A Gas Concentration Prediction Method Driven by a Spark Streaming Framework

Yuxin Huang, Jingdao Fan, Zhenguo Yan *, Shugang Li and Yanping Wang

Abstract: In the traditional coal-mine gas-concentration prediction process, problems such as low timeliness of data and low efficiency of the prediction model in learning data features result in low accuracy of the final prediction. To solve these problems, a gas-concentration prediction method driven by the Spark Streaming framework is proposed. In this research study, the Spark Streaming framework, autoregressive integrated moving average (ARIMA) model and support vector machine (SVM) model are used to construct a new prediction model called the SPARS model. The Spark Streaming framework is used to process large batches of real-time streaming data in a short period of time, and the model can be used to intermittently update and optimize the prediction model so that the model can fully learn the characteristics of the data. At the same time, the advantages of the ARIMA model and SVM model for processing linear data and nonlinear data are combined to improve the model’s prediction efficiency and fully reflect the timeliness of gas prediction. Finally, the proposed prediction model is verified using gas data collected on site. The optimal learning time for the SPARS model in predicting this set of data is determined, and a comparative analysis of the prediction results obtained from the ARIMA, SVM and other models fully confirms that high-precision prediction results can be obtained using the SPARS model. The proposed model can be used to realize scientific and accurate real-time prediction and analyses of coal-mine gas concentrations and provides a new idea for realizing real-time and accurate gas prediction in coal mines.

Keywords: spark streaming; ARIMA; SVM; SPARS model; real-time

1. Introduction

Coal has been China’s main and basic energy source for a long period of time. Experts predict that China’s coal production capacity will reach 4 billion tons in 2030 and 3.4 billion tons in 2050. Therefore, the dominant position of coal in China’s energy consumption structure will not change for a long period of time into the future [1,2]. In China, with the continuous depletion of shallow coal resources, the current coal mining situation has required continually extending mining operations deeper underground, and the risk of gas disasters has also increased significantly, which seriously affects the safety of coal mining. Gas accidents have long comprised the most fatal accidents in coal mines. According to statistics, in the eight years from 2014 to 2021, there were 1947 accidents in China and a total of 3472 fatalities. Among them, there were 189 gas accidents, accounting for 10% of all accidents, but the death toll reached 989, accounting for 28%, which is close to the total death toll [3]. Therefore, accurate and efficient gas prediction and early warning are of great significance for the safe mining of coal.

In recent years, with the rapid development of information technology and automatic control, safety monitoring and monitoring in the process of coal mining has gradually developed in the direction of intelligence. In traditional monitoring, the monitoring system has been unable to meet the needs of coal mine development, and more achievements have been made in mine gas prediction and early warning research [4,5].
constructed a gas outburst early warning system based on the analysis of the abnormal characteristics of gas emissions, and the variance, peak difference and fluctuation of the slope of the gas emission data for the excavation face were determined by analysing gas emission amounts. The early warning level was determined based on various indicators, which ensured the safe operation at the excavation face of the coal mine. Huang et al. [7] established a multifactor coupling relationship analysis model for the outliers discarded from coal mining face gas data during the daily gas-prediction process and established an early warning level for gas anomaly analysis by analysing the association rules for multidimensional gas outliers to effectively improve the effectiveness of coal-mine gas early warning systems. Liang et al. [8] constructed a bidirectional gated recurrent unit neural network gas concentration prediction model based on the adaptive estimated maximum (Adamax) optimization algorithm for predicting coal-mine gas concentrations. The results show that the optimization algorithm has higher accuracy and an improved prediction effect than other prediction algorithms. Zhang et al. [9] constructed a coal mine ventilation system (CVS) safety prediction and early warning system. This system has high prediction accuracy, can accurately reflect the rationality of the underground mining process and has good utilization value. Xu et al. [10] constructed a gas concentration prediction algorithm based on the superposition model to determine the optimal position for the model by optimizing the parameters. Finally, the prediction results of the algorithm were verified by experimental simulations for more accuracy and the prediction accuracy improved. Jia et al. [11] proposed a prediction model for coal-mine gas concentrations based on gated regression units (GRUs). This model can make full use of the time series characteristics of gas data to predict gas concentrations with high accuracy, which improves the validity and accuracy for gas prediction.

