealing with imbalanced datasets.
- The increased scores were obtained before and after the balancing of the database.
- The most reliable model was the Random Forest algorithm consisting of five attributes and trained in a dataset composed of 340 instances, from which the number of synthetic instances was 162. The classifier obtained 77.46% accuracy.
- The maximum percentage increase due to SMOTE occurred in the J48 algorithm consisting of 7 attributes. The increase in accuracy, K-statistic, AUC and F-Measure was 14%, 25.5%, 10.2% and 13.9% respectively.
- In general SMOTE increased the overall metrics by 5-10%, but most importantly improved and smoothed the within class classification metrics of the algorithms up to 30%.
- The addition of synthetic instances was carried out in very small steps with rates of 5-10%, however the results of the evaluation metrics showed great sensitivity per step of the process especially in the J48 and Random Forest algorithms. This nonlinearity that is reflected on the above diagrams, reveals the lack of sufficient number of training data and indicate the need for enriching the rockburst database.
- Even though SMOTE managed to increase the within-class performance of the algorithms, it failed to enhance the classification accuracy of the Class “Heavy Rockburst”.

The incorporation of the SMOTE technique can be a useful tool to have more balanced databases, an element of key importance in making accurate prognosis, especially in cases of geotechnical character where the data are hard to find. Of course, this is something that can be used for the final optimization of the ML algorithms accuracy and additional research should be made to further enhance their ability to make accurate predictions. To further facilitate this the authors as already mentioned include the database of this paper as supplement data for other researchers to pursue solutions to better describe the rockburst phenomenon with the use of ML.

Declarations

Funding: None

Conflicts of interest/Competing interests: None

Availability of data and material: Available after request

Code availability: Not applicable

REFERENCES

Adoko AC, Gokceoglu C, Wu L, Zuo QJ (2013) Knowledge-based and data-driven fuzzy modeling for rockburst prediction. Int J Rock Mech Min Sci 61:86–95

Afraei S, Shahriar K, Madani SH, (2018) Statistical assessment of rock burst potential and contributions of considered predictor variables in the task. Tunn Undergr Sp Tech 72:250–271

Andrieux P, Blake W, Hedley DGF, Nordlund E, Phipps D, Simser B, Swan G (2013) Rockburst case histories: 1985, 1990, 2001 & 2013. CAMIRO Mining Division for the Deep Mining Research Consortium, Sudbury

Bai MZ, Wang LJ, Xu ZY (2002) Study on a neural network model and its application in predicting the risk of rock burst. China Saf. Sci. J., 12(4), 65–69

Blake W, Hedley DGF (2003) Rockbursts case studies from North American hard-rock mines. Littleton, CO: Society for Mining, Metallurgy, and Exploration

Board M, Damjanac B, Pierce M (2007) Development of a methodology for analysis of instability in room and pillar mines. Proceedings of the Fourth International Seminar on Deep and High Stress Mining. Perth, Australia

Brady BHG, Brown ET (2004) Rock mechanics for underground mining. Kluwer Academy Publishers, Dordrecht.

Cai Si-jing, Zhang Lu-hua, Zhou Wen-lue (2005) Research on prediction of rock burst in deep hard-rock mines [J]. Journal of Safety Science and Technology, 1(5): 17−20 (in Chinese)

Cai W, Dou L, Cao A, Gong S, Li Z (2014) Application of seismic velocity tomography in underground coal mines: A case study of Yima mining area, Henan, China. Journal of Applied Geophysics, 109, 140–149

Cai W, Dou L, Gong S, Li Z, Yuan S (2014) Quantitative analysis of seismic velocity tomography in rock burst hazard assessment. Natural Hazards, 75(3), pp. 1-13

Cai W, Dou L, Li Z, Liu J, Gong S, He J (2014) Microseismic multidimensional information identification and spatio-temporal forecasting of rock burst: a case study of Yima Yuejin coal mine, Henan, China. Chinese J. Geophys. Ed. 57, 2687–2700

Cao A, Dou L, Cai W, Gong S, Liu S, Jing (2015) Case study of seismic hazard assessment in underground coal mining using passive tomography. International Journal of Rock Mechanics and Mining Sciences, 78, 1–9

Cao A, Dou L, Cai W, Gong S, Liu S, Zhao Y (2016) Tomographic imaging of high seismic activities in underground island longwall face. Arabian Journal of Geosciences, 9(3)

Castro LM, Bewick, Carter (2012) An Overview of Numerical Modelling Applied to Deep Mining. Innovative Numerical Modeling in Geomechanics

Chawla N (2004) Data Mining for Imbalanced Datasets: An Overview. Data Mining and Knowledge Discovery Handbook pp 853-867

Chawla N, Bowyer K, Hall L, Kegelmeyer P (2000) SMOTE: Synthetic MinorityOver-sampling TEchnique. International Conference of Knowledge Based Computer Systems, pp. 46–57. National Center for Software Technology, Mumbai, India,Allied Press

Chen HJ, Li NH, Ni DX, Shang YQ (2003) Prediction of rockburst by artificial neural network, Chin J Rock Mech Eng 22:762–768

Chen ZH, Tang CA, Huang RQ (1997) A double rock sample model for rockbursts. International Journal of Rock Mechanics and Mining Sciences, 34(6), 991–1000

Cheng Y., Jiang F., Zou Y. (2009). Research on inversion high mining pressure distribution and technology of preventing dynamic disasters by MS monitoring in longwall face. Journal of Coal Science and Engineering (China), 15(3), 252–257

Cho SH, Kaneko K (2004) Influence of the applied pressure waveform on the dynamic fracture processes in rock. International Journal of Rock Mechanics and Mining Sciences, 41(5), 771–784

Cook NGW (1963) The basic mechanics of rockburst. Journal of South African Institute of Mining and Metallurgy, 64: 71–81

Cook NGW, Hoek E, Pretorius JPG, Ortlepp WD, Salmon MDG (1966) Rock mechanics applied to the study of rockbursts. I.S. Afr. Inst. of Min. Metall., 66, pp.435-528

Dehghan Banadaki MM, Mohanty B (2012) Numerical simulation of stress wave induced fractures in rock. International Journal of Impact Engineering, 40-41, 16–25

Diederichs M (2018) Early assessment of dynamic rupture hazard for rockburst risk management in deep tunnel projects. Journal of the Southern African Institute of Mining and Metallurgy, 118(3):193-204

Ding XD, Wu JM, Li J, Liu CJ (2003) Artificial neural network for forecasting and classification of rockbursts. J. Hohai Univ. (Nat. Sci.), 31(4), 424–427

Dong LJ, Li XB, Peng K (2013) Prediction of rockburst classification using random forest. Trans Nonferrous Meterol Soc China 23(2):472–477

Dou L (2018) Comprehensive early warning of rockburst utilizing microseismic multi-parameter indices. International Journal of Mining Science and Technology Volume 28, Issue 5

Dou LM, Cai W, Gong SY, Han RJ, Liu J (2014) Dynamic risk assessment of rockburst based on the technology of seismic computed tomography detection. J China Coal Soc, 39 (2), pp. 238-244

Du Zi-jian, Meng-guo XU, LIU, Zhen-ping WU, Xuan (2006) Laboratory integrated evaluation method for engineering wall rock rock-burst. [J]Gold, 11(27): 26 −30. (in Chinese)

Durrheim RJ, Roberts MKC, Haile AT, Hagan TO, Jager AJ, Handley MF, Spottiswoode SM, Ortlepp WD (1998) Factors influencing the severity of rockburst damage in South African gold mines. JS Afr. Inst. Min. Metall. 98 (2), 53–57

Faradonbeh RS, Taheri A (2018) Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques. Comput Eng

Feng T, Xie XB, Wang WX (2000) Brittleness of rocks and brittleness indexes for describing rockburst proneness. Min. Metall. Eng. 20 (4): 18–19

Feng XT, Wang, LN (1994) Rockburst prediction based on neural networks. Trans. Nonferrous Met. Soc. China, 4(1), 7–14

Fernandez A, Garcia S, Herrera F, Chawla N (2018) SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research 61 (2018) 863-905

Gao F, Kaiser PK, Stead D, Eberhardt E, Elmo D (2019) Numerical simulation of strainbursts using a novel initiation method Computers and Geotechnics, 106, 117–127

Ge QF, Feng XT (2008) Classification and prediction of rockburst using AdaBoost combination learning method. Rock Soil Mech 29:943–948

Ghasemi E, Gholizadeh H, Adoko AC (2019) Evaluation of rockburst occurrence and intensity in underground structures using decision tree approach. Engineering with Computers

Gill DE, Aubertin M, Simon R (1993) A practical engineering approach to the evaluation of rockburst potential. Proc., 3rd Int. Symp. on Rockbursts and Seismicity in Mines, Rotterdam, Netherlands, 63–68

Gong F, Li X (2007) A distance discriminant analysis method for prediction of possibility and classification of rockburst and its application. Chinese Journal of Rock Mechanics and Engineering 26(5):1012-1018

Gong FQ, Li XB, Zhang W (2010) Rockburst prediction of underground engineering based on Bayes discriminant analysis method. Rock Soil Mech 31(Suppl. 1):370–377

Gong SY (2010) Research and application of using mine tremor velocity tomography to forecast rockburst danger in coal mine. China University of Mining and Technology, Xuzhou

Gu ST, Wang CQ, Jiang BY, Tan YL, Li NN (2012) Field test of rockburst danger based on drilling pulverized coal parameters. Disaster Adv 5:237– 240

Hall MA (1999) Correlation-Based Feature Selection for Machine Learning; The University of Waikato: Hamilton, New Zealand

He BG, Zelig R, Hatzor YH, Feng XT (2016) Rockburst Generation in Discontinuous Rock Masses. Rock Mechanics and Rock Engineering, 49(10), 4103–4124

He M, Sousa L, Faramarzi L (2012) Rockburst Process Evaluation Using Experimental and Artificial Intelligence Techniques. Conference: 1st Iranian Mining Technologies Conference, Volume: 24p

He X, Chen W, Nie B, Mitri H (2011) Electromagnetic emission theory and its application to dynamic phenomena in coal-rock. International Journal of Rock Mechanics and Mining Sciences, 48(8), 1352–1358

Heal D, Potvin Y, Hudyma M (2006) Evaluating rockburst damage potential in underground mining. In: Yale, D.P. et al. (Eds.), Proceedings of 41st U.S. Symposium on Rock Mechanics (USRMS). USA, Curran Associates, Colorado School of Mines, pp. 322–329

Hedley DGF (1992) Rockburst handbook for Ontario hardrock mines. Ottawa: Canada Centre for Mineral and Energy Technology

Hoek E, Brown ET (1980) Underground Excavations in Rock. Institution of Mining & Metallurgy, London.

