# of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br> **GDD-K-Means Text Clustering Algorithm Based on Grid**<br> **Filtering Distance and Density of Outliers**<br>
Yao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>
School re and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br> **-Means Text Clustering Algorithm Based on Grid**<br> **Filtering Distance and Density of Outliers**<br>
Yao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br> *Mathematics and Data Sci* ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br> **Yext Clustering Algorithm Based on (Distance and Density of Outliers)<br>
Yao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>** *Manghang* **Mang<sup>\*</sup>, Xiuwen Qi<br>** *Manghang, Changji College, Chan* 59-0620) Vol. 2 No. 3, 2024<br>**Algorithm Based on Grid<br>ensity of Outliers**<br>, Xiuwen Qi<br>, Xiuwen Qi<br>*gji College, Changji, Xinjiang, China*<br>tuthor. *School of Mathematics and Density of Outliers*<br> *School of Mathematics and Data Science, Changji College, Changji, Xinjiang, China*<br> *School of Mathematics and Data Science, Changji College, Changji, Xinjiang, China*<br> *Sc* Fineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br> **Clustering Algorithm Based on Gric<br>
<b>tance and Density of Outliers**<br>
Vang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br> *Mang*, Bin Wang<sup>\*</sup>, Xiuwen Qi<br> *Mang, Bin Wang\*, Xiuwen Qi*<br> *Mang, China*

Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 20<br> **Abstract: In the era of big data, fully mining**<br> **Abstract: In the era of big data, fully mining**<br> **Abstract: In the era of big data, f Iournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024**<br> **ADD-K-Means Text Clustering Algorithm Based**<br> **Filtering Distance and Density of Outliers**<br>
Yao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>
Schoo **GDD-K-Means Text Clustering Algorithm Based or**<br>Filtering Distance and Density of Outliers<br>Xao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>Xao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>Xao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>Xao Wang*i College, Changji, Xi* **GDD-K-Means Text Clustering Algorithm**<br>
Filtering Distance and Density of O<br>
Yao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>
School of Mathematics and Data Science, Changji College, Chan<br>
\*Corresponding Author.<br>
Abstract: In the era of **Filtering Distance and Density of Outli**<br>
Yao Wang, Bin Wang<sup>\*</sup>, Xiuwen Qi<br>
School of Mathematics and Data Science, Changji College, Changji,<br>
\*Corresponding Author.<br>
\*Corresponding Author.<br>
Abstract: In the era of big da **Example 11 Set of Mathematics and Data Science, Changi College, Chang<br>** *unsupervised of Mathematics and Data Science, Changji College, Chang***<br>** *unsupervised to the era of big data, fully mining* **technology application<br>
<b>a learning brownthenomiature School of Mathematics and Data Science, Changji College, Changji, \*Corresponding Author.**<br> **learning the case of big data, fully mining** technology application and utilizing the value of big dat **Example 18 School of Mathematics and Data Science, Changji College, Changji, Xinji,<br>
\*Corresponding Author.**<br>
\*Corresponding Author.<br> **Abstract: In the era of big data, fully mining** technology application have<br>
and utili School of Mathematics and Data Science, Changji College, Changji, Xinjia<br>
\*Corresponding Author.<br>
\*Corresponding Author.<br> **Abstract:** In the era of big data, fully mining technology application have a<br>
and utilizing the va Schoot of Mathematics and Data Science, Changli College, Changli, Xinjuan,<br>
\*Corresponding Author.<br> **Abstract:** In the era of big data in line with<br>
the requirements of big data strategy plays a<br>
the requirements of big da The era of big data, fully mining<br>
technology application<br>
and utilizing the value of big data in line with<br>
the requirements of big data strategy plays a<br>
environment, it becom<br>
significant role in social development.<br>
to Abstract: In the era of big data, fully mining<br>
and utilizing the value of big data in line with<br>
the requirements of big data strategy plays a<br>
environment, it becomes particular<br>
tine requirements of big data strategy pl Abstract: In the era of big data, fully mining<br>
and utilizing the value of big data in line with<br>
the requirements of big data strategy plays a<br>
environment, it becomes particular<br>
the requirements of big data strategy pla **Example 12**<br> **ABSTRACE:** In the era of big data, tuny mining<br>
and utilizing the value of big data in line with<br>
the requirements of big data strategy plays a<br>
environment, it becomes p<br>
significant role in social developm and unizing the value of orgidaa in line with charactery can distribe requirements of big data strategy plays a environment, it becomes particular significant role in social development. Consaize the effective minimal proc Ine requirements of organa strategy prays a<br>
significant role in social development. to realize the effective mining<br>
Clustering algorithm can effectively partition<br>
massive data and grasp the munlabeled data sets through **Supering algorithm can serve in social and the serve the entire the substrate and graphs unalabeled data sets through unsupervised behind the data. In real life, learning process, and traditional K-Means amount of unlabel CURE THE SET AND THE CONSERGATE CONSERGATE CONSERGATE CONSERGATE CONSERGATE CONSERVIDE INTERPRETATION IN THE LIGE INTERPRETATION IN THE STATE AND ABSOLUTION A SURFACT IN THE SET AND A SURFACT AND A SURFACT AND A SURFACT A unancied data sets through unsupervised** benomd the data. In real line<br>algorithm is still the most widely used mine unsupervised data through<br>algorithm at present. By studying and to obtain great value has become<br>learning **incerting** process, and **reading the most wilely used** initial mine using and algorithm is still the most widely used in incertigated data through learning various improved algorithms of topic [1]. **traditional K-Means cl combining the most where the most where** the memberous and the learning various improved algorithms of toolical granit and the obtain great value has beed learning various improved algorithms of topic [1].<br> **traditional K Example 18 and Solution** and the number<br> **Solutional K-Means clustering algorithms** of topic [1].<br> **Interactional K-Means clustering algorithms**, The clustering algorit<br> **this paper has optimized the problems such as** uns **Example 12 Consumer the consumption and Experimentional K-Means clustering algorithm**<br> **this paper has optimized the problems such as** unsupervised data set well an<br> **unsatisfactory clustering results caused by** effecti **Traditional K-Weans custering argorinm**, The custering argorinm<br> **this paper has optimized the problems such as** unsupervised data set well<br> **unsatisfactory clustering results caused by** effectively, which may<br> **butters** this paper has optimized the problems such as<br>
unstablection<br>
outliers and disadvantages of initial center<br>
outliers and disadvantages of initial center<br>
point affecting initial partitioning. Good<br>
information clusters. Th **clustering algorithms** and disadvantage of initial center information accumulation<br> **opoint affecting initial partitioning.** Good information accumulation<br> **results** have been obtained. Firstly, the grid cluster analysis **Multiers and disadvantages of initial center**<br> **Means contract in the propertion in the properties are results have been obtained. Firstly, the grid cluster analysis is the p<br>
<b>Means cluster analysis is the p**<br> **Means dif bend and the set of the matter of the set of Fraction and LOF detection method of objects, where objects the may evaluation and density are used to are similar to each remove outliers. Then, the randomness of different from objects initial center selection is better intering and LOF detection method of** objects, where objects withm<br> **indexerging distance and density are used to**<br> **increase the similar to each other chindren are interior extent in<br>
<b>initial center selection is better The complement is the calculation is the randomness of** different from objects in the randomness of different from objects combining the "max-min principle" with the high intrinsic consisten strategy of maximum weight, an **reduced.** For the max-film principle with the minder information of clusters is determined according to the 2<sup>1</sup>. As a classical clustering algorithms, the proposed GDD-K-<br>
Means algorithm is that compared with the currently popular **Strategy of maximum weight, and the number**<br> **Griducers is determined according to the**<br> **Griducers is determined according to the**<br> **BWP index.** Experimental results have shown Means algorithm is wide<br> **that compared wit** that compared with the currently po<br>clustering algorithms, the proposed Gl<br>Means clustering algorithm has acl<br>better results in different data sets, an<br>accuracy and F-number and other eval<br>indexes are improved to a certain

**Centers**

2959-0620) Vol. 2 No. 3, 2024<br> **g Algorithm Based on Grid**<br> **Density of Outliers**<br>
ng<sup>\*</sup>, Xiuwen Qi<br>
nangji College, Changji, Xinjiang, China<br>
ng Author.<br>
technology application have also brought severe<br>
challenges to data 2959-0620) Vol. 2 No. 3, 2024<br> **g Algorithm Based on Grid**<br> **Density of Outliers**<br>
ng<sup>\*</sup>, Xiuwen Qi<br>
nangji College, Changji, Xinjiang, China<br>
ng Author.<br>
technology application have also brought severe<br>
challenges to dat **g Algorithm Based on Grid**<br> **Density of Outliers**<br>
ng<sup>\*</sup>, Xiuwen Qi<br>
nangji College, Changji, Xinjiang, China<br>
ng Author.<br>
technology application have also brought severe<br>
challenges to data mining technology. In such an **g Algorithm Based on Grid**<br> **Density of Outliers**<br>
ng<sup>\*</sup>, Xiuwen Qi<br> *aangji College, Changji, Xinjiang, China*<br> *ng Author.*<br>
technology application have also brought severe<br>
challenges to data mining technology. In such **Density of Outliers**<br>
ng<sup>\*</sup>, Xiuwen Qi<br>
nanggii College, Changji, Xinjiang, China<br>
ng Author.<br>
technology application have also brought severe<br>
challenges to data mining technology. In such an<br>
environment, it becomes par **Defisity of Outfler's**<br> **ng**<sup>\*</sup>, **Xiuwen Qi**<br> *dangji College, Changji, Xinjiang, China*<br>
technology application have also brought severe<br>
challenges to data mining technology. In such an<br>
environment, it becomes particul **ng<sup>\*</sup>, Xiuwen Qi**<br>
atanggii College, Changji, Xinjiang, China<br>
ag Author.<br>
technology application have also brought severe<br>
challenges to data mining technology. In such an<br>
environment, it becomes particularly important<br> **ng<sup>\*</sup>, Xiuwen Qi**<br>aangji College, Changji, Xinjiang, China<br>g Author.<br>technology application have also brought severe<br>challenges to data mining technology. In such an<br>environment, it becomes particularly important<br>to reali **ng**, **Xiuwen Qi**<br> *nangji College, Changji, Xinjiang, China*<br> *tg Author.*<br>
technology application have also brought severe<br>
challenges to data mining technology. In such an<br>
environment, it becomes particularly importan nangji College, Changji, Xinjiang, China<br>tig Author.<br>technology application have also brought sechallenges to data mining technology. In suce<br>environment, it becomes particularly impo<br>to realize the effective mining and ap *ng Author.*<br>technology application have also brought severe<br>challenges to data mining technology. In such an<br>environment, it becomes particularly important<br>to realize the effective mining and application of<br>massive data a technology application have also brought severe<br>challenges to data mining technology. In such an<br>environment, it becomes particularly important<br>to realize the effective mining and application of<br>massive data and grasp the technology application have also brought severe<br>challenges to data mining technology. In such an<br>environment, it becomes particularly important<br>to realize the effective mining and application of<br>massive data and grasp the technology application have also brought severe<br>challenges to data mining technology. In such an<br>environment, it becomes particularly important<br>to realize the effective mining and application of<br>massive data and grasp the challenges to data mining technology. In such an<br>environment, it becomes particularly important<br>to realize the effective mining and application of<br>massive data and grasp the mystery and mystery<br>behind the data. In real lif