In fact, in the process of gas prediction, the training set’s data need to show high timeliness. If the data’s timeliness is insufficient, the accuracy of the prediction results will be reduced, thus affecting safe coal mining [12–14]. The above research still shows a certain lag in the forecast of data in the process of research on gas prediction and early warning. There is still room for improvement in the operational efficiency of related algorithms. Ensuring both the timeliness of data extraction and the timeliness of algorithm learning rules in the prediction process is a research direction that needs urgent attention. In this paper, a gas prediction and early warning model is proposed based on the Spark Streaming framework. This model is capable of processing large batches of real-time streaming data in a short period of time. The autoregressive integrated moving average (ARIMA) model and support vector machine (SVM) model are combined to process linear and nonlinear data, respectively. While improving the efficiency of the forecast model, this model fully reflects the timeliness of the forecast data. This model lays a certain foundation for the real-time monitoring and prediction of gas in the intelligent construction of coal mines.

2. Materials and Methods
2.1. Spark Streaming Framework

Spark Streaming was developed on the basis of the Spark platform. A distributed stream-computing data-processing framework based on the discrete stream (DStream) model can process massive data in batches in a short period of time and has the advantages of high fault tolerance, scalability, high traffic and low latency [15,16]. Spark Streaming can split the real-time streaming data according to a certain time interval, pass it to the Spark Engine and finally obtain batches of results. The essence of Spark Streaming is to divide the collected data into DStreams and convert each DStream into resilient distributed datasets (RDDs), which are stored in memory, and finally stored in external devices. DStream represents a continuous data stream and is part of the Spark Streaming framework. It consists of a continuous sequence of RDD sets, and each RDD contains a data stream at a certain time interval. The Spark Streaming flowchart is shown in Figure 1, and the Spark Streaming architecture diagram is shown in Figure 2.
Vapnik in 1995, SVM was originally used to solve linearly separable problems and later extended to regression problems \[22,23\]. Assuming the training set \(\{x_i, y_i\}_{i=1}^{l} \in (x \times y)^l\), where \(x_i \in x = \mathbb{R}^n\) is the input and \(y_i \in y = \mathbb{R}\) is the output, the SVM model can be expressed as follows:

\[
\min_{w,b,\xi} \left[ \frac{1}{2} ||w||_2^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^+) \right]
\]
The time series monitoring system has the characteristics of streaming data, and a continuous change in the data saved in the monitoring system can be used as the training dataset, the real-time data saved in the monitoring system can be used as the training dataset, the real-time transmission data can also be used. However, the real-time transmission dataset from the monitoring system has the characteristics of streaming data, and a continuous change in the

Both linear and nonlinear trends are observed in gas-concentration time-series data [24,25]. Considering that the ARIMA model has unique advantages when dealing with linear data, it can fully capture the linear part in the time series, and SVM has outstanding performance when analysing and predicting nonlinear data [26,27]. Therefore, the ARIMA model is used to process the historical data for the one-dimensional gas time series and obtain the corresponding linear prediction results and residual series. Then, SVM is used to further analyse and predict the nonlinear factors in the residual series on the panel data affecting the gas time series. Finally, the analysis and prediction results for the two models are combined to obtain the final prediction result for the target gas time-series data. The principle of the gas concentration prediction framework is shown in Figure 3. The time series $Y = \{y_k, k = 1, 2, \cdots, N\}$ consists of two parts: a linear part and a nonlinear part, i.e., $y_k = \tilde{y}_k + nl_k$. First, the one-dimensional gas data are processed by the ARIMA model, and the time series $\tilde{y}_k$ and residual series $nl_k$ of the linear prediction result are obtained. Second, by further processing the residual time series, a set of time series $nl_k$ of nonlinear prediction results is obtained. The final combination of linear and nonlinear results is the final time series forecast value $\hat{y}_k = \tilde{y}_k + nl_k$.

