Spatial Data Mining for Earthquake Significance Classification: Exploring Geospatial Insights

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology

In

Computer Science and Engineering

School of Engineering and Sciences

Submitted by

Gollapalli Ganga Srinivas **(AP21110010262)**
Gadde Madhukar Sai Babu **(AP21110010277)**
Gadde Maruti Mahesh **(AP21110010289)**
Uppuluri Bogesh **(AP21110010309)**

Under the Guidance of

Dr. Rajiv Senapati

SRM University–AP

Neerukonda, Mangalagiri, Guntur

Andhra Pradesh – 522 240

Nov, 2023

Certificate

Date: 28-Nov-23

This is to certify that the work present in this Project entitled "**Spatial Data Mining for Earthquake Significance Classification: Exploring Geospatial Insights**" has been carried out by Gollapalli Ganga Srinivas (AP21110010262), Gadde Madhukar Sai Babu (AP21110010277), Gadde Maruti Mahesh (AP21110010289), Uppuluri Bogesh (AP21110010309) under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in **School of Engineering and Sciences**.

Supervisor

Rajiv Senapati.

(Signature)

Dr. Rajiv Senapati

Designation,

Affiliation.

Acknowledgements

We would like to express our gratitude to everyone who has played a role in the completion of this project. Their dedication, work, patience and guidance have been instrumental in achieving this milestone. A special mention goes to our mentor, who has been by our side since the beginning offering advice and unwavering support. We are also deeply thankful to our friends and all those who have contributed to this project. Without their contribution we would not have reached our end goal. The progress of the project has been remarkable thanks to the efforts of everyone involved. We recognize that a project of this scale could not have succeeded without the assistance of individuals. Therefore we extend our gratitude to all those who have supported us directly or indirectly. Thank you all for your support, throughout this journey. In conclusion we sincerely appreciate every individual who has contributed to making this project a reality. Without their assistance it would not have been possible for us to bring it to completion.We are incredibly grateful for the support we have received. We extend our heartfelt appreciation to everyone who has played a vital role in contributing to our cause.

Table of Contents

Abstract

Spatial Data Mining is an area within the field of data mining that focuses on extracting patterns and information from geographical datasets. These datasets include attributes like latitude, longitude, population density and location. In this research we specifically explore Spatial Data Mining in relation to classifying the significance of earthquakes. We highlight the role that spatial data plays in decision making across domains.We conducted an evaluation of ten classification algorithms. Found that the Random Forest algorithm stood out for its accuracy, robust F1 score and well balanced precision and recall rates. As a result it emerged as the choice for classifying earthquakes as either "significant" or "not significant." The decision matrix is a tool that evaluates the performance of this algorithm, by considering metrics like accuracy, precision, recall and the F1 score. It provides insights into how the algorithm is performing.The outcomes of our research have implications for disaster response and earthquake forecasting. They provide insights that can inform decision making processes. Our study also emphasizes the importance of selecting algorithms to address classification challenges effectively. Through data mining techniques real world issues can be effectively tackled. Furthermore our paper offers a perspective on data mining by discussing its significance, applications and challenges. Data mining, a part of the data science field, plays a role in uncovering hidden patterns, relationships and anomalies within the data. Spatial data mining further refines this by focusing on information that is linked to locations. These two elements work together to provide the basis for data-driven decision-making. Data mining functions at the nexus of machine learning, artificial intelligence, and statistics.Encompasses techniques that cover different types of data. It serves as a core principle in data science bringing together fields like computer science, statistics and domain specific expertise in a way. During the phase of data analysis analysts employ methods such as visualizing the data and using techniques to gain an initial understanding.We highlight how spatial data mining goes beyond being a tool; it serves as a key to uncovering hidden insights, within maps, locations and geographic information. Its influence extends across domains, such, as planning, healthcare management and conservation efforts. Even though there are challenges in working with datasets and ensuring privacy, spatial data mining continues to be essential, in the age of data decision making. This study enhances our comprehension of the significance of data mining by highlighting its capacity to facilitate decision making processes.