Hosseini N, Oraee K, Shahriar K, Goshtasbi K (2011) Studying the stress redistribution around the longwall mining panel using passive seismic velocity tomography and geostatistical estimation. Arabian Journal of Geosciences, 6(5), 1407–1416

Hu L, Feng X, Xiao Y, Wang R, Feng G, Yao Z, Niu W, Zhang W (2019) Effects of structural planes on rockburst position with respect to tunnel cross-sections: a case study involving a railway tunnel in China. Bulletin of Engineering Geology and the Environment

Jia YP, Lu Q, Shang YQ (2013) Rockburst prediction using particle swarm optimization algorithm and general regression neural network. Chinese Journal of Rock Mechanics and Engineering, 32(2), 343-348

Jia YR, Fan ZQ (1991) Hydraulic underground cavern medium of rockburst mechanism and criterion. Water Power, (6), 30–34

Jiang Q, Feng XT, Xiang TB, Su GS (2010) Rockburst characteristics and numerical simulation based on a new energy index: a case study of a tunnel at 2,500 m depth. Bulletin of Engineering Geology and the Environment, 69(3), 381

Jiang Y, Pan Y, Jiang F, Dou L, Ju Y (2014) State of the art review on mechanism and prevention of coal bumps in China. Journal of China Coal Society, 39 (2), pp. 205-213, (in Chinese)

Jiang LF (2008) Study on prediction and prevention of rockburst in Anlu tunnel. Master's thesis, Southwest Jiaotong Univ., Chendu, China

Kaiser P, McCreath D, Tannant D (1996) Rockburst support handbook Geomechanics Research Centre. Laurentian University

Kang Y (2006) Research on relevant problems about failure mechanism of surrounding rock in deep buried tunnel. Ph.D. thesis, Chongqing Univ., Chongqing, China, 118–120

Khademian Z (2016) Studies of Seismicity Generated by Unstable Failures around Circular Excavations. Conference: 50th US Rock Mechanics/Geomec Symposium, At Houston, TX

Khademian Z, Ozbay U (2019) Modeling violent rock failures in tunneling and shaft boring based on energy balance calculations. Tunnelling and Underground Space Technology, 90, 62–75

Kidybinski A (1981) Bursting liability indices of coal. Int. J. Rock Mech. Mining Sci. Geomech. Abstr, 18(4), 295–304

Li N, Feng X, Jimenez R (2017) Predicting rock burst hazard with incomplete data using Bayesian networks. Tunn Undergr Sp Tech 61:61–70

Li N, Jimenez R (2018) A logistic regression classifier for long-term probabilistic prediction of rock burst hazard. Nat Hazards 90:197–215

Li TZ, Li YX, Yang XL (2017) Rock burst prediction based on genetic algorithms and extreme learning machine. J Central South Univ 24:2105–2113

Li DQ, Wang LG (2009) Theory and technology of the large-scale mining in hard-rock and deep mine-A case study of Dongguashan copper mine. Beijing: Metallurgical Industry Press

Li L (2009) Study on scheme optimization and rockburst prediction in deep mining in Xincheng gold mine. Ph.D. thesis, Univ. of Science and Technology, Beijing

Li XF, and Xie CJ (2005) Research on prevention of rock burst in deep high-stress area of Fankou deposit. Mining Res. Dev., 25(1), 76–79

Liang ZY (2004) Study on the prediction and prevention of rockburst in the diversion tunnel of Jinping II hydropower station. Master's thesis, Chengdu Univ. of Technology, Chendu, China, 61–62

Lin Y, Zhou K, Li J (2018) Application of cloud model in rock burst prediction and performance comparison with three machine learning algorithms. IEEE Access 6:30958–30968

Liu JH, Di MH, Guo XS, Jiang FX, Sun GJ, Zhang ZW (2014) Theory of coal burst monitoring using technology of vibration field combined with stress field and its application. Journal of the China Coal Society, vol. 39, no. 2, pp. 353–363

Liu Z, Shao J, Xu W, Meng Y (2013) Prediction of rock burst classification using the technique of cloud models with attribution weight. Nat Hazards 68:549–568

Liu JP (2011) Studies on relationship between Microseism time-space evolution and ground pressure activities in deep mine. Ph.D. thesis, Northeastern Univ., Shenyang, China

Manouchehriana A, Cai M (2018) Numerical modeling of rockburst near fault zones in deep tunnels. Tunnelling and Underground Space Technology 80 (2018) 164–180

Martin C, Kaiser P, McCreath D (1999) Hoek-Brown parameters for predicting the depth of brittle failure around tunnels. Can. Geotech. J., 36(1), 136–151

McGaughey WJ (2019) Data-driven geotechnical hazard assessment: practice and pitfalls. J Wesseloo (ed.), Proceedings of the First International Conference on Mining Geomechanical Risk, Australian Centre for Geomechanics, Perth, pp. 219-232

Mitri HS, Hassani FP, Kebbe R (1993) A strain energy approach for the prediction of rockburst potential in underground hard rock mines. Proc., 1st Canadian Symp. Numerical Modelling Applications in Mining and Geomechanics, McGill Univ., Montréal, 228–239

Mitri HS, Tang B, Simon R (1999) FE modelling of mining-induced energy release and storage rates. The South African Institute of Mining and Metallurgy, SA ISSN 0038–223X/3.00

Mitri HS (1996) Study of rockburst potential at Sigma Mine. Val d'Or. Quebec using numerical modelling. Final Report. p. 1 29

Morgenroth J, Khan UT, Perras MA (2019) An Overview of Opportunities for Machine Learning Methods in Underground Rock Engineering Design. Geosciences 9(12), 504

Muller W (1991) Numerical simulation of rock bursts. Elsevier, Mining Science and Technology, Volume 12, Issue 1, January 1991, pp 27-42

Neyman B, Szecowka Z, Zuberek W (1972) Effective methods for fighting rock burst in polish collieries. In:Proceedings of the 5th international strata control conference. pp 1–9

Ortlepp WD, Stacey TR (1994) Rockburst mechanisms in tunnels and shafts. Tunnelling and Underground Space Technology, 9(1): 59–65

Peng Z, Wang YH, Li TJ (1996) Griffith theory and rock burst of criterion. Chinese J. Rock Mech. Eng., 15, 491–495

Peng Q, Qian AG, Xiao Y (2010) Research on Prediction System for Rockburst Based on Artificial Intelligence Application Methods. Journal of Sichuan University 42: 18–24

Poeck EC, Khademian Z, Garvey R, Ozba U (2016) Modeling unstable rock failures in underground excavations. Conference: Rock Mechanics and Rock Engineering: From the Past to the Future, Volume: Pages 505–509

Pu Y, Apel DB, Lingga B (2018) Rockburst prediction in kimberlite using decision tree with incomplete data. J Sustain Min 17:158–165

Qin SW, Chen JP, Wang Q (2009) Research on rockburst prediction with extenics evaluation based on rough set. Proc., RaSiM7 (2009): Controlling Seismic Hazard and Sustainable Development of Deep Mines, C. A. Tang, ed., Rinton Press, Princeton, NJ, 937–944

Qinghua X, Jianguo L, Shenxiang L, Bo G (2016) A new method for calculating energy release rate in tunnel excavation subjected to high in situ stress. Perspectives in Science, 7, 292–298

Qiu D, Chen J, Xu Q (2019) Dynamic responses and damage forms analysis of underground large scale frame structures under oblique SV seismic waves. Soil Dynamics and Earthquake Engineering, 117, 216–220

Qiu SL, Feng XT, Zhang CQ, Wu WP (2011) Development and validation of rockburst vulnerability index (RVI) in deep hard rock tunnels. Chin. J. Rock Mech. Eng. 30 (6), 1126–1141

Raffaldi MJ, Chambers DJA, Johnson JC (2017) Numerical study of the relationship between seismic wave parameters and remotely triggered rockburst damage in hard rock tunnels. Proceedings of the Eighth International Conference on Deep and High Stress Mining, Australian Centre for Geomechanics, Perth, pp. 373-386

Ribeiro Sousa L, Miranda T, Leal e Sousa R, Tinoco J (2017) The Use of Data Mining Techniques in Rockburst Risk Assessment. Engineering, 3(4), 552–558

Russenes BF (1974) Analysis of rock spalling for tunnels in steep valley sides. M.Sc. thesis, Norwegian Institute of Technology, Trondheim, Norway, 247

Ryder JA (1988) Excess shear stresses in the assessment of geologically hazardous situations. Journal of South Africa Institute of Mining and Metallurgy, pp. 27-39

Salamon MDG (1983) Rockburst hazard and the fight for its alleviation in South Africa. Symposium Papers, Rockbursts: Prediction and Control, Institute of Mining and Metallurgy, London, pp. 11–36.

Salamon MDG (1984) Energy considerations in rock mechanics: fundamental results. I. S.Afr. Inst. Min. Metall.. Vol. 84, No. 8, pp.233-246.

Sharan SK (2007) A finite element perturbation method for the prediction of rockburst. Computers and Structures 85 (2007) 1304–1309

Shepherd J, Rixon L, Griffiths L (1981) Outbursts and geological structures in coal mines: a review. Int J Rock Mech Min Sci Geomech Abstracts, 267–83. Elsevier

Shirani Faradonbeh R, Taheri A, Ribeiro e Sousa L, Karakus M (2020) Rockburst assessment in deep geotechnical conditions using true-triaxial tests and data-driven approaches. International Journal of Rock Mechanics and Mining Sciences, 128

Singh SP (1987) The influence of rock properties on the occurrence and control of rockbursts. Mining Sci. Technol., 5(1), 11–18

Su GS, Zhang XF, Yan LB (2008) Rockburst prediction method based on case reasoning pattern recognition. J. Min. Saf. Eng. 25 (1), 63–67.

Su GS, Zhang Y, Chen GQ (2010) Identify rockburst grades for Jinping hydropower station using Gaussian II process for binary classification. Proc., 2010 Int. Conf. on Computer, Mechatronics, Control and Electronic Engineering (CMCE 2010), Vol. 2, IEEE Press, NJ, 364–367

Sun J, Lang J, Fujita H, Li H (2018) Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging

Sun Y, Wong AKC, Kamel MS (2009) Classification of imbalanced data: a review. International Journal of Pattern, International Journal of Pattern Recognition and Artificial Intelligence Vol. 23, No. 4 (2009) 687–719 c World Scientific Publishing Company

Sun HF, Li SC, Qiu DH, Zhang LW, Zhang N (2009) Application of extensible comprehensive evaluation to rockburst prediction in a relative shallow chamber. Proc., RaSiM7 (2009): Controlling Seismic Hazard and Sustainable Development of Deep Mines, C. A. Tang, ed., Rinton Press, Princeton, NJ, 777–784

Tajdus K, Flisiak J, Cala M (1997) Estimation of rockburst hazard basing on 3D stress field analysis. Rockbursts and Seisrnicity in Mines, Gibowicz & Lasocki(ed.), Balkema, Rotterdam, ISBN 9054 108908, pp272-277

Tan YA (1992) Rockbursting characteristics and structural effects of rock mass. International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 29(6)

Tianwei L, Hongwei Z, Sheng L, Jun H, Weihua S, Batugin AC, Guoshui T (2015) Numerical Study on 4-1 Coal Seam of Xiaoming Mine in Ascending Mining. Scientific World Journal, 2015, 1-4

Tsangaratos P, Ilia I. (2014). A supervised machine learning spatial tool for detecting terrain deformation induced by landslide phenomena. Proceedings of the 10th International Congress of the Hellenic Geographical Society 22-24 October 2014, Thessaloniki, Greece

Turchaninov IA, Markov GA, Gzovsky MV, Kazikayev DM, Frenze UK, Batugin SA, Chabdarova UI (1972) State of stress in the upper part of the Earth's crust based on direct measurements in mines and on tectonophysical and seismological studies. Physics of the Earth and Planetary Interiors, 6(4), 229–234

Vardar O, Zhang C, Canbulat I, Hebblewhite B (2019) Numerical modelling of strength and energy release characteristics of pillar-scale coal mass. Journal of Rock Mechanics and Geotechnical Engineering 11 935-943

Vatcher J, McKinnon SD, Sjöberg J (2014) Mine-scale numerical modelling, seismicity and stresses at Kiirunavaara Mine, Sweden. Proceedings of the Seventh International Conference on Deep and High Stress Mining, Australian Centre for Geomechanics, Perth, pp. 363-376

Wang C (2018) Predicting Model of Rockburst Based on Nondeterministic Theory, Evolution, Monitoring and Predicting Models of Rockburst. pp.149-161

Wang JA, Park HD (2001) Comprehensive prediction of rockburst based on analysis of strain energy in rocks. Tunnelling and Underground Space Technology, 16(1), 49–57

Wang GY, Zhang SX, Ren GF (2005) Analysis and prediction of rock burst in deep mining of Tonglushan copper-iron ore. Mining Saf. Environ. Prot., 32(5), 20–22

Wang JL, Chen JP, Yang J, Que JS (2009) Method of distance discriminant analysis for determination of classification of rockburst. Rock Soil Mech., 30(7), 2203–2208

Wang XF, Li XH, Gu YL, Jin XG, Kang Y, Li DX (2004) Application of BP neural network into prediction of rockburst in tunneling. Proc., 2004 Int. Symp. on Safety Science and Technology, China Science Press, Shanghai, China, 617–621

Wang YC, Shang YQ, Sun HY, Yan XS (2010) Research and application of rockburst intensity prediction model based on entropy coefficient and ideal point method. J. China Coal Soc., 35(2), 218–221

Wang YH, Li WD, Li QG (1998) Method of fuzzy comprehensive evaluations for rockburst prediction. Chinese Journal of Rock Mechanics and Engineering. 15, 493-50

Weng L, Huang LQ, Taheri A, Li XB (2017) Rockburst characteristics and numerical simulation based on a strain energy density index: a case study of a roadway in Linglong gold mine. China. Tunn Undergr Space Technol 69:223–232