\[
\begin{align*}
\begin{cases}
\text{s.t. } & -\{w, \phi(x_j)\} - b \leq \epsilon + \xi_i \\
& \{w, \phi(x_j)\} - b \leq -\epsilon - \xi_i^* \\
& \xi_i \geq 0, \xi_i^* \geq 0, i = 1, \cdots, l
\end{cases}
\end{align*}
\]

Among these expressions, the dual problem is described as follows.

\[
\max_{\alpha, \alpha^*} \left\{ \frac{1}{2} \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \left( a_i - a_i^* \right) K(x_i, x_j) + \epsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \epsilon \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \right\}
\]

\[
\begin{align*}
\text{s.t. } & \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 \\
& \alpha_i, \alpha_i^* \in (0, C), i = 1, \cdots, l
\end{align*}
\]

In the above formula, $\xi_i$ is the slack variable $\xi_i \geq 0$, $\xi_i^* \geq 0$, $i = 1, \cdots, l$ and $C$ is the penalty parameter.

Then, the solution is given by the following.

\[
f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x_i, x) - b
\]
data stream leads to the continuous updating of the prediction model. Streaming data are a set of large, fast and consecutively arriving sequences of data [28,29]. Therefore, the use of conventional model prediction will lead to long modelling times, which will indirectly lead to poor timeliness for the forecast data and a low utilization value.

2.3. SPARS Model

In this section, a parallel prediction method combining Spark Streaming with the ARIMA-SVM combination algorithm is proposed and is named the SPARS algorithm. It can be used to quickly analyse and predict streaming data. At the same time, it combines the ARIMA model and SVM model to process linear and nonlinear data, respectively, which improves the timeliness of traditional model prediction and improves accuracies.

To perform predictive modelling for the gas concentration data stream, a distributed stream processing framework called Spark Streaming is used to build a real-time gas-concentration prediction system based on the ARIMA-SVM model. The real-time data generated by the gas monitoring source are sent to Spark Streaming through the stream generator. In the sliding window calculation provided by Spark Streaming, the time window is used to divide the original DStream into data RDDs with specified time slices. DStream, which is a part of Spark Streaming, can perform stream data processing and batch processing at the same time, meeting the requirements of various processing types such as dataset extraction, machine learning model training and model application. RDD based on the window length is the basic unit of the prediction model, and the window length can be determined according to the rate of streaming data and modelling complexity. The structure of the real-time gas-concentration prediction system based on Spark Streaming is shown in Figure 4. The Hadoop distributed file system (HDFS) is used for data storage. Alternatively, the data can be sent directly to Spark Streaming [30–32]. Coal-mine gas and related sensors transmit real-time monitoring data to the system through the network, and data streams can be written to distributed storage and combined with Spark MLlib to build predictive models, which can be dynamically updated with data streams. Finally, the real-time prediction of gas concentration is carried out by using the constructed ARIMA-SVR prediction algorithm.

![Spark Streaming sliding window operation diagram](image)

**Figure 4.** Gas framework structure diagram of Spark Streaming.
3. Experiment

3.1. Data Sources

Gas data were collected from the 802 working face of a mine in Shaanxi for experimental analysis. A KG9001C sensor was used to collect the gas concentration, and the sensor concentration measurement range was (0–100)% CH\(_4\). The sensor measurement error was (0–1)% CH\(_4\) \(\leq\) 0.1% CH\(_4\), (1–2)% CH\(_4\) \(\leq\) 0.2% CH\(_4\), (2–4)% CH\(_4\) \(\leq\) 0.3% CH\(_4\), (4–10)% CH\(_4\) \(\leq\) 1% CH\(_4\), and (10–100)% CH\(_4\) \(\pm\) 10% CH\(_4\). The sampling rate for collecting gas data was 0.2/s. A total of 2880 sets of gas data were collected for 4 h of gas measurement sequences. Some of the original data are provided in Table 1. The original gas sequence is shown in Figure 5. In the collected data, the maximum value is 0.82%; the minimum value is 0.02%; the mean value is 0.17%; the standard deviation is 0.14.

