SMU Data Science Review

Volume 7 | Number 3 Article 1

2023

Predicting Land Reclamation of Bond Released Surface Mines

Kendall Scott
Southern Methodist University, kendallscott@smu.edu

Austin Webb
austinwebb@mail.smu.edu

Tadd Backus
Southern Methodist University, tbackus@mail.smu.edu

Robert Slater
SMU, rslater@mail.smu.edu

Predicting Land Reclamation of Bond Released Surface Mines in Southern and Central Appalachia

Kendall Scott[1],Tadd Backus[1],Austin Webb[1], Robert Slater[2]
[1]Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
[2]Adjunct Lecturer at Southern Methodist University,
Dallas, TX 75275 USA
[3]Appalachian Voices, 589 West King Street,
Boone, NC 28607 USA

tbackus@smu.edu[1], kendallscott@smu.edu[1], austinwebb@smu.edu[1]

Abstract. Accurately measuring the recovery of released surface mines in the United States poses crucial challenges. This study aims to develop a prediction of land classification, that considers various environmental and coal mine variables. By utilizing this prediction, the researchers and environmentalists (specifically Appalachian Voices, the group heading this research) can better understand the relevant factors for successful reclamation. Efficient management of mine recovery is essential for environmental sustainability, regulatory compliance, and resource utilization. This study focuses on the Appalachian Forest area, which risks becoming a net carbon source (a place that emits more carbon than it absorbs) due to mine recovery. Machine and deep learning methods will be employed using Dynamic World land classification probabilities to identify areas requiring intervention and to provide ongoing insight into released mine conditions. The findings enable decision-making for prioritized reclamation and restoration measures.

1 Introduction

Appalachia, a cultural and geographic region in the central and southern sections of the United States, confronts formidable challenges marked by severe economic hardship and environmental concerns resulting from coal mining and logging activities. This region, rich in natural resources, holds significant potential for addressing global warming challenges. Given its history of facing prejudice and environmental adversity, the health of the area and its population is a vital indicator for analyzing global warming and the intersection of environmental health with historically underprivileged populations.

This research project focuses on released surface mines in the US and the crucial challenges of accurately measuring and predicting recovery time. This involves accounting for various environmental and mine-specific variables to identify factors essential for coal mine recovery. Subsequently, the study aims to develop a recovery score, a predictive tool for accurately estimating the time required for mine recovery. By analyzing factors critical to reclamation and providing estimates for reclamation timelines, this investigation will enhance the understanding of successful reclamation and restoration processes, supporting the efforts of organizations such as Appalachian Voices.

Efficiently managing the recovery of released surface mines is of utmost importance due to its immediate implications on environmental sustainability, US regulatory compliance, and resource utilization. Accurate prediction models developed within this project are imperative for informed decision-making regarding these valuable natural resources. Moreover, the research acknowledges the Appalachian forests' significance as a net carbon sink, potentially at risk of becoming a net carbon source. The recovery of released surface mines threatens these forests, contributing greenhouse gases in the form of coal mine methane, further exacerbating global warming [1]. The research endeavors to create methods to classify and comprehend the factors that facilitate recovery to provide insights to support effective advocacy for remediation.

Addressing climate change requires a focus on carbon methane emissions from coal mines. Appalachian Voices requires an efficient method to gather information on released mines. Utilizing data from Dynamic World, an ever-updating tool that classifies land cover data into one of nine categories based on Sentinel-2 satellite imagery will aid the organization in identifying crucial factors for recovering released mines and assessing areas at risk. Employing a combination of unsupervised machine learning methods, the study aims to classify released coal mines based on factors considering the mined land's state, business practices, and the health of the surrounding area. The research aims to utilize Dynamic World data to provide ongoing insight into the state of released coal mines, thereby identifying environmental risks and effective interventions for recovery. This will be achieved through land cover classification based on Sentinel-2 satellite imagery and new statistical forecasting models created and built by the researchers.

The overarching goal is to contribute to the sustainable management and reclamation of released surface mines in Appalachia. By bridging the gap between environmental concerns, social equity, and economic vitality, the study seeks to offer practical solutions to complex challenges faced by the region. Understanding the factors that facilitate successful mine recovery will not only aid in reducing the environmental impact but also create opportunities for community development and job creation in an area historically burdened by economic hardship. Additionally, this work aligns with global efforts to combat climate change, as effective reclamation of surface mines can mitigate greenhouse gas emissions and preserve the ecological integrity of Appalachian forests, a critical component of the region's unique ecosystem.

This approach takes a multi-faceted perspective to explore the recovery of released surface mines in the Appalachian region. By developing predictive tools and utilizing innovative data analysis methods, the study aims to provide insights into effective reclamation processes, environmental risks, and advocacy for remediation. Through this research, the critical importance of balancing environmental health, social equity, and economic prosperity will be underscored, offering valuable information for decision-makers, stakeholders, and environmental organizations (specifically Appalachian Voices) invested in the sustainable future of Appalachia and beyond.

This research project aims to develop a code driven solution to enhance the understanding of released surface mine recovery in the Appalachian region. The code will integrate various environmental and mine specific variables to create a predictive tool for classifying tree classification. By doing so, this project will address the pressing need for informed decision-making regarding the reclamation of valuable natural resources, which plays a pivotal role in environmental sustainability and resource utilization.

The significance of using a code-based approach for this project lies in its ability to assist organizations like the one partnered with (Appalachian Voices) by efficiently gathering data on released mines. Utilizing Dynamic World, a versatile tool for classifying land cover data based on satellite imagery, the code will assist in identifying key factors for mine recovery and assessing areas at risk. Employing various unsupervised machine learning models and methods, this project seeks to classify released coal mines based on critical factors such as water quality, max temperature, precipitation, and many others. Ultimately, the project will harness data's power to offer ongoing insights into the state of released coal mines, thereby enabling the identification of environmental risks and effective interventions for recovery.

Utilizing a machine learning approach for this research project aligns with the overarching goal of contributing to the sustainable management and reclamation of released surface mines in the Appalachian region. It emphasizes the importance of environmental concerns by providing practical solutions to the complex challenges faced by this region. This project's purpose is to support the understanding of factors that facilitate successful mine recovery, thus reducing environmental impact and creating opportunities for growth and recovery. The research project’s outcomes will be integral to global efforts to combat climate change because effective reclamation of surface mines can mitigate greenhouse gas emissions and preserve the ecological integrity of the Appalachian forests.