Wiles TD (2002) Loading system stiffness—A parameter to evaluate rockburst potential. Proc. 1st Int. Seminar on Deep and High Stress Mining, Australian Centre for Geomechanics, Perth, Australia, 10

Witten I, Frank E (2005) Data Mining: Practical Machine Learning Tools and Techniques. Second Edition 9The Morgan Kaufmann Series in Data Mining Systems

Wu De-Xing, Yang Jian (2005) Prediction and countermeasure for rockburst in Cangling mountain highway tunnel. [J] Chinese Journal of Rock Mechanics and Engineering, 24(21): 3965−3971

Wu S, Wu Z, Zhang C (2019) Rock burst prediction probability model based on case analysis. Tunnelling and Underground Space Technology, 93

Xia BW (2006) Study on prediction and forecast of geologic disaster in highway tunnel construction. Master's thesis, Chongqing Univ., Chongqing, China

Xiao XP (2005) A study on the prediction and prevention of rockburst traffic tunnel of Jinping II hydropower station. Master's thesis, Chengdu Univ. of Technology, Chengdu, China

Xu J, Jiang J, Xu N, Liu Q, Gao Y (2017) A new energy index for evaluating the tendency of rockburst and its engineering application. Eng. Geol. 230, 46–54

Xu MG, Du ZJ, Yao GH, Liu ZP (2008) Rockburst prediction of Chengchao iron mine during deep mining. Chinese J. Rock Mech. Eng., 27(S1), 2921–2928

Yang JL, Li XB, Zhou ZL, Lin Y (2010) A fuzzy assessment method of rock-burst prediction based on rough set theory. Metal Mine, 6, 26–29

Yi YL, Cao P, Pu CZ (2010) Multi-factorial comprehensive estimation for Jinchuan's deep typical rockburst tendency. Sci. Technol. Rev., 28(2), 76–80

Yu H, Liu H, Lu X, Liu H (2009) Prediction method of rock burst proneness based on rough set and genetic algorithm. Journal of Coal Science and Engineering (China), 15(4), 367–373.

Yu XZ (2009) Highway tunnel geological disaster prediction and the development of treatment measures database management system. Master's thesis, Chongqing Univ., Chongqing, China

Zhang C, Feng XT, Zhou H, Qiu S, Wu W (2012) Case histories of four extremely intense rockbursts in deep tunnels. Rock Mech Rock Eng;45(3):275–88

Zhang CQ, Yu J, Chen J, Lu JJ, Zhou H (2016) Evaluation method for potential rockburst in underground engineering. Rock Soil Mech. 37 (S1), 341–349

Zhang J, Fu B (2008) Rockburst and Its criteria and control. Chinese Journal of Rock Mechanics and Engineering

Zhang Xuan-Zhuang (2005) Prediction of rock burst at underground works based on artificial neural network. [J] Yangtze River, 36(5): 17−18. (in Chinese)

Zhang CQ, Zhou H, Feng XT (2011) An index for estimating the stability of brittle surrounding rock mass: FAI and its engineering application. Rock Mech. Rock Eng., 44(4), 401–414

Zhang LW, Zhang DY, Qiu DH (2010) Application of extension evaluation method in rockburst prediction based on rough set theory. J. China Coal Soc., 35(9), 1461–1465

Zhang LX, Li CH (2009) Study on tendency analysis of rockburst and comprehensive prediction of different types of surrounding rock. RaSiM7 (2009): Controlling Seismic Hazard and Sustainable Development of Deep Mines, C. A. Tang, ed., Rinton Press, Princeton, NJ, 1451–1456

Zhang YL, Liu X, Hu ZQ (2007) Rock burst forecast based on artificial neural network in underground engineering. Hunan Nonferrous Metal, 23(3), 1–4

Zhang ZL (2002) Study on rockburst and large deformation of Xuefeng mountain tunnel of Shaohuai highway. Master's thesis, Chengdu Univ. of Technology, Chengdu, China

Zhao HB (2005) Classification of rockburst using support vector machine. Rock Soil Mech 26(4):642–644

Zhao XF (2007) Study on the high geo-stress and rockburst of the deep-lying long tunnel. Master's thesis, North China Univ. of Water Resources and Electric Power, Zhengzhou, China

Zhou J, Li X, Shi X (2012) Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. Safety Science, 50(4), 629–644

Zhou J, Shi XZ, Dong L, Hu HY, Wang HY (2010) Fisher discriminant analysis model and its application for prediction of classification of rockburst in deep buried long tunnel. J Coal Sci Eng (China) 16(2):144–149

Zhou J, Li X, Mitri H (2016) Classification of Rockburst in Underground Projects: Comparison of Ten Supervised Learning Methods. J Comput. Civ. Eng., 04016003

Zhou KP, Lin Y, Deng HW, Li JL, Liu CJ (2016) Prediction of rock burst classification using cloud model with entropy weight. Trans Nonferrous Meterol Soc China 26:1995–2002

Zhu Q, Lu W, Sun J, Luo Y, Chen M (2009) Prevention of rockburst by guide holes based on numerical simulations. Mining Science and Technology (China), 19(3), 346–351

Zhu YH, Liu XR, Zhou JP (2008) Rockburst prediction analysis based on v-SVR algorithm. J Chin Coal Soc 33:277–281

List of figures

Techniques	Attributes	Data	Accuracy (%)	References
Artificial neural networks	σθ, σc, σt, Wet	18	72.2	Chen (2003)
Artificial neural networks	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	50-67.5	Zhou (2016)
Decision tree	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	56.3–60.9	Zhou (2016)
Gradient boosting machine	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	61-76.6	Zhou (2016)
Naïve Bayes	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	53.9–67.2	Zhou (2016)
k-nearest neighbors	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	53.2–67.2	Zhou (2016)
Random forest	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	55.9–76.6	Zhou (2016)
Support vector machine	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	51.7–67.2	Zhou (2016)
Fisher linear discriminant analysis	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	48.4–55.9	Zhou (2016)
Quadratic discriminant analysis	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	48.4–60.9	Zhou (2016)
Partial least-squares discriminant analysis	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	246	45.3–57.5	Zhou (2016)
Decision tree-based C4.5 algorithm	σθ, σc, σt, Wet	134	81.48	Faradonbeh (2018)
Gene expression programming	σθ, σc, σt, Wet	134	85.16	Faradonbeh (2018)
Artificial neural networks	σθ, σc, σt, Wet	134	85.19	Faradonbeh (2018)
Bayes discriminant analysis	σθ/σc, σc/σt, Wet	21	100	Gong (2010)
Fisher linear discriminant analysis	σθ/σc, σc/σt, Wet	15	100	Zhou (2010)
Support vector machine	σθ, σc, σt, Wet	16	100	Zhao (2010)
Support vector machine	σθ, σc, σt, σθ/σc, σc/σt, Wet	45	93.75	Zhu (2008)
Support vector machine (optimized by GSM)	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	132	66.67–88.9	Zhou (2012)
Support vector machine (optimized by GA)	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	132	66.67-80	Zhou (2012)
Support vector machine (optimized by PSO)	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	132	66.67-90	Zhou (2012)
Adaptive neuro fuzzy inference system	σθ, σc, σt, σθ/σc, σc/σt, Wet	174	66.5–95.6	Adoko (2013)
Adaptive boosting	σθ, σθ/σc, σc/σt, Wet	36	87.8–89.9	Ge and Feng (2008)
Random forest	σθ, σc, σt, Wet	46	100	Dong (2013)
Bayesian network	σθ, σc, σt, σθ/σc, σc/σt, Wet	246	53.95	Lin (2018)
k-nearest neighbors	σθ, σc, σt, σθ/σc, σc/σt, Wet	246	50	Lin (2018)
Random forest	σθ, σc, σt, σθ/σc, σc/σt, Wet	246	60.53	Lin (2018)

Cloud model	σθ, σc, σt, σθ/σc, σc/σt, Wet	246	71.05	Lin (2018)
Cloud model	σθ, σc, σt, σθ/σc, σc/σt, Wet	164	90-94.1	Li (2013)
Cloud model	σθ, σc, σt, σθ/σc, σc/σt, Wet	209	76.4–82	Zhou (2016)
Bayesian network	H, σθ, σc, σt, Wet	135	91.75	Li (2017)
Decision tree-based ID3 algorithm	σθ/σc, σc/σt, Wet	132	73–93	Pu (2018)
Genetic algorithms and extreme learning machine	σθ, σc, σt, Wet	30	100	Li (2017)
Logistic regression	H, σθ, σc, σt, Wet	135	80.2–90.9	Li and Jimenez (2018)
Logistic regression	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	188	88.3	Afraei (2018)
Probit regression	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	188	87.77	Afraei (2018)
Ordinal regression	H, σθ, σc, σt, σθ/σc, σc/σt, Wet	188	60.64	Afraei (2018)
Decision Tree	σθ/σc, σc/σt , Wet	174	90.23	Ghasemi (2019)

Table 1: Different ML models for rockburst prediction including the principle parameters (attributes) used, the total database cases used, as well as the attained accuracy prediction.

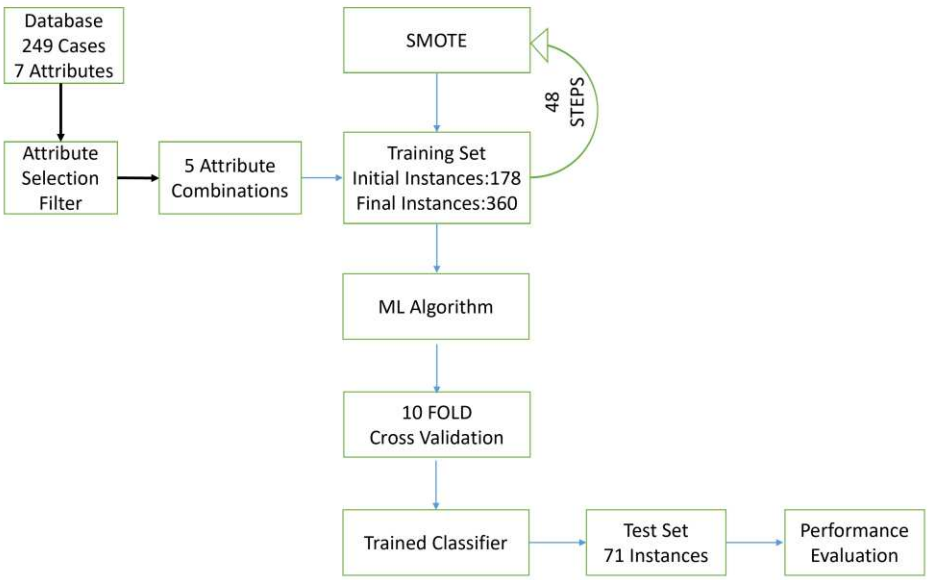

Figure 1: Stages of the proposed ML methodology to assess rockburst intensity

Rockburst Intensity	**Description / Consequence**
None	No sound of rock burst and absence of rock burst activities
Low	May cause loosening of a few fragments. The surrounding rock will be deformed, cracked or rib spalled. There would be a weak sound, but no ejection phenomenon

Moderate	Spalling and falls of thin rock fragments. The surrounding rock will be deformed and fractured; there may be a considerable number of rock chip ejections and loose and sudden destructions, accompanied by crisp crackling and often presented in the local cavern of surrounding rock
High	Loosening and falls, often as violent detachment of fragments and platy blocks. The surrounding rock will be bursting severely and suddenly thrown out or ejected into the tunnel, accompanied by strong bursts and roaring sound, and will expand rapidly to the deep surrounding rock

Table 2: Rockburst Intensity Classification (Zhou 2012)