clustering algorithms, the proposed GDD-K-<br>
Means clustering algorithm has achieved<br>
and strong compatibility.<br>
better results in different data sets, and the research direction for be<br>
accuracy and F-number and other eval Means clustering algorithm has achieved<br>
and strong compatibility. It is<br>
actuared at sets, and the research direction for big d<br>
actuared and the realuation<br>
indexes are improved to a certain extent, and<br>
the calculation decutiver and the model in different data sets, and the research direction for big decutation increasing. However, K-Mean individually has limitations, such as the alcelation time complexity is effectively set the number o accuracy and F-number and other evaluation<br>
index are improved to a certain extent, and<br>
internet internet internet internet internet internet internet internet<br>
internet calculation time complexity is effectively<br>
set the indexes are improved to a certain extent, and<br>
the calculation time complexity is effectively<br>
set the number of categories in<br>
when they do not know enough<br>
Keywords: Data Mining; K-Means Algorithm;<br>
Another limitation i the calculation time complexity is effectively<br>
reduced.<br>
difficult for users to give<br>
Keywords: Data Mining; K-Means Algorithm;<br>
Another limitation is that d<br>
Grid Filtering Outlier; Number of Class<br>
initial center point environment, it becomes particularly important<br>to realize the effective mining and application of<br>massive data and grasp the mystery and mystery<br>behind the data. In real life, there is a large<br>amount of unlabeled data, and to realize the effective mining and application of<br>massive data and grasp the mystery and mystery<br>behind the data. In real life, there is a large<br>amount of unlabeled data, and how to deeply<br>mine unsupervised data through b massive data and grasp the mystery and mystery<br>behind the data. In real life, there is a large<br>amount of unlabeled data, and how to deeply<br>mine unsupervised data through big data strategy<br>to obtain great value has become a behind the data. In real life, there is a large<br>amount of unlabeled data, and how to deeply<br>mine unsupervised data through big data strategy<br>to obtain great value has become an important<br>The clustering algorithm deals with amount of unlabeled data, and how to deeply<br>mine unsupervised data through big data strategy<br>to obtain great value has become an important<br>topic [1].<br>The clustering algorithm deals with the<br>unsupervised data set well and m mine unsupervised data through big data strategy<br>to obtain great value has become an important<br>topic [1].<br>The clustering algorithm deals with the<br>unsupervised data set well and mining the data<br>effectively, which makes the to obtain great value has become an important<br>topic [1].<br>The clustering algorithm deals with the<br>unsupervised data set well and mining the data<br>effectively, which makes the discrete<br>information accumulate into valuable<br>inf topic [1].<br>The clustering algorithm deals with the<br>unsupervised data set well and mining the data<br>effectively, which makes the discrete<br>information accumulate into valuable<br>information clusters. The clustering generated by The clustering algorithm deals with the<br>unsupervised data set well and mining the data<br>effectively, which makes the discrete<br>information accumulate into valuable<br>cluster analysis is the process of grouping data<br>objects, wh unsupervised data set well and mining the data<br>effectively, which makes the discrete<br>information accumulate into valuable<br>information clusters. The clustering generated by<br>cluster analysis is the process of growing data<br>ob effectively, which makes the discrete<br>information accumulate into valuable<br>information clusters. The clustering generated by<br>cluster analysis is the process of grouping data<br>redisting the objects within the same cluster<br>ar mformation accumulate into valuable<br>information clusters. The clustering generated by<br>cluster analysis is the process of grouping data<br>objects, where objects within the same cluster<br>are similar to each other but significan information clusters. The clustering generated by<br>cluster analysis is the process of grouping data<br>objects, where objects within the same cluster<br>are similar to each other but significantly<br>different from objects in other cluster analysis is the process of grouping data<br>objects, where objects within the same cluster<br>are similar to each other but significantly<br>different from objects in other clusters. These<br>clusters are collections of data o objects, where objects within the same cluster<br>are similar to each other but significantly<br>different from objects in other clusters. These<br>clusters are collections of data objects that have<br>high intrinsic consistency in fe are similar to each other but significantly<br>different from objects in other clusters. These<br>clusters are collections of data objects that have<br>high intrinsic consistency in features, but exhibit<br>significant differences bet different from objects in other clusters. These<br>clusters are collections of data objects that have<br>high intrinsic consistency in features, but exhibit<br>significant differences between different clusters.<br>[2]. As a classical clusters are collections of data objects that have<br>high intrinsic consistency in features, but exhibit<br>significant differences between different clusters.<br>[2]. As a classical clustering algorithm, the K-<br>Means algorithm is high intrinsic consistency in features, but exhibit<br>significant differences between different clusters.<br>[2]. As a classical clustering algorithm, the K-<br>Means algorithm is widely popular for its<br>concise ideas and easy impl signiticant differences between different clusters.<br>[2]. As a classical clustering algorithm, the K-Means algorithm is widely popular for its concise ideas and easy implementation, a wide<br>range of application scenarios, ea [2]. As a classical clustering algorithm, the K-<br>Means algorithm is widely popular for its<br>concise ideas and easy implementation, a wide<br>range of application scenarios, easy to operate,<br>and strong compatibility. It is stil Means algorithm is widely popular for its<br>concise ideas and easy implementation, a wide<br>range of application scenarios, easy to operate,<br>and strong compatibility. It is still a valuable<br>presearch direction for big data ana concise ideas and easy implementation, a wide<br>range of application scenarios, easy to operate,<br>and strong compatibility. It is still a valuable<br>research direction for big data analysis and<br>processing. However, K-Means algo range of application scenarios, easy to operate,<br>and strong compatibility. It is still a valuable<br>research direction for big data analysis and<br>processing. However, K-Means algorithm also<br>has limitations, such as the algori and strong compatibility. It is still a valuable<br>research direction for big data analysis and<br>processing. However, K-Means algorithm also<br>has limitations, such as the algorithm needs to<br>set the number of categories in adva research direction for big data analysis and<br>processing. However, K-Means algorithm also<br>has limitations, such as the algorithm needs to<br>set the number of categories in advance, and it is<br>difficult for users to give approp processing. However, K-Means algorithm also<br>has limitations, such as the algorithm needs to<br>set the number of categories in advance, and it is<br>difficult for users to give appropriate values<br>when they do not know enough abo has limitations, such as the algorithm needs to<br>set the number of categories in advance, and it is<br>difficult for users to give appropriate values<br>when they do not know enough about the data.<br>Another limitation is that the set the number of categories in advance, and it is<br>difficult for users to give appropriate values<br>when they do not know enough about the data.<br>Another limitation is that the randomness of The<br>initial center point of the al ITEN 114<br>
Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
required to be high. Researchers also failed to the initial division, and achiev<br>
screen the noise effectively, resulting in the By comparing w ITEN 14<br>
Sournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
required to be high. Researchers also failed to the initial division, and achie<br>
screen the noise effectively, resulting in the By comparing wit 114 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0<br>
required to be high. Researchers also failed to the initial division, and a<br>
screen the noise effectively, resulting in the By comparing with three c<br>
no Journal of Intelligence and Knowledge Engineering (ISSN: 2959-062<br>
required to be high. Researchers also failed to the initial division, and achiever<br>
screen the noise effectively, resulting in the By comparing with three ITENT 114<br>
114 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
114 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
114 required to be high. Researchers also failed to the initial d IT4 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
required to be high. Researchers also failed to the initial division, and achie<br>
screen the noise effectively, resulting in the By comparing with thr 114 Journal of Intelligence and Knowledge Engineering (ISSN: 29<br>
required to be high. Researchers also failed to the initial division, an-<br>
sicrecen the noise effectively, resulting in the By comparing with three<br>
noise p 114 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-062<br>
required to be high. Researchers also failed to the initial division, and achieveren the noise effectively, resulting in the By comparing with three cl I14 Iournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620<br>
required to be high. Researchers also failed to the initial division, and achie<br>
screen the noise effectively, resulting in the By comparing with thre Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
required to be high. Researchers also failed to the initial division, and achiev<br>
screen the noise effectively, resulting in the By comparing with three Juanual of Intelligence and Knowledge Engineering (ISSN: 29)<br>
required to be high. Researchers also failed to the initial division, an<br>
screen the noise effectively, resulting in the By comparing with thre<br>
noise point bei 114 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620<br>
required to be high. Researchers also failed to the initial division, and achies<br>
sereen the noise effectively, resulting in the By comparing with thr Fractrical to be high. Researchers also failed to the initial division, and ach<br>serect the noise effectively, resulting in the By comparing with three classes<br>point being taken as the initial center, four data sets, this p required to be high. Researchers also failed to the initial division, and achies<br>screen the noise effectively, resulting in the By comparing with three class<br>invision and increase in the stop and long traversing and long screen the noise effectively, resulting in the By comparing with thre<br>noise point being taken as the initial center, four data sets, this p<br>which may lead to inaccurate clustering and long<br>calculation time [4]. Researcher noise point being taken as the initial center, tour data sets, this pay<br>which may lead to inaccurate clustering and long proposed algorithm signi<br>calculation time [4]. Researchers considered the clustering quality and acc which may lead to inaccurate clustering and long<br>calculation time [4]. Researchers considered the clustering quality and accuracy<br>influence of density and combined with the the density maximum-minimum principle", thus eff calculation time [4]. Researchers considered the clustering quality and influence of density and combined with the "maximum-minimum principle", thus effectively avoiding local optimality, but data in low-<br>density regions influence of density and combined with the<br>
"maximum-minimum principle", thus effectively<br>
2. Algorithm Introduction<br>
avoiding local optimality, but data in low-<br>
density regions may be mistaken for outliers [5].<br>
2.1 Int "maximum-minimum principle", thus effectively<br>avoiding local optimality, but data in low-<br>density regions may be mistaken for outliers [5]. **2.1 Introduction to K-Means**<br>Ahmad W, et al. combined canopy algorithm In the cl avoiding local optimality, but data in low-<br>density regions may be mistaken for outliers [5]. **2.1 Introduction to K-Mean**<br>Ahmad W, et al. combined canopy algorithm<br>with density, although it could effectively D is divided density regions may be mistaken for outliers [5].<br>
A **Almad With** density, although it could effectively D is divided into several disjoin<br>
a Number data with density, although it could effectively D is divided ano severa Ahmad W, et al. combined canopy algorithm<br>
with density, although it could effectively D is divided into severa<br>
process low-density region data and iteration technology, a<br>
automatically determine the number of class div with density, although it could effectively D is divided into several disp<br>process low-density region data and iteration technology, and t<br>automatically determine the number of class divided according to the di<br>centers, i process low-density region data and iteration technology, an<br>automatically determine the number of class divided according to the<br>centers, it would only stop after taversing all sample and the cluster<br>data points, failing automatically determine the number of class<br>
cartern constants, fit would only stop after traversing all<br>
carterns, it would only stop after traversing all<br>
sample and the cluster<br>
effect and the accuracy of the number of centers, it would only stop atter traversing all sample and the cluster cent<br>data points, failing to consider the clustering iteration until convergence. T<br>effect and the accuracy of the number of centers, minimize the sq data points, tailing to consider the clustering<br>
effect and the accuracy of the number of centers, minimize the square error<br>
and lacking in the processing of noise points and<br>
outliers, making it easy to miss key informa effect and the accuracy of the number of centers,<br>
minimize the square error,<br>
and lacking in the processing of noise points and<br>
function is:<br>
coutiers, making it easy to miss key information<br>
[6]. Depth-based methods ca and lacking in the processing of noise points and<br>
tunction is:<br>
coutliers, making it easy to miss key information<br>
[6]. Depth-based methods can solve this problem<br>
by mapping data points into space, assuming that<br>
the da outliers, making it easy to miss key information<br>
[6]. Depth-based methods can solve this problem<br>
by mapping data points into space, assuming that<br>
the data points will be wrapped layer by layer<br>
Traditional K-Means clus [6]. Depth-based methods can solve this problem<br>by mapping data points into space, assuming that<br>the data points will be wrapped layer by layer<br>from the inside out, and the more data points in defects. Cluster k nee<br>the o by mapping data points into space, assuming that<br>
the data points will be wrapped layer by layer<br>
from the instituted out, and the more data points in defects. Cluster k needs<br>
the outer layer will be defined as more abno the data points will be wrapped layer by layer Traditional K-Means of<br>
the outer layer vill be defined as more abormal, its value is difficult to<br>
but the operation is not practical in high-<br>
given data set cannot lot<br>
dim from the inside out, and the more data points in<br>
the oter layer will be defined as more abnormal,<br>
the studie is difficult to estimate.<br>
but the operation is not practical in high-<br>
given data st cannot be deter-<br>
dimensi the outer layer will be defined as more abnormal, its value is difficult to estimate but the operation is not practical in high-<br>dimensional data. Researchers have proposed into several classes in advance many detection me but the operation is not practical in high-<br>
given data set cannot be dimensional data. Researchers have proposed<br>
into several classes in adva<br>
many detection methods based on density-<br>
clustering, such as LOF, INFLO, INS dimensional data. Researchers have proposed into several classes in advance.<br>
many detection methods based on density<br>
elagrithm is optimized accordinated accordinated which have high effectiveness and strong intuitive and many detection methods based on density-<br>
clustering, such as LOF, INFLO, INS, etc.,<br>
the distance between cl<br>
which have high effectiveness and strong intuitive and simplification and will be widely used [7]. This increas clustering, such as LOF, INFLO, INS, etc., the distance between cluster<br>which have high effectiveness and strong intuitive and simple, is extrem<br>simplification and will be widely used [7]. This increasing the spacing betwe which have high effectiveness and strong intuitive and simple, is extre<br>simplification and will be widely used [7]. This increasing the spacing betwee<br>algorithm identifies outlier subsets by evaluating the clustering effe simplification and will be widely used [ $\frac{1}{l}$ ]. This increasing the spacing betwee<br>the degree of abnormality of each data point in the tightness within each<br>the degree of abnormality of each data point in the tightnes algorithm identifies outlier subsets by evaluating<br>the clustering effect ca<br>the dataset, they also have some shortcoming<br>that cannot be ignored. This algorithm identifies well improves the ran<br>and determines a subset of ou the degree of abnormality of each data point in the tightness within each c<br>the dataset, they also have some shortcoming discrimination between different<br>that cannot be ignored. This algorithm identifies well improves the the dataset, they also have some shortcomings<br>
and determining that cannot be ignored. This algorithm identifies<br>
and the degree of abnormality of each data point in<br>
After determining the i<br>
the dataset, and usually selec that cannot be ignored. This algorithm identities<br>
and determines a subset of outliers by evaluating<br>
the data scale of aboromlily of each data point and After determining the<br>
the data set, and usually selects several dat and determines a subset of outliers by evaluating<br>
the degree of abnormality of each data point in After determining the initial p<br>
the dataset, and usually selects several data are consistent with the or<br>
points with larg the degree of abnormality of each data point in<br>the dataset, and usually selects several data are con<br>points with large outlier value as outlier points.<br>This method, which uses outlier factors to implement<br>determine outlie the dataset, and usually selects several data<br>
reaconsistent with the<br>
points with large outlier value as outlier points.<br>
This method, which uses outlier factors to implement for K-Means+<br>
determine outlier subsets, has h points with large outlier value as outlier points.<br>
This method, which uses outlier factors to implement for K-Means-<br>
determine outlier subsets, has high detection<br>
efficiency in small-scale data sets where the data, so i This method, which uses outlier factors to implement for K-Means<br>determine outlier subsets, has high detection it on large data, and it<br>efficiency in small-scale data sets where the data, so its applicability<br>number of ou determine outlier subsets, has high detection it on large data, an efficiency in small-scale data sets where the data, so its applicabile number of outliers is known. However, most [8]. Noise points in loutlier detection a number of outliers is known. However, most [8]. Noise points in low-doutlier detection algorithms do not pre-set the likely to be selected as cl specific number of outliers during execution, the data belonging to such dete

ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa

ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significa divided according to the simulation of the initial division, and achieved good results.<br>By comparing with three classical algorithms in four data sets, this paper verifies that the proposed algorithm significantly improves the initial division, and achieved good results.<br>By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significantly improves the<br>clustering quality and accuracy.<br> By comparing with three classical algorithms in<br>four data sets, this paper verifies that the<br>proposed algorithm significantly improves the<br>clustering quality and accuracy.<br>2. **Algorithm Introduction**<br>In the classics k-mea four data sets, this paper verifies that the<br>proposed algorithm significantly improves the<br>clustering quality and accuracy.<br>2. **Algorithm Introduction**<br>**2.1 Introduction to K-Means** ++ **Algorithm**<br>In the classics k-means proposed algorithm significantly improves the<br>clustering quality and accuracy.<br>2. **Algorithm Introduction**<br>2.1 **Introduction to K-Means** ++ **Algorithm**<br>In the classics k-means algorithm, the sample so<br>D is divided into se ssics k-means algorithm, the sample set<br>led into several disjoint subsets through<br>technology, and the cluster class is<br>ccording to the distance between the<br>nd the cluster center point, and the<br>until convergence. The final k-means algorithm, the sample set<br>to several disjoint subsets through<br>nology, and the cluster class is<br>ding to the distance between the<br>he cluster center point, and the<br>convergence. The final goal is to<br>square error, and **2.1 Introduction to K-Means ++ Algorithm**<br>In the classics k-means algorithm, the sample set<br>D is divided into several disjoint subsets through<br>iteration technology, and the cluster class is<br>divided according to the dista **2.1 Introduction to K-Means**  $++$  **Algorithm**<br>
In the classics k-means algorithm, the sample set<br>
D is divided into several disjoint subsets through<br>
divided according to the distance between the<br>
sample and the cluster c In the classics *K*-means algorithm, the sample set<br>
D is divided into several disjoint subsets through<br>
iteration technology, and the cluster class is<br>
divided according to the distance between the<br>
sample and the cluste

$$
E = \sum_{i=1}^{K} \sum_{x=C_i} ||x - \mu_i||^2
$$
 (1)