Abbreviations

SVM	Support Vector Machine
KNN	K-Nearest Neighbors
ADA BOOST	Adaptive Boosting
NB	Naive Bayes
DT	Decision Tree
LR	Logistic Regression
RF	Random Forest

List of Tables

List of Figures

List of Equations

Introduction

In this era of information it has become crucial to extract knowledge from complex and extensive datasets. Data mining, a part of the data science field, plays a role in uncovering hidden patterns, relationships and anomalies within the data. Spatial data mining further refines this by focusing on information that is linked to locations. These two elements work together to provide the basis for data-driven decision-making. Data mining functions at the nexus of machine learning, artificial intelligence, and statistics.Encompasses techniques that cover different types of data. It serves as a core principle in data science bringing together fields like computer science, statistics and domain specific expertise in a way. During the phase of data analysis analysts employ methods such as visualizing the data and using techniques to gain an initial understanding. This stage acts as a lens through which potential patterns and anomalies are identified for investigation. The process then progresses into feature selection and engineering where relevant attributes are chosen to enhance analyses. Spatial data mining zooms in on locations within datasets, with the goal of discovering valuable insights.Spatial data covers formats like maps, GPS coordinates and satellite imagery. It is utilized in fields such, as logistics, environmental science, urban planning, transportation and public health. Data mining is a tool with applications in domains, like healthcare, finance, retail, manufacturing, education, environmental science and more.It influences patient diagnosis in healthcare, guides stock market analysis in finance, optimizes marketing strategies in retail, refines manufacturing processes, improves education by predicting student success, assists environmental scientists in understanding climate change, optimizes transportation systems, aids in scientific research, enhances network security, influences pharmaceutical sales, informs insurance decisions, empowers research across disciplines, and plays a pivotal role in IoT applications.In contrast, spatial data mining leaves an indelible mark on areas such as urban planning, natural resource management, epidemiology, agriculture, emergency response, crime analysis, environmental monitoring, geological exploration, public health planning, climate change modeling, transportation, natural resource exploration, agricultural yield prediction, tourism and location-based services, wildlife conservation, infrastructure planning, transportation safety, historical preservation, satellite imagery analysis, and market analysis. These applications capitalize on spatial data's unique attributes to make more informed decisions, protect the environment, and ensure public safety.In contrast, spatial data mining is attuned to data rich in spatial or geographic components, delving deep into the spatial context to unveil patterns influenced by proximity, distance, and spatial relationships. While both fields employ similar algorithms, their application diverges based on the type of data under analysis.In conclusion, data mining and spatial data mining stand as the pillars of insight extraction from the ever-expanding sea of data. Their rich tapestry of applications stretches across a plethora of domains, guiding decisions, driving innovation, and unlocking the potential of data-driven solutions. While data mining illuminates patterns and relationships spanning a multitude of data types,spatial data mining hones in on geographic intricacies, propelling more informed decision-making in a spatial context. Together, they represent the twin engines of progress, steering us toward a smarter, more interconnected future.

- This research underscores the importance of algorithm selection in solving classification problems, highlighting the impact of data-driven insights on safety and resource optimization.
- Random Forest exhibited the highest accuracy (0.898) and a well-balanced F1 score (0.75) for classifying earthquake significance, making it an efficient choice among the evaluated algorithms.
- The study underscores the broader applications of spatial data mining in real-world scenarios, ranging from disaster management to resource allocation, showcasing its versatility and impact in diverse fields.

Methodology

2.1 Dataset Description:

The dataset provides information about earthquake events which consists of various attributes. These attributes include the timestamp **(`time`**) of each earthquake, its geographical coordinates in terms of latitude and longitude **(`latitude` and `longitude`)**, and the depth **(`depth`)** at which the earthquake originated. The dataset also records the earthquake's magnitude **(`mag`)** and the type of magnitude measurement method employed **(`magType`**). Information about the seismic monitoring network, the number of reporting seismometer stations **(`nst`)**, the gap between these reporting stations **(`gap`)**, and the closest distance to a reporting station **(`dmin`)** is also available. Additionally, the dataset contains details about the signal quality represented as the root mean square **(`rms`)**. Each earthquake event is uniquely identified by an `id`. The dataset provides location descriptions **(`place`)** and categorizes earthquake events by type **(`type`)**. Geospatial accuracy is accounted for with attributes such as horizontal error **(`horizontalError`)**, depth error **(`depthError`)**, and magnitude error **(`magError`)**. The dataset also captures the number of stations contributing to magnitude information **(`magNst`)**. The reporting status of each earthquake event **(`status`)** and the sources of location **(`locationSource`)** and magnitude information **(`magSource`)** are included. This dataset enables in-depth analysis and modeling of earthquake events with a wide range of attributes and information sources.

2.2 Problem Statement:

To classify earthquakes into two categories: **"significant"** and **"not significant."** The significance is determined based on the magnitude **(`mag`)** of the earthquake. Earthquakes with a magnitude greater than or equal to 5.0 are considered **"significant,"** while those below 5.0 are labeled as **"not significant."**

2.3 Algorithms Used:

2.3.1 Logistic Regression:

Logistic regression is a technique that assists in assessing the probability of an event falling into a category. In this scenario we employ regression analysis to estimate the likelihood of an earthquake being significant, by taking factors into account. The mathematical equation, for regression can be expressed in the manner:

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 \cdot x_1+\beta_2 \cdot x_2+\ldots+\beta_n \cdot x_n)}} \tag{1}$$

Let:

p represents the probability of an earthquake being significant.