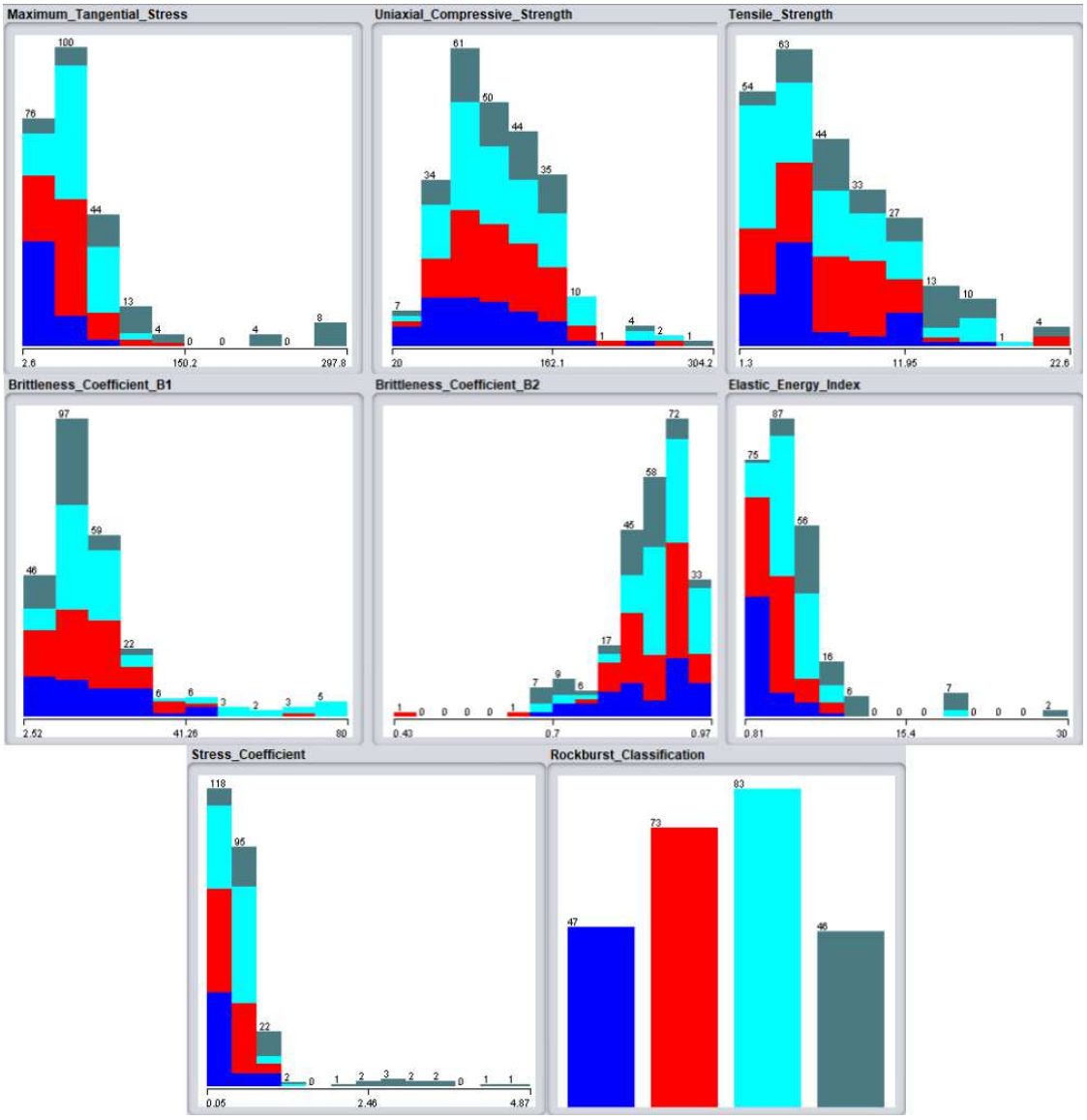

Figure 2: Data Visualization in terms of rockburst intensity and attribute distribution

	Maximum Tangential Stress (σθ)	Uniaxial Compressive Strength (σc)	Tensile Strength (σt)	Stress Coefficient (σθ/σc)	Brittleness Coefficient B1 (σc/σt)	Brittleness Coefficient B2 ((σc-σt)/(σc+σt))	Elastic Energy Index (Wet)

Minimum	2.6	20	1.3	0.052	2.52	0.432	0.81
Maximum	297.8	304.2	22.6	4.874	80	0.975	30
Mean	58.37	114.084	7.501	0.583	20.301	0.872	5.205
Standard Deviation	54.053	44.794	4.433	0.67	14.02	0.071	4.134

Table 3: Basic Statistical Information of the Input Attributes

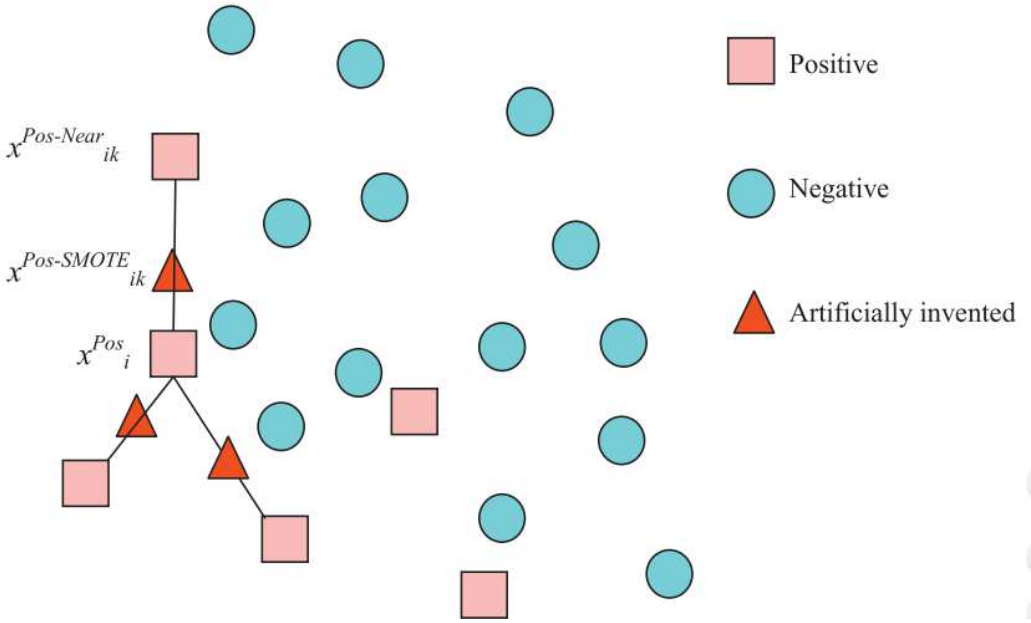

Figure 3: SMOTE illustration (Sun 2018)

No			MIN	START	MAX		MIN	START	MAX		MIN	START	MAX		MIN	START	MAX		MIN	START	MAX
7	J48	ACC	59,15	60,56	69,01	KNN	60,56	60,56	67,61	L.R	49,30	57,75	59,15	N.B	54,93	57,75	61,97	R.F	66,20	73,24	76,06
		%	-2		14		0		12		-17		2		-5		7		-11		4
		k	0,44	0,47	0,59		0,46	0,46	0,57		0,32	0,43	0,45		0,40	0,43	0,49		0,54	0,64	0,68
		%	-5		26		0		23		-32		5		-7		15		-17		6
		F-M	0.59	0.60	0.69		0.61	0.61	0.67		0.48	0.58	0.59		0.53	0.58	0.62		0.66	0.73	0.76
		%	-2.5		13.9		0		9.8		-21.9		0		-9.8		5.7		-11.6		4
		AUC	0,72	0,78	0,86		0,81	0,84	0,86		0,76	0,79	0,79		0,78	0,80	0,81		0,86	0,89	0,90
		%	-8		10		-4		2		-3		2.2		-3		0		-3		1
6		ACC	54,93	63,38	71,83		60,56	64,79	71,83		49,30	57,75	57,75		54,93	60,56	64,79		63,38	74,65	76,06
		%	-15		13		-7		11		-17		0		-10		7		-18		2
		k	0,40	0,51	0,62		0,47	0,52	0,62		0,32	0,43	0,43		0,41	0,47	0,53		0,51	0,66	0,68
		%	-28		22		-10		20		-31		0		-15		13		-29		3
		F-M	0.53	0.63	0.71		0.59	0.65	0.71		0.48	0.58	0.58		0.52	0.60	0.64		0.62	0.75	0.76
		%	-17.6		13.7		-10.1		8.9		-21.9		0		-16.1		6.1		-20.3		1.9
		AUC	0,72	0,80	0,84		0,83	0,85	0,85		0,76	0,78	0,78		0,81	0,82	0,85		0,88	0,89	0,91
		%	-10		5		-2		1		-2		1		-1		4		-2		2
5		ACC	57,75	64,79	70,42		64,79	67,61	73,24		47,89	54,93	56,34		57,61	64,79	70,42		64,79	73,24	77,46
		%	-12		9		-4		8		-15		3		-12		9		-0,13		0,06

	k	0,43	0,52	0,60	0,53	0,56	0,64	0,31	0,39	0,42	0,41	0,52	0,60	0,53	0,64	0,70
	%	-23		14	-7		13	-24		8	-27		16	-0,20		0,09
	F-M	0.58	0.64	0.71	0.64	0.65	0.74	0.47	0.55	0.58	0.58	0.65	0.70	0.64	0.73	0.77
	%	-9.9		10	-2		12.6	-16.4		4.7	-13.2		7.2	-14.9		5.5
	AUC	0,74	0,81	0,83	0,84	0,85	0,88	0,76	0,76	0,78	0,80	0,84	0,84	0,87	0,89	0,91
	%	-10		2	-1		4	0		3	-4		0	-0,02		0,02
4	**ACC**	47,89	69,01	69,01	64,79	69,01	69,01	50,70	57,75	57,75	59,15	66,20	69,01	59,15	74,65	74,65
	%	-44		0	-7		0	-14		0	-12		4	-0,26		0,00
	k	0,30	0,58	0,58	0,53	0,58	0,58	0,35	0,43	0,44	0,46	0,54	0,58	0,45	0,66	0,66
	%	-95		0,00	-9		0	-24		3	-18		7	-0,46		0,00
	F-M	0.48	0.68	0.68	0.64	0.69	0.69	0.49	0.58	0.58	0.57	0.66	0.69	0.58	0.74	0.74
	%	-41.7		0	-7.8		0	-18.8		0	-15.5		3.9	-27.4		0
	AUC	0,70	0,80	0,81	0,82	0,86	0,86	0,77	0,77	0,78	0,82	0,83	0,84	0,84	0,90	0,90
	%	-13		1	-4		0	0		1	-2		0	-0,06		0,01
3	**ACC**	60,56	66,20	70,42	60,56	61,97	67,61	53,52	56,34	57,75	57,75	66,20	69,01	60,56	71,83	74,65
	%	-9		6	-2		9	-5		3	-15		4	-0,19		0,04
	k	0,47	0,54	0,60	0,47	0,49	0,57	0,31	0,41	0,42	0,44	0,54	0,58	0,47	0,62	0,66
	%	-16		11	-4		17	-30		3	-24		7	-0,32		0,06
	F-M	0.60	0.66	0.7	0.61	0.62	0.67	0.48	0.57	0.58	0.57	0.66	0.69	0.60	0.66	0.74
	%	-10.6		5.5	-2.5		7.7	-19.3		2.1	-15.5		3.9	-9.1		12.6
	AUC	0,75	0,82	0,85	0,81	0,82	0,84	0,76	0,78	0,78	0,82	0,82	0,83	0,86	0,88	0,89
	%	-9		3	-1		3	-2		1	0		1	-0,02		0,02

Table 4: SMOTE Effect on the Evaluation Metrics of ML Algorithms

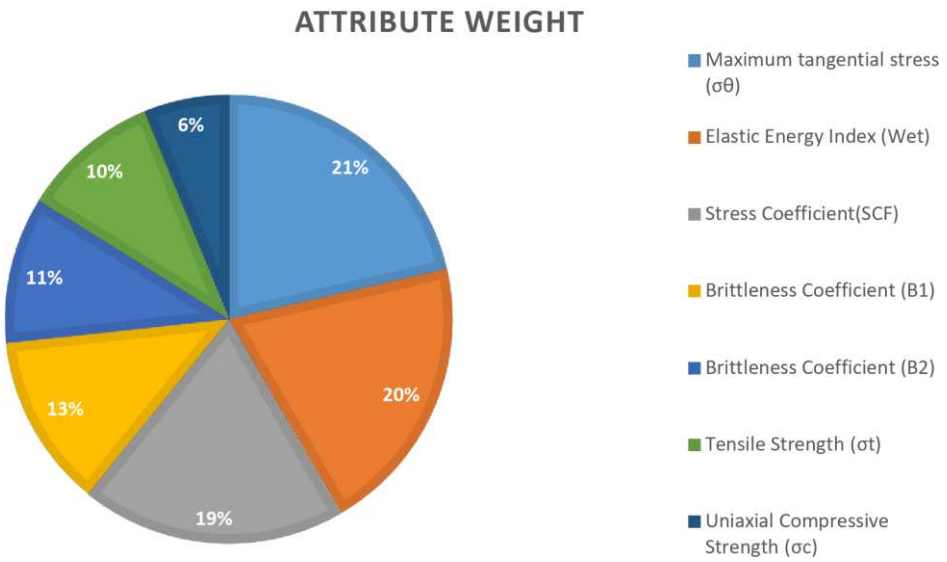

Figure 4: Attribute weight to the overall rockburst intensity

Algorithm	Class	Before SMOTE		After SMOTE		SMOTE Effect %	
		TP Rate	F-Measure	TP Rate	F-Measure	TP Rate	F-Measure
J48 (6 attributes)	None	0.929	0.813	1	0.875	+7.6%	+7.6%
	Low	0.476	0.541	0.619	0.634	+30.0%	+17.2%
	Moderate	0.619	0.565	0.714	0.682	+15.3%	+20.7%