Table 1. Part of the original data record.

| Time             | Gas Concentration/% | Time             | Gas Concentration/% |
|------------------|---------------------|------------------|---------------------|
| 8 July 2021 15:00:00 | 0.13                | 8 July 2021 15:00:40 | 0.14               |
| 8 July 2021 15:00:05 | 0.13                | 8 July 2021 15:00:45 | 0.12               |
| 8 July 2021 15:00:10 | 0.13                | 8 July 2021 15:00:50 | 0.13               |
| 8 July 2021 15:00:15 | 0.14                | 8 July 2021 15:00:55 | 0.14               |
| 8 July 2021 15:00:20 | 0.15                | 8 July 2021 15:01:00 | 0.15               |
| 8 July 2021 15:00:25 | 0.14                | 8 July 2021 15:01:05 | 0.14               |
| 8 July 2021 15:00:30 | 0.14                | 8 July 2021 15:01:10 | 0.14               |
| 8 July 2021 15:00:35 | 0.13                | 8 July 2021 15:01:15 | 0.14               |

Figure 5. Original sequence diagram of gas data collection.

3.2. Prediction of the Gas Concentration by the SPARS Model

The dataset collected for the first 3 h is used as the training dataset of the model for the estimation of the model parameters. The dataset for the final 1 h is used as the test set for the model to test the fitness of the prediction model. To simulate the dynamic gas data flow monitored by the coal-mine gas sensor, the training set’s gas concentration sequence data are also sent to Spark Streaming at a rate of 0.2/s through a transmission control protocol (TCP) socket to simulate the real-time dynamic data flow for the gas concentration. Spark Streaming utilizes the built-in data stream source socketTextStream to receive gas concentration stream data as the input source. When predicting the gas concentration, the length and sliding distance of the Spark Streaming stream window can be specified according to the requirements of the actual application, and the training data’s length and update cycle for the stream regression model can be determined. To verify the real-time performance and accuracy for the flow regression, the gas concentration data are predicted and analysed by setting different model update cycles. In the experiment, the length of the flow regression window is 1 min, and the model update period is set to 10 s, 20 s, 30 s, 40 s, 50 s and 60 s for the real-time prediction of the gas concentration. The experiment was carried out six times to simulate the prediction accuracy for gas concentration data under
different model update cycles. The prediction results are shown in Figures 6–11. In these figures, real represents the real value of the gas concentration, and predict represents the predicted value. It can be concluded that the degree to which the mean value predicted by the model is close to the mean value of the true value at different update times is 60 s, 50 s, 40 s, 30 s, 20 s, and 10 s from high to low. Figure 12 shows a comparison of the fit between the actual value and the predicted value. It can be observed from the figure that when the update time is 50 s or 60 s, the fitting degrees of the predicted value and the real value are higher. In general, the longer the model update time, the larger the training set data flow, the more fully the model is trained, and the higher the prediction accuracy. In contrast, the shorter the model update time, the smaller the training set data flow and the more insufficient the model training process, resulting in lower prediction accuracy.

![Figure 6](image1.png)  
**Figure 6.** Gas concentration prediction of the model with an update period of 10 s.

![Figure 7](image2.png)  
**Figure 7.** Gas concentration prediction of the model with an update period of 20 s.

![Figure 8](image3.png)  
**Figure 8.** Gas concentration prediction of the model with an update period of 30 s.
Figure 6. Gas concentration prediction of the model with an update period of 10 s.

Figure 7. Gas concentration prediction of the model with an update period of 20 s.

Figure 8. Gas concentration prediction of the model with an update period of 30 s.

Figure 9. Gas concentration prediction of the model with an update period of 40 s.

Figure 10. Gas concentration prediction of the model with an update period of 50 s.