2 Literature Review

The literature review focuses on four principal areas: Environmental Impact of Mining, Machine Learning and Interpretation, Mining Practices and Ethics, and Geospatial Analysis.

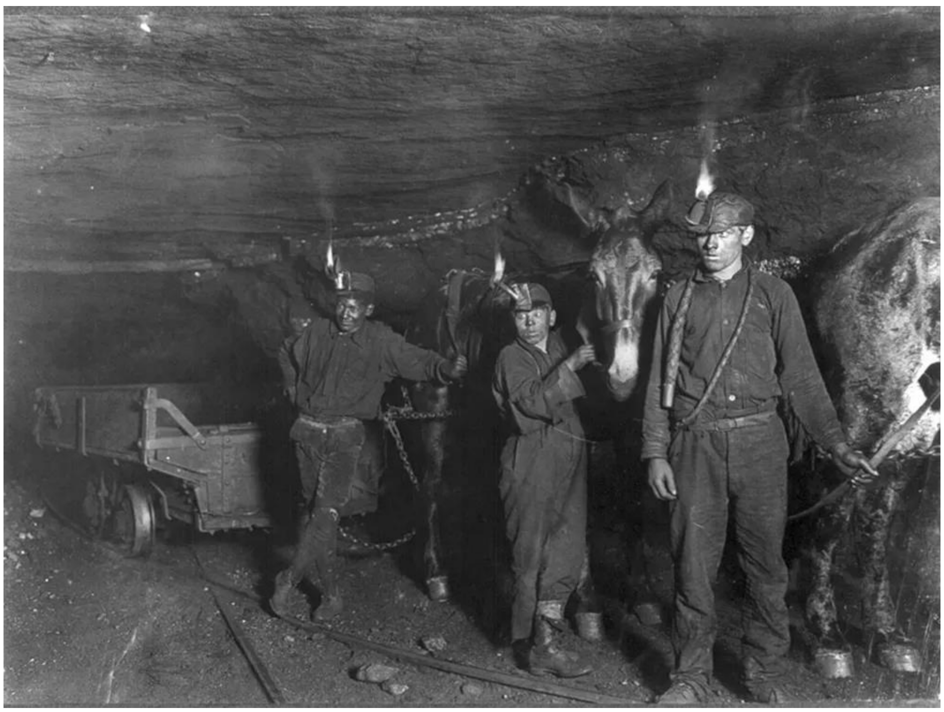

Fig. 1. Child coal miners with mules in Gary, West Virginia in 1908. Working conditions were brutal for coal miners, and unionization was violently suppressed. [2]

2.1 Environmental Impact of Mining

The environmental impact of mining, particularly surface mining in the Appalachian region of the United States, has been a subject of worry, concern, and study. Researchers have attempted to quantify the potential impact of future mining in the area and its impact on the environment. By employing various scenarios based on historical data, valuable information for accurate predictions has been uncovered.

Mountaintop removal coal mining with valley fills (MTMVF) has become the predominant method for coal extraction in Central Appalachia, particularly in Kentucky, Tennessee, and West Virginia [3]. The process of MTMVF generates significant amounts of waste rock due to it involving blasting, excavating bedrock, and extracting coal seams. This waste rock is then deposited in headwater valleys or buried in streams under layers of spoil that can reach up to around 200 meters in depth [4]. The resulting rock spoil is then pushed to adjacent valleys, burying headwater streams. MTMVF can severely impact and impede vegetation, surface topography, and subsurface structure, causing native forests to struggle to re-establish, leading to altered landscapes and shifts from forests to grasslands.

With approximately 25% of the coal production in the United States happening in the Appalachian area, the environmental impact of surface mining is of utmost importance in this area. Massive negative impacts on the environment can be observed, whether aquatic impact (including aquatic animal life), forest impact, surface temperature, or microbial communities. Mining has been identified as the leading force of landscape changes in the Appalachian region, and its diffusion was the greatest cause of net forest loss over the last four decades [5]. Efforts are needed to manage the intensity of human-caused deforestation through surface mining. These efforts will increase communication and discussion as they become more prominent in land-management strategies.

Findings from previous studies identify potential sources of errors to avoid misinterpretation of land change analysis results. In one study, three land cover classifications consisting of five classes (developed areas, barren land, water, low vegetation, and forest) were derived from spatial resolution imagery [6]. This allows for better understanding of how areas can be accurately predicted and classified after MTMVF. Improvements to these findings are paramount as environmentalists and others seek to make surface mine reclamation the main objective when discussing mining.

2.2 Machine Learning and Interpretation

In the realm of surface mine reclamation, the incorporation of machine learning has emerged as a groundbreaking tool for interpreting and analyzing complex environmental data. By harnessing the power of advanced algorithms and computational models, valuable insights can be gleaned into the intricate processes involved with restoring disrupted landscapes. With an increasing focus on sustainable land management practices, machine learning techniques offer a promising avenue to enhance the efficiency and accuracy of surface mine reclamation.

Many different machine learning models can and may be implemented in this effort to understand land reclamation better; however, one of the key drivers to deciding which to use is the interpretability of the model. This can be partly due to their nature of over-parameterization, which involves so many different parameters with hundreds of layers that it is often difficult to understand and interpret [7]. Understanding the model itself is equally as important as understanding its output. It is important to understand what part of the input pattern is responsible for a particular class being predicted. Still, it is also important to focus on understanding which internal features computed by the model are responsible for a particular class [8].

Research shows that integrating gradient boosting with a neural decision tree aims to leverage bagging and boosting techniques [9]. Tree-based models perform with more robustness and interpretability (which is extremely important) due to their hierarchical structure, while deep neural networks excel with high-dimensional data. By combining both of these approaches, researchers hope to improve the performance of classification tasks. A recently proposed algorithm known as the Tree Alternating Optimization (TAO) algorithm can help researchers better learn trees that are both highly accurate and interpretable [10]. This algorithm allows trees to mimic the part of the neural network they replaced. This is a large step in better understanding different machine learning models.

Machine learning has been identified as a valuable tool in evaluating the sustainable development level of coal enterprises. One study shows that establishing a multi-layer forward neural network model based on the error backpropagation algorithm (BP Algorithm) is proposed to evaluate this sustainable development level [11]. The study suggests that the neural network can overcome the limitations of traditional sustainable development evaluation models by considering the coal mine as a complex man-machine-environment-management system. By analyzing these details and numerous other factors, an improvement in the safety production level of the coal mine industry could occur.