X represents the vector of input features (latitude, longitude, depth, etc.).

β represents the vector of coefficients for each feature.

β0 represents the intercept term.

The logistic regression model computes the probability as:

The natural logarithm's base is e, or roughly 2.71828.

Algorithm: Logistic Regression
Accuracy: 0.8637814827953175
Classification Report:

	precision	recall	f1-score	support
False	0.88	0.96	0.92	2180
True	0.78	0.55	0.65	639
accuracy			0.86	2819
macro avg	0.83	0.75	0.78	2819
weighted avg	0.86	0.86	0.85	2819

Figure 1

2.3.2 Decision Tree:

Decision trees are tree structures that we use to make decisions and classify tasks. In this problem we can employ a decision tree to create rules based on features and classify earthquakes as either significant or not.

Algorithm: Decision Tree
Accuracy: 0.8456899609790706
Classification Report:

	precision	recall	f1-score	support
False	0.90	0.90	0.90	2180
True	0.66	0.66	0.66	639
accuracy			0.85	2819
macro avg	0.78	0.78	0.78	2819
weighted avg	0.85	0.85	0.85	2819

Figure 2

2.3.3 Random Forest:

The algorithm Random Forest combines decision trees to increase classification accuracy and consistency. A Random Forest is made up of several decision trees that are often made using bagging techniques.

Algorithm: Random Forest
Accuracy: 0.8978361120964882
Classification Report:

	precision	recall	f1-score	support
False	0.91	0.96	0.94	2180
True	0.84	0.68	0.75	639
accuracy			0.90	2819
macro avg	0.88	0.82	0.84	2819
weighted avg	0.89	0.90	0.89	2819

Figure 3

2.3.4 Gradient Boosting:

Gradient Boosting is a method that constructs a series of decision trees one after another, with the goal of enhancing the classification of instances that were previously classified incorrectly.

$$\mathbf{F(x)=F\ previous\ (x)+\alpha \times Tree\ k\ (x)} \tag{2}$$

where F previous is the previous model, α is the learning rate, and Tree k is the new tree.

```
Algorithm: Gradient Boosting
Accuracy: 0.8914508691025186
Classification Report:
              precision    recall  f1-score   support

       False       0.90      0.96      0.93      2180
        True       0.84      0.65      0.73       639

    accuracy                           0.89      2819
   macro avg       0.87      0.81      0.83      2819
weighted avg       0.89      0.89      0.89      2819
```

Figure 4

2.3.5 SVM (Support Vector Machine):

Support Vector Machines (SVMs) are a kind of machine learning algorithm that seeks to identify the hyperplane that can effectively classify instances into classes.

Hyperplane Equation: In a binary classification problem, the equation of the hyperplane can be represented as:

$$\mathbf{p * x + q = 0} \tag{3}$$

Here, "p" is the weight vector that determines the orientation of the hyperplane. "x" is the feature vector of an instance. "q" is the bias term.

```
Algorithm: SVM
Accuracy: 0.871940404398723
Classification Report:
              precision    recall  f1-score   support

       False       0.87      0.97      0.92      2180
        True       0.86      0.52      0.65       639

    accuracy                           0.87      2819
   macro avg       0.87      0.75      0.79      2819
weighted avg       0.87      0.87      0.86      2819
```

Figure 5

2.3.6 K-Nearest Neighbors :

The K Nearest Neighbors (KNN) algorithm is a classification method that assigns a label to an instance by considering the class of its k neighbors.

Mathematical Formula:

Let x be the instance to be classified.

Let x1, x2, ..., x_k be k nearest neighbors of x.

Let y1, y2, ..., y_k are the labels of the k nearest neighbors.

The class with the highest sum of indicator functions id given by

$$\mathbf{y = argmax\ (sum\_i(1[y\_i = c]))} \tag{4}$$

```
Algorithm: K-Nearest Neighbors
Accuracy: 0.8793898545583541
Classification Report:
              precision    recall  f1-score   support

       False       0.90      0.95      0.92      2180
        True       0.79      0.63      0.70       639

    accuracy                           0.88      2819
   macro avg       0.85      0.79      0.81      2819
weighted avg       0.87      0.88      0.87      2819
```

Figure 6

2.3.7 Gaussian Naive Bayes:

Naive Bayes is a technique that employs probability and Bayes theorem to determine the likelihood of an earthquake falling into a category by analyzing the distribution of its features.

Mathematical Formula:

$$\mathbf{P(S \mid T) = (P(T \mid S) * P(S)) / P(T)} \tag{5}$$

Where:

P(S | T) -The probability that the earthquake is considered "significant" based on the given features X.