efficiency in small-scale data sets where the data, so its applicability to lar<br>number of outlier ais known. However, most [8]. Noise points in low-density<br>specific number of outliers during execution,<br>the data belonging t D is alvided according to the distance between the divided accordination<br>teration technology, and the cluster class is<br>divided according to the distance between the<br>sample and the cluster center point, and the<br>iteration u iteration technology, and the cluster class is<br>divided according to the distance between the<br>sample and the cluster center point, and the<br>iteration until convergence. The final goal is to<br>minimize the square error, and th divided according to the distance between the<br>sample and the cluster center point, and the<br>trattion until convergence. The final goal is to<br>trantion is:<br> $E = \sum_{i=1}^{K} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>Traditional K-Means clustering al sample and the cluster center point, and the<br>iteration until convergence. The final goal is to<br>minimize the square error, and the objective<br>function is:<br> $E = \sum_{i=1}^{K} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>Traditional K-Means clustering al iteration until convergence. Ine mail goal is to<br>minimize the square error, and the objective<br>function is:<br> $E = \sum_{i=1}^{K} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>Traditional K-Means clustering algorithm has<br>defects. Cluster k needs to be set minimize the square error, and the objective<br>function is:<br> $E = \sum_{i=1}^{K} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>Traditional K-Means clustering algorithm has<br>defects. Cluster k needs to be set in advance, and<br>its value is difficult to estima Function is:<br>  $E = \sum_{i=1}^{K} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>
Traditional K-Means clustering algorithm has<br>
defects. Cluster k needs to be set in advance, and<br>
its value is difficult to estimate. In many cases, a<br>
given data set cann  $E = \sum_{i=1}^{n} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>
Traditional K-Means clustering algorithm has<br>
defects. Cluster k needs to be set in advance, and<br>
its value is difficult to estimate. In many cases, a<br>
given data set cannot be determin  $E = \sum_{i=1}^{n} \sum_{x=C_i} ||x - \mu_i||^2$  (1)<br>
Traditional K-Means clustering algorithm has<br>
defects. Cluster k needs to be set in advance, and<br>
its value is difficult to estimate. In many cases, a<br>
given data set cannot be determin Fraditional K-Means clustering algorithm has<br>defects. Cluster k needs to be set in advance, and<br>its value is difficult to estimate. In many cases, a<br>given data set cannot be determined to cluster<br>into several classes in a Traditional K-Means clustering algorithm has<br>defects. Cluster k needs to be set in advance, and<br>its value is difficult to estimate. In many cases, a<br>given data set cannot be determined to cluster<br>into several classes in ad detects. Cluster k needs to be set in advance, and<br>its value is difficult to estimate. In many cases, a<br>given data set cannot be determined to cluster<br>into several classes in advance. The K-Means ++<br>algorithm is optimized Its value is difficult to estimate. In many cases, a<br>given data set cannot be determined to cluster<br>into several classes in advance. The K-Means ++<br>algorithm is optimized accordingly. Maximizing<br>the distance between cluste given data set cannot be determined to cluster<br>into several classes in advance. The K-Means ++<br>algorithm is optimized accordingly. Maximizing<br>the distance between cluster centers, although<br>intuitive and simple, is extremel into several classes in advance. The K-Means ++<br>algorithm is optimized accordingly. Maximizing<br>the distance between cluster centers, although<br>intuitive and simple, is extremely effective. By<br>increasing the spacing between algorithm is optimized accordingly. Maximizing<br>the distance between cluster centers, although<br>intuitive and simple, is extremely effective. By<br>increasing the spacing between cluster centers,<br>the clustering effect can be im the distance between cluster centers, although<br>intuitive and simple, is extremely effective. By<br>increasing the spacing between cluster centers,<br>the clustering effect can be improved, ensuring<br>the tightness within each clus mututive and simple, is extremely effective. By<br>increasing the spacing between cluster centers,<br>the clustering effect can be improved, ensuring<br>the tightness within each cluster and the<br>discrimination between different clu increasing the spacing between cluster centers,<br>the clustering effect can be improved, ensuring<br>the tightness within each cluster and the<br>discrimination between different clusters, and it<br>well improves the randomness of th the clustering effect can be improved, ensuring<br>the tightness within each cluster and the<br>discrimination between different clusters, and it<br>well improves the randomness of the initial<br>center point selection of K-Means algo the tightness within each cluster and the<br>discrimination between different clusters, and it<br>well improves the randomness of the initial<br>center point selection of K-Means algorithm.<br>After determining the initial point, the discrimination between different clusters, and it<br>well improves the randomness of the initial<br>center point selection of K-Means algorithm.<br>After determining the initial point, the rest parts<br>are consistent with the origina well improves the randomness of the initial<br>center point selection of K-Means algorithm.<br>After determining the initial point, the rest parts<br>are consistent with the original clustering<br>algorithm. Researchers although it is center point selection of K-Means algorithm.<br>After determining the initial point, the rest parts<br>are consistent with the original clustering<br>algorithm. Researchers although it is easy to<br>implement for K-Means++, it is not After determining the initial point, the rest parts<br>are consistent with the original clustering<br>algorithm. Researchers although it is easy to<br>implement for K-Means++, it is not easy to use<br>it to n large data, and it requir are consistent with the original clustering<br>algorithm. Researchers although it is easy to<br>implement for K-Means++, it is not easy to use<br>it on large data, and it requires traversal of all<br>data, so its applicability to larg algorithm. Researchers although it is easy to<br>implement for K-Means++, it is not easy to use<br>it on large data, and it requires traversal of all<br>data, so its applicability to large data is limited<br>[8]. Noise points in low-d implement for K-Means++, it is not easy to use<br>it on large data, and it requires traversal of all<br>data, so its applicability to large data is limited<br>[8]. Noise points in low-density regions are more<br>likely to be selected it on large data, and it requires traversal of all<br>data, so its applicability to large data is limited<br>[8]. Noise points in low-density regions are more<br>likely to be selected as clustering centers, so that<br>the data belongi data, so its applicability to large data is limited [8]. Noise points in low-density regions are more likely to be selected as clustering centers, so that the data belonging to such centers is too small, the ansishility of [8]. Noise points in low-density regions are more likely to be selected as clustering centers, so that the data belonging to such centers is too small, and the possibility of change in the subsequent K-Means algorithm iter

Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3,<br>good clustering effect. For a data set composed generated by canopy. H<br>of two clusters, one of which contains a noise initial values such as Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed generated by canopy. Howe<br>of two clusters, one of which contains a noise initial values Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed generated by canopy. However<br>of two clusters, one of which contains a noise initial va Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed generated by canopy. Howe<br>of two clusters, one of which contains a noise initial values Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed generated by canopy. Howeve<br>of two clusters, one of which contains a noise initial valu Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed<br>generated by canopy. However,<br>of two clusters, one of which contains a noise initial va Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed<br>of two clusters, one of which contains a noise initial values such as the initia<br>point, Lournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>good clustering effect. For a data set composed<br>generated by canopy. However<br>of two clusters, one of which contains a noise initial val **FREE 100**<br> **Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024**<br> **good clustering effect.** For a data set composed<br>
of two clusters, one of which contains a noise initial values such as Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
good clustering effect. For a data set composed enerated by canopy. Howe<br>
of two clusters, one of which contains a noise initial value Journal of Intelligence and Knowledge Engineerin<br>good clustering effect. For a data set comport<br>of two clusters, one of which contains a n<br>point, the K-Means++ algorithm will divide<br>data set into two cases. In the first ca good clustering effect. For a data set composed<br>
of two clusters, one of which contains a noise<br>
initial values such as the initial<br>
point, the K-Means++ algorithm will divide the<br>
point and Canopy area size has a g<br>
data of two clusters, one of which contains a noise<br>
mutial values such as the in<br>
point, the K-Means++ algorithm will divide the<br>
point and Canopy area size h<br>
data set into two cases. In the first case, the noise<br>
points will point, the K-Means ++ algorithm will divide the<br>point and Canopy ar<br>data set into two cases. In the first case, the noise on the clustering qu<br>points will be classified into a single class, and<br>algorithm class in the secon data set into two cases. In the first case, the noise<br>
on the clustering quality when<br>
points will be classified into a single class, and<br>
the remaining points will be classified into a<br>
class. In the second case, the nois

# **Introduction**

points will be classitied into a single class, and<br>
the remaining points will be classified into a<br>
classified into a<br>
classified into a class with the surrounding data<br>
classified into a class with the surrounding data<br> the remanning points will be classified into a<br>
class. In the second case, the noise points will be<br>
classification of the surrounding data<br>
points, resulting in the misclassification of the Many researchers have dorigina class. In the second case, the noise points will be<br>
classified into a class with the surrounding data<br>
points, resulting in the misclassification of the Many researchers have<br>
original data points. The correct clustering classified into a class with the surrounding data<br>
points, resulting in the misclassification of the Many researchers have different<br>
original data points. The correct clustering result<br>
is not obtained.<br>
2.2 Canopy + K-M points, resulting in the misclassification of the Many researchers have different<br>original data points. The correct clustering result outliers according to different<br>is not obtained.<br>**2.2 Canopy** + **K-Means Algorithm** d original data points. The correct clustering result<br>
is not obtained.<br> **Cao about the correct clustering result**<br> **Cao about the constrained a** the most<br> **Canopy** algorithm pre-classifies data on the basis<br>
currence canop 1s not obtained.<br>
2.2 Canopy + K-Means Algorithm deviates so badl<br>
Introduction<br>
2.2 Canopy + K-Means Algorithm deviates so badl<br>
5. Canopy algorithm pre-classifies data on the basis<br>
3. mechanisms. The algorithm, and can 2.2 Canopy + K-Means Algorithm duiters: an outlier represents a<br>
Introduction<br>
Introduction<br>
Canopy algorithm pre-classifies data on the basis<br>
suspected to be generated<br>
of K-Means algorithm, and can be approximated dete 2.2 Canopy + K-Means Algorithm deviates so badly from other control<br>action suspected to be generated of K-Means algorithm, and can be approximated detection algorithm calculat<br>by the number of large circles generated by d **Introduction**<br>
suspected to be generat<br>
Canopy algorithm pre-classifies data on the basis<br>
of K-Means algorithm, and can be approximated detecton algorithm calculat<br>
by the number of large circles generated by dispersion Canopy algorithm pre-classities data on the basis mechanisms. The density-base<br>of K-Means algorithm, and can be approximated detection algorithm calculate<br>by the number of large circles generated by compy when the number

of K-Means algorithm, and can be approximated<br>
by the number of large circles generated by dispersion obtained by c<br>
canopy when the number of cluster centers<br>
cluster centers<br>
cannot be determined artificially [9]. Canop by the number of large circles generated by<br>
cannot be determined artificially [9]. Canopy density of an object P with<br>
cannot be determined artificially [9]. Canopy density. [11]. The following<br>
processes data sets throu canopy when the number of cluster centers<br>
density of an object P with<br>
cannot be determined artificially [9]. Canopy<br>
density. [11]. The followin<br>
determined thresholds 11 and t2(e.g., Figure 1. (1) Reach distance<br>
Effec cannot be determined artificially [9]. Canopy density. [11]. The followin<br>processes data sets through two artificially involved:<br>determined thresholds t1 and t2(e.g., Figure 1. (1) Reach distance<br>chaotic data into several processes data sets through two artificially involved:<br>
determined thresholds t1 and t2(e.g, Figure 1. (1) Reach distance<br>
Effect of canopy classification), which can sort The KTH reachable distance<br>
chaotic data into sev determined thresholds t1 and t2(e.g., Figure 1. (1) Reach distance<br>
Effect of canopy classification), which can sort in the KTH reachable distance<br>
chaotic data into several data piles with certain point P is defined as f Effect of canopy classification), which can sort<br>
chaotic data into several data piles with certain<br>
rules. The algorith flow is as follows:<br>
and (1) Determine the two thresholds tl and t2<br>
(1) Determine the two threshold chaotic data into several data piles with certain<br>
point P is defined as follows:<br>
rules. The algorithm flow is as follows:<br>  $RD_K(p, o) = mc$ <br>
(11) Determine the two thresholds t1 and t2<br>
(t1>t2). (2) Select a data at random fr rules. The algorithm flow is as follows:<br>
(1) Determine the two thresholds t1 and t2<br>
(1)->2). C) Select a data tradical transform from the data<br>
set and calculate the distance between this data<br>
P and O. distance indicat (1) Determine the two thresholds t1 and t2 - distance (t1>t2). (2) Select a data at random from the data Where: d (p, o) represents t and calculate the distance eversent this data per and calculate the distance incidents (1)-t2). (2) Select a data at random from the data<br>
set and calculate the distance between this data<br>
paral O. distance indicates the set and canopy id ff there is no canopy at present, this<br>
point is directly used as the set and calculate the distance between this data P and O. dis<br>and canopy (if there is no canopy at present, this KTH point<br>point is directly used as the canopy center point). distance rep<br>(3) If this distance is less than



2959-0620) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorith  $2959-0620$ ) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algor 2959-0620) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorit 2959-0620) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorit 2959-0620) Vol. 2 No. 3, 2024<br>
115<br>
generated by canopy. However, the selection of<br>
initial values such as the initial Canopy center<br>
point and Canopy area size has a great influence<br>
on the clustering quality when the alg 2959-0620) Vol. 2 No. 3, 2024<br>
2959-0620) Vol. 2 No. 3, 2024<br>
2015<br>
2016 generated by canopy. However, the selection of<br>
initial values such as the initial Canopy center<br>
point and Canopy area size has a great influence<br>
o 2959-0620) Vol. 2 No. 3, 2024<br>
115<br>
generated by canopy. However, the selection of<br>
initial values such as the initial Canopy center<br>
point and Canopy area size has a great influence<br>
on the clustering quality when the alg 2959-0620) Vol. 2 No. 3, 2024<br>
115<br>
generated by canopy. However, the selection of<br>
initial values such as the initial Canopy center<br>
point and Canopy area size has a great influence<br>
on the clustering quality when the alg

2959-0620) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorith 2959-0620) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorit 2959-0620) Vol. 2 No. 3, 2024 115<br>generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorit 2009 0020) 00121601, 1020<br>
generated by canopy. However, the selection of<br>
initial values such as the initial Canopy center<br>
point and Canopy area size has a great influence<br>
any philed in practice [10].<br>
2.3 Local Anomaly generated by canopy. However, the selection of<br>initial values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorithm is<br>applied in practice [10].<br>**2.** mttal values such as the initial Canopy center<br>point and Canopy area size has a great influence<br>on the clustering quality when the algorithm is<br>applied in practice [10].<br>**2.3 Local Anomaly Detection Method of LOF**<br>**Based o** point and Canopy area size has a great influence<br>on the clustering quality when the algorithm is<br>applied in practice [10].<br>2.3 Local Anomaly Detection Method of LOF<br>Based on Density<br>Many researchers have different definiti on the clustering quality when the algorithm is<br>applied in practice [10].<br>2.3 Local Anomaly Detection Method of LOF<br>Based on Density<br>Many researchers have different definitions of<br>outliers according to different detection applied in practice [10].<br> **2.3 Local Anomaly Detection Method of LOF**<br> **Based on Density**<br>
Many researchers have different definitions of<br>
outliers according to different detection methods.<br>
Researchers gave the most clas **2.3 Local Anomaly Detection Method of LOF**<br>**Based on Density**<br>Many researchers have different definitions of<br>outliers according to different detection methods.<br>Researchers gave the most classic definition of<br>outliers: an involved: **Based on Density**<br>Many researchers have different definitions of<br>outliers according to different detection methods<br>Researchers gave the most classic definition of<br>outliers: an outlier represents a data point that<br>deviate Many researchers have different definitions of<br>outliers according to different detection methods.<br>Researchers gave the most classic definition of<br>outliers: an outlier represents a data point that<br>deviates to bed if from o outliers according to different detection methods.<br>Researchers gave the most classic definition of<br>outliers: an outlier represents a data point that<br>deviates so badly from other data points that it is<br>suspected to be gene deviates so badly from other data points that it is<br>suspected to be generated by different<br>mechanisms. The density-based LOF anomaly<br>detection algorithm calculates the degree of<br>dispersion obtained by comparing the local<br> suspected to be generated by different<br>mechanisms. The density-based LOF anomaly<br>detection algorithm calculates the degree of<br>dispersion obtained by comparing the local<br>density of an object P with the surrounding<br>density.

$$
D_K(p, o) = max{k
$$
  
- distance(o), d(p, o) (2)