Machine learning can play a role in environmental monitoring and management in mining operations. By analyzing data from various sensors and remote sensing technologies, such as satellite imagery, machine learning models can help monitor the impact of mining activities on the surrounding ecosystem. These models can then identify patterns and trends in vegetation health, soil quality, water quality, etc. By utilizing these machine learning techniques and models, mining companies can make more informed decisions to minimize their ecological footprint and ensure compliance with environmental regulations.

2.3 Mining Practices and Ethics

Coal remains the dominant energy source and is expected to continue playing a crucial role for the foreseeable future, emphasizing the pressing need for the coal industry to prioritize sustainable and secure development [12]. Some research highlights the high-risk nature of coal mining due to factors such as gas outbursts, complex working environments, and constant changes in the industry. It is important to emphasize the significance of safety and the need to follow ethical guidelines. Some find it safer, quicker, and easier to step around ethical lines to accomplish their tasks, which is why focusing on appropriate mining practices and ethics is vital.

Fig. 2. The Kayford Mine near Charleston, West Virginia [13]

Many argue that there is a moral relationship between present and future generations in the context of environmental ethics. It poses questions about the obligations the current generation owes to the future and how present actions should be influenced by their impact on future generations. This moral obligation is evaluated using terms familiar in Western ethical thought, namely "rights" and "obligations." However, relying on rights and obligations to determine the moral relationship to the future may result in proposed rights that are not practical for present decision-making. Researchers have proposed adopting a perspective rooted in "virtue ethics" to understand this relationship with future generations [14]. By focusing on present virtues and considering our vision of well-being, people of the current generation can navigate the moral complexities of their actions without attempting to predict the preferences of future generations.

Another study focused on predicting methane concentrations and dispersion in coal mines, directly affecting mining practices and ethics. Methane is a hazardous gas in coal mines and poses significant health concerns for miners. The study can enhance mining practices and safety by developing a model that quantifies the influence of numerous factors on methane dispersion. The main objective of the model is to quantify the factors' relative influences on methane dispersion in coal mines [15]. The researchers align the study with the principles of promoting the well-being and safety of workers in the mining industry. Implementing the findings from these studies demonstrates a commitment to ethical conduct by prioritizing the safety of the miners. The results from multiple models all produced comparable results, leading to the discovery that air velocity is the most significant factor in affecting methane dispersion [16].

Coal's massive and ubiquitous use necessitates focusing on sustainable and safe development in the coal industry. To address the environmental impact that coal mining has and promote sustainability, efforts are being made to explore cleaner and more efficient technologies. Machine learning models can assist in these efforts by monitoring and identifying potential violations, working conditions, and other human rights violations. On top of coal mining's working side, ethical questions must be addressed from a local viewpoint. A more inclusive and comprehensive approach can be achieved and applied by involving local communities, environmental organizations, and other entities heavily involved in the aftermath of surface mining.

2.4 Geospatial Analysis

Geospatial analysis involves examining and interpreting spatial data to gain insights into various phenomena related to the earth's surface. In the context of this paper, the phenomena looked at relate to the reclamation of surface mines. One study uses geospatial analysis to assess groundwater potential in a specific district in India. By layering multiple themes representing groundwater influencing factors, such as geology precipitation, the identification of areas with high and low groundwater potential can be identified. The predictive performance of these models was then compared using various statistical performance metrics to construct a groundwater potential zoning map [17]. A map was henceforth created to detail the findings of high and low potential for groundwater.

Accurate classification of downhole exploration data is crucial for geological modeling and predicting mining outputs. Traditional manual interpretation of this data is time-consuming and subjective. However, by applying machine learning techniques that can then be fed into geospatial analysis, this can be automated and improve accuracy. These techniques can be used to analyze mineral groups and stratigraphy to classify rock types in each deposit [18]. By integrating multiple spatial datasets and applying advanced analytical techniques, geospatial analysis enables the identification of patterns, relationships, and potential areas for intervention.

In environmental sustainability and management, geospatial analysis assists in monitoring and evaluating natural resources, ecosystems, and environmental changes. Researchers can use satellite imagery to track deforestation (and the opposite – forest reclamation), land degradation, and general changes in land coverage over time. By analyzing these patterns, geospatial analysis can benefit in assessing the impact of human activities, specifically coal mining, and support the development of effective mitigation strategies. In addition, geospatial analysis can help provide valuable insights into the pre-mining conditions of the area where surface mining occurs. This information forms the basis for developing effective reclamation plans and strategies.

Satellite imagery can help researchers identify areas with insufficient vegetation, which could be a cause for additional observation or assistance. Vegetation health is a key indicator of reclamation status, and geospatial analysis can help assess the health of the vegetation through different areas of time. Previous research leveraging Google Earth has successfully detected habitat loss from background changes between images [19].

3 Methods

3.1 Data Collection

Data collection for this research project was of utmost importance and was a pivotal component of the entire process. These valuable data sources offer insights into environmental conditions, land use, and changes in landscape, all of which are essential in assessing released surface mine reclamation.

Data was also provided by Appalachian Voices containing information on coal mines throughout multiple states in the Appalachian region. This data is from the Office of Surface Mining Reclamation and Enforcement. This data includes information on mines such as operation status, bond status, owner, different permit dates, the type of mine, and geospatial data. The distribution of surface mines across West Virginia, Kentucky, and Tennessee are shown in the map below (Figure 3).