Figure 11. Gas concentration prediction of the model with an update period of 60 s.
Figure 10. Gas concentration prediction of the model with an update period of 50 s.

Figure 11. Gas concentration prediction of the model with an update period of 60 s.

Figure 12. Comparison of gas concentration prediction results for different model update periods.

4. Discussion

In the above discussion, it can be observed that, with the expansion of the model update period, the model’s ability to predict gas is gradually enhanced, the fit between the predicted value and the actual value is enhanced, and the predicted value becomes more accurate. However, if the update time of the model becomes infinitely long, it cannot meet the requirements of real-time processing of data streams. Based on the principle of real-time data-stream processing, the prediction accuracy should be relatively high when the model’s update time is relatively short [33,34]. Therefore, the above results need to be further discussed. To further determine the corresponding relationship between the model update period and the prediction results, the root mean square error (RMSE) is used to evaluate the model [35,36]. The precision RMSE of the statistical model is the square root of the squared error and the mean, and its formula is expressed as follows.

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{n}} \]  

(9)

In the above prediction process for the test set, when the model update time is 60 s, the model is updated 60 times in total. By analogy, when the model update time is 50 s, 40 s, 30 s, 20 s, and 10 s, the model is updated 72 times, 90 times, 120 times, 180 times, and 360 times, respectively. The individual calculation results of RMSE for each update model are shown in Table 2, and the changes in the RMSE for the model under different update cycles are shown in Figure 13. The lower abscissa in this figure represents the number of model updates. The ordinate is divided into six groups, all of which represent the RMSE value, and the upper abscissa represents the update time for the six groups of models. The lower abscissa and ordinate constitute the change trend for the model RMSE value under different update periods, and the upper abscissa and ordinate constitute the trend diagram of the overall change in the RMSE value for each model during the 1 h prediction process. It can be observed from the figure that when the model’s update time is 50 s, the minimum value is 0.01211; when the update time is 10 s, the maximum value is 0.02617. It can be concluded that in the overall prediction process, when the model update time is 50 s, the prediction accuracy is higher.
Table 2. The overall value of the RMSE of the model under different update cycles.

| Model Update Time | Number of Updates | Overall Value RMSE |
|-------------------|-------------------|--------------------|
| 10 s              | 360               | 0.02617            |
| 20 s              | 180               | 0.02412            |
| 30 s              | 120               | 0.02147            |
| 40 s              | 90                | 0.0156             |
| 50 s              | 72                | 0.01211            |
| 60 s              | 60                | 0.01212            |

Figure 13. Changes in the RMSE of the model under different update cycles.

To further explore the superiority of SPARS model prediction, the above data are predicted by the ARIMA model, SVM model and ARIMA-SVM-combined model. The first 3 h of the gas data are set as the training set, and the final 1 h of the gas data are set as the test set. The prediction results are compared with the real value data series and the prediction results from the SPARS model, and the results are shown in Figure 14. It can be observed from this figure that the prediction effect of the ARIMA model and the SVM model is relatively poor because the ARIMA model can show a good degree of fit for the linear data part, and the prediction result is poor for the nonlinear part of the data. The SVM model shows a good fit for the nonlinear part of the data, and the prediction results for the linear part are relatively poor. The combined ARIMA-SVM model shows relatively good prediction results. In this model, the advantages of linear data and nonlinear data prediction from the above two models are integrated; thus, a good prediction trend is achieved. However, the prediction trend for the SPARS model is the closest to the real value sequence because the model is updated every 50 s according to the time series' characteristics. Compared with the non-real-time prediction model, the SPARS model can be used to more accurately capture the time series characteristics of each time period to
achieve accurate prediction. The prediction accuracy of the above four prediction models is quantitatively analysed to obtain the maximum error, minimum error and average error of the four prediction models, as shown in Table 3. The error value for the SPARS model is the smallest; thus, it has the highest prediction accuracy.