P(T | S) -The chances of observing the features T when the earthquake is deemed "significant".

P(S) is the prior probability of an earthquake being "significant."

```
Algorithm: Gaussian Naive Bayes
Accuracy: 0.8389499822632139
Classification Report:
              precision    recall  f1-score   support

       False       0.89      0.90      0.90      2180
        True       0.65      0.63      0.64       639

    accuracy                           0.84      2819
   macro avg       0.77      0.76      0.77      2819
weighted avg       0.84      0.84      0.84      2819
```

Figure 7

2.3.8 Neural Network:

Neural networks, a type of network used for classification consist of interconnected nodes called neurons. Mathematically the process, at each node can be described as follows:

For a hidden layer node h_j :

$$\mathbf{h_j = Activation(W_{j1} * x\_1 + W_{j2} * x\_2 + ... + W_{jn} * x\_n + b\_j)}$$

For the output layer node o_k :

$$\mathbf{o_k = Activation(W_{k1} * h\_1 + W_{k2} * h\_2 + ... + W_{km} * h\_m + b\_k)}$$

In this context let's consider _ji as the significance of the link connecting the i th node in the layer, with the j th node in the layer. Meanwhile x_i refers to input feature i. Well as b_k represent biases, for the hidden layers j th node and output layers k th node, respectively.

```
Algorithm: Neural Network
Accuracy: 0.8726498758424973
Classification Report:
              precision    recall  f1-score   support

       False       0.87      0.98      0.92      2180
        True       0.87      0.51      0.65       639

    accuracy                           0.87      2819
   macro avg       0.87      0.75      0.78      2819
weighted avg       0.87      0.87      0.86      2819
```

Figure 8

2.3.9 AdaBoost (Adaptive Boosting):

Adaptive Boosting, or AdaBoost, is a machine learning technique that builds a powerful classifier by combining several classifiers. AdaBoost is an ensemble learning algorithm that iteratively combines weak learners, giving misclassified cases increasing weights. A weighted sum of weak learners makes up the final prediction, which improves model performance by concentrating on hard-to-classify data points.

```
Algorithm: AdaBoost
Accuracy: 0.8797445902802412
Classification Report:
              precision    recall  f1-score   support

       False       0.90      0.95      0.92      2180
        True       0.80      0.63      0.70       639

    accuracy                           0.88      2819
   macro avg       0.85      0.79      0.81      2819
weighted avg       0.87      0.88      0.87      2819
```

Figure 9

2.3.10 Bagging:

Bagging, which stands for Bootstrap Aggregating, is a technique used in learning. It involves combining the predictions from models that have been trained on subsets of the data. The main goal of bagging is to reduce variability and enhance the stability and accuracy of predictions by averaging or voting on the outputs of these models.

```
Algorithm: Bagging
Accuracy: 0.8914508691025186
Classification Report:
              precision    recall  f1-score   support

       False       0.91      0.96      0.93      2180
        True       0.82      0.67      0.74       639

    accuracy                           0.89      2819
   macro avg       0.86      0.81      0.83      2819
weighted avg       0.89      0.89      0.89      2819
```