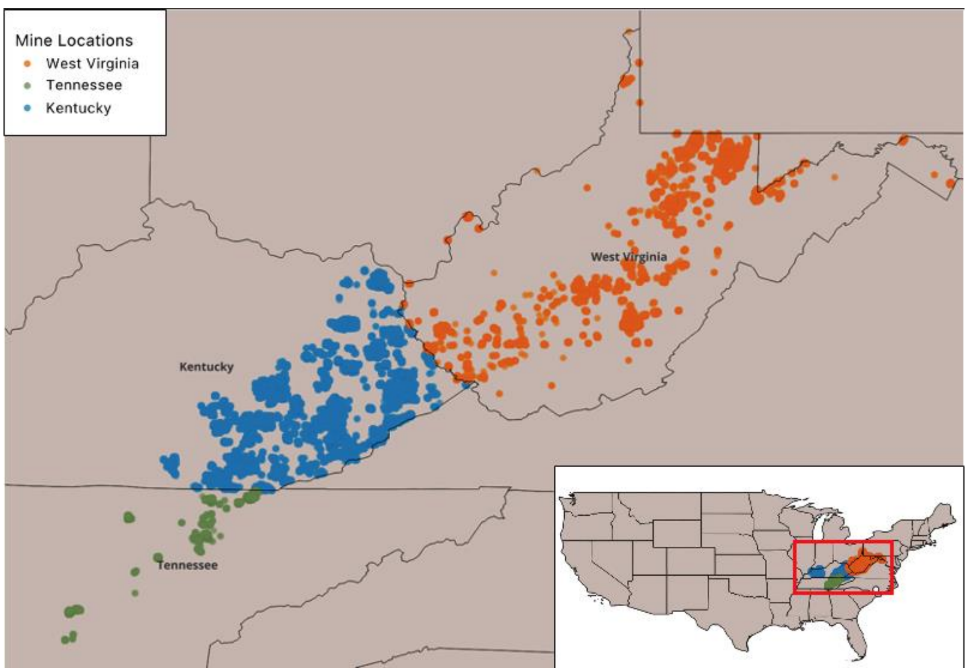

Fig. 3. Plot of mines by location

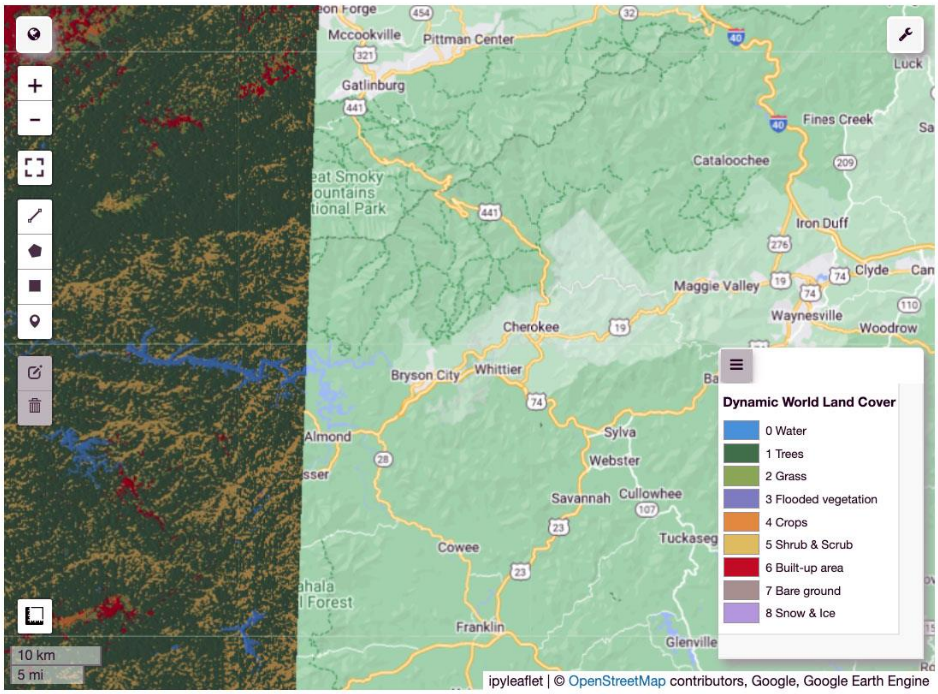

Fig. 4. Classification of land type by Dynamic World dataset in Tennessee

Additional data will be collected from Dynamic World, which is a near real-time land cover dataset, which includes AI (Artificial Intelligence) generated probabilities for different classifications (Figure 4): Water, Trees, Grass, Flooded vegetation, Crops, Shrub & Scrub, Built Area, Bare ground, Snow & Ice. This data set will provide the current and historical land cover for the surface mines in question. Previous research has utilized Google Earth engine to determine yearly surface mining on active mines across Central Appalachia [20]. The difference in our methodology is we well be leveraging the predictions available in Dynamic World, instead of exporting raw images for classification later. This will save processing/computational time dealing with a large number of image files. Water Quality will be leveraged to estimate the crucial factors for mine reclamation. Surface mining has previously had associations with higher toxins in water quality, which has implications for the survival of over 50 federally protected species in Central Appalachia [21].

Dynamic World's continuous data updates are not only convenient, but they are also essential for this project. Timely access to geographical data is paramount in understanding and addressing the intricate environmental challenges in the Appalachian region. Preparing the data will leverage the Google Earth Engine API to join in multiple datasets, so that land cover and water quality can be associated with the mines at that location. The various categories of land cover will be changed into averages of total mine individually.

Mining data from the Google Earth Engine API, while a very powerful and useful approach, presented several complexities. The first being the obvious integration of various datasets from diverse sources, each with its own data structure and format. The process of harmonizing these diverse datasets into one coherent and standardized format can be very intricate and time consuming. Additionally, working with large-scale geospatial data often includes handling substantial amounts of data and information, which can put a strain on computational resources. The Google Earth Engine data was pulled at a resolution of 200 meters (scale), due to the large number of observations that needed to be pulled. The GEO Json shape was converted into a randomized series of latitude and longitude, from within the perimeter of each mine.

Google Earth Engine's API also has some time constraints in terms of the availability of data, so each data point may have irregular results over time. This is true for the various types of Image Libraries/bands (various classifications available on Google Earth Engine). The image library sources for Maximum Temperature (Maximum temperature in kelvin) and Precipitation (Precipitation amount daily in mm) were University of Idaho Gridded Surface Meteorological Dataset [22]. The image library source for Water Quality (Normalized difference chlorophyll index) is the MultiSpectral Instrument, Level-2A Dataset [23]). These features offer insights into the climatic conditions of the region, which could ultimately influence and explain the recovery rate of these mines. Including these features allows for a more comprehensive understanding of the multifaceted factors affecting mine recovery.

By integrating multiple and diverse datasets, the project can establish a holistic view of the factors affecting mine reclamation in this Appalachian region. This multidimensional approach enables the development of predictive models that account for various environmental and climatic variables. The importance of these features lies in their ability to shed light on the intricate and delicate relationship between environmental factors and mine recovery.

3.2 Data Integration

Once all data from the various data sources was properly combined and cleaned, a data preprocessing step was conducted. This step is responsible for handling encoding, scaling, and ensuring the data is in a suitable format for the models to properly work. The first step in the preprocessing process was to separate the numeric features from the dataset. Part of this also included excluding certain variables that would cause data leakage if they remained within the dataset (the other land type probabilities like water, grass, etc.). Scaling these numeric features was then conducted, which transforms the values into a common range between zero and one. This is done to prevent some variables from dominating the learning process due to their larger magnitude. Neural networks perform better with scaled inputs, enabling more stable and efficient convergence during training.