![Figure 14. Comparison of prediction results of different models.](image)

**Table 3.** Comparison of the errors in the prediction results using different models.

| Model        | Maximum Error/% | Minimum Error/% | Average Error/% |
|--------------|-----------------|-----------------|-----------------|
| SPARS        | 0.0189          | 0.0070          | 0.0124          |
| ARIMA        | 0.0270          | 0.0078          | 0.0162          |
| SVM          | 0.0326          | 0.0127          | 0.0217          |
| ARIMA-SVM    | 0.0208          | 0.0094          | 0.0145          |

5. Conclusions

In this paper, a new method for predicting coal-mine gas concentrations is studied. The problems of gas concentration data hysteresis and low efficiency of the prediction model when learning data features in the gas prediction process are mainly solved using traditional algorithms. In this research, Spark Streaming and the ARIMA-SVM algorithm were applied and combined into a parallel prediction method called the SPARS model. This model can be used to quickly process daily monitoring gas concentration flow data, and the advantages of the ARIMA model and the SVM model in predicting linear and nonlinear data, respectively, are combined in the model. While improving the timeliness of traditional model prediction, the prediction accuracy is also improved when using this model. The prediction performance of the SPARS model is validated by using examples. First, the model update cycle is analysed, and the RMSE is used to evaluate the prediction performance of the model under different update cycles. It is found that the prediction accuracy of the SPARS model in the process of predicting gas concentrations does not improve with the continuous expansion of the model’s update cycle. It is concluded that when the model update period is 50 s, the best prediction effect of the model with the smallest RMSE value is obtained. Then, the accuracy of the SPARS model prediction is verified. Comparing the prediction results of the ARIMA model, SVM model and ARIMA-SVM-combined model, the smallest prediction error and the highest prediction accuracy are achieved using the SPARS model. The proposed model can be used to realize high-precision real-time predictions of coal-mine gas concentrations. This model can be used to help provide a safe and reliable working environment for underground operators while ensuring the safe, stable and efficient mining of coal mines. It also provides a new idea for coal mines with respect to achieving real-time and accurate gas prediction and lays a
foundation for intelligent coal-mine gas prediction and early warning. In the next step of this research, the real-time prediction and early warning analysis of gas concentrations under the influence of various factors will be considered to further improve the timeliness and accuracy of coal-mine gas-concentration prediction.

**Author Contributions:** Conceptualization, Y.H. and J.F. Methodology, Y.H., J.F., Z.Y. and S.L. Validation, Y.H., J.F., Z.Y., S.L. and Y.W. Theoretical analysis, Y.H. and Y.W. Data curation, Y.H. Writing—original draft preparation, J.F., S.L. and Z.Y. Writing—review and editing, Y.H. and Y.W. Project administration, J.F. Funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of Shannxi, grant number 2019JLZ-08.

**Acknowledgments:** We thank the National Natural Science Foundation of Shannxi for its support of this study. We thank the academic editors and anonymous reviewers for their kind suggestions and valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Abbreviations**

| Abbreviation | Description                                      |
|--------------|--------------------------------------------------|
| ARIMA        | Autoregressive integrated moving average         |
| SVM          | Support vector machine                           |
| SPARS        | Spark Streaming-autoregressive integrated moving average-support vector machine |
| CVS          | Coal mine ventilation system                     |
| RDD          | Resilient distributed dataset                    |
| HDFS         | Hadoop distributed file system                    |
| TCP          | Transmission control protocol                     |
| RMSE         | Root mean square error                            |

**References**

1. Cong, Y.; Zhao, X.; Tang, K.; Wang, G.; Hu, Y.; Jiao, Y. FA-LSTM: A Novel Toxic Gas Concentration Prediction Model in Pollutant Environment. *IEEE Access* 2022, 10, 1591–1602. [CrossRef]

2. Zhang, Y.; Guo, H.; Lu, Z.; Zhan, L.; Hung, P.C. Distributed gas concentration prediction with intelligent edge devices in coal mine. *Eng. Appl. Artif. Intell.* 2020, 92, 103643. [CrossRef]

3. Wang, X.Q.; Xu, N.K.; Meng, X.R.; Chang, H.Q. Prediction of Gas Concentration Based on LSTM-LightGBM Variable Weight Combination Model. *Energies* 2022, 15, 827. [CrossRef]

4. Li, Y.; Yang, K.; Qin, R.; Yu, Y. Technical system and prospect of safe and efficient mining of coal and gas outburst coal seams. *Coal Sci. Technol.* 2020, 48, 167–173.