Figure 10

Discussion

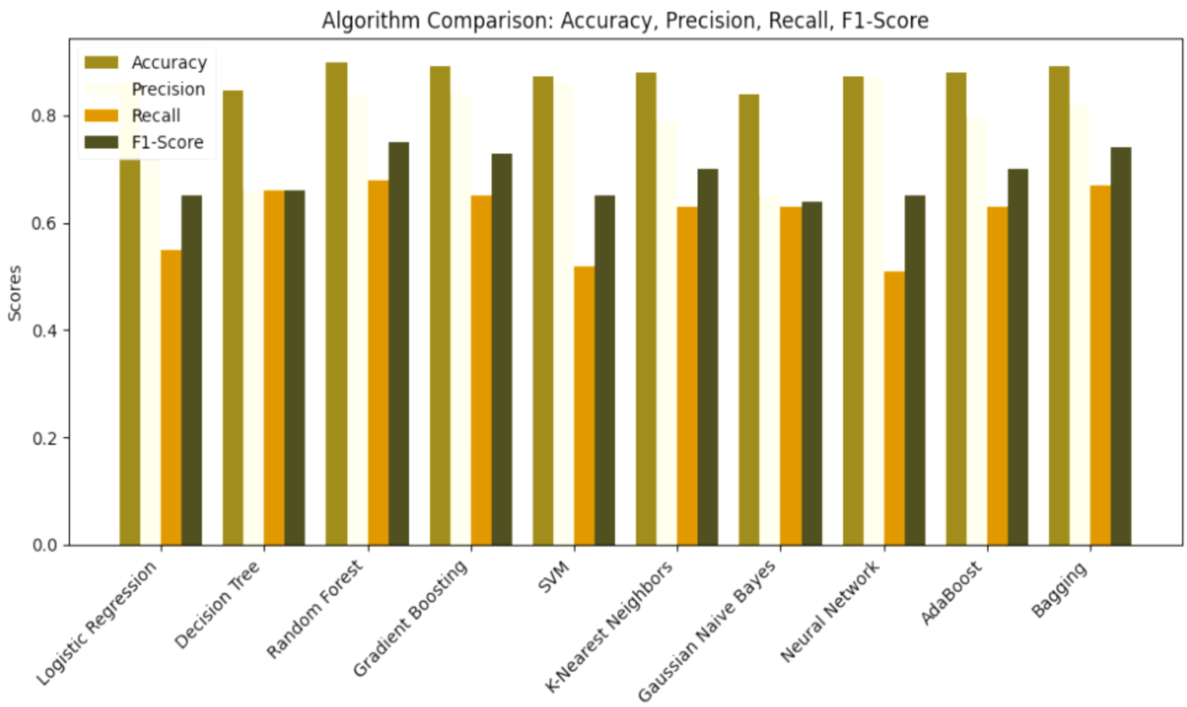

Figure 11

Observation:

Among the models that were evaluated it is worth noting that Random Forest performed well with an accuracy of 0.898 indicating strong overall performance. It achieved a F1 Score of 0.75 which demonstrates precision and recall. Gradient Boosting closely follows with an accuracy of 0.891 showcasing precision but a lower recall of 0.65 resulting in an F1 Score of 0.73. K Nearest Neighbors also demonstrated performance, with an accuracy of 0.879 and a balanced F1 Score of 0.70.It is interesting to observe that the Neural Network showed a accuracy of 0.864 but it struggled with recall at only 0.51 suggesting some difficulty in identifying positive instances accurately.On the hand the Decision Tree, SVM (Support Vector Machine) and Gaussian Naive Bayes models displayed performance levels with accuracies ranging from 0.839 to 0.846 and F1 Scores between 0.64 to 0.66.These metrics offer information, about the advantages and limitations of each model aiding in determining their suitability for tasks within a given context.

Algorithm	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.864	0.78	0.55	0.65
Decision Tree	0.846	0.66	0.66	0.66
Random Forest	0.898	0.84	0.68	0.75
Gradient Boosting	0.891	0.84	0.65	0.73
SVM	0.872	0.86	0.52	0.65
K-Nearest Neighbors	0.879	0.79	0.63	0.70
Gaussian naive bayes	0.839	0.65	0.63	0.64
Neural Network	0.873	0.87	0.51	0.65
ADA Boost	0.880	0.80	0.63	0.70
Bagging	0.891	0.82	0.67	0.74

Table 1

Concluding Remarks

The study presented here focuses on the field of Spatial Data Mining on classifying earthquake significance using important attributes, like magnitude. Spatial data mining is crucial for uncovering patterns and insights from data enabling informed decision making across different industries.This research extensively evaluates ten classification algorithms to classify earthquakes as "significant" or "not significant." The decision matrix showcases the performance of each algorithm based on metrics such as accuracy, precision, recall and the F1 score. Among these algorithms Random Forest stands out as the performer with its accuracy, strong F1 score and well balanced precision and recall rates. These impressive results indicate that Random Forest is the choice for classifying earthquake significance.In summary this study highlights how spatial data mining effectively addresses real world challenges and emphasizes the importance of selecting the algorithm to solve classification problems.The findings of this research clearly show that they can greatly influence areas, like disaster response and forecasting earthquakes. These insights will help us make choices based on data ensuring safety and efficient use of resources.

Future Work

In future,it would be valuable to investigate how well the Random Forest algorithm can adapt to geographic regions and seismic characteristics. To enhance the models capabilities we could also consider incorporating real time data streams and taking into account the dynamics of seismic events. Additionally it would be worth exploring how integrating spatial attributes, like geological features could improve the accuracy of classification. Working closely with seismologists and disaster response teams would provide insights for refining the model to suit geographical contexts. Moreover we should explore the benefits of combining classification algorithms through ensemble methods to create a more robust earthquake significance classification system. Finally it is important to assess how well the model scales for datasets and its practical applicability, in real time decision making scenarios when deployed in real time.