	Heavy	0.6	0.667	0.6	0.72	0%	+7.9%
KNN (5 attributes)	None	0.857	0.8	0.857	0.889	0%	+11.1%
	Low	0.571	0.6	0.762	0.78	+33.5%	+30.0%
	Moderate	0.667	0.636	0.714	0.652	+7%	+2.5%
	Heavy	0.667	0.714	0.6	0.643	-11.0%	-11.0%
Logistic Regression (7 attributes)	None	0.643	0.72	0.714	0.769	+11.0%	+6.8%
	Low	0.667	0.571	0.667	0.583	0%	2.1%
	Moderate	0.381	0.4	0.381	0.4	0%	0%
	Heavy	0.667	0.714	0.667	0.714	0%	0%
Naïve Bayes (5 attributes)	None	0.714	0.8	0.929	0.813	+30.1%	+1.6%
	Low	0.667	0.622	0.619	0.684	-7.8%	+9.9%
	Moderate	0.619	0.578	0.667	0.651	+7.8%	+12.6%
	Heavy	0.6	0.667	0.667	0.69	+11.1%	+3.5%
Random Forest (5 attributes)	None	0.929	0.897	1	0.903	+7.6%	+0.7%
	Low	0.571	0.632	0.619	0.703	+8.4%	+12.2%
	Moderate	0.857	0.706	0.905	0.792	+5.6%	+12.2%
	Heavy	0.6	0.75	0.6	0.692	0%	-8.3%

Table 5: Comparison of Classification Performance in Terms of Within-Class Classification Metrics