5. Wang, G.; Yang, S.; Zhang, S.; Liu, Y. Status and prospect of coal mine gas drainage and utilization technology in Xinjiang Coal Mining Area. *Coal Sci. Technol.* 2020, 48, 154–161.

6. Zhang, J.; Ai, Z.; Guo, L.; Cui, X. Research of Synergy Warning System for Gas Outburst Based on Entropy-Weight Bayesian. *Int. J. Comput. Intell. Syst.* 2021, 14, 376–385. [CrossRef]

7. Huang, Y.; Fan, J.; Yan, Z.; Li, S.; Wang, Y. Research on Early Warning for Gas Risks at a Working Face Based on Association Rule Mining. *Energies* 2021, 14, 6889. [CrossRef]

8. Liang, R.; Chang, X.; Jia, P.; Xu, C. Mine Gas Concentration Forecasting Model Based on an Optimized BiGRU Network. *ACS Omega* 2020, 5, 28579–28586. [CrossRef]

9. Zhang, X.Q.; Cheng, W.M.; Zhang, Q.; Yang, X.X.; Du, W.Z. Partition airflow varying features of chaos-theory-based coalmine ventilation system and related safety forecasting and forewarning system. *Int. J. Min. Sci. Technol.* 2017, 27, 269–275.

10. Xu, Y.; Meng, R.; Zhao, X. Research on a Gas Concentration Prediction Algorithm Based on Stacking. *Sensors* 2021, 21, 1597. [CrossRef]

11. Jia, P.; Liu, H.; Wang, S.; Wang, P. Research on a Mine Gas Concentration Forecasting Model Based on a GRU Network. *IEEE Access* 2020, 8, 38023–38031. [CrossRef]

12. Wang, K.; Zhang, J.; Cai, B.; Yu, S. Emission factors of fugitive methane from underground coal mines in China: Estimation and uncertainty. *Appl. Energy* 2019, 250, 273–282. [CrossRef]

13. Zhao, B.; Cao, J.; Sun, H.; Wen, G.; Dai, L.; Wang, B. Experimental investigations of stress-gas pressure evolution rules of coal and gas outburst: A case study in Dingji coal mine, China. *Energy Sci. Eng.* 2019, 8, 61–73. [CrossRef]

14. Lu, Z.; Zhu, X.; Wang, H.; Li, Q. Mathematical modeling for intelligent prediction of gas accident number in Chinese coal mines in recent years. *J. Intell. Fuzzy Syst.* 2018, 35, 2649–2655. [CrossRef]
15. Xiao, W.; Hu, J. SWEclat: A frequent itemset mining algorithm over streaming data using Spark Streaming. J. Supercomput. 2020, 76, 7619–7634. [CrossRef]

16. Lee, Y.; Song, S. Distributed Indexing Methods for Moving Objects based on Spark Stream. Int. J. Contents 2015, 11, 69–72. [CrossRef]

17. Wang, Y.; Guo, Y. Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. China Commun. 2020, 17, 205–221. [CrossRef]

18. Wang, J.; Sun, L.; Zhao, H.; Wang, Y. ARIMA-BP integrated intelligent algorithm for China’s consumer price index forecasting and its applications. J. Intell. Fuzzy Syst. 2016, 31, 2187–2193. [CrossRef]

19. Svetunkov, I.; Boylan, J.E. State-space ARIMA for supply-chain forecasting. Int. J. Prod. Res. 2020, 58, 818–827. [CrossRef]