References

1) B. He, Y. Cui, J. Chen and P. Xie, "A spatial data mining method for mineral resources potential assessment," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 96-99, doi: 10.1109/ICSDM.2011.5969012.
2) Y. Wang and X. Chen, "Study on land use of changping district with spatial data mining method," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 218-222, doi: 10.1109/ICSDM.2011.5969035.
3) V. Bogorny and S. Shekhar, "Spatial and Spatio-temporal Data Mining," 2010 IEEE International Conference on Data Mining, Sydney, NSW, Australia, 2010, pp. 1217-1217, doi: 10.1109/ICDM.2010.166.
4) J. Lin, C. Chen, X. Wu, J. Wu and W. Wang, "GeoKSGrid: A geographical knowledge grid with functions of spatial data mining and spatial decision," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 121-126, doi: 10.1109/ICSDM.2011.5969017.
5) Jiandong Tu, Chongcheng Chen, Hongyu Huang and Xiaozhu Wu, "A visual multi-scale spatial clustering method based on graph-partition," Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., Seoul, Korea (South), 2005, pp. 4 pp.-, doi: 10.1109/IGARSS.2005.1525214.
6) K. Sravanthi and K. S. Rajan, "Mining Spatial Co-occurrence of Drought Events from Climate Data of India," 2010 IEEE International Conference on Data Mining Workshops, Sydney, NSW, Australia, 2010, pp. 106-112, doi: 10.1109/ICDMW.2010.142.
7) J. Peng, K. Qin, H. Yang and L. Cheng, "The study on the exploratory spatial data mining method based on partial random walk and its application in GPS TEC analysis," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 112-115, doi: 10.1109/ICSDM.2011.5969015.
8) M.Bendeck Ache and M. -T. Kechadi, "Distributed clustering algorithm for spatial data mining," 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), Fuzhou, China, 2015, pp. 60-65, doi: 10.1109/ICSDM.2015.7298026.
9) Hai-Bin Sun and Wen-Hui Li, "Qualitative spatial relationships cleaning for spatial data mining,"2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 2005, pp.1851-1857 Vol. 3, doi: 10.1109/ICMLC.2005.1527247.

10) Hailiang Jin and Baoliang Miao, "The research progress of spatial data mining technique," 2010 3rd International Conference on Computer Science and Information Technology, Chengdu, China, 2010, pp. 81-84, doi: 10.1109/ICCSIT.2010.5564659.
11) P. Dai, C. Jing, M. Du and W. Zhou, "A method based on spatial analyst to detect hot spot of urban component management events," 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), Fuzhou, China, 2015, pp. 55- 59, doi: 10.1109/ICSDM.2015.7298025.
12) T. S. Korting, L. M. G. Fonseca, M. I. S. Escada, F. C. da Silva and M. P. dos Santos Silva, "GeoDMA - A Novel System for Spatial Data Mining," 2008 IEEE International Conference on Data Mining Workshops, Pisa, Italy, 2008, pp. 975-978, doi: 10.1109/ICDMW.2008.22.
13) D. -H. Bae, J. -H. Baek, H. -K. Oh, J. -W. Song and S. -W. Kim, "SD-Miner: A spatial data mining system," 2009 IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 2009, pp. 803-807, doi: 10.1109/ICNIDC.2009.5360924.
14) D. Liu, D. Cao, S. Yin, Y. Li and H. Lian, "Water inrush dangerousness evaluation of coal floor from SDM," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 87-89, doi: 10.1109/ICSDM.2011.5969010.
15) V. Putrenko, N. Pashvnska and S. Nazarenko, "Data Mining of Network Events with Space-Time Cube Application," 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 2018, pp. 79-83, doi: 10.1109/DSMP.2018.8478437.
16) Yuanzhi Zhang, Xie Kunqing, Ma Xiujun, Xu Dan, Cai Cuo and Tang Shiwei, "Spatial data cube: provides better support for spatial data mining," Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., Seoul, 2005, pp. 4 pp.-, doi: 10.1109/IGARSS.2005.1525227.
17) X. Qiang, Y. Wei and Z. Hanfei, "Application of Visualization Technology in Spatial Data Mining," 2010 International Conference on Computing, Control and Industrial Engineering, Wuhan, China, 2010, pp. 153-157, doi: 10.1109/CCIE.2010.157.
18) He YueShun and Xu Wei, "A study of spatial data mining architecture and technology," 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, 2009, pp. 163-166, doi: 10.1109/ICCSIT.2009.5234974.
19) W. Sun and S. Wang, "A new data mining method for early warning landslides based on parallel coordinate," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 66-70, doi: 10.1109/ICSDM.2011.5969006.