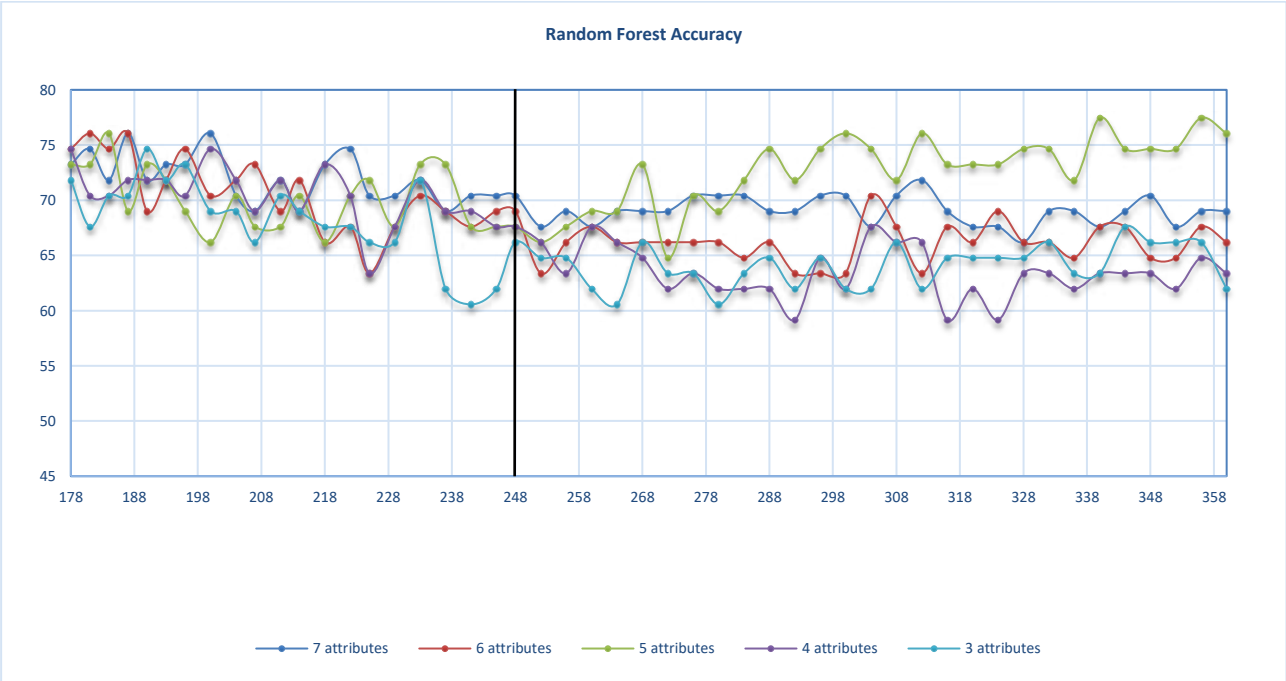

Figure 5: Accuracy attained by the Random Forest ML

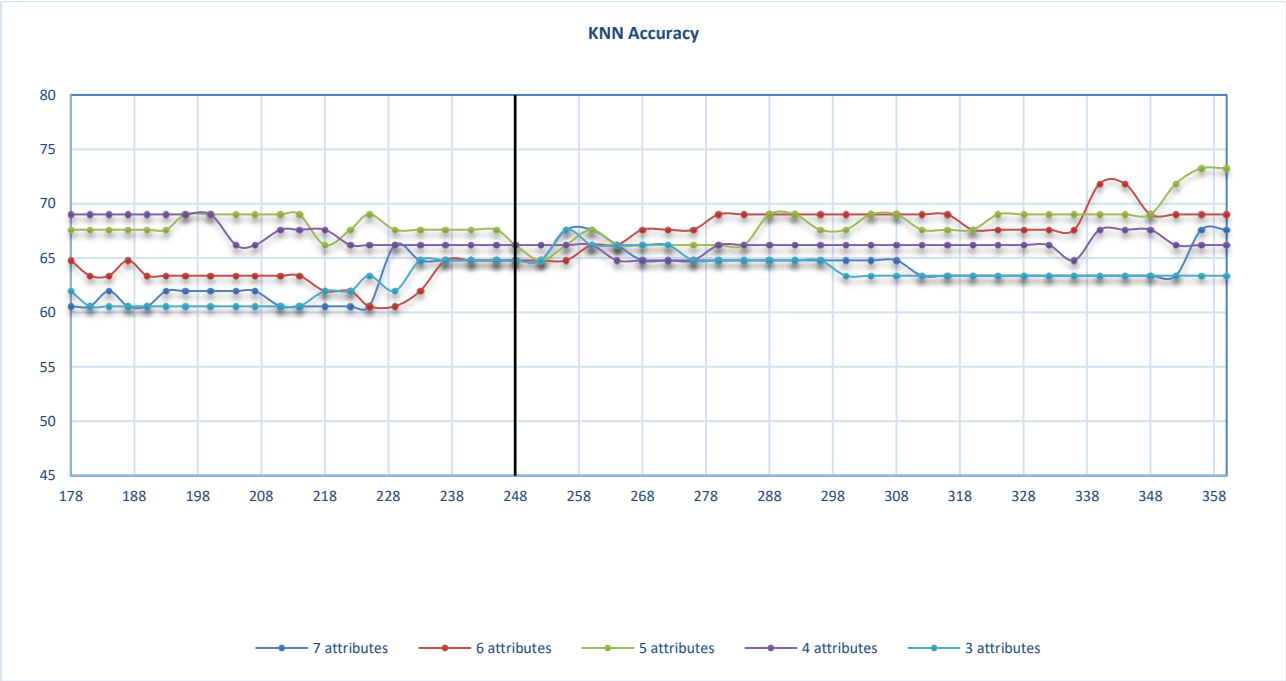

Figure 6: Accuracy attained by the KNN ML model

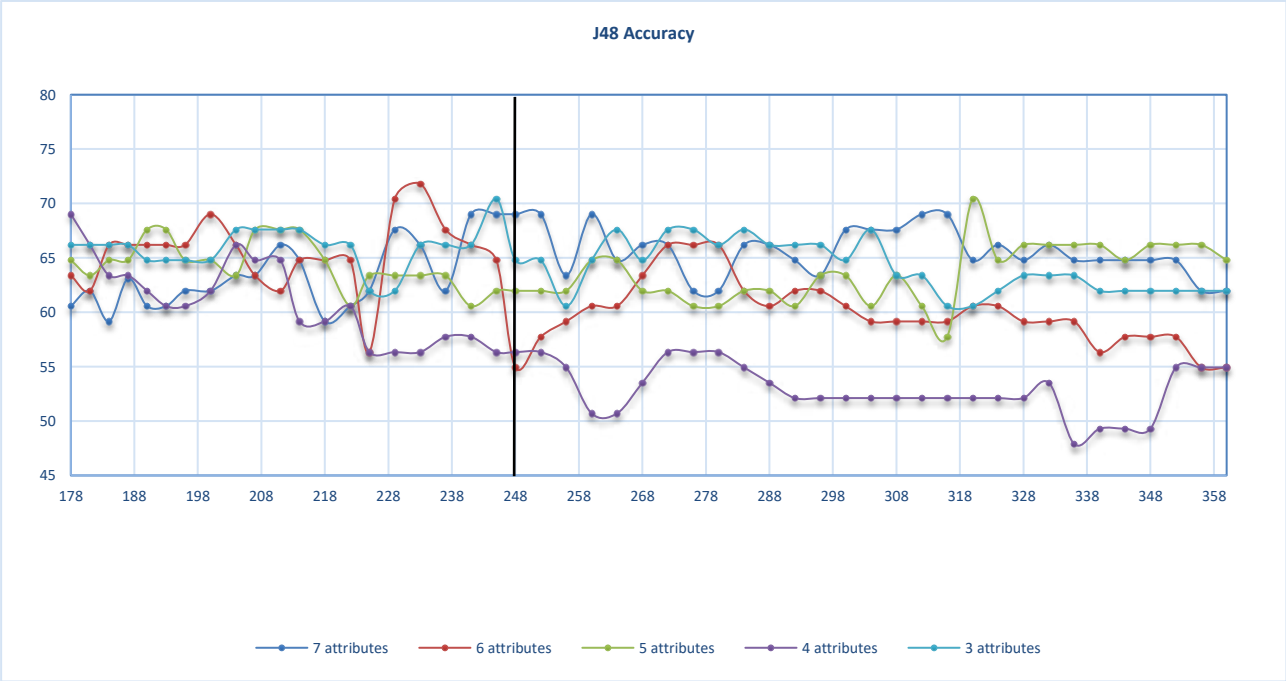

Figure 7: Accuracy attained by the J48 ML model

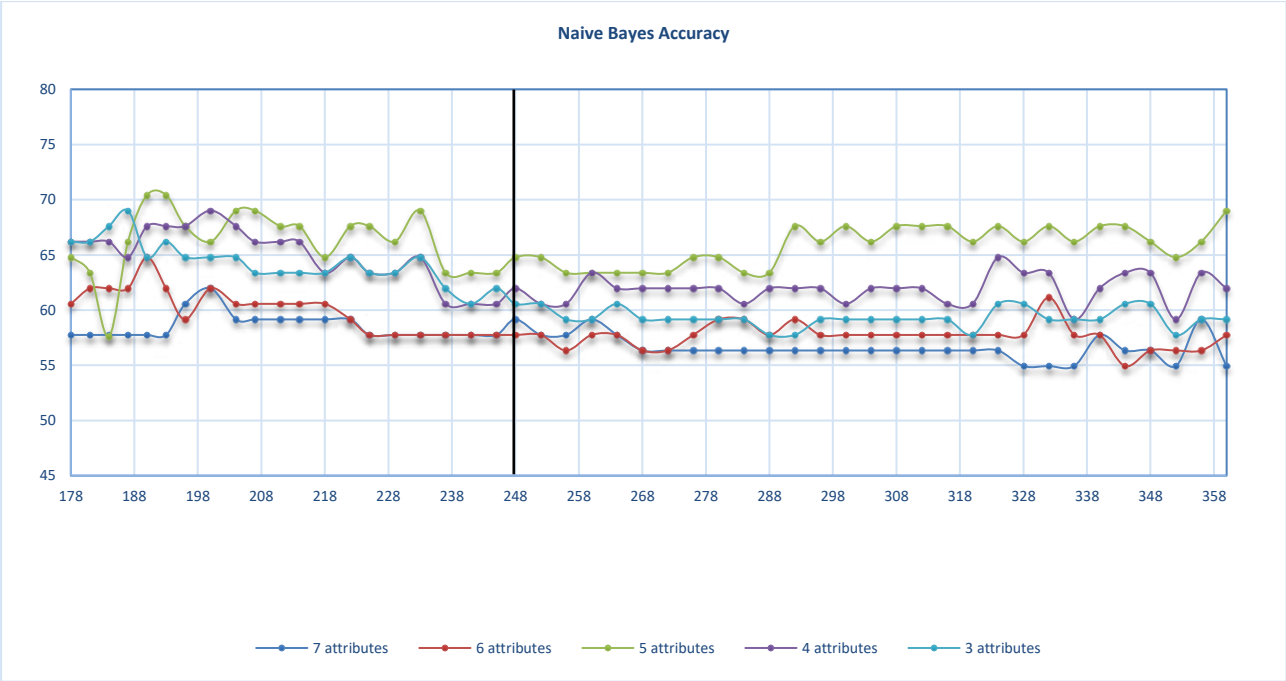

Figure 8: Accuracy attained by the Naïve Bayes ML model

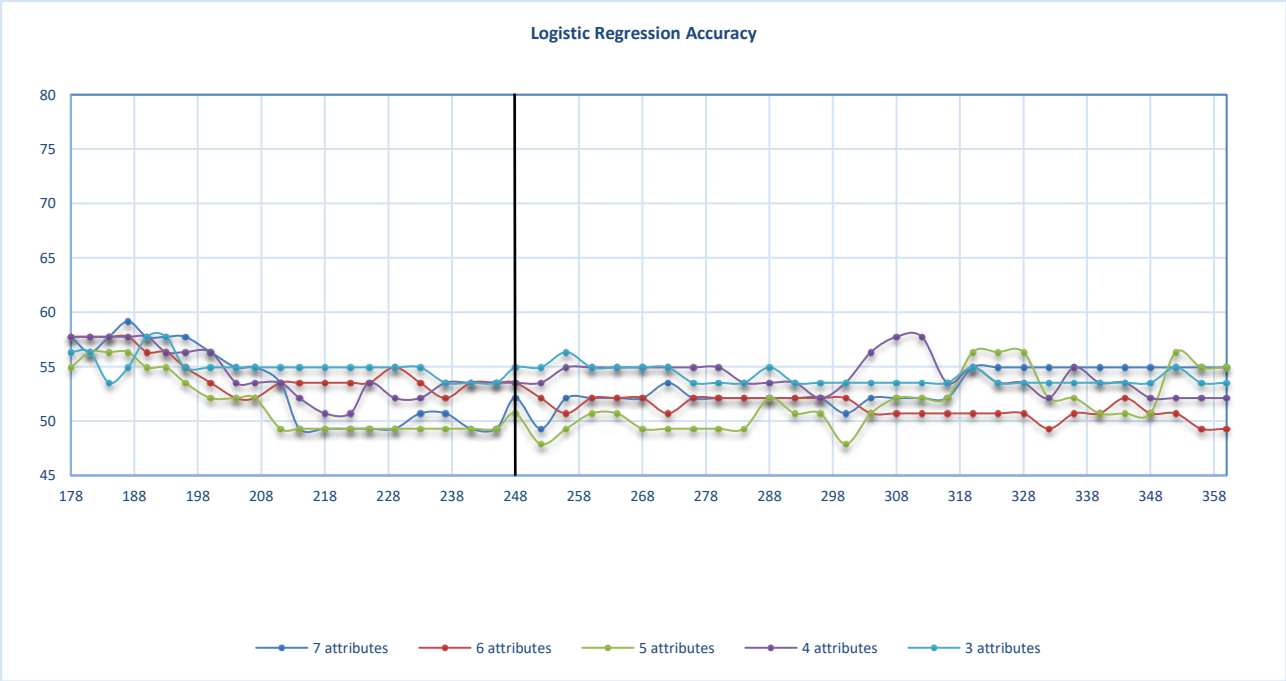