Once this scaling process is completed, these numeric features are once again added to the original dataset. The next part of the preprocessing step is to one-hot encode categorical variables. The final step in this section of the project is to impute the missing values. The mean and the mode were used to impute missing values, and decisions were based on percentage of values in the respective columns (see Table 1).

Table 1. Table that shows the percentage of the mode in the columns with missing data

Feature	**Mode**	**Percentage of the Mode in the Column**
highwall	0	55.92
steep_slope	0	63.36
contour	0	48.27
post_smcra	0	38.35
coalmine_op_status	4	30.19
Number_company_to_permit_id	1	38.21
contact	3	55.97

Data preparation steps as outlined previously are fundamental steps in machine learning, as the quality and format of the input data greatly impacts the performance and reliability of a deep learning model, in this case RNN and Random Forest.

Once the data preparation steps were completed, the data was split into the 80/20 train and test splits. Stratification was done by the mine id variable, so that the models are safeguarded against predicting on mines that have not been seen by the model. This was important to implement, because the model needed to have at least one row of data on a specific mine location, in order to be able to predict further for different years at that location. This method would not be appropriate to use for extrapolation outside of the 10,221 mines currently present in our dataset.

Table 2. Percentage of mines by state

State	**Number of Mines**	**Percentage of Mines**
West Virginia	1,661	16%
Tennessee	135	1%

Kentucky	8,425	82%

The geographic location of the dataset was heavily skewed toward mines in Kentucky, representing 82% of the data (Table. 2).

Table 3. Overview of mining data variables

Variable	Description
FeatCLS	Feature Class, which categorizes geographic features.
Mine Status	The current status of the mine, indicating whether it is active, abandoned, or in some other state.
Permit Type	Describes the type of permit associated with mining activities.
Type Flag	An indicator that provides additional information about the feature or data.
Region des	For Kentucky, a categorical variable for the region the mine was present in.
Quad Description	Quadrangle Description, which refers to the geographic location of a specific quadrangle on a map.
Coalmine op status	Numeric flag indicating the coal mine operating status.
Inspectable unit status	Numeric flag indicating the inspectable unit status.
Permit approval year	Year the permit for a coal mine was approved.
Edit date	Year of last edit for coal mine

4 Results

The results of this model are being leveraged to provide insight on released surface mines in Appalachia, which can be leveraged to find locations which need support and advocacy to recover. Below is a discussion of the different models leveraged, RNN with LSTM layers and Random Forest.

4.1 RNN with LSTM layers

The RNN LSTM "long short-term memory" has the benefit of utilizing weights within the model, which allows it to preserve memory and predict on time series data. Multiple Long Short-Term Memory (LSTM) layers are employed to handle the sequential nature of this data. LSTMs are designed to remember information over an extended period, making them well-suited for this modeling complex, time-related patterns within this dataset [24]. Hyperparameter selection was done utilizing a recurrent drop out of 0.7, which reduces the risk of overfitting. Between each LSTM and Dense Layer, there was an additional .15 drop out added. Early Stopping on the validation data was utilized as well, with a patience of 5, in order to limit the number of epochs for the model and reduce the chances of overfitting. Results of R2, MAE, and MSE were utilized from the test dataset (rows unseen by the model during training), to ensure the model generalizes well.

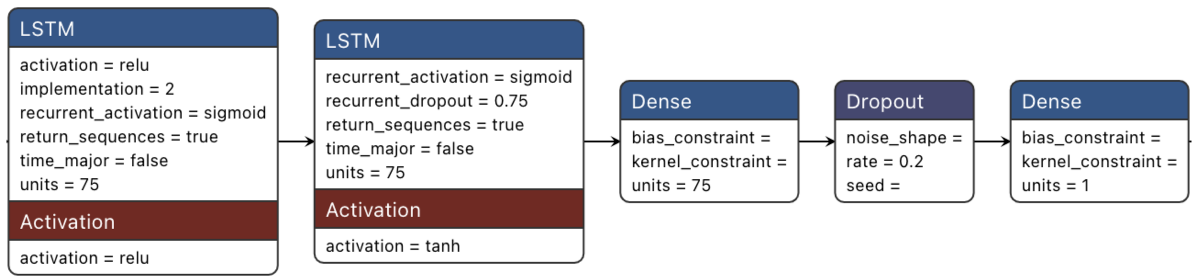

Fig 6. LSTM model architecture.

Utilizing the variables available, we built the RNN model (Fig. 6) utilized early stopping criteria for the validation loss, with a patience of 5, recurrent dropout at 0.75, and a 0.2 dropout, with a sequence length of 4.

Table 5. Results for RNN

Scoring Metric	Scores on Validation Dataset
R2	60.1%
MSE	.007
MAE	.05

4.2 Random Forest

The Random Forest model, with criterion of "squared-error," had the following scores on the hold-out dataset: mean squared error of .007, mean absolute error of .05, and R2 of 64.7%. This model was leveraged to look at feature importance, and the Google Earth Engine variables were in the top results for the model (Water Quality, Max Temperature, and Precipitation). This provides an important insight for future enhancements on this project, as there is a large library of different metrics available from Google Earth Engine libraries and adding more variables into the coal mine dataset via the API, could provide additional insights.