20) L. Wang, X. Bao and L. Cao, "Interactive Probabilistic Post-Mining of User-Preferred Spatial Co-Location Patterns," 2018 IEEE 34th International Conference on Data Engineering (ICDE), Paris, France, 2018, pp. 1256-1259, doi: 10.1109/ICDE.2018.00124.
21) M. Ji, F. Jin, X. Zhao, B. Ai and T. Li, "Mine geological hazard multi-dimensional spatial data warehouse construction research," 2010 18th International Conference on Geoinformatics, Beijing, China, 2010, pp. 1-5, doi: 10.1109/GEOINFORMATICS.2010.5567673.
22) B. Qiu, C. Zeng, R. Long, C. Chen and X. Tu, "Quantifying spatial heterogeneity of Coniferous trees in ATM, CASI and Eagle airborne images," Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 2011, pp. 198-203, doi: 10.1109/ICSDM.2011.5969031.
23) A. A. S. Gunawan, A. N. Falah, A. Faruk, D. S. Lutero, B. N. Ruchjana and A. S. Abdullah, "Spatial data mining for predicting of unobserved zinc pollutant using ordinary point Kriging," 2016 International Workshop on Big Data and Information Security (IWBIS), Jakarta, Indonesia, 2016, pp. 83-88, doi: 10.1109/IWBIS.2016.7872894.
24) Z. Cheng, J. Wang, N. Yen and Y. Wu, "Allocation of Resources after Disaster Based on Big Data from SNS and Spatial Scan," 2016 International Conference on Advanced Cloud and Big Data (CBD), Chengdu, China, 2016, pp. 334-334, doi: 10.1109/CBD.2016.066.
25) L. Xu, Q. Qi and Z. Wang, "Regional Ecological Environment Spatial Data Mining Based on GIS and RS: A Case Study in Zhangjiajie, China," 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, China, 2008, pp. 266-270, doi: 10.1109/FSKD.2008.526.
26) Z. Shunzhi, H. Wenxing, W. Qunyong and L. Maoqing, "Research on data mining model of GIS-based urban underground pipeline network," 2009 IEEE International Conference on Control and Automation, Christchurch, New Zealand, 2009, pp. 1515-1520, doi: 10.1109/ICCA.2009.5410340.
27) Q. Zhu, "Improvement of spatial data clustering algorithm in city location," 2016 Eighth International Conference on Advanced Computational Intelligence (ICACI), Chiang Mai, Thailand, 2016, pp. 108-111, doi: 10.1109/ICACI.2016.7449812.
28) W. Peng and Z. Lei, "Mining Association Patterns between Forest and Influencing Factors Based on Spatial Data Handling and Statistical Techniques," 2010 WASE International Conference on Information Engineering, Beidai, China, 2010, pp. 41-45, doi: 10.1109/ICIE.2010.18.
29) Jong-Sheng Cherng and Mei-Jung Lo, "A hypergraph based clustering algorithm for spatial data sets," Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 2001, pp. 83-90, doi: 10.1109/ICDM.2001.989504.

30) A. Tripathy, S. Das and P. K. Patra, "An intelligent approach for mining frequent patterns in spatial database system using SQL," 2012 International Conference on Power, Signals, Controls and Computation, Thrissur, India, 2012, pp. 1-6, doi: 10.1109/EPSCICON.2012.6175236.
31) Z. -q. Wang, H. -b. Chen and H. -q. Yu, "Spatial Co-Location Rule Mining Research in Continuous Data," 2006 International Conference on Machine Learning and Cybernetics, Dalian, China, 2006, pp. 1362-1367, doi: 10.1109/ICMLC.2006.258705.
32) P. Kamakshi, "A survey on privacy issues and privacy preservation in spatial data mining," 2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014], Nagercoil, India, 2014, pp. 1759-1762, doi: 10.1109/ICCPCT.2014.7054961.
33) L. Xia and P. Hu, "Data Mining for Spatial Relations Based on Zero Initialization," 2008 International Conference on Advanced Language Processing and Web Information Technology, Dalian, China, 2008, pp. 380-386, doi: 10.1109/ALPIT.2008.24.
34) B. M. Varghese, A. Unnikrishnan and P. Jacob, "Information Content Extraction on Quad Trees for Active Spatial Image Clustering," 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, USA, 2009, pp. 306-310, doi: 10.1109/CSIE.2009.638.
35) Y. Zhan and H. Yang, "Extracting Hazardous Geology Information Based on Spatial Data Mining," 2009 Second International Conference on Intelligent Networks and Intelligent Systems, Tianjian, China, 2009, pp. 514-517, doi: 10.1109/ICINIS.2009.137.
36) L. Yu and F. Bian, "An Incremental Data Mining Method for Spatial Association Rule in GIS Based Fireproof System," 2007 International Conference on Wireless Communications, Networking and Mobile Computing, Shanghai, China, 2007, pp. 5983-5986, doi: 10.1109/WICOM.2007.1467.
37) Jing Cheng, Lei Liu, Yong Gao, Dan Jiang, Jiajun Liu and Haoran Li, "The Implementation of Spatial Co-location Mining and Its Application for POIs in Beijing," 2014 22nd InternationalConference on Geoinformatics, Kaohsiung, 2014, pp.1-4, doi: 10.1109/GEOINFORMATICS.2014.6950815.
38) S. Bi, S. Xue, Y. Xu and A. Pei, "Spatial Data Mining in Settlement Archaeological Databases Based on Vector Features," 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, China, 2008, pp. 277-281, doi: 10.1109/FSKD.2008.490.
39) J. S. Yoo and D. Boulware, "Incremental and parallel spatial association mining," 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2014, pp. 75-76, doi: 10.1109/BigData.2014.70.