Frame dianomology and the set of the set of the set of the set of point of the set of obtained by comparing the local an object P with the surrounding and object P with the surrounding definitions are stance eachable dist ithm calculates the degree of<br>
ined by comparing the local<br>
object P with the surrounding<br>
The following definitions are<br>
ce<br>
ce<br>
aable distance from point O to<br>
d as follows:<br>  $\chi(p, o) = max\{k$ <br>  $-$  distance(o),  $d(p, o)\}$  (2) mechanisms. The density-based LOP anomaly<br>detection algorithm calculates the degree of<br>dispersion obtained by comparing the local<br>density of an object P with the surrounding<br>density. [11]. The following definitions are<br>in detection algorithm calculates the degree of<br>dispersion obtained by comparing the local<br>density of an object P with the surrounding<br>density. [11]. The following definitions are<br>involved:<br> $(1)$  Reach distance<br>The KTH reach dispersion obtained by comparing the local<br>density of an object P with the surrounding<br>density. [11]. The following definitions are<br>involved:<br> $(1)$ Reach distance<br> $(1)$ Reach distance<br> $\frac{B_K(p, o) = max\{k}{o}$ <br> $\frac{distance(o), d(p, o)\}$ <br>Whe density of an object P with the surrounding<br>density. [11]. The following definitions are<br>involved:<br>(1) Reach distance<br>point P is defined as follows:<br> $RD_K(p, o) = max\{k$ <br> $- distance(o), d(p, o)\}$ <br>Where: d (p, o) represents the distance betwe density. [11]. The following definitions are<br>involved:<br>(1) Reach distance<br>The KTH reachable distance from point O to<br>point P is defined as follows:<br> $R D_K(p, o) = max\{k$ <br> $- distance(o), d(p, o)\}$ <br>Where: d (p, o) represents the distance betwe mvolved:<br>
(1) Reach distance<br>
The KTH reachable distance from point O to<br>
point P is defined as follows:<br>  $RD_K(p, o) = max\{k$ <br>  $-$  distance  $(o)$ ,  $d(p, o)$ } (2)<br>
Where: d (p, o) represents the distance between<br>
P and O. distance (1) Reach distance<br>
The KTH reachable distance from point O to<br>
point P is defined as follows:<br>  $RD_K(p, o) = max\{k$ <br>  $- distance(o), d(p, o)\}$ <br>
Where: d (p, o) represents the distance between<br>
P and O. distance indicates the distance from t The KIH reachable distance from point O to<br>point P is defined as follows:<br> $RD_K(p, o) = max\{k$ <br> $- distance(o), d(p, o)\}$  (2)<br>Where: d (p, o) represents the distance between<br>P and O. distance indicates the distance from the<br>KTH point of point point P is defined as follows:<br>  $RD_K(p, o) = max\{k$ <br>  $- distance(o), d(p, o)\}$  (2)<br>
Where: d (p, o) represents the distance between<br>
P and O. distance indicates the distance from the<br>
KTH point of point P, excluding point P; K-<br>
distance re  $RD_K(p, o) = max\{R$ <br>  $- distance(o), d(p, o)\}$ <br>
Where: d (p, o) represents the distance between<br>
P and O. distance indicates the distance from the<br>
KTH point of point P, excluding point P; K-<br>
distance represents the k th distance, and the KIT point of point 1, executing point 1, K-<br>
distance represents the k th distance, and the K<br>
distance of P is also the distance from the K th<br>
point of P, excluding the point P. By definition,<br>
the KTH reachable distanc distance represents the k in distance, and the K th<br>point of P, excluding the point P. By definition,<br>the KTH reachable distance from point O to<br>point P is at least the KTH distance from O, or<br>the true distance between OP

$$
LRD_k(p) = \frac{1}{\left(\frac{\sum_{o \in N_K(P)} reach - dist_k(p, o)}{|N_K(P)|}\right)}
$$
(3)

ccluding the point P. By definition,<br>
cchable distance from point O to<br>
least the KTH distance from O, or<br>
ce between OP's.<br>
chability density<br>
accessible density of point P is<br>
s follows:<br>
1<br>  $\frac{\sum_{o \in N_K(P)} reach - dist_k(p, o)}{|N_K(P)|}$ In urstance of P is also the ustance Hom the K the<br>point of P, excluding the point P. By definition,<br>the KTH reachable distance from point O to<br>point P is at least the KTH distance from O, or<br>the true distance between OP' bom of the KTH reachable distance from point O to<br>the KTH reachable distance from point O to<br>point P is at least the KTH distance from O, or<br>the true distance between OP's.<br>(2) Local reachability density<br>The locally acces number of K field points of P; local reachability<br>density of point P is at least the KTH distance from O, or<br>the true distance between OP's.<br>(2) Local reachability density of point P is<br>represented as follows:<br> $LRD_k(p) = \frac{1$ point 1 is at least the KTH ustance Hom O, or<br>the true distance between OP's.<br>(2) Local reachability density<br>The locally accessible density of point P is<br>represented as follows:<br>LRD<sub>k</sub>(p) =  $\frac{1}{\sqrt{\frac{\sum_{o \in N_K(P)} reach - dist_k(p, o)}{|N_K$ (2) Local reachability density<br>
The locally accessible density of point<br>
represented as follows:<br>
LRD<sub>k</sub>(p) =  $\frac{1}{\sqrt{\sum_{o \in N_K(P)} reach - dist_k(p, o)}}$ <br>
Where: Reach-distance (p) represents<br>
relative distance between point P and K fiel (2) Local reachability density<br>
The locally accessible density of point P is<br>
represented as follows:<br>
LRD<sub>k</sub>(p) =  $\frac{1}{\left(\frac{\sum_{o \in N_K(P)} reach - dist_k(p, o)}{|N_K(P)|}\right)}$  (3)<br>
Where: Reach-distance (p) represents the<br>
relative distance bet The locally accessible delistry of point 1 is<br>represented as follows:<br> $LRD_k(p) = \frac{1}{\left(\frac{\sum_{o \in N_K(P)} reach - dist_k(p, o)}{|N_K(P)|}\right)}$  (3)<br>Where: Reach-distance (p) represents the<br>relative distance between point P and K field; K-<br>distance nei

$$
LOF_{K}(P) = \frac{\sum_{O \in N_{K}(P)} \frac{lrd(o)}{lrd(p)}}{|N_{K}(P)|} = \frac{\sum_{O \in N_{K}(P)} lrd_{K}(o)}{|N_{K}(P)|} / lrd_{K}(p)
$$
(4)

116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density value of  $LOF_K(P)$  approaches 1, indicating that correspondingly larger. Only p 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density value of  $LOF_K(P)$  approaches 1, indicating that correspondingly larger. Once t 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620<br>
LOF represents the local outlier factor. The increase, and the density value of  $LOF_K(P)$  approaches 1, indicating that correspondingly larger. O<br>
poin 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density value of  $LOF_K(P)$  approaches 1, indicating that correspondingly larger. On poi 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density through<br>
value of  $LOF_K(P)$  approaches 1, indicating that correspondingly large 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density throught P is close to its domain density value and impossible to ensure that 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density through<br>
value of  $LOF_R(P)$  approaches 1, indicating that correspondingly large 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
LOF represents the local outlier factor. The increase, and the density three of  $LOF_K(P)$  approaches 1, indicating that correspondingly larger. Other LOF represents the local outlier factor. The increase, and the density through<br>
LOF represents the local outlier factor. The increase, and the density three<br>
value of LOF<sub>K</sub>(P) approaches 1, indicating that corresponding 116 Journal of Intelligence and Knowledge Engineering<br>
LOF represents the local outlier factor. The increase, and<br>
value of  $LOF_K(P)$  approaches 1, indicating that corresponding<br>
point P is close to its domain density valu 116 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vo<br>
LOF represents the local outlier factor. The increase, and the density thres<br>
value of  $LOF_R(P)$  approaches 1, indicating that correspondingly lar LOF represents the local outlier factor. The increase, and the density the value of  $LOF_K(P)$  approaches 1, indicating that correspondingly larger. Ot point P is close to its domain density value and impossible to ensure th value of  $LOF_K(P)$  approaches 1, indicating that<br>point P is close to its domain density value and<br>belongs to a cluster. The more its ratio is less<br>than 1, the higher the density of the point P is<br>defined as a dense point. T point *P* is close to its domain density value and<br>
belongs to a cluster. The more its ratio is loss of filtering effect of<br>
than 1, the higher the density of the point *P* is in the loss of filtering effect of<br>
than 1, t belongs to a cluster. The more its ratio is less<br>
than 1, the higher the density of the point P is<br>
than its domain density, and the point P is<br>
then its domain density threshold<br>
defined as a dense point. The greater the

than 1, the higher the density of the point P is<br>
than its domain density, and the point P is<br>
defined as a dense point. The greater the ratio is<br>
functional relationship with the<br>
than 1, the lower the density of point P than its domain density, and the point P is<br>
Ihen the density threshold of the<br>
defined as a dense point. The greater the ratio is<br>
the density threshold fu<br>
than 1, the lower the density of point P is<br>
set, and the densi defined as a dense point. The greater the ratio is<br>
than 1, the lower the density of point P is<br>
set, and the density threshol<br>
compared to its domain density, and point P is<br>
as:<br>
defined as a dispersion point [12].<br> **3.** than 1, the lower the density of point P is<br>
set, and the density thresho<br>
compared to its domain density, and point P is<br>
defined as a dispersion point [12].<br> **3. GDD-K-Means Clustering Algorithm**<br> **Improved Based on** compared to its domain density, and point P is<br>defined as a dispersion point [12].<br>
3. **GDD-K-Means** Clustering Algorithm<br>
Improved Based on Traditional Algorithm<br>
3.1 Outlier Removal<br>
5.1 Outlier Removal<br>
Firstly, a grid defined as a dispersion point [12].<br> **3.** GDD-K-Means Clustering Algorithm<br>
Improved Based on Traditional Algorithm<br>  $\mathbf{3}$  is the threshold of mess<br>  $\mathbf{3}$ .**1 Outlier Removal**<br>  $\mathbf{3}$  a grid-based outlier filtering 3. **GDD-K-Means Clustering Algorithm**  $\beta = \sqrt{\frac{3}{10}}$ <br> **Improved Based on Traditional Algorithm** Where:  $\beta$  is the threshold of<br>
is the size of the data set. The<br>
Firstly, a grid-based outlier filtering method is<br>
the me **3. GDD-K-Means Clustering Algorithm**<br> **Improved Based on Traditional Algorithm**<br> **IMEGEF 10** is the size of the data set. The m<br> **3.1 Outlier Removal**<br> **Einzyly, a grid-based outlier filtering method is**<br> **Einzyly, a g** Improved Based on Traditional Algorithm<br>
is the size of the algos of the data set. T<br>
Firstly, a grid-based outlier filtering method is<br>
the mesh density threshold fu<br>
subset through grid filtering. The algorithm<br>
density 3.1 Outlier Removal<br>
is used to initially screen the candidate outlier<br>
in the mesh density threshold<br>
subset through grid filtering. The algorithm<br>
sused to initially screen the candidate outlier<br>
mesh density threshold **3.1 Outlier Removal**<br>
Firstly, a grid-based outlier filtering method is reasonably by the mesh division<br>
use to initially screen the candidate outlier mesh density threshold funct<br>
subset through grid filtering. The algo Firstly, a grid-based outlier filtering method is<br>
used to initially screen the candidate outlier<br>
mesh density threshold can<br>
subset through grid filtering. The algorithm<br>
focuses on considering the density threshold of<br> used to initially screen the candidate outlier<br>subset through grid filtering. The algorithm<br>from emsile are the mesh is dense.<br>focuses on considering the density threshold of the mesh is dense.<br>the grid distribution of da subset through grid tiltering. The algorithm<br>focuses on considering the density threshold of the mesh is dense is decomes<br>the grid distribution of data points in the global<br>range to determine whether there is an outlier. T tocuses on considering the density threshold of<br>the grid distribution of data points in the global<br>range to determine whether there is an outlier. The liminary subset of candidate<br>The density threshold is taken as a filter the grid distribution of data points in the global<br>
multe technine whether there is an outlier.<br>
The density threshold is taken as a filter, and the obtained, and then a density<br>
data set with a density less than the thres range to determine whether there is an outlier.<br>
The density threshold is taken as a filter, and the<br>
data set with a density less than the threshold is<br>
taken as the candidate outlier subset. This stage in the  $\varepsilon$ <br>
can The density threshold is taken as a filter, and the<br>data set with a density less than the threshold is<br>taken as the candidate outlier subset. This stage in the a<br>can effectively reduce the amount of the density<br>computatio data set with a density less than the threshold is<br>
the data set of decect outlier<br>
taken as the candidate outlier subset. This stage<br>
computation to a certain extent. Then the density points per unit area. The mo<br>
comput taken as the candidate outlier subset. This stage<br>
can effectively reduces the amount of given the density is represented by<br>
computation to a certain extent. Then the density points per unit area. The more<br>
computation t can effectively reduce the amount of<br>
computation to a certain extent. Then the density is represented by<br>
outlier detection method is used to determine points per unit area, the greater<br>
more accurate abnormal data point computation to a certain extent. Then the density<br>
outlier detection method is used to determine<br>
more accurate abormal data points. The the greater the probability<br>
more accurate abormal data points. The grid point. On t outlier detection method is used to determine pound at position of accurate abovement and the performance of the algorithm is improved point. On the contrary, the more effectively, and the time complexity of the become a

more accurate abnormal data points. The unegreated the probability of dependent of the algorithm is improved point. On the contrary, the manifestive dialgorithm is related. In the grid filtering stage, classification rend performance of the algorithm is improved point. On the contrary, the effectively, and the time complexity of the become a noise point (e.g. desirfication rendering).<br>
the data set is scanned and each data point is mapped effectively, and the time complexity of the second a loose por<br>algorithm is reduced. In the grid filtering stage, classification render<br>in data set is scanned and each data point is<br>mapped to the corresponding grid cell t parameter setting<br>
data set into each grid, and set the<br>
of data sets as<br>
2, n3, ..., n<sub>n</sub>}, The number of grids is<br>
the number of grids and the number<br>
are mutually dependent. There is a<br>
relationship between the size of setting<br>
to each grid, and set the<br>
data sets as<br>
3, The number of grids is<br>
c of grids and the number<br>
ly dependent. There is a<br>
between the size of the<br>
f the grid, and the mesh<br>
n be defined as:<br>  $\left[\frac{3 + |N|^{1/4}}{1}\right]$ mapped to the corresponding grid cell to<br>
complete the mapping task.<br>
(1)Grid step parameter setting<br>
Divide the data set into each grid, and set the<br>
number of grids as<br>
N = { $n1$ ,  $n2$ ,  $n3$ , ...,  $n_n$ }, The number of g complete the mapping task.<br>
(1)Grid step parameter setting<br>
Divide the data set into each grid, and set the<br>
number<br>
of data sets as<br>  $N = {n1, n2, n3, ..., n_n}$ , The number of grids is<br>  $m * m$ , and the number of grids and the numb (1) Grid step parameter setting<br>
Divide the data set into each grid, and set the<br>
number<br>
of data sets as<br>  $N = {n1, n2, n3, ..., n_n}$ , The number of grids is<br>  $m * m$ , and the number of grids and the number<br>
of datasets are mutuall

$$
m = [IN]^{1/3} + [N]^{1/4}
$$
 (5)

ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate out ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate out ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate out ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate out ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate out ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate ou ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate ou ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate ou as: lingly larger. Otherwise, it is<br>
to ensure that the non-dense grid data<br>
egrated as candidate outliers, resulting<br>
s of filtering effect of grid filtering.<br>
lensity threshold of the grid will have a<br>
relationship with th density threshold will be<br>rger. Otherwise, it is<br>that the non-dense grid data<br>candidate outliers, resulting<br>mg effect of grid filtering.<br>shold of the grid will have a<br>p with the size of the data<br>rreshold function is defin ge Engineering (ISSN, 2959-0020) vol. 2 1No. 3, 2024<br>increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate increase, and the density threshold will be<br>correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate outliers, resulting<br>in the loss of filtering effect of correspondingly larger. Otherwise, it is<br>impossible to ensure that the non-dense grid data<br>will be integrated as candidate outliers, resulting<br>in the loss of filtering effect of grid filtering.<br>Then the density threshold impossible to ensure that the non-dense grid data<br>will be integrated as candidate outliers, resulting<br>in the loss of filtering effect of grid filtering.<br>Then the density threshold of the grid will have a<br>functional relat

$$
\beta = \left[ \frac{|N|^{1/3} + |N|^{1/4}}{3} \right] \tag{6}
$$

will be integrated as candidate outliers, resulting<br>in the loss of filtering effect of grid filtering.<br>Then the density threshold of the grid will have a<br>functional relationship with the size of the data<br>set, and the dens in the loss of filtering effect of grid filtering.<br>Then the density threshold of the grid will have a<br>functional relationship with the size of the data<br>set, and the density threshold function is defined<br>as:<br> $\beta = \left[\frac{|M|^{1/$ Then the density threshold of the grid will have a<br>functional relationship with the size of the data<br>set, and the density threshold function is defined<br>as:<br> $\beta = \left[\frac{|N|^{1/3} + |N|^{1/4}}{3}\right]$  (6)<br>Where:  $\beta$  is the threshold functional relationship with the size of the data<br>set, and the density threshold function is defined<br>as:<br> $\beta = \left[\frac{|N|^{1/3} + |N|^{1/4}}{3}\right]$  (6)<br>Where:  $\beta$  is the threshold of mesh density and N<br>is the size of the data set. set, and the density threshold function is defined<br>as:<br><br> $\beta = \left[\frac{|N|^{1/3} + |N|^{1/4}}{3}\right]$  (6)<br>Where:  $\beta$  is the threshold of mesh density and N<br>is the size of the data set. The mesh step size and<br>the mesh density threshold as:<br>  $\beta = \left[\frac{|N|^{1/3} + |N|^{1/4}}{3}\right]$  (6)<br>
Where:  $\beta$  is the threshold of mesh density and N<br>
is the size of the data set. The mesh step size and<br>
the mesh density threshold are calculated<br>
reasonably by the mesh division

 $\beta = \left[\frac{|N|^{1/3} + |N|^{1/4}}{3}\right]$  (6)<br>Where:  $\beta$  is the threshold of mesh density and N<br>is the size of the data set. The mesh step size and<br>the mesh density threshold are calculated<br>reasonably by the mesh division function  $\beta = \frac{1}{\sqrt{3}}$  (6)<br>
Where:  $\beta$  is the threshold of mesh density and N<br>
is the size of the data set. The mesh step size and<br>
the mesh density threshold are calculated<br>
reasonably by the mesh division function and the<br>
me Where:  $\beta$  is the threshold of mesh density and N<br>is the size of the data set. The mesh step size and<br>the mesh density threshold are calculated<br>treasonably by the mesh division function and the<br>mesh density threshold fun Where:  $\beta$  is the threshold of mesh density and N<br>is the size of the data set. The mesh step size and<br>the mesh density threshold are calculated<br>reasonably by the mesh division function and the<br>mesh density threshold can is the size of the data set. The mesh step size and<br>the mesh density threshold are calculated<br>reasonably by the mesh division function and the<br>mesh density threshold function, and the mesh<br>density threshold can be used to the mesh density threshold are calculated<br>reasonably by the mesh division function and the<br>mesh density threshold function, and the mesh<br>density threshold can be used to judge whether<br>the mesh is dense.<br>In the technique o reasonably by the mesh division function and the<br>mesh density threshold function, and the mesh<br>density threshold can be used to judge whether<br>the mesh is dense.<br>The technique of grid filtering outliers, a<br>preliminary subse mesh density threshold function, and the mesh<br>density threshold can be used to judge whether<br>the mesh is dense.<br>In the technique of grid filtering outliers, a<br>preliminary subset of candidate outliers has been<br>obtained, an density threshold can be used to judge whether<br>the mesh is dense.<br>In the technique of grid filtering outliers, a<br>preliminary subset of candidate outliers has been<br>obtained, and then a density-based detection<br>method is use



than that formed by the field of P2 points.<br>Taking the ratio of the number of fields to the Divide the data set into each grid, and set the<br>
number<br>  $N = {n1, n2, n3, ..., n_n}$ , The number of grids is<br>  $N = {n1, n2, n3, ..., n_n}$ , The number of grids is<br>
financial relationship dependent. There is a<br>
financial relationship between number<br>  $N = {n1, n2, n3, ..., n_n}$ , The number of grids is<br>  $N = {n1, n2, n3, ..., n_n}$ , The number of grids is<br>  $m * m$ , and the number of grids and the number<br>
functional relationship between the size of the<br>
functional relationship betwe N = {n1, n2, n3, ..., n<sub>n</sub>}, The number of grids is<br>
m \* m, and the number of grids and the number<br>
of datasets are mutually dependent. There is a<br>
functional relationship between the size of the<br>
data set and the size of m \* m, and the number of grids and the number<br>of datasets are mutually dependent. There is a<br>functional relationship between the size of the Figure 2. Canopy Classifica<br>partitioning function can be defined as:<br> $m = [M_1^{1/3$ of datasets are mutually dependent. There is a<br>
functional relationship between the size of the<br>
data set and the size of the grid, and the mesh<br>
partitioning function can be defined as:<br>  $m = [N1^{1/3} + N1^{1/4}]$  (5) than tha functional relationship between the size of the<br>
data set and the size of the grid, and the mesh<br>
partitioning function can be defined as:<br>  $m = [N_1^{1/2} + |N_1^{1/4}]$  (5) than that formed by the field of PI poi<br>
Where: N is The ration of the ration of the algorithm and improves the efficiency of the algorithm and improves the field of P1 points is much larger than that formed by the field of P1 points is much larger than that formed by the f Figure 2. Canopy Classification Rendering<br>Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>than that formed by the field of P2 poi Figure 2. Canopy Classification Rendering<br>Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>than that formed by the field of P2 poi Figure 2. Canopy Classification Rendering<br>Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>than that formed by the field of P2 poin **Example 12**<br> **Example 2. Canopy Classification Rendering**<br> **Example 2. Canopy Classification Rendering**<br>
As shown in Figure 2, the area of the unit circle<br>
formed by the field of P1 points is much larger<br>
than that formed **Example 12**<br> **Example 2. Canopy Classification Rendering**<br> **Example 2.** Canopy Classification Rendering<br>
formed by the field of PI points is much larger<br>
than that formed by the field of P2 points.<br>
Taking the ratio of th Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>than that formed by the field of P2 points.<br>Taking the ratio of the number of fie Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>than that formed by the field o Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>funn that formed by the field of PI points is much larger<br>Taking the ratio of the Figure 2. Canopy Classification Rendering<br>As shown in Figure 2, the area of the unit circle<br>formed by the field of PI points is much larger<br>than that formed by the field of P2 points.<br>Taking the ratio of the number of fie Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024 117

Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the judgment basis of the density between data<br>points.<br>3.2 Center Point Selection Method Where: p indicates the number of Indicates th points.

**3.2 Center Point Selection Method**<br> **3.3 Center Point Selection Method**<br> **5.4 Center Point Selection Method**<br> Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the judgment basis of the density between data<br>points.<br> $W_P = \sum_{q=1}^{m} \frac{range -}{ran}$ <br>3.2 Center Point Selection Method<br>In the traditional k-Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3,<br>the judgment basis of the density between data<br>points.<br>**3.2 Center Point Selection Method** Where: p indicates the 1<br>In the traditional k-Me Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the judgment basis of the density between data<br>points.<br> $W_P = \sum_{q=1}^{m} \frac{range - d(p) - d(p)}{range}$ <br>3.2 Center Point Selection Method<br>In the traditi Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the judgment basis of the density between data<br>points.<br>**2.2 Center Point Selection Method**<br>**3.2 Center Point Selection Method**<br>**1.1 th** Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3,  $\hat{p}$ <br>
the judgment basis of the density between data<br>
points.<br> **3.2 Center Point Selection Method**<br>
In the traditional k-Means algorithm Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the judgment basis of the density between data<br>points.<br> $W_P = \sum_{q=1}^{m} \frac{range - \frac{1}{q}}{randg}$ <br>2.2 Center Point Selection Method<br>In the tradit Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2the judgment basis of the density between data  $W_P = \sum_{q=1}^{m} \frac{range}{r}$ <br>points.  $W_P = \sum_{q=1}^{m} \frac{range}{r}$ <br>3.2 Center Point Selection Method In Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>the judgment basis of the density between data<br>points.<br> $W_P = \sum_{q=1}^{m} \frac{range - a}{range - a}$ <br>**3.2 Center Point Selection Method**<br>In the tradition **Samma or intelligential controlling results.** We points.<br>
the judgment basis of the density between data<br>
points.<br> **3.2 Center Point Selection Method**<br>
In the traditional k-Means algorithm, an initial points; m represent the judgment basis of the density between data<br>
points.<br>  $W_P = \sum_{q=1}^{n} \frac{range - d_1}{range - d_2}$ <br>
and  $\sum_{r=1}^{n} \frac{range - d_2}{range - d_1}$ <br>
Tange  $\sum_{r=1}^{n} \frac{range - d_1}{range - d_2}$ <br>  $\sum_{r=1}^{n} \frac{range - d_2}{range - d_1}$ <br>  $\sum_{r=1}^{n} \frac{range - d_2}{map - d_2}$ <br>  $\sum_{r=$ points.<br>  $W_P = \sum_{q=1}^{\infty} \frac{1}{\tan q}$ <br>
3.2 Center Point Selection Method<br>
In the traditional k-Means algorithm, an initial<br>
cluster center is randomly selected to determine<br>
cluster center is randomly selected to determine<br> **3.2 Center Point Selection Method**<br>In the traditional k-Means algorithm, at<br>cluster center is randomly selected to de<br>an initial partition, and then the clu<br>performed by iterative technology and th<br>partition is continuou on is continuously optimized. Ho<br>lection of the initial center has a<br>nce on the result, and the eff<br>on of the initial value plays a cruci<br>clustering result. Therefore, we will<br>nter points successively according<br>mum-minimu inuously optimized. However, calce<br>the initial center has a great follower, calce<br>net result, and the effective<br>nitial value plays a crucial role<br>result. Therefore, we will select<br>successively according to the<br>num princip initial center has a great<br>
sult, and the effective<br>
value plays a crucial role<br>
Therefore, we will select<br>
essively according to the<br>
principle", which is shown<br>
Where: x re<br>
Range repre<br>  $d_{i1}, d_{i2}$ ]<br>  $> \theta * ||z_1 - z_2||$ <br> an initial partition, and then the cluster is<br>
performed by iterative technology and the initial<br>
performed by iterative technology and the initial<br>
the vector space. Adopts the<br>
partition is continuously optimized. Howev mindent of the initial value plays a crucial role<br>since the centering result. Therefore, we will select<br>the center points successively according to the<br>"maximum-minimum principle", which is shown Where: x repres<br>as follow partition is continuously optimized. However, calculation method, the c<br>the selection of the initial center has a great follows:<br>influence on the result, and the effective<br>selection of the initial value plays a crucial ro the selection of the initial center has a great<br>
influence on the result, and the effective<br>
selection of the initial value plays a crucial role<br>
in the clustering result. Therefore, we will select<br>
the center points succ

$$
d_{l} = \max_{i} \left[ \min [d_{i1}, d_{i2}] \right] > \theta * ||z_{1} - z_{2}||
$$
 (7)

Where:  $\theta$  is the selected scale coefficient, passed<br>by  $d_i = min[d_{i1}, d_{i2}]$ ,  $i = 1, 2, ..., N$  The<br>minimum value between samples is obtained,<br>the larger the weight va<br>minimum value between samples is obtained,<br>the fuger the weig ta point with the largest index value of<br>point is used as the clustering center<br>perimental data for pre-classification. T<br>tion methods of the distance in t<br>hm all use Euclidean distance, and<br>of the data point is continuo Figure 1. The cluster center, the closer<br>
is used as the clustering center of<br>
is used as the clustering center of<br>
intil data for pre-classification. The<br>
methods of the distance in this<br>
use Euclidean distance, and the where the distance between samples is expressed<br>
by  $d_{ij} = ||x_i - z_j||_p = 1.2$ .<br>
The data point with the largest index value of the<br>
calue, the farther away<br>
cluster center, the closer the<br>
centeral point is used as the cluste by  $d_{ij} = ||x_i - z_j||_i = 1.2$ .<br>
The data point with the largest index value of the cluster center, the closer the cl<br>
echral point is used as the clustering center of<br>
the experimental data for pre-classification. The differen The data point with the largest index value of the<br>
the central point is used as the clustering center of<br>
the capcimental data for pre-classification. The difference between the two cluster<br>
calculation methods of the di

$$
\varepsilon = \frac{1}{K} \sum_{P_{I \in \mathcal{C}}} d(P_i, P_{K-nearrst(i)}) \tag{8}
$$

central point is used as the clustering center of<br>
calculation methods of the distance, and the interiore between the two clustering algorithm all use Euclidean distance, and the clustering effect. The time<br>
radius of the the experimental data for pre-classification. The<br>
calculation methods of the distance in this<br>
algorithm all use Euclidean distance, and the<br>
algorithm all use Euclidean distance, and the mainly determined by the number<br> calculation methods of the distance in this<br>algorithm all use Euclidean distance, and the<br>algorithm all use Euclidean distance, and the mainly determined by the<br>adaptively calculated according to the greedy<br>strategy. The algorithm all use Euclidean distance, and the<br>
mainly determined by the nun<br>
adaptively calculated according to the greedy<br>
strategy. The calculation formula is as follows:<br>  $\epsilon = \frac{1}{K} \sum_{p_{\text{i}\in\mathcal{C}}} d(P_{i}, P_{K-nearrst}(i))$  (8) radius of the data point is continuously and<br>adaptively calculated according to the greedy<br>strategy. The calculation formula is as follows:<br>time performance of the alg<br> $\epsilon = \frac{1}{K} \sum_{P \in \mathcal{C}} d(P_i, P_{K-nearrst}(i))$  (8) improved b adaptively calculated according to the greedy<br>
strategy. The calculation formula is as follows:<br>  $\epsilon = \frac{1}{K} \sum_{P_{i\in C}} d(P_i, P_{K-nearrst(i)})$  (8) improved by effectively selective<br>
where:  $P_{K-nearrst(i)}$  represents the K points There is strategy. The calculation formula is as follows:<br>  $\epsilon = \frac{1}{K} \sum_{P_{i\in\mathcal{L}}} d(P_{i}, P_{K-nearrst(i)})$  (8) improved by effectively select<br>
where:  $P_{K-nearrst(i)}$  represents the K points<br>
There is no cluster center point index<br>
Where:  $P_{$  $\epsilon = \frac{1}{K} \sum_{P_{\text{IEC}}} d(P_i, P_{K-nearrst(i)})$  (8) improved<br>
where:  $P_{K-nearrst(i)}$  represents the K points<br>
nearest to the point; In general, the value of K is<br>
4 in the two-dimensional spatial cluster, and in<br>
other cases the value of mensional spatial cluster, and in<br>
value of  $[n/25]$  in the data set is<br>
(n is the total number of data<br>
is rounded down). The weight<br>
is calculated according to the<br>
in the object and the data object of<br>
m, and the weight Where:  $P_{K-nearrst(i)}$  represents the K points<br>
4 in the two-dimensional spatial cluster, and in<br>
4 in the two-dimensional spatial cluster, and in<br>
4 in the two-dimensional spatial cluster, and in<br>
4 in the two-dimensional sp where:  $P_K$ -*nearest*(i) represents the value of  $\lfloor n/25 \rfloor$  in the data set, and the value of  $\lfloor n/25 \rfloor$  in the data set, and the value of  $\lfloor n/25 \rfloor$  in the data set, and the value of  $\lfloor n/25 \rfloor$  in the data set, which is more complete between the object and the object is calculated according to the value of  $\lfloor n/25 \rfloor$  in the  $\lfloor n/25 \rfloor$  in the data set, which is more complete. The number of data is the value of  $\lfloor n/25 \rfloor$  in the  $\lfloor n/25 \rfloor$  in the data setmearest to the point; in general, the value of K is<br>
4 in the two-dimensional spatial cluster, and in<br>
other cases the value of  $[n/25]$  in the data set is<br>
taken. Where (n is the total number of data within a given rang<br> ct is calculated according to the<br>veen the object and the data object q<br>main, and the weight is processed to<br>ral point index of each data, the<br>ormula is as follows:<br> $C_P = W_P * \theta_P$  (9)<br>epresents the distance between the<br>and t calculated according to the<br>
ne object and the data object q<br>
int index of each data, the<br>
is as follows:<br>  $= W_P * \theta_P$  (9)<br>
intial cent<br>
the center point i closest to<br>
no formula is as follows:<br>  $\min_{1 \le i \le k} d(i, p)$  (10)<br>  $\lim$ taken. Where (n is the total number of data<br>
samples and [ ] is rounded down). The weight<br>
which is more conducive<br>
of the object is calculated according to the<br>
the object is calculated according to the<br>
Therefore, selec samples and [ I is rounded down). The weight<br>of the object is calculated according to the distance between the object and the data object q<br>in the  $\varepsilon$ -domain, and the weight is processed to<br>weight as the initial<br>calcula of the object is calculated according to the<br>distance between the object and the data object q<br>in the  $\varepsilon$ -domain, and the weight is processed to<br>get the central point index of each data, the<br>calculation formula is as fo