Table 6. Results for Random Forest

Scoring Metric	Scores on Validation Dataset
R2	64.7%
MSE	.007
MAE	.05

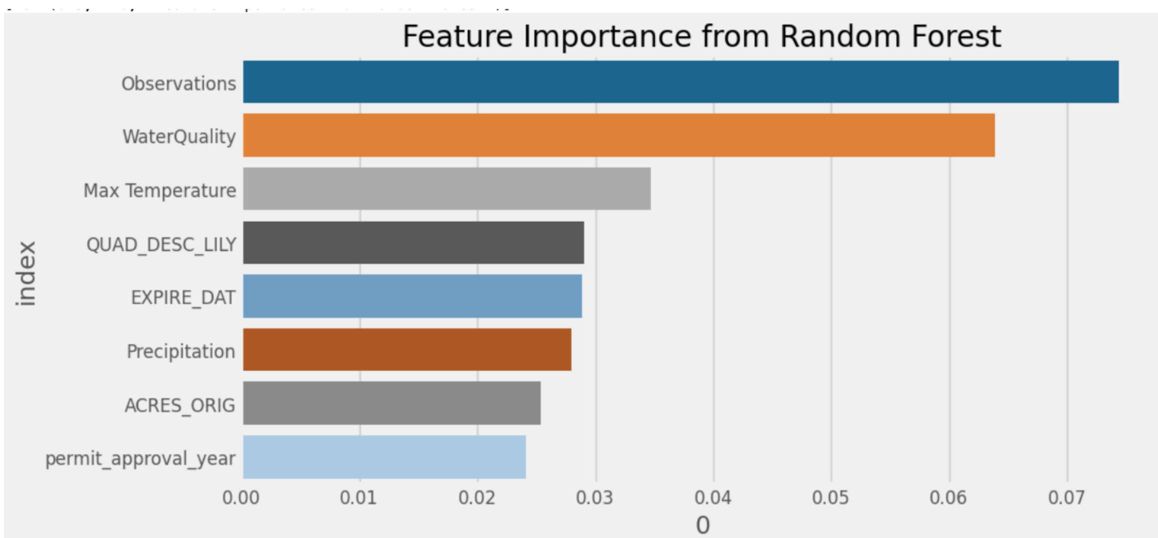

Fig. 7. Feature importance from Random Forest Model

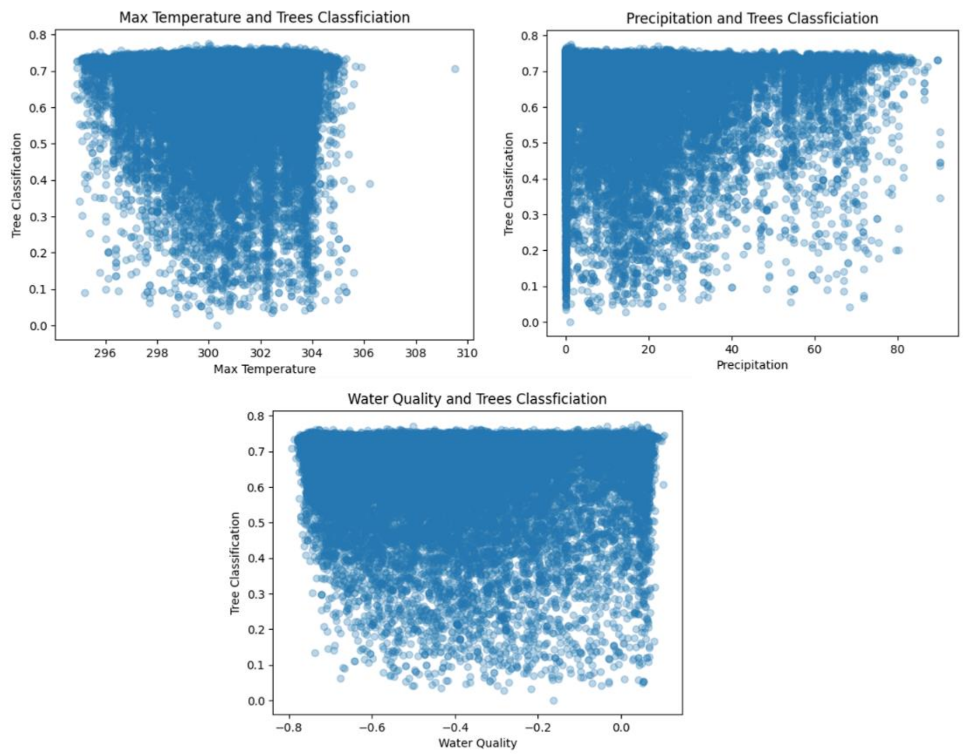

Fig. 8. Tree Classification vs. Max Temperature, Precipitation, and Water Quality.

Looking closer at the features selected by Random Forest, lower max temperatures are associated with more observations of tree classification. The relationship between Precipitation and Tree Classification, where there are more observations of trees in areas with higher precipitation. For Water Quality (Normalized difference chlorophyll index [23]), the lower values have a larger distribution in tree classification.

5 Discussion

5.1 Environmental Factors Impacting Land Recovery

Understanding the intricate relationship between environmental factors and land degradation or recovery mining areas is paramount for sustainable resource management.

One of the key environmental factors that significantly influences recovery in mining areas is maximum temperature. Elevated maximum temperatures can have multiple and substantial effects on mining sites. High temperatures can exacerbate water scarcity, affecting the availability of water resources for restoration and reclamation efforts. This extreme heat can also lead to increased evaporation rates which could potentially reduce the success of vegetation growth and soil restoration.

Another crucial environmental factor influencing land recovery in mining areas is water quality. Water quality plays a pivotal role in reclamation efforts. Poor water quality, which often results from mining activities, can hinder the establishment of healthy ecosystems during the recovery process. Contaminated water can harm aquatic life and vegetation, which in turn slows down the overall recovery of the area. There is a relationship between Water quality and the average tree classification, based on the importance of the feature in Random Forest model.

5.2 Ethical considerations in Surface Mining Reclamation

Surface mine reclamation is a layered process with complex ethical considerations that traverse the domains of the environment, society, and the economy. The ethical dimensions of surface mine reclamation underscore its pivotal role in mitigating the environmental and social impacts of mining activities. Within this context, there are several ethical aspects that warrant in depth analysis:

Environmental stewardship, at the heart of surface mine reclamation, represents a moral and ethical obligation to protect and restore natural ecosystems. It represents a dedicated commitment to mitigating the environmental impacts of mining activities.

One of the focal points of environmental stewardship in reclamation is the concept of rewilding. Rewilding is the principle that involves rehabilitating mined lands to their pre-mining state, or in cases where this is not feasible, to ecosystems with similar characteristics. Rewilding accentuates the ethical responsibility to minimize ecological disruption and support the recovery of native plants and wildlife. This approach not only safeguards the biodiversity of the region but also helps reestablish the intricate ecological relationships that sustain these healthy ecosystems.

Another crucial aspect of environmental stewardship is the sustainable management of mined land. Ethical mine reclamation encompasses strategies that promote the prudent use of natural resources. It involves reducing waste and managing land in a way that ensures long-term ecological health and productivity. By prioritizing sustainability, the reclamation process serves as a model for responsible environmental practices. It aligns with the global mandate to conserve resources and reduce the ecological impacts of resource extraction.

Environmental stewardship in surface mine reclamation is underpinned by a deep sense of responsibility towards the environment and future generations. It embodies the ethical commitment to leave behind a world that is ecologically sound, supporting a balanced coexistence between humans and the natural world. This stewardship philosophy recognizes that the choices made during reclamation reverberate far beyond the present, shaping the legacy left to posterity.

The Surface Mining Control and Reclamation Act (SMCRA) of 1977 stands as a landmark piece of legislation in the United States, that tackles the ethical, social, and environmental issues associated with surface mining activities. The primary and main objective of the SMRCA is to strike a balance between the economic benefits of mining and the necessity to protect the environment affected by the mining operations [25].