Figure 9: Accuracy attained by the Logistic Regression ML model

Figures

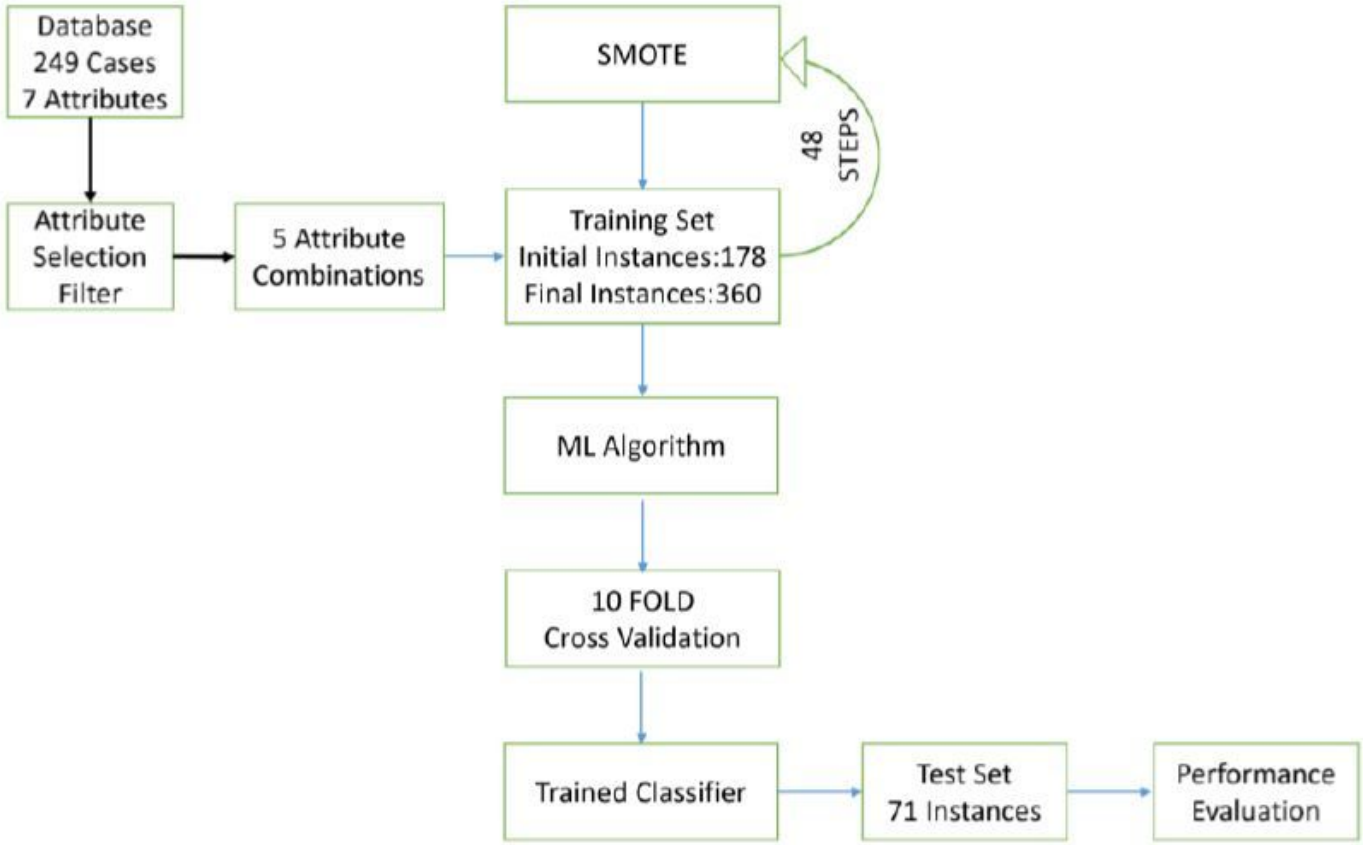

Figure 1

Stages of the proposed ML methodology to assess rockburst intensity

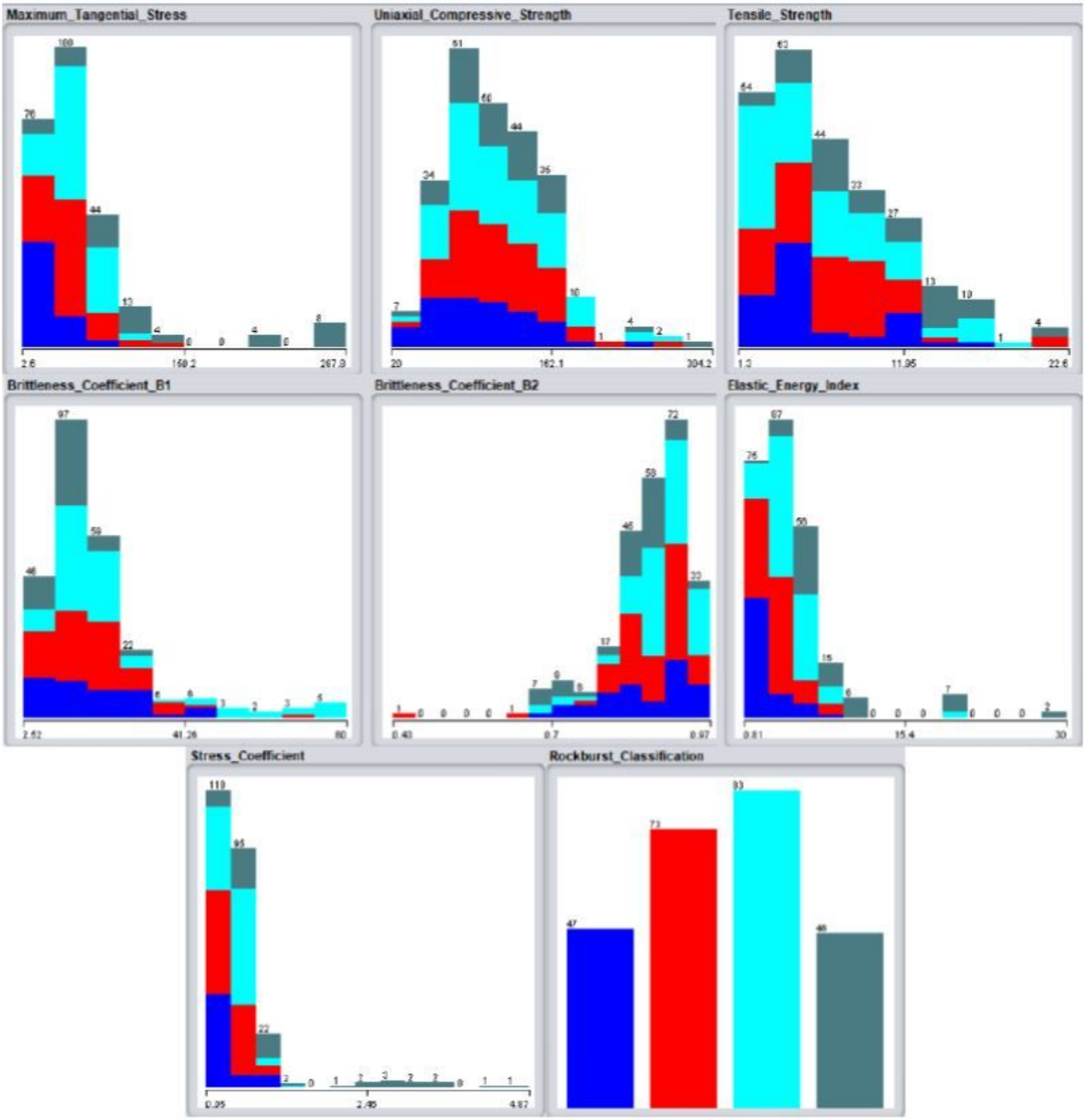

Figure 2

Data Visualization in terms of rockburst intensity and attribute distribution

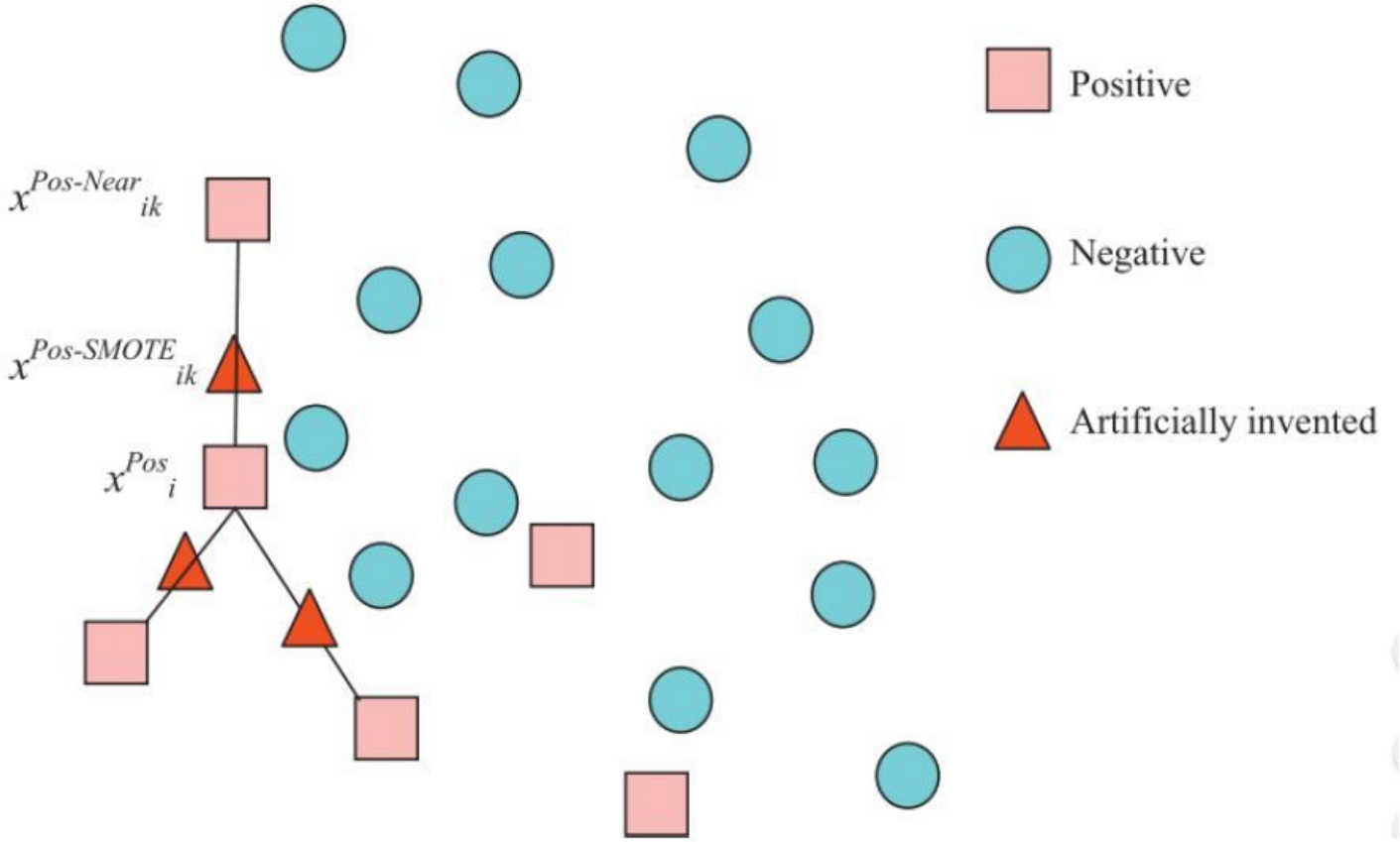

Figure 3

SMOTE illustration (Sun 2018)