$$
C_P = W_P * \theta_P \tag{9}
$$

$$
\theta_p = \min_{1 \le i \le k} d(i, p) \tag{10}
$$

20) Vol. 2 No. 3, 2024 117  
\n
$$
W_P = \sum_{q=1}^{m} \frac{range - d(p,q)}{range}
$$
 (11)  
\np indicates the number of current center  
\nm represents the number of objects of the  
\nect in the field of  $\varepsilon$ - of the data point p.  
\ntts the dimension size of the dataset in  
\ntor space. Adopts the Euclidean distance  
\nion method, the calculation formula is as

2959-0620) Vol. 2 No. 3, 2024 117<br>  $W_P = \sum_{q=1}^{m} \frac{range - d(p,q)}{range}$  (11)<br>
Where: *p* indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$  – of the data po 2959-0620) Vol. 2 No. 3, 2024 117<br>  $W_P = \sum_{q=1}^{m} \frac{range - d(p,q)}{range}$  (11)<br>
Where: p indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$  – of the data poin

2959-0620) Vol. 2 No. 3, 2024 117<br>  $W_P = \sum_{q=1}^{m} \frac{range - d(p, q)}{range}$  (11)<br>
Where: p indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$ − of the data poin 2959-0620) Vol. 2 No. 3, 2024 117<br>  $W_P = \sum_{q=1}^{m} \frac{range - d(p, q)}{range}$  (11)<br>
Where: p indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$  – of the data poi 2959-0620) Vol. 2 No. 3, 2024 117<br>  $W_P = \sum_{q=1}^{m} \frac{range - d(p,q)}{range}$  (11)<br>
Where: p indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$  – of the data poin 2959-0620) Vol. 2 No. 3, 2024 117<br>  $W_P = \sum_{q=1}^{m} \frac{range - d(p,q)}{range}$  (11)<br>
Where: *p* indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$  – of the data po follows:

$$
\text{range} = \sqrt{\sum_{z=0}^{x} \left\| \max_{z} - \min_{z} \right\|^2} \qquad (12)
$$

dimension range of the dataset. The contribution<br>degree of each data point in the  $\varepsilon$ -field of the selection of the initial value plays a crucial role<br>
in the clustering result. Therefore, we will select<br>
"maximum-minimum principle", which is shown<br>
"maximum-minimum principle", which is shown<br>  $d_l = \max \left[ \min \left[ d_{i1}, d_{i2} \$ in the clustering result. Therefore, we will select<br>
the center points successively according to the<br>
"maximum-minimum principle", which is shown<br>  $d_l = \max \left[ \min [d_{l1}, d_{l2}] \right]$ <br>  $\downarrow = \max \left[ \min [d_{l1}, d_{l2}] \right]$ <br>
Where:  $\theta$  is th the center points successively according to the<br>
"maximum-minimum principle", which is shown<br>
as follows:<br>  $d_l = \max \left[\min [d_{i1}, d_{i2}] \right]$  (7)<br>
Where: *x* represents the modul<br>  $d_l = \max \left[\min [d_{i1}, d_{i2}] \right]$  (7)<br>
Where:  $\theta$  is the "maximum-minimum principle", which is shown<br>
as follows:<br>
as follows:<br>  $d_l = \max \left[ \min [d_{i1}, d_{i2}] \right]$  (7)  $\theta * ||z_1 - z_2||$  (7) degree of each data point in t<br>
Where: 0 is the selected scale coefficient, passed<br>
by  $d_i = \min [d_{i1}, d$ algorithm all use Euclidean distance, and the  $d_l = \max \left[\min[d_{i1}, d_{i2}]\right]$  (7) dimension range of the dataset<br>  $\forall \theta \in \mathbb{R}$ . Where:  $\theta$  is the selected scale coefficient, passed<br>
by  $d_i = \min[d_{i1}, d_{i2}], i = 1, 2, ..., N$  The contribution degree range is [<br>
minimum value between **EXECUTE:**  $\theta$  is the selected scale coefficient, passed by  $d_i = min[d_{i1}, d_{i2}], i = 1, 2, ..., N$  The contribution degree range minimum value between samples is obtained, the larger the weight value, where the distance between samp e: *p* indicates the number of current center<br>
; m represents the number of objects of the<br>
bject in the field of  $\varepsilon$  – of the data point *p*.<br>
ects the dimension size of the dataset in<br>
cetor space. Adopts the Euclidea es the number of current center<br>the number of objects of the<br>field of  $\varepsilon$  − of the data point *p*.<br>mension size of the dataset in<br>Adopts the Euclidean distance<br>d, the calculation formula is as<br> $\sum_{z=0}^{x} \left\|\max_{z} - \min_{z}\$ We =  $\frac{q}{q=1}$  range (11)<br>
Where: p indicates the number of current center<br>
points; m represents the number of objects of the<br>
data object in the field of  $\varepsilon$  – of the data point p.<br>
It reflects the dimension size of Where: *p* indicates the number of current center<br>points; m represents the number of objects of the<br>data object in the field of  $\varepsilon$  – of the data point *p*.<br>It reflects the dimension size of the dataset in<br>the vector sp where: *p* indicates the number of current center<br>points; m represents the number of objects of the<br>data object in the field of  $\varepsilon$  – of the data point *p*.<br>It reflects the dimension size of the dataset in<br>the vector sp points; m represents the number of objects of the<br>data object in the field of  $\varepsilon$  – of the data point *p*.<br>It reflects the dimension size of the dataset in<br>the vector space. Adopts the Euclidean distance<br>calculation met data object in the field of  $\varepsilon$ - of the data point *p*.<br>It reflects the dimension size of the dataset in<br>the vector space. Adopts the Euclidean distance<br>calculation method, the calculation formula is as<br>follows:<br> $\text{range} =$ It reflects the dimension size of the dataset in<br>the vector space. Adopts the Euclidean distance<br>calculation method, the calculation formula is as<br>follows:<br> $\text{range} = \sqrt{\sum_{z=0}^{x} ||\text{max}_z - \text{min}_z||}$  (12)<br>Where: x represents the the vector space. Adopts the Euclidean distance<br>calculation method, the calculation formula is as<br>follows:<br> $\text{range} = \sqrt{\sum_{z=0}^{x} ||\text{max}_z - \text{min}_z||^2}$  (12)<br>Where: x represents the dimension of the data;<br>Range represents the mo calculation method, the calculation formula is as<br>follows:<br><br> $\text{range} = \sqrt{\sum_{z=0}^{x} ||\text{max}_z - \text{min}_z||^2}$  (12)<br>Where: x represents the dimension of the data;<br>Range represents the modulus of the whole<br>dimension range of the datas follows:<br>
range =  $\sqrt{\sum_{z=0}^{x} ||max_z - min_z||}$  (12)<br>
Where: x represents the dimension of the data;<br>
Range represents the modulus of the whole<br>
dimension range of the dataset. The contribution<br>
degree of each data point in the range =  $\left|\sum_{z=0}^{x} \left\|\max_{z} - \min_{z}\right\|\right|^2$  (12)<br>Where: x represents the dimension of the data;<br>Range represents the modulus of the whole<br>dimension range of the dataset. The contribution<br>degree of each data point in the  $\v$ range =  $\sum_{z=0} \|\text{max}_z - \text{min}_z\|$  (12)<br>Where: x represents the dimension of the data;<br>Range represents the modulus of the whole<br>dimension range of the dataset. The contribution<br>degree of each data point in the  $\varepsilon$ - fiel Where:  $x$  represents the dimension of the data;<br>Range represents the modulus of the whole<br>dimension range of the dataset. The contribution<br>degree of each data point in the  $\varepsilon$ — field of the<br>point is greater the closer Where:  $x$  represents the dimension of the data;<br>Range represents the modulus of the whole<br>dimension range of the dataset. The contribution<br>dimension range of the dataset. The contribution<br>point is greater the closer the Where:  $x$  represents the dimension of the data;<br>
Range represents the modulus of the whole<br>
dimension range of the dataset. The contribution<br>
degree of each data point in the  $\varepsilon$ - field of the<br>
point is greater the clo Range represents the modulus of the whole<br>dimension range of the dataset. The contribution<br>degree of each data point in the  $\varepsilon$ -- field of the<br>point is greater the closer the point is, and the<br>contribution degree range dimension range of the dataset. The contribution<br>degree of each data point in the  $\varepsilon$ -- field of the<br>point is greater the closer the point is, and the<br>contribution degree range is [0,1]. In summary,<br>the larger the weigh degree of each data point in the  $\varepsilon$ -- field of the<br>point is greater the closer the point is, and the<br>contribution degree range is [0,1]. In summary,<br>the larger the weight value, the more data around<br>value, the more den point is greater the closer the point is, and the contribution degree range is [0,1]. In summary, the larger the weight value, the more data around the object point, the more dense. The larger the dauster center way from t contribution degree range is [0,1]. In summary,<br>the larger the weight value, the more data around<br>the object point, the more dense. The larger the<br>cluster center, the closer the cluster clustering.<br>Chuster center, the clo the larger the weight value, the more data around<br>the object point, the more dense. The larger the<br>value, the farther away from the generated<br>cluster center, the closer the cluster clustering.<br>The larger the central index the object point, the more dense. The larger the value, the farther away from the generated cluster center, the closer the cluster clustering. The larger the central index obtained by multiplication, the higher the degree value, the farther away from the generated<br>cluster center, the closer the cluster clustering.<br>The larger the central index obtained by<br>multiplication, the higher the degree of<br>difference between the two clusters, the bett cluster center, the closer the cluster clustering.<br>The larger the central index obtained by<br>multiplication, the higher the degree of<br>difference between the two clusters, the better<br>the clustering effect. The time consumpt The larger the central index obtained by<br>multiplication, the higher the degree of<br>difference between the two clusters, the better<br>the clustering effect. The time consumption is<br>mainly determined by the number of iteration

multiplication, the higher the degree of difference between the two clusters, the better the clustering effect. The time consumption is mainly determined by the number of iterations, and the number of iterations of K-neans difference between the two clusters, the better<br>the clustering effect. The time consumption is<br>mainly determined by the number of iterations,<br>and the number of iterations of K-means<br>algorithm can be effectively reduced an the clustering effect. The time consumption is<br>mainly determined by the number of iterations,<br>and the number of iterations of K-means<br>algorithm can be effectively reduced and the<br>time performance of the algorithm can be<br>i manly determined by the number of iterations,<br>and the number of iterations of K-means<br>algorithm can be effectively reduced and the<br>time performance of the algorithm can be<br>time performance of the algorithm can be<br>center o and the number of iterations of K-means<br>algorithm can be effectively reduced and the<br>time performance of the algorithm can be<br>improved by effectively selecting the clustering<br>Center of the center point index.<br>There is no algorithm can be effectively reduced and the<br>time performance of the algorithm can be<br>improved by effectively selecting the clustering<br>conter of the center point index.<br>There is no cluster center point at the beginning<br>of time performance of the algorithm can be<br>improved by effectively selecting the clustering<br>center of the center point index.<br>There is no cluster center point index of<br>data<br>cannot be calculated due to the lack of  $\theta$ <br>param mproved by effectively selecting the clustering<br>center of the center point index.<br>There is no cluster center point at the beginning<br>of clustering, and the center point index of data<br>cannot be calculated due to the lack of center of the center point index.<br>There is no cluster center point at the beginning<br>of clustering, and the center point index of data<br>cannot be calculated due to the lack of  $\theta$ <br>parameter. The more times a data point,<br>wh There is no cluster center point at the beginning<br>of clustering, and the center point index of data<br>cannot be calculated due to the lack of  $\theta$ <br>parameter. The more times a data point appears<br>within a given range, the den of clustering, and the center point index of data<br>cannot be calculated due to the lack of  $\theta$ <br>parameter. The more times a data point appears<br>within a given range, the denser the data point,<br>which is more conducive to the cannot be calculated due to the lack of  $\theta$ <br>parameter. The more times a data point appears<br>within a given range, the denser the data point,<br>thich is more conducive to the convergence of<br>the objective function as the clus parameter. The more times a data point appears<br>within a given range, the denser the data point,<br>which is more conducive to the convergence of<br>the objective function as the cluster center point.<br>Therefore, selecting the poi within a given range, the denser the data point,<br>which is more conducive to the convergence of<br>the objective function as the cluster center point.<br>Therefore, selecting the point with the largest<br>weight as the initial cente which is more conducive to the convergence of<br>the objective function as the cluster center point.<br>Therefore, selecting the point with the largest<br>weight as the initial center point is also<br>conducive to improving the tightn the objective tunction as the cluster center point.<br>Therefore, selecting the point with the largest<br>weight as the initial center point is also<br>conducive to improving the tightness within the<br>cluster, and conforms to the id