The SMCRA's ethical significance stems from its commitment to environmental stewardship and the responsible utilization of natural resources. One of its primary ethical tenets is the duty to protect the environment. The SMCRA embodies the belief that natural landscapes and ecosystems are valuable and deserve safeguarding. The act obligates mine operators to minimize disturbances and restore mined lands, reflecting a moral duty to mitigate the ecological footprint of mining activates.

Violations to this act, and general violations on mining land may impact the recovery process, which can present many ethical and practical challenges. These violations can be harmful to the environment and undermine the reclamation efforts. Upfront, these violations can cause a general disruption in the planned reclamation process, often leading to the need for additional resources and time to rectify the damage done. The ethical implications of these violations are clear – as they represent a breach of commitment to environmental stewardship and regulatory compliance.

The impact of violations on mining land not only causes harm in the restoration process but also causes damage to the life surrounding the mining area. Certain violations can lead to water contamination, which poses a threat to aquatic life and local residents who depend on these water sources. Air quality can be compromised by unauthorized activities on mining sites, which could affect the health of living things near and around.

Violations in general can undermine the credibility and trust of mining companies, regulatory agencies, and reclamation organizations. When these violations occur and go unaddressed, it erodes the public trust and raises concerns about the effectiveness of regulations. Ethical mine reclamation practices entail accountability and transparency when working. This is essential to ensure that it fulfills its ethical duty to protect the environment.

For the communities of people around these coal mines, there is a significant employment gap when coal mines close, which leads to significant increases in unemployment in the region." Coal industry employment fell by around 54 percent between 2005 and 2020" [26]. The economic impact of coal mining closures flags an important consideration for people in the region, which could be a motivating factor to ignore environmental impacts in the region, in favor of economic growth.

5.3 Environemtal Degradation Challenges

The extensive environmental degradation caused by mining activities is a formidable challenge in surface mine reclamation. When mining operations begin, the removal of topsoil is often one of the first steps, resulting in the loss of the most fertile and ecologically rich layer of soil. This depletion disrupts the natural nutrient cycles and can have long lasting effects on soil productivity [27]. Disruption of ecosystems is another critical aspect of environmental degradation. Mining operations frequently entail the removal of forests, wetlands, and other natural habitats. This not only results in the immediate displacement of wildlife but also disrupts the intricate ecological relationships that sustain these ecosystems.

The results obtained through this project hold significance due to their diverse implications for the environment and society. These results provide insights into the future of surface mine reclamation in the Appalachian region, which can be used for further research in other regions around the world. By accurately predicting recovery for released surface mine areas, this research enables a greater level of insight into the ecological disruptions caused by mining activities.

5.4 Analysis from Model Results

Some of the company names showed up in the top 20 results for feature importance from Random Forest. For this project, any specifics on mine locations will be generalized to explain the overall and overarching insights, while more specific information will be shared with App Voices. There were a number of companies that showed up as significant for the Random Forest model, and when we did research we found that the companies listed as significant in the Random Forest model, also had a history (Environmental Protection Agency investigation and closure for permit violations) indicating that the feature importance from the model is helpful to identify which locations may be at risk from not following best practices for environmental safety. While listing details on the company name is outside of the scope of this project; this information will be provided to App Voices.

Mines per permit id was another variable that showed as important in the Random Forest model. Mines per permit id is the number of mines that are associated with any given permit id. While the averages for 1-3 have an average 0.66 probability for being classified as a tree, while the higher (4-6) range had an average of 0.6 for tree classification.

Google Earth Engine has a wide variety of metrics available from different libraries, and because of the feature importance results, adding in more explanatory variables would likely lead to more accurate predictions. The time to code, construct, query, and save the data is computationally expensive, but these initial results indicate that effort has potential to improve and create a more accurate prediction. Specifically, the University of Idaho Gridded Surface Meteorological Dataset [22], has a few variables (Burning index, Wind velocity at 10m, Humidity, etc.), that were not leveraged in this project due to time and computational limitations, that may lead to an improvement in predictions for coal mines.

Further research may involve expanding the model to predict other reclamation related factors and using this predictive framework as a decision support tool for land reclamation efforts in surface mining. As for this project, only the trees variable was predicted.

6 Conclusion

The results of this model indicate that it is possible to get statistically significant predictions on tree classification from Google Earth Engine (utilizing data about coal mines) and which environmental factors influence tree coverage. There are considerations that have not been included in this paper, and which offer an opportunity for future enhancement. There is a dataset which offers information on violations (Office of Surface Mining Reclamation and Enforcement's Applicant/Violator System), which was not available before the completion of this project. This dataset could illuminate differences in coal mine recovery due to a mine's compliance with government guidelines.

While the variables pulled from Google Earth Engine's API were valuable to help predict tree classification, it is a time-consuming process to query 10,221 different mines. Each mine has a perimeter and each query to the API was a randomized sample of a variety of latitudes and longitudes throughout that perimeter. Establishing a public repository with information on coal mines is worth considering in this context. Considerations for this would include creating an established framework (sampling size and timeframes), so that the work of getting details on a coal mine could be distributed, and future research could continue to layer in important environmental factors.

Leveraging feature importance from Random Forest was a valuable tool for garnering insight into specific features, where it was possible to identify specific companies that had historic issues with environmental agencies of various forms. These locations would be important to prioritize for intervention in any reclamation work.

Acknowledgments.

This research owes its success to the unwavering support and mentorship Dr. Jacquelyn Cheun-Jensen, as well as the entire esteemed faculty of the Data Science Graduate Program at Southern Methodist University.