40) Q. Ding, Q. Ding and W. Perrizo, "PARM—An Efficient Algorithm to Mine Association Rules From Spatial Data," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 38, no. 6, pp. 1513-1524, Dec. 2008, doi: 10.1109/TSMCB.2008.927730.
41) A. M. Sainju, D. Aghajarian, Z. Jiang and S. Prasad, "Parallel Grid-Based Colocation Mining Algorithms on GPUs for Big Spatial Event Data," in IEEE Transactions on Big Data, vol. 6, no. 1, pp. 107-118, 1 March 2020, doi: 10.1109/TBDATA.2018.2871062.
42) M. Zhou, T. Ai, G. Zhou and W. Hu, "A Visualization Method for Mining Colocation Patterns Constrained by a Road Network," in IEEE Access, vol. 8, pp. 51933-51944, 2020, doi: 10.1109/ACCESS.2020.2980168.
43) Y. Huang, L. Zhang and P. Zhang, "A Framework for Mining Sequential Patterns from Spatio-Temporal Event Data Sets," in IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 4, pp. 433-448, April 2008, doi: 10.1109/TKDE.2007.190712.
44) S. Gao, Y. Chen, K. Li, Y. Li, J. Yu and R. Rao, "Mapping Opencast Iron Mine and Mine Solid Waste Based on a New Spectral Index From Medium Spatial Resolution Satellite Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 7788-7798, 2021, doi: 10.1109/JSTARS.2021.3098801.
45) B. G. Peter and J. P. Messina, "Errors in Time-Series Remote Sensing and an Open Access Application for Detecting and Visualizing Spatial Data Outliers Using Google Earth Engine," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 4, pp. 1165-1174, April 2019, doi: 10.1109/JSTARS.2019.2901404.
46) Z. Jiang, "A Survey on Spatial Prediction Methods," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 9, pp. 1645-1664, 1 Sept. 2019, doi: 10.1109/TKDE.2018.2866809.
47) M. J. Zaki, "Scalable algorithms for association mining," in IEEE Transactions on Knowledge and Data Engineering, vol. 12, no. 3, pp. 372-390, May-June 2000, doi: 10.1109/69.846291.
48) J. Wan, P. Nan, Q. Guo and Q. Wang, "Multi-mode radar signal sorting by means of spatial data mining," in Journal of Communications and Networks, vol. 18, no. 5, pp. 725-734, October 2016, doi: 10.1109/JCN.2016.000102.
49) B. G. Peter and J. P. Messina, "Errors in Time-Series Remote Sensing and an Open Access Application for Detecting and Visualizing Spatial Data Outliers Using Google Earth Engine," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 4, pp. 1165-1174, April 2019, doi: 10.1109/JSTARS.2019.2901404.
50) Y. Huang, S. Shekhar and H. Xiong, "Discovering colocation patterns from spatial data sets: a general approach," in IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 12, pp. 1472-1485, Dec. 2004, doi: 10.1109/TKDE.2004.90.

51) E. M. Knorr and R. T. Ng, "Finding aggregate proximity relationships and commonalities in spatial data mining," in IEEE Transactions on Knowledge and Data Engineering, vol. 8, no. 6, pp. 884-897, Dec. 1996, doi: 10.1109/69.553156.
52) Z. Cai, J. Wang and K. He, "Adaptive Density-Based Spatial Clustering for Massive Data Analysis," in IEEE Access, vol. 8, pp. 23346-23358, 2020, doi: 10.1109/ACCESS.2020.2969440.
53) P. C. Smits and A. Friis-Christensen, "Resource Discovery in a European Spatial Data Infrastructure," in IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 1, pp. 85-95, Jan. 2007, doi: 10.1109/TKDE.2007.250587.
54) Y. Ge, Z. Yao and H. Li, "Computing Co-Location Patterns in Spatial Data with Extended Objects: A Scalable Buffer-Based Approach," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 2, pp. 401-414, 1 Feb. 2021, doi: 10.1109/TKDE.2019.2930598.
55) Z. Jiang, "A Survey on Spatial Prediction Methods," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 9, pp. 1645-1664, 1 Sept. 2019, doi: 10.1109/TKDE.2018.2866809.