20. Dawoud, I.; Kaçiranlar, S. An optimal k of kth MA-ARIMA models under a class of ARIMA model. Commun. Stat. Theory Methods 2017, 46, 5754–5765. [CrossRef]

21. Xu, D.; Zhang, Q.; Ding, Y.; Zhang, D. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ. Sci. Pollut. Res. 2022, 29, 4128–4144. [CrossRef] [PubMed]

22. Bhandari, S.; Zhao, H.P.; Kim, H.; Khan, P.; Ullah, S. Packet Scheduling Using SVM Models in Wireless Communication Networks. J. Internet Technol. 2019, 20, 1505–1512.

23. Jung, C.; Shen, Y.; Jiao, L. Learning to Rank with Ensemble Ranking SVM. Neural Process. Lett. 2015, 42, 703–714. [CrossRef]

24. Zhang, T.; Song, S.; Li, S.; Ma, L.; Pan, S.; Han, L. Research on Gas Concentration Prediction Models Based on LSTM Multidimensional Time Series. Energies 2019, 12, 161. [CrossRef]

25. Liang, Y.Q.; Guo, D.Y.; Huang, Z.F.; Jiang, X.H. Prediction model for coal-gas outburst using the genetic prediction pursuit method. Int. J. Oil Gas Coal Technol. 2017, 16, 271–282. [CrossRef]

26. Lim, S.; Yun, H. Forecasting Tanker Indices with ARIMA-SVM Hybrid Models. Korean J. Financ. Eng. 2018, 17, 79–98.

27. Ordoñez, C.; Lasheras, F.S.; Roca-Pardiñas, J.; Juez, F.J.D.C. A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines. J. Comput. Appl. Math. 2019, 346, 184–191. [CrossRef]

28. Chen, L.; Wang, E.; Feng, J.; Kong, X.; Li, X.; Zhang, Z. A dynamic gas emission prediction model at the heading face and its engineering application. J. Nat. Gas Sci. Eng. 2016, 30, 228–236. [CrossRef]

29. Zhao, X.; Sun, H.; Cao, J.; Ning, X.; Liu, Y. Applications of online integrated system for coal and gas outburst prediction: A case study of Xinjing Mine in Shanxi, China. Energy Sci. Eng. 2020, 8, 1980–1996. [CrossRef]

30. Tutak, M.; Brodny, J. Predicting Methane Concentration in Longwall Regions Using Artificial Neural Networks. Int. J. Environ. Res. Public Health 2019, 16, 1406. [CrossRef]

31. Eckhoff, R.K. Testing of dust clouds for the electrostatic-spark ignition hazard in industry. Need for a modified approach? J. Loss Prev. Process Ind. 2021, 70, 104405. [CrossRef]

32. Prats, D.B.; Portella, F.A.; Costa, C.H.A.; Berral, J.L. You Only Run Once: Spark Auto-Tuning From a Single Run. IEEE Trans. Netw. Serv. Manag. 2020, 17, 2039–2051. [CrossRef]

33. Zheng, T.; Chen, G.; Wang, X.; Chen, C.; Wang, X.; Luo, S. Real-time intelligent big data processing: Technology, platform, and applications. Sci. China Inf. Sci. 2019, 62, 82101. [CrossRef]

34. Guo, Z.; Zhang, Y.; Lv, J.; Liu, Y.; Liu, Y. An Online Learning Collaborative Method for Traffic Forecasting and Routing Optimization. IEEE Trans. Intell. Transp. Syst. 2021, 22, 6634–6645. [CrossRef]

35. Ouyang, Q.; Lv, Y.B.; Ma, J.H.; Li, J. An LSTM-Based Method Considering History and Real-Time Data for Passenger Flow Prediction. Appl. Sci. 2020, 10, 3788. [CrossRef]

36. Liew, J.; Göçmen, T.; Lio, W.H.; Larsen, G.C. Streaming dynamic mode decomposition for short-term forecasting in wind farms. Wind Energy 2022, 25, 719–734. [CrossRef]