Copyright @ STEMM Institute Press http://www.stemmpress.com

ITEN 18<br>
118 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-062)<br>
clustering effect. According to the central point not participate in the subsection.<br>
index, it can be seen that the greater the density<br>
aro Index, it can be seen that the greater the density<br>index, it can be seen that the greater the density<br>index, it can be seen that the greater the density<br>around the cluster central point, the greater the density<br>weight of t ITE 18 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
clustering effect. According to the central point in the subsequendex, it can be seen that the greater the density selection. Compare the change<br> Journal of Intelligence and Knowledge Engineering (ISSN: 295<br>clustering effect. According to the central point is not participate in the<br>index, it can be seen that the greater the density selection. Compare the e<br>around th Journal of Intelligence and Knowledge Engineering (ISSN: 2959-06<br>
clustering effect. According to the central point in the participate in the subs<br>
index, it can be seen that the greater the density selection. Compare the 118 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>clustering effect. According to the central point and the subseque<br>index, it can be seen that the greater the density selection. Compare the change<br>ar Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>clustering effect. According to the central point and point intervalses<br>index, it can be seen that the greater the density selection. Compare the change<br>a Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) v<br>clustering effect. According to the central point anticipate in the subseque<br>index, it can be seen that the greater the density selection. Compare the ITENT 18<br>
118 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
clustering effect. According to the central point into participate in the subsequender, it can be seen that the greater the density selecti ITEN 118<br>
118 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-06<br>
clustering effect. According to the central point to participate in the subs<br>
index, it can be seen that the greater the density<br>
around the c I18 Journal of Intelligence and Knowlect Clustering effect. According to the central point index, it can be seen that the greater the density around the cluster central point. If the "max-min principle" is applied to the c clustering effect. According to the central point<br>
index, it can be seen that the greater the density<br>
around the cluster central point, the greater the of BWP index of data<br>
weight of the data point. If the "max-min pre-c clustering effect. According to the central point<br>index, it can be seen that the greater the density<br>according to HOBO around the cluster central point, the greater the change<br>of BWP index of data point<br>weight of the data mate the cluster central point in greater the density<br>
are around the cluster central point, the greater the of BWP index of data points be<br>
are weight of the data point. If the "max-min pre-classification. If the mean<br>
pr able to the data point. The section of the cluster central point. If the "max-min pre-classification. If the r<br>principle is applied to the central point index, index increases, this point<br>the larger the weight of a data po

weight of the data point. If the "max-min pre-classification. If the mean<br>principle" is applied to the central point index, index increases, this point will<br>greater the distance from other central points, the clustering ce principle" is applied to the central point index, index increases, this point<br>the larger the weight of a data point and the clustering center, and<br>greater the distance from other central points, the clustering center will the larger the weight of a data point and the clustering center, and the greater the distance from other central points, the clustering center will cause compare a greater the interest of the set manually be new cluster ce greater the distance from other central points, the<br>greater the possibility of this point becoming a<br>new clustering center point<br>new clustering between the cluster centres, the better caregory of the center point<br>clusterin greater the possibility of this point becoming a<br>new cluster central point, the greater the category of the center point<br>clustering between the cluster centers, the better Therefore, the center point<br>the clustering result. new cluster central point, the greater the category of the center point clustering between the cluster centers, the better Therefore, the center point in the clustering result.<br>
3.3 Selection of K Value of Cluster Number<br> clustering between the cluster centers, the better<br>
the clustering result.<br> **as 3.3 Selection of K Value of Cluster Number**<br>
or no data points exist, the<br>
After the data is pre-classified by the central<br>
point will stop.<br> the clustering result.<br> **3.3 Selection of K Value of Cluster Number**<br>
After the data is pre-classified by the central<br>
point index, that is, all data points will be<br>
classified into the category of the nearest central<br>
poi 3.3 Selection of K Value of Cluster Number<br>
After the data is pre-classified by the central<br>
point vill stop.<br>
After the data is pre-classified by the central<br>
point vill stop.<br>
classified into the category of the nearest 3.3 Selection of K Value of Cluster Number<br>
or no data points exist, the select<br>
After the data is pre-classified by the central<br>
point index, that is, all data points will select<br>
classified into the category of the near After the data is pre-classified by the central point will stop.<br>
point index, that is, all data points will be<br>
classified into the category of the nearest central<br>
point. The K-means algorithm belongs to the **Algorithm** point index, that is, all data points will be<br>classified into the category of the nearest central<br>3.4 Basic Ideas of GDD-K-Me<br>class. The K-means algorithm belongs to the **Algorithm**<br>unsupervised clustering method, and the classitied into the category of the nearest central<br>point. The K-means algorithm belongs to the<br>unsupervised clustering method, and the cluster<br>number k needs to be set manually, but it is<br>difficult to estimate. Therefore m to divide a given data set into several<br>
ries. This paper will refer to a new<br>
the introduction of BWP index and the<br>
mation of canopy algorithm, the algorithm<br>
utomatically determine the number of<br>
s. The average value ect K clustering centers. outliers are<br>
ion of BWP index and the with the la<br>
by algorithm, the algorithm<br>
etermine the number of selected st<br>
value of BWP index is point inde<br>
generated,<br>  $\sum_{i=1, j \in j}^{n} \frac{b(j,i)-w(j,i)}{b(j,i)+$ number k needs to be set manually, but it is preliminary screening of<br>difficult to estimate. Therefore, it is an urgent<br>data whose density is<br>problem to divide a given data set into several<br>thustening index BWP proposed i difficult to estimate. Therefore, it is an urgent<br>
data whose density is le<br>
problem to divide a given data set into several<br>
categories. This paper will refer to a new the classical density-based<br>
clustering index BWP pr problem to divide a given data set into several<br>
threshold is divided into c<br>
categories. This paper will refer to a new<br>
the classical density-bas<br>
clustering index BWP proposed in the literature<br>
to automatically select categories. This paper will refer to a new the classical density-base<br>clustering index BWP proposed in the literature used to accurately remove<br>to automatically select K clustering centers. outliers are removed from<br>Throu

$$
\overline{BWP}(j,i) = \frac{1}{n} \sum_{i=1, j \in j}^{n} \frac{b(j,i) - w(j,i)}{b(j,i) + w(j,i)} \qquad (13) \qquad \text{d}s
$$

clustering index BWP proposed in the literature<br>
to automatically select K clustering centers.<br>
Through te introduction of BWP index and the with the largest weight is selected<br>
combination of canopy algorithm, the algori to automatically select K clustering centers. outliers are removed from<br>
Through the introduction of BWP index and the with the largest weight is<br>
combination of canopy algorithm, the algorithm center point, and then<br>
cal Through the introduction of BWP index and the<br>
combination of canopy algorithm, the algorithm<br>
center point, and then n<br>
center point, and then n<br>
can automatically determine the number of<br>
selected successively accord<br>
c combination of canopy algorithm, the algorithm<br>
centre point, and then n ce<br>
calculated so the average value of BWP index is point index. When  $n+2$  cell<br>
calculated as follows:<br>  $\overline{BWP}(j, i) = \frac{1}{n} \sum_{i=1, j \in j}^{n} \frac{b(j,i)$ can automatically determine the number of selected successively accordical<br>classes. The average value of BWP index is point index. When n+2 calculated as follows:<br>calculated as follows: generated, the average value<br> $\overline{B$ classes. The average value of BWP index is point index. When n+2 calculated as follows:<br>  $\overline{BWP}(j, i) = \frac{1}{n} \sum_{i=1, j \in j}^{n} \frac{b(j,i) - w(j,i)}{b(j,i) + w(j,i)}$  (13) decreases, then the selection is under points are obtained. Firm the calculated as follows:<br> **EWP(j, i)** =  $\frac{1}{n} \sum_{i=1, j \in j}^{n} \frac{b(j,i) - w(j,i)}{b(j,i) + w(j,i)}$  (13) decreases, then the selection is<br>
where: n represents the size of the dataset, center point is used as the interpolism are obtained follows: , to be the minimum value of the avera<br>
, to be the minimum value of the avera<br>
of samples from this sample to every oth<br>
and define the intra-class distance<br>
i of class j, to be the average value of t<br>
ce between this da  $\begin{aligned}\n &\text{if } \mathbf{r} \text{ is a simple to every other} \\
 &\text{if } \mathbf{r} \text{ is a simple to every other} \\
 &\text{if } \mathbf{r} \text{ is a positive, and } \mathbf{r} \text{ is a$ value of samples from this sample to every other<br>
class, and define the intra-class distance of<br>
class, and define the intra-class distance of<br>
distance between this data object and other data<br>
objects of class j. The cal class, and define the intra-class distance of<br>
object i of class j, to be the average value of the<br>
distance between this data object and other data<br>
objects of class j. The calculation formula is as<br>  $\text{GDD-K-Means algorithm}$  algori

$$
b(j, i) = \min_{1 \le c \le k, c \ne j} \left(\frac{1}{n}\right)^n \sum_{p=1}^c \|x_p(c)\| \tag{14}
$$
  
- x.(j) \|2

$$
w(j, i) = \left(\frac{1}{n_j - 1}\right)^2 \sum_{p=1, p \neq i}^{j} \|x_p(j) - x_i(j)\|^2 \tag{15}
$$

given to class j, to be the average variate<br>stance between this data object and other<br>jects of class j. The calculation formulal<br>llows:<br> $b(j, i) = \min_{1 \leq c \leq k, c \neq j} \left(\frac{1}{n}\right)^n \sum_{p=1}^c \|x_p(c) - x_i(j)\|^2$ <br> $w(j, i) = \left(\frac{1}{n_j - 1}\right)^2 \$ ween this data object and other da<br>
ass j. The calculation formula is<br>  $=\min_{1 \leq c \leq k, c \neq j} \left(\frac{1}{n}\right)^n \sum_{p=1}^c \|x_p(c)$ <br>  $-x_i^{(j)}\|^2$ <br>  $\frac{1}{n_j - 1} \sum_{p=1, p \neq i}^{j} \|x_p(j) - x_i^{(j)}\|^2$  (1<br>
of BWP index determines wheth<br>
next clu ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the not participate in the subsequent center point<br>of BWP index of data points before and after the<br>of BWP index of data points before and after the<br>pre-classification. If the mean value of BWP<br>index increases, this point will not participate in the subsequent center point<br>selection. Compare the change of the mean value<br>of BWP index of data points before and after the<br>pre-classification. If the mean value of BWP<br>index increases, this point will 3.4 Basic Ideas of GDD-K-Means Clustering<br>pre-classification. If the mean value of BWP<br>index increases, this point will be used as a new<br>clustering center, and the generation of<br>clustering center will cause changes of data

index increases, this point will be used as a new<br>clustering center, and the generation of<br>clustering center will cevantly be divided into the<br>points, which will eventually be divided into the<br>Category of the center point

## **Algorithm**

clustering center, and the generation of<br>clustering center will cause changes of data<br>points, which will eventually be divided into the<br>category of the center point closest to itself.<br>Therefore, the center point index need clustering center will cause changes of data<br>points, which will eventually be divided into the<br>category of the center point iclosest to itself.<br>Therefore, the center point index needs to be<br>updated every time a new center points, which will eventually be divided into the<br>category of the center point closest to itself.<br>Therefore, the center point index needs to be<br>updated every time a new center point is<br>generated. If the BWP indicator becom category of the center point closest to itself.<br>Therefore, the center point index needs to be<br>updated every time a new center point is<br>generated. If the BWP indicator becomes smaller<br>or no data points exist, the selection Theretore, the center point index needs to be<br>updated every time a new center point is<br>generated. If the BWP indicator becomes smaller<br>or no data points exist, the selection of the center<br>point will stop.<br>3.4 Basic Ideas o updated every time a new center point is<br>generated. If the BWP indicator becomes smaller<br>or no data points exist, the selection of the center<br>point will stop.<br>3.4 Basic Ideas of GDD-K-Means Clustering<br>Algorithm<br>Firstly, th generated. If the BWP indicator becomes smaller<br>or no data points exist, the selection of the center<br>point will stop.<br>3.4 Basic Ideas of GDD-K-Means Clustering<br>Algorithm<br>Firstly, the grid filtering method is used for<br>prel or no data points exist, the selection of the center<br>point will stop.<br>3.4 Basic Ideas of GDD-K-Means Clustering<br>Algorithm<br>Firstly, the grid filtering method is used for<br>preliminary screening of the data set, and the<br>data **3.4 Basic Ideas of GDD-K-Means Clustering**<br>**Algorithm**<br>**Algorithm**<br>**Firstly, the grid filtering method is used for**<br>**Firstly, the grid filtering method is used for**<br>**preliminary screening of the data set, and the**<br>data w **3.4 Basic Ideas of GDD-K-Means Clustering Algorithm**<br>**Algorithm**<br>**Firstly, the grid filtering method is used for preliminary screening of the data set, and the data whose density is less than a specific threshold is divi 3.4 Basic Ideas of GDD-K-Means Clustering**<br>**Algorithm**<br>**Firstly, the grid filtering method is used for**<br>**preliminary screening of the data set, and the**<br>**data** whose density is less than a specific<br>**threshold** is divided Algorithm<br>Firstly, the grid filtering method is used for<br>preliminary screening of the data set, and the<br>data whose density is less than a specific<br>threshold is divided into candidate subsets. Then,<br>the classical density-b Firstly, the grid filtering method is used for<br>preliminary screening of the data set, and the<br>data whose density is less than a specific<br>threshold is divided into candidate subsets. Then,<br>the classical density-based LOF a preliminary screening of the data set, and the<br>data whose density is less than a specific<br>threshold is divided into candidate subsets. Then,<br>the classical density-based LOF algorithm is<br>used to accurately remove the outlie data whose density is less than a specific<br>threshold is divided into candidate subsets. Then,<br>the classical density-based LOF algorithm is<br>used to accurately remove the outliers. When the<br>outliers are removed from the dat threshold is divided into candidate subsets. Then,<br>the classical density-based LOF algorithm is<br>used to accurately remove the outliers. When the<br>outliers are removed from the data set, the point<br>with the largest weight is the classical density-based LOF algorithm is<br>used to accurately remove the outliers. When the<br>outliers are removed from the data set, the point<br>with the largest weight is selected as the initial<br>center point, and then n ce butters are removed from the data set, the point<br>with the largest weight is selected as the initial<br>center point, and then n center points are<br>selected successively according to the central<br>point index. When  $n+2$  center p center point, and then n center points are<br>selected successively according to the central<br>point index. When  $n+2$  center points are<br>generated, the average value of BWP index<br>decreases, then the selection is stopped and  $n$ selected successively according to the central<br>point index. When  $n+2$  center points are<br>generated, the average value of BWP index<br>decreases, then the selection is stopped and  $n+1$ <br>center point is used as the initial clu point index. When n+2 center points are<br>generated, the average value of BWP index<br>decreases, then the selection is stopped and n+1<br>center points are obtained. Finally, the generated<br>center point is used as the initial clus generated, the average value of BWP index<br>decreases, then the selection is stopped and n+1<br>center points are obtained. Finally, the generated<br>center point is used as the initial clustering<br>center to execute the k-means clu

 $\sum_{n=1}^{\infty} |x_p^{(c)}(14)|$  Anaconda3&Spyder3 as development decreases, then the selection is stopped and n+1<br>center points are obtained. Finally, the generated<br>center point is used as the initial clustering<br>center to execute the k-means clustering<br>algorithm, and the final clusterin center points are obtained. Finally, the generated<br>center point is used as the initial clustering<br>center to execute the k-means clustering<br>algorithm, and the final clustering result is<br>obtained to end the operation.<br>4. **Ex** center point is used as the initial clustering<br>center to execute the k-means clustering<br>algorithm, and the final clustering result is<br>obtained to end the operation.<br>4. Experimental Results and Analysis<br>4.1 Experimental Env argorium, and the mail custering result is<br>obtained to end the operation.<br>4. Experimental Results and Analysis<br>4.1 Experimental Environment<br>of K-means, k-means ++, canopy + K-means,<br>GDD-K-Means algorithm is: Core i3720M<br>(1 obtained to end the operation.<br>
4. Experimental Results and Analysis<br>
4.1 Experimental Environment<br>
The experimental environment building platform<br>
of K-means, k-means ++, canopy + K-means,<br>
GDD-K-Means algorithm is: Core **4. Experimental Results and Analysis**<br> **4.1 Experimental Environment**<br>
The experimental environment building platform<br>
of K-means, k-means  $++$ , canopy  $+$  K-means,<br>
GDD-K-Means algorithm is: Core i3720M<br>
(1.80GHz) proces