References

1. Elliott Campbell, J., Fox, J. F., & Acton, P. M. (2012). Terrestrial carbon losses from mountaintop coal mining offset regional forest carbon sequestration in the 21st century. Environmental Research Letters, 7(4), 1–6. https://doi.org/10.1088/1748-9326/7/4/045701
2. Library of Congress, Child coal miners with mules in Gary, West Virginia in 1908. Working conditions were brutal for coal miners, and unionization was violently suppressed, digital image, accessed 13 October 2023, https://www.smithsonianmag.com/history/forgotten-matewan-massacre-was-epicenter-20th-century-mine-wars-180963026
3. Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., & Dou, D. (2022). Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. Knowledge and Information Systems, 64(12), 3197–3234. https://doi.org/10.1007/s10115-022-01756-8
4. Hada, S. S., Carreira-Perpiñán, M. Á., & Zharmagambetov, A. (2023). Sparse oblique decision trees: a tool to understand and manipulate neural net features. Data Mining and Knowledge Discovery. https://doi.org/10.1007/s10618-022-00892-7
5. Sachdeva, S., & Kumar, B. (2021). Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in Dholpur (Rajasthan), India. Stochastic Environmental Research and Risk Assessment, 35(2), 287–306. https://doi.org/10.1007/s00477-020-01891-0
6. Dong, M., Yao, L., Wang, X., Benatallah, B., Zhang, S., & Sheng, Q. Z. (2021). Gradient Boosted Neural Decision Forest. IEEE Transactions on Services Computing, 1–1. https://doi.org/10.1109/TSC.2021.3133673
7. International journal of surface mining, reclamation and environment. (1994). A.A. Balkema.
8. Pericak, A. A., Thomas, C. J., Kroodsma, D. A., Wasson, M. F., Ross, M. R. V., Clinton, N. E., Campagna, D. J., Franklin, Y., Bernhardt, E. S., & Amos, J. F. (2018). Mapping the yearly extent of surface coal mining in Central Appalachia using Landsat and Google Earth Engine. PloS One, 13(7), e0197758–e0197758. https://doi.org/10.1371/journal.pone.0197758
9. Lima, A. T., Mitchell, K., O'Connell, D. W., Verhoeven, J., & Van Cappellen, P. (2016). The legacy of surface mining: Remediation, restoration, reclamation and rehabilitation. Environmental Science & Policy, 66, 227–233. https://doi.org/10.1016/j.envsci.2016.07.011

10. Li, X., Chen, W., Cheng, X., & Wang, L. (2016). A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. Remote Sensing (Basel, Switzerland), 8(6), 514–. https://doi.org/10.3390/rs8060514
11. Cribari, V., Strager, M. P., Maxwell, A. E., & Yuill, C. (2021). Landscape Changes in the Southern Coalfields of West Virginia: Multi-Level Intensity Analysis and Surface Mining Transitions in the Headwaters of the Coal River from 1976 to 2016. Land (Basel), 10(7), 748–. https://doi.org/10.3390/land10070748
12. Bai, G., & Xu, T. (2022). Coal Mine Safety Evaluation based on Machine Learning: A BP neural network model. Computational Intelligence and Neuroscience, 2022, 1–9. https://doi.org/10.1155/2022/5233845
13. Dimick, Dennis, The Kayford Mine near Charleston, West Virginia, digital image, accessed 13 October 2023, https://e360.yale.edu/features/a-troubling-look-at-the-human-toll-of-mountaintop-removal-mining
14. Gaba, Jeffrey M., Environmental Ethics and Our Moral Relationship to Future Generations: Future Rights and Present Virtue (April 9, 1999). Columbia Journal of Environmental Law Vol. 24, p. 249, 1999, Available at SSRN: https://ssrn.com/abstract=1586949
15. Karacan, C. O., & Warwick, P. D. (2019). Assessment of coal mine methane (CMM) and abandoned mine methane (AMM) resource potential of longwall mine panels; example from Northern Appalachian Basin, USA. International Journal of Coal Geology, 208, 37–53. https://doi.org/10.1016/j.coal.2019.04.005
16. Mishra, D. P., Panigrahi, D. C., Kumar, P., Kumar, A., & Sinha, P. K. (2021). Assessment of relative impacts of various geo-mining factors on methane dispersion for safety in gassy underground coal mines: an artificial neural networks approach. Neural Computing & Applications, 33(1), 181–190. https://doi.org/10.1007/s00521-020-04974-9
17. Mueller, C., & Mueller, C. (2019). Mining goes digital: proceedings of the 39th International Symposium "Application of Computers and Operations Research in the Mineral Industry" (APCOM 2019), June 4-6, 2019, Wroclaw, Poland (C. Mueller, Ed.). Taylor & Francis. https://doi.org/10.1201/9780429320774
18. Tutak, M., & Brodny, J. (2019). Forecasting Methane Emissions from Hard Coal Mines Including the Methane Drainage Process. Energies (Basel), 12(20), 3840–. https://doi.org/10.3390/en12203840
19. Evans, M. J., & Malcom, J. W. (2021). Supporting habitat conservation with automated change detection in Google Earth Engine. Conservation Biology, 35(4), 1151–1161. https://doi.org/10.1111/cobi.13680
20. Pericak, A. A., Thomas, C. J., Kroodsma, D. A., Wasson, M. F., Ross, M. R. V., Clinton, N. E., Campagna, D. J., Franklin, Y., Bernhardt, E. S., & Amos, J. F. (2018). Mapping the yearly extent of surface coal mining in Central Appalachia using Landsat and Google Earth Engine. PloS One, 13(7), e0197758–e0197758. https://doi.org/10.1371/journal.pone.0197758

21. Evans, M. J., Kay, K., Proctor, C., Thomas, C. J., & Malcom, J. W. (2021). Linking mountaintop removal mining to water quality for imperiled species using satellite data. PloS One, 16(11), e0239691–e0239691. https://doi.org/10.1371/journal.pone.0239691
22. Abatzoglou J. T., Development of gridded surface meteorological data for ecological applications and modelling, International Journal of Climatology. (2012) doi:10.1002/joc.3413
23. Muduli, P.R., Kumar, A., Kanuri, V.V. et al. Water quality assessment of the Ganges River during COVID-19 lockdown. Int. J. Environ. Sci. Technol. 18, 1645–1652 (2021). https://doi.org/10.1007/s13762-021-03245-x
24. Tf.keras.layers.LSTM : tensorflow V2.14.0. TensorFlow. (n.d.). https://www.tensorflow.org/api_docs/python/tf/keras/layers/LSTM
25. Dunlap, L C, & Lyon, J S. Effectiveness of the Surface Mining Control and Reclamation Act: reclamation or regulatory subversion. United States.
26. Ross, M. R., McGlynn, B. L., & Bernhardt, E. S. (2016). Deep impact: Effects of mountaintop mining on surface topography, bedrock structure, and downstream waters. Environmental Science & Technology, 50(4), 2064–2074. https://doi.org/10.1021/acs.est.5b04532
27. Macdonald, S. E., Landhäusser, S. M., Skousen, J., Franklin, J., Frouz, J., Hall, S., Jacobs, D. F., & Quideau, S. (2015). Forest restoration following surface mining disturbance: Challenges and solutions. New Forests, 46(5–6), 703–732. https://doi.org/10.1007/s11056-015-9506-4