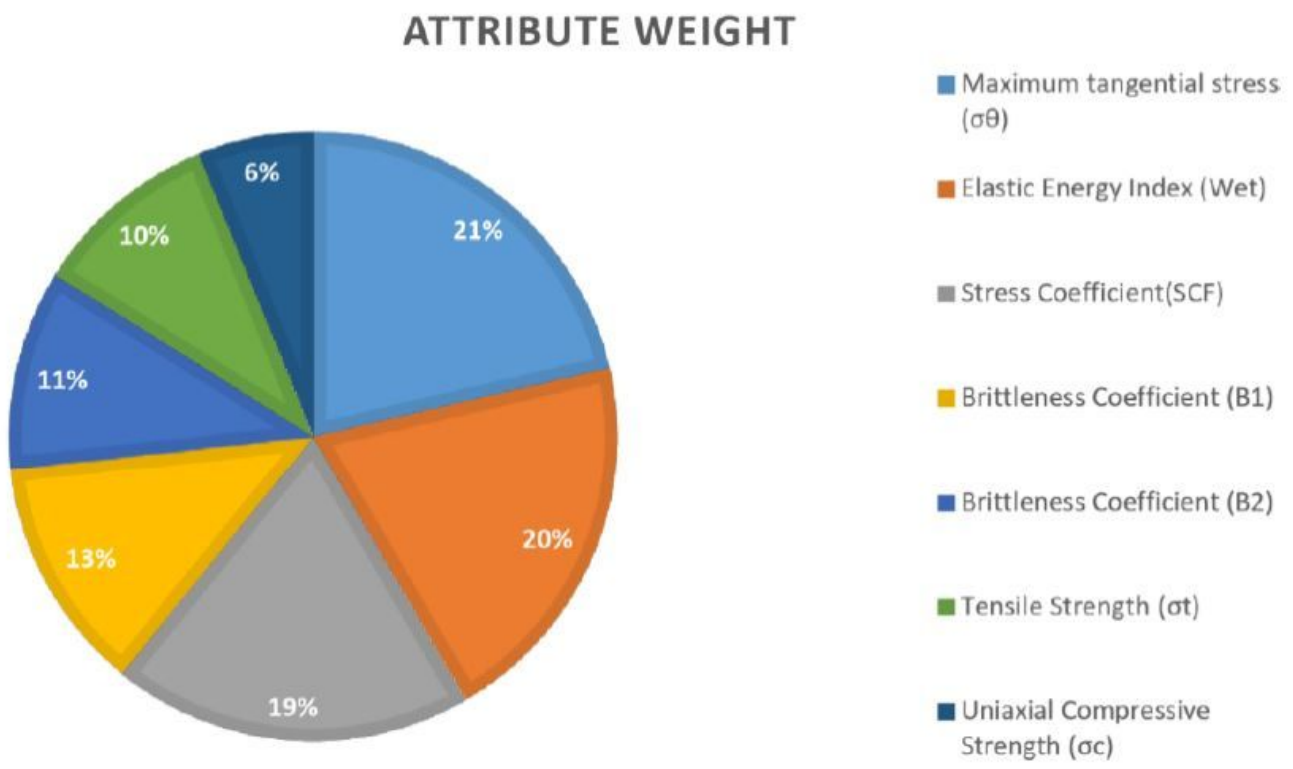

Figure 4

Attribute weight to the overall rockburst intensity

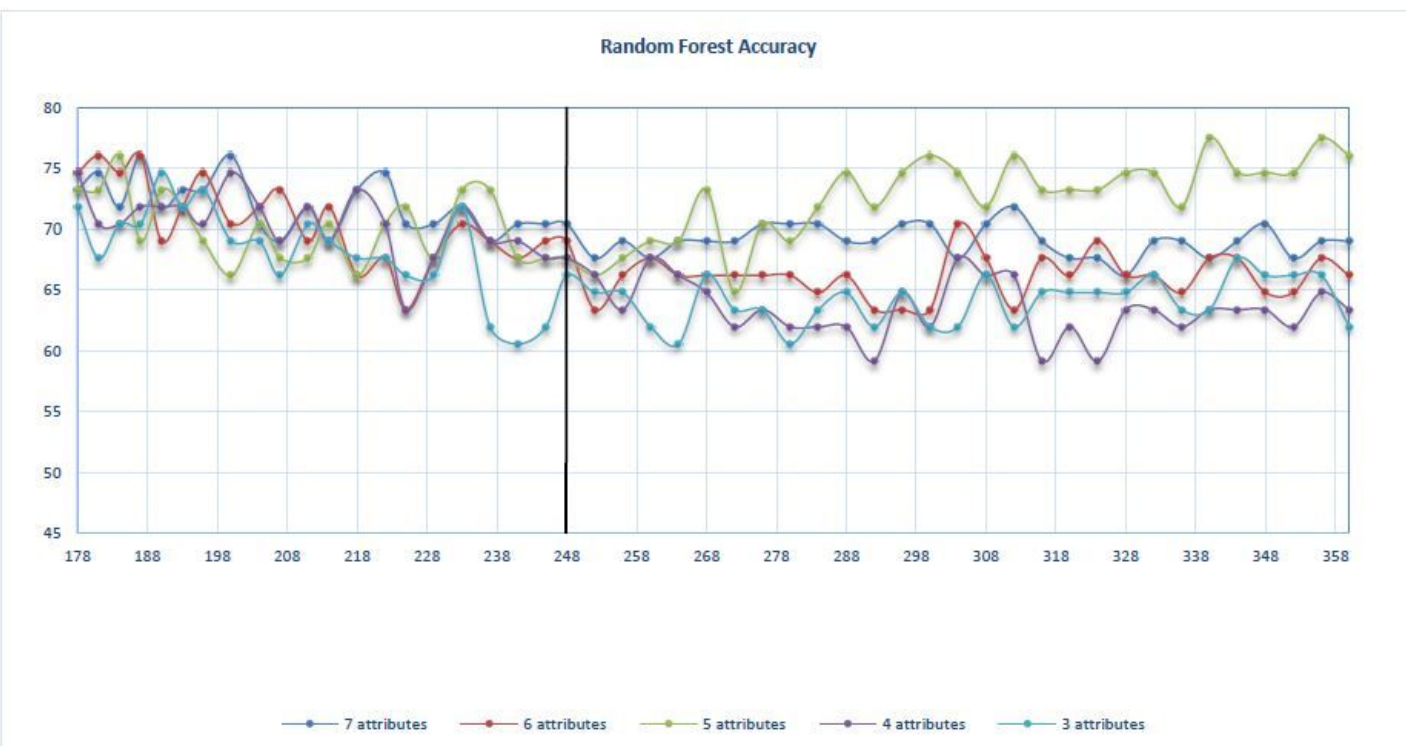

Figure 5

Accuracy attained by the Random Forest ML

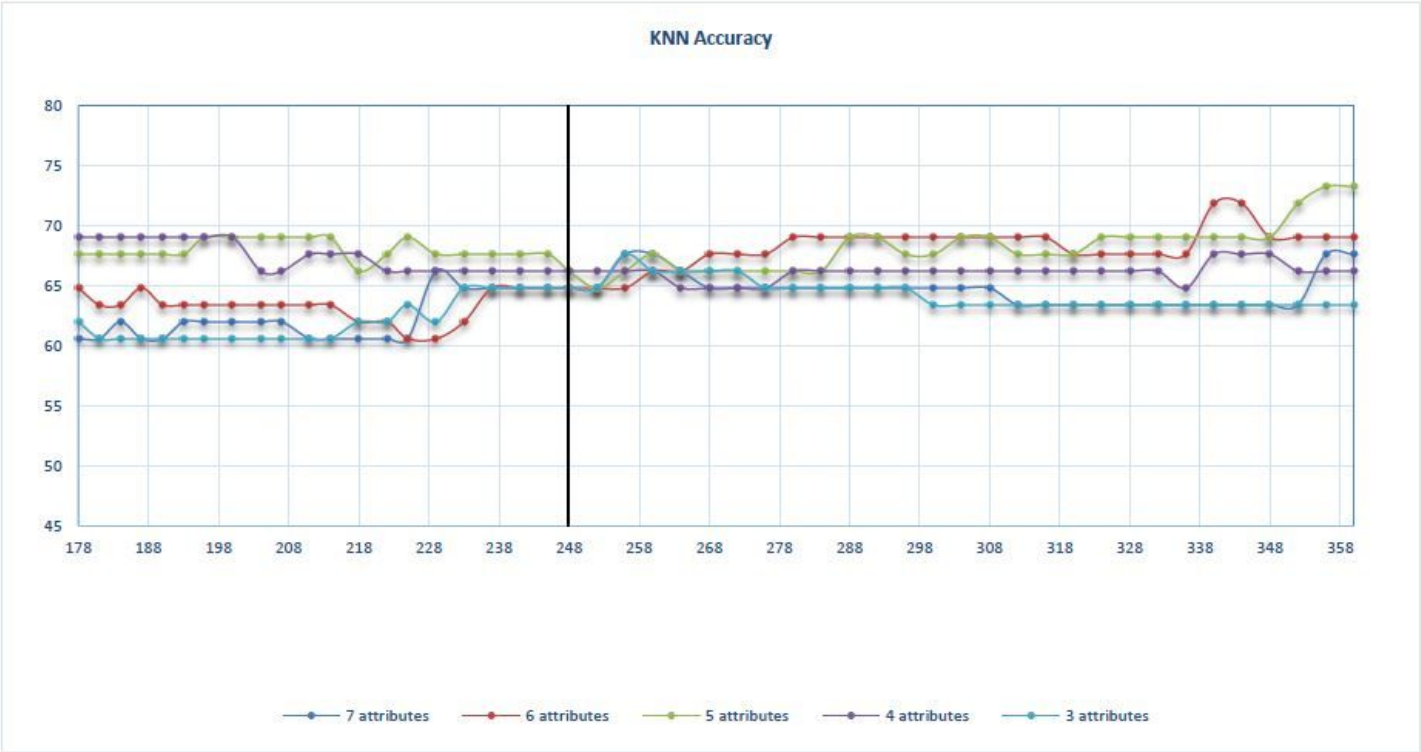

Figure 6

Accuracy attained by the KNN ML model

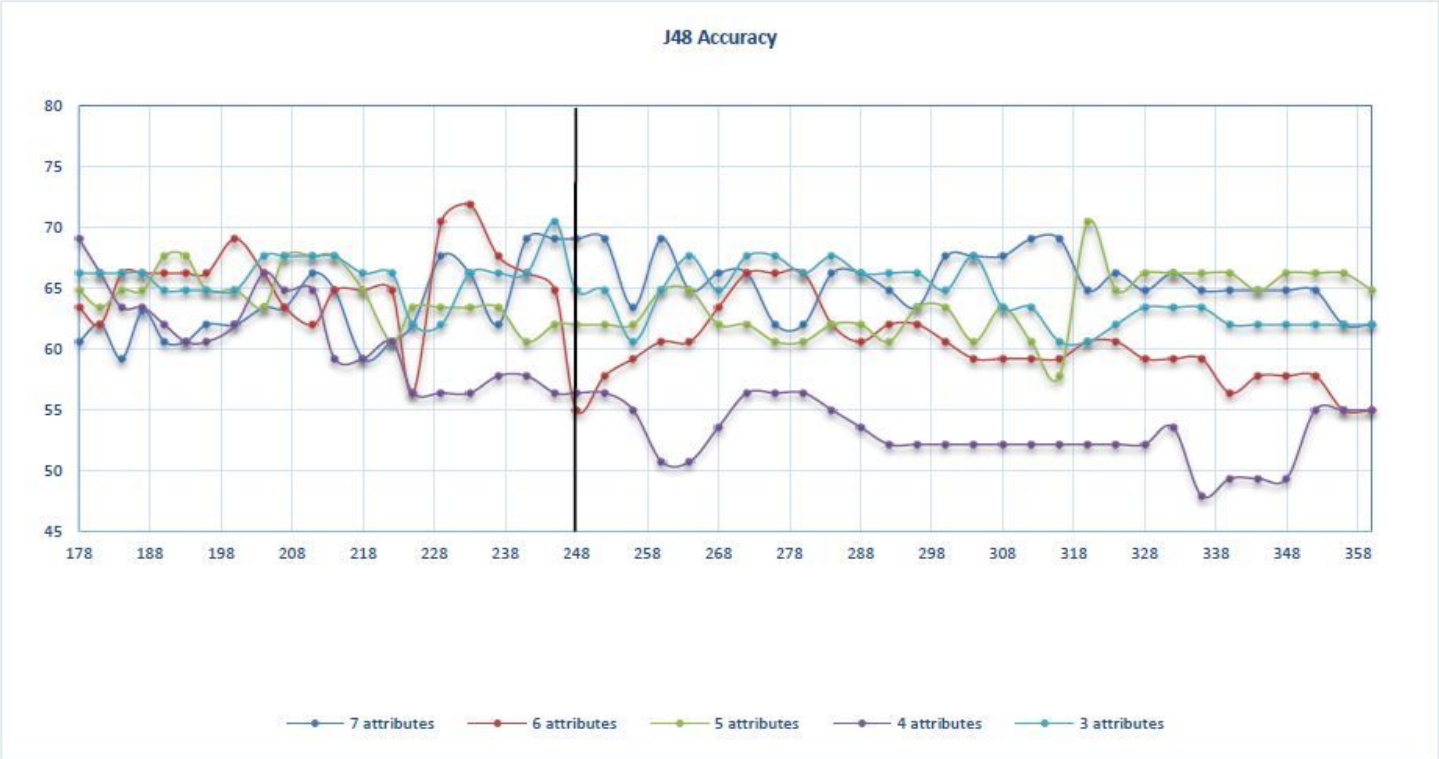

Figure 7

Accuracy attained by the J48 ML model

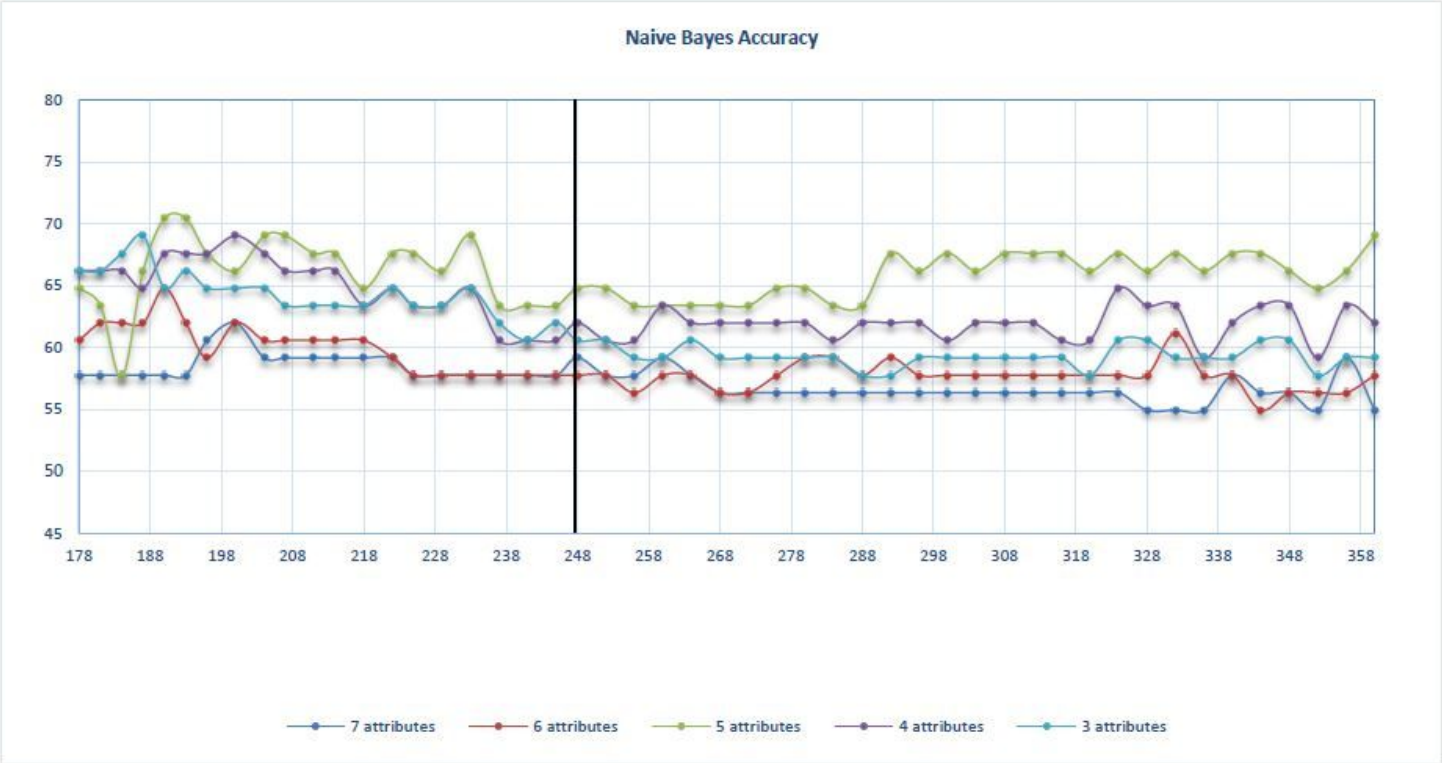

Figure 8

Accuracy attained by the Naïve Bayes ML model

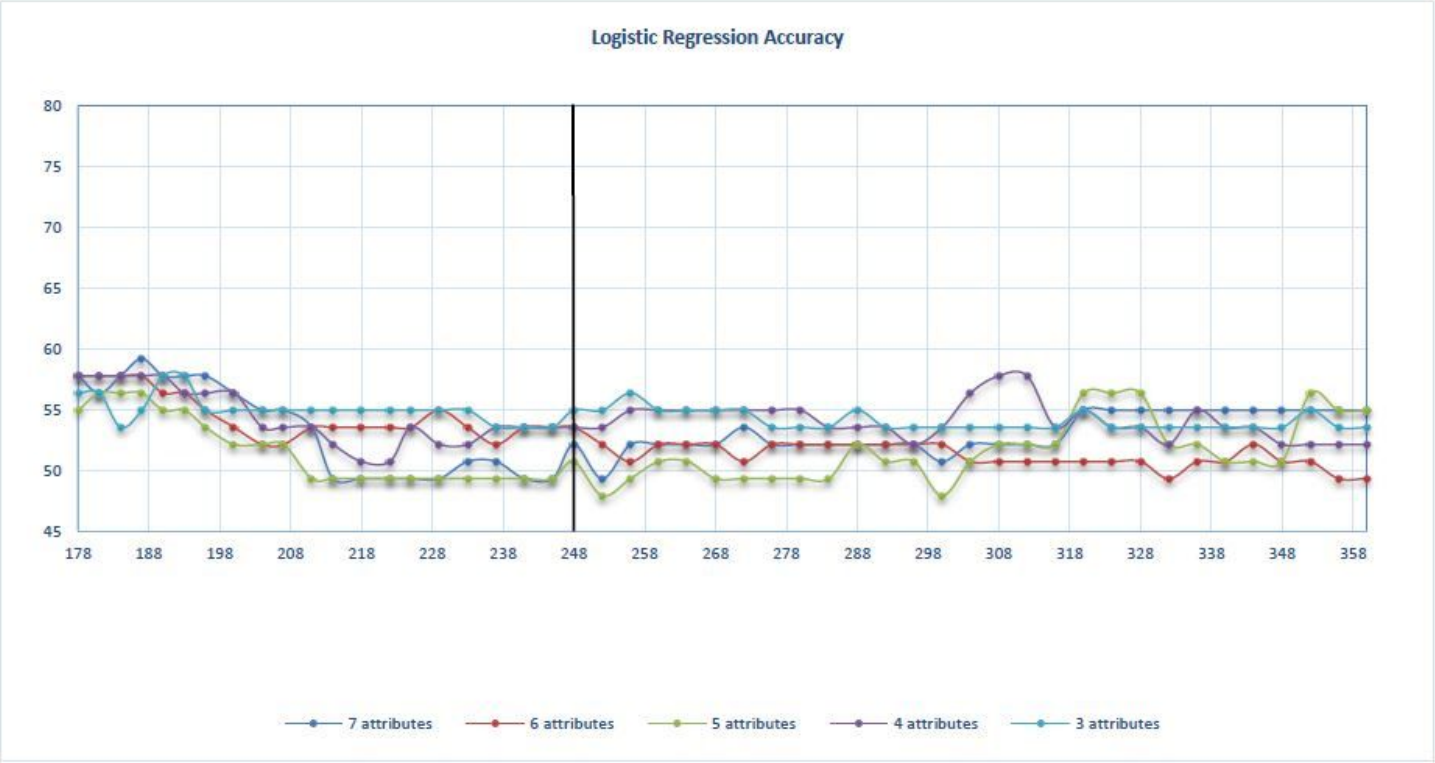

Figure 9

Accuracy attained by the Logistic Regression ML model