Let us a verify the experimental environment<br>
(1.80GHz) processor, The<br>  $-\frac{x_i}{n}\sum_{p=1}^{n} ||x_p(c) - x_i(t)||^2$ <br>
(1.80GHz) processor, The<br>  $-\frac{x_i}{n}||x_p(t) - x_i(t)||^2$ <br>
(1.80GHz) processor, The<br>  $-\frac{x_i}{n}||x_p(t) - x_i(t)||^2$ <br>
(1.80GHz) processor object i of class j, to be the average value of the<br>
distance between this data object and other data<br>
of K-means, k-means ++, ca<br>
objects of class j. The calculation formula is as<br>  $\frac{1}{2} \int_{1 \leq c \leq k, c \neq j} \left( \frac{1}{n} \right$ distance between this data object and other data<br>
objects of class j. The calculation formula is as<br>  $b(j,i) = \min_{1 \leq i \leq k, c \neq j} \left(\frac{1}{n}\right)^n \sum_{p=1}^c ||x_p(c)$ <br>  $w(j,i) = \left(\frac{1}{n_j - 1}\right)^2 \sum_{p=1}^j ||x_p(j) - x_i(j)||^2$ <br>  $w(j,i) = \left(\frac{1}{n_j - 1}\right$ 4. Experimental Results and Analysis<br>
4.1 Experimental Environment<br>
The experimental environment building platform<br>
of K-means, k-means  $++$ , canopy  $+$  K-means,<br>
GDD-K-Means algorithm is: Core i3720M<br>
(1.80GHz) processor, **4.1 Experimental Environment**<br>The experimental environment building platform<br>of K-means, k-means  $++$ , canopy  $+$  K-means,<br>GDD-K-Means algorithm is: Core i3720M<br>(1.80GHz) processor, The algorithms are<br>developed using pyth 4.1 Experimental Environment<br>The experimental environment building platform<br>of K-means, k-means  $++$ , canopy  $+$  K-means,<br>GDD-K-Means algorithm is: Core i3720M<br>(1.80GHz) processor, The algorithms are<br>developed using python The experimental environment building platform<br>of K-means, k-means  $++$ , canopy  $+$  K-means,<br>GDD-K-Means algorithm is: Core i3720M<br>(1.80GHz) processor, The algorithms are<br>developed using python3.6 language and<br>Anaconda3&Sp of K-means, k-means  $++$ , canopy  $+$  K-means,<br>GDD-K-Means algorithm is: Core i3720M<br>(1.80GHz) processor, The algorithms are<br>developed using python3.6 language and<br>Anaconda3&Spyder3 as development tools.<br>**4.2 Experimental D** 

Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>performance of the algorithm in different data candidate subset of the data s<br>scales and different data structures. The data will<br>be ve Sournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>performance of the algorithm in different data<br>scales and different data structures. The data will<br>be verified by selecting a synthetic be verified by selecting a synthetic leads of the algorithm in different data<br>scales and different data structures. The data will<br>be verified by selecting a synthetic UCI data set.<br>the shold of grid density. This a databas Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>performance of the algorithm in different data candidate subset of the data set<br>scales and different data structures. The data will<br>be Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
performance of the algorithm in different data<br>
scales and different data structures. The data will<br>
be verified by selecting a synthe Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 N<br>
performance of the algorithm in different data candidate subset of<br>
scales and different data structures. The data will<br>
be verified by selecti Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 20<br>
performance of the algorithm in different data<br>
scales and different data structures. The data will<br>
be verified by selecting a synthet Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 202-<br>performance of the algorithm in different data candidate subset of the dat<br>scales and different data structures. The data will calcula Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
performance of the algorithm in different data<br>
scales and different data structures. The data will<br>
calculates the number of grid<br>
b Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
performance of the algorithm in different data<br>
scales and different data structures. The data will<br>
be verified by selecting a synth Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
performance of the algorithm in different data candidate subset of the data secales and different data structures. The data will<br>
let Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No.<br>
performance of the algorithm in different data<br>
candidate subset of t<br>
scales and different data structures. The data will<br>
calculates the nu performance of the algorithm in different<br>scales and different data structures. The dat<br>be verified by selecting a synthetic UCI dat<br>It is a database proposed by the Universi<br>California for machine learning.<br>experimental From the different data structures. The data will<br>
verified by selecting a synthetic UCI data set.<br>
were trieshold of grid<br>
sa a database proposed by the University of follows:<br>
ifornia for machine learning. The **Table 2.** Experimental data sets factor clear enassinearions<br>so the quality of clustering can be directly<br>observed. Experimental data will select Seeds<br>Wine two data sets of different data size<br>detection, Dataset1, Dataset2 are used





enceution, Datasett, Datasetz are used to the density offficient four data sets. Table 1 describes the density coefficient density the deform of the density coefficient of the density of the Datasets Samples Attributes Cat represent ure nout data sets. Table 1 describes ure<br>data sets.<br>data sets.<br>Table 1. Experimental Data Description<br>data sets.<br>Table 1. Experimental Data Description<br>data sets.<br>Through given the density thressets<br> $\frac{3}{210}$  of the feature is at the same end of the total and the set and the set and the set and the to the them the distingent than the density threshold, we defined a dense subset. If the d<br> **Datasets** Samples Attributes Categorie The transfer of the date asset.<br>
Table 1. Experimental Data Description<br>
Datasets Samples Attributes Categories<br>
Seeds 210 7 3<br>
Seeds 210 7 3<br>
The mean subset, which dates uses the whole through seeds that are to be used a **Table 1. Experimental Data Description**<br>
Datasets Samples Attributes Categories a dense subset, which does<br>
Seeds 210 7 3 candidate subset. Through grid<br>
set is initially screened to effect<br>
Different feature values in a **Exerces Exerces E** Seeds 210  $\frac{3}{4}$  candidate subset. Inrough gr<br>
wine 178 13 4 candidate subset. Inrough gr<br>
and most dense part of the data<br>
data set to be used as much<br>
Different feature values in a data set often have<br>
different dime **Example 18** and the most discussing the Min-Max Scaling most discussion of the data set part of the second and different feature values in a data set often have data set to be used as different dimensions. When different 4.3 Data Processing<br>
most dense part of the<br>
different feature values in a data set often have<br>
different dimensions. When different features<br>
points are more accuracy<br>
are listed together, small data in absolute values<br>
i **4.3 Data Processing**<br> **Example 1.4 Confident** intensions. When different features through density-based detection<br>
different dimensions. When different features points are more accuracy<br>
of the listed together, small d Different feature values in a data set often have<br>
different dimensions. When different features points are more accurately<br>
are listed together, small data in absolute values improves the accuracy of the<br>
are listed in t different dimensions. When different features points are more are listed together, small data in absolute values improves the accura will be ignored by big data in data mining reduces the running processing due to the dif are listed together, small data in absolute values<br>
process improves the accuracy of<br>
will be ignored by big data in data mining reduces the running tim<br>
processing due to the different expression ways<br>
of the features th

algorithm, the impact of carry out left<br>iminated. The data were method effect<br>ne Min-Max Scaling method. sets. Because<br>with a feature, by traversing the sparse g<br>ture vector, Max and Min are feature of th<br>-min is used as that each feature reasonably participates in the<br>execution of the algorithm, the impact of carr<br>dimension is eliminated. The data were metl<br>normalized using the Min-Max Scaling method. sets.<br>For the data set with a featur

$$
x_{normalization} = \frac{x_i - \text{Min}(x_i)}{\text{Max}(x_i) - \text{Min}(x_i)} \qquad (16) \qquad \text{shortence}
$$

For the data set with a feature, by traversing<br>
each data in the feature vector, Max and Min are<br>
recorded, and max-min is used as the base (that<br>
is, Min=0, Max=1) for data normalization cluster center, and the rem-<br>
is, each data in the feature vector, Max and Min are<br>
recorded, and max-min is used as the base (that<br>
is, Min=0, Max=1) for data normalization cluster center, and the removal<br>
is, Min=0, Max=1) for data normalization cluster recorded, and max-min is used as the base (that density, the is, Min=0, Max=1) for data normalization cluster cervicessing. The calculation formula is as follows: selection of  $x_i - \text{Min}(x_i)$  the selection of  $x_i - \text{Min}(x_i)$  processing. The calculation formula is as follows: selection of the cluster<br>  $x_{normalization} = \frac{x_i - \text{Min}(x_i)}{\text{Max}(x_i) - \text{Min}(x_i)}$  (16) shortend the execution<br>
the selection of<br>
the selection of<br>
the selection of<br>
the selection of<br>
the  $x_{normalization} = \frac{1}{\text{Max}(xi) - \text{Min}(xi)}$  (16) shortened the execution time<br>thereby significantly impreced. (e.g., Figure 3. Cluster<br>for the execution in the stream of the sets).<br>In order to perform effective cluster evaluation<br>on clu **4.4 Experimental Evaluation**<br>
In order to perform effective cluster evaluation<br>
commonly used evaluation criteria in data<br>
mining field: BWP index, Rand index, contour<br>
coefficient, recall rate, accuracy rate and F-<br>
mea Filtering algorithm screens the preliminary  $\frac{12}{100}$ <br>
Commonly used evaluation criteria in data<br>  $\frac{1}{200}$ <br>  $\frac$ 

## **Results**

2959-0620) Vol. 2 No. 3, 2024 119<br>candidate subset of the data set in Table 2, and<br>calculates the number of grid divisions and the<br>threshold of grid density. The results are as<br>follows:<br>**Table 2. Experimental Data Statisti** 2959-0620) Vol. 2 No. 3, 2024 119<br>
candidate subset of the data set in Table 2, and<br>
calculates the number of grid divisions and the<br>
threshold of grid density. The results are as<br>
follows:<br> **Table 2. Experimental Data St** 2959-0620) Vol. 2 No. 3, 2024 119<br>
candidate subset of the data set in Table 2, and<br>
calculates the number of grid divisions and the<br>
threshold of grid density. The results are as<br>
follows:<br> **Table 2. Experimental Data St** follows: **Table 2. Experimental Data Statistics**<br> **Table 2. Experim** 

: 2959-0620) Vol. 2 No. 3, 2024 119			
candidate subset of the data set in Table 2, and			
calculates the number of grid divisions and the			
threshold of grid density. The results are as			
follows:			
<b>Table 2. Experimental Data Statistics</b>			
Datasets		Number of Grid density	The threshold
	meshing	threshold	is partially
			reached
Seeds	$9*9$	3	5
Wine	$9*9$	3	
If the density coefficient is lower than the			
density threshold, we define it as a candidate			
anomaly subset. If the density coefficient is			
higher than the density threshold, we define it as			
a dense subset, which does not fit into the			

a dense subset, which does not fit into the candidate subset. Through grid filtering, the data set is initially screened to effectively remove the So the quality of clustering can be directly<br>
Datasets<br>
Wine two data sets of different data size<br>
Wine two data sets of different data size<br>
detection, Dataset1, Dataset2 are used to<br>
The density coefficient<br>
basic chara Wine two data sets of different discusses<br>
Wine two data sets of different data size<br>
detection, Dataset1, Dataset2 are used to<br>
represent the four data sets. Table 1 describes the<br>
data sets.<br> **Table 1. Experimental Data** whe two causes of uniterint data size<br>
detection, Dataset1, Dataset2 are used to Wine 9% and<br>
represent the four data sets. Table 1 describes the<br>
density threshold, we defin<br>
density the shold, we defin<br>
data sets.<br> **Tab** 2959-0620) Vol. 2 No. 3, 2024 119<br>
candidate subset of the data set in Table 2, and<br>
calculates the number of grid divisions and the<br>
threshold of grid density. The results are as<br>
follows:<br> **Table 2. Experimental Data St** candidate subset of the data set in Table 2, and<br>calculates the number of grid divisions and the<br>threshold of grid density. The results are as<br>follows:<br>**Table 2. Experimental Data Statistics**<br> $\begin{array}{r} \n\text{Table 2. Experimental Data Statistics} \\
\text{Substituting the$ canonical subset of the data set in Table 2, and<br>calculates the number of grid divisions and the<br>threshold<br>of grid density. The results are as<br>follows:<br>**Table 2. Experimental Data Statistics**<br> $\begin{array}{r} \n\text{Table 2. Experimental Data Statistics} \\
\text{The threshold} \\$ calculates the number of grid divisions and the<br>threshold of grid density. The results are as<br>follows:<br>Table 2. Experimental Data Statistics<br>meshing threshold is partially<br>meshing threshold is partially<br>seeds  $9*9$  3 5<br>Wi threshold of grid density. The results are as<br>follows:<br>**Table 2. Experimental Data Statistics**<br>Datasets Number of Grid density  $\begin{bmatrix} \text{The threshold} \\ \text{is partially} \\ \text{reached} \end{bmatrix}$ <br>signarially<br>seeds  $9*9$  3 5<br>The density coefficient is lowe Table 2. Experimental Data Statistics<br>
Datasets<br>
Number of Grid density<br>
meshing<br>
meshing<br>
threshold<br>
set is partially<br>
reached<br>
Seeds<br>  $9*9$ <br>  $3$ <br>  $5$ <br>
The density coefficient is lower than the<br>
density threshold, we def Table 2. Experimental Data statistics<br>
Datasets Number of Grid density The threshold<br>
meshing threshold is partially<br>
reached<br>
Seeds  $9*9$  3 5<br>
The density coefficient is lower than the<br>
density threshold, we define it as Datasets Number of Grid density<br>
meshing threshold is partially<br>
seeds  $9*9$  3 5<br>
The density coefficient is lower than the<br>
density threshold, we define it as a candidate<br>
anomaly subset. If the density coefficient is<br>
h Datasets meshing threshold reached<br>
Seeds  $9*9$  3 5<br>
Wine  $9*9$  3 5<br>
If the density coefficient is lower than the<br>
density threshold, we define it as a candidate<br>
anomaly subset. If the density coefficient is<br>
higher than Fraction Seeds  $9*9$  3 5<br>
Wine  $9*9$  3 5<br>
If the density coefficient is lower than the<br>
density threshold, we define it as a candidate<br>
anomaly subset. If the density coefficient is<br>
higher than the density threshold, we Seeds  $9*9$  3 5<br>
Wine  $9*9$  3 5<br>
If the density coefficient is lower than the<br>
density threshold, we define it as a candidate<br>
anomaly subset. If the density coefficient is<br>
higher than the density threshold, we define it **EXECUTE:** The density coefficient is lower than the density threshold, we define it as a candidate anomaly subset. If the density coefficient is higher than the density threshold, we define it as a dense subset, which do It the density coefficient is lower than the<br>density threshold, we define it as a candidate<br>anomaly subset. If the density coefficient is<br>higher than the density threshold, we define it as<br>a dense subset, which does not fi density threshold, we define it as a candidate<br>anomaly subset. If the density coefficient is<br>higher than the density threshold, we define it as<br>a dense subset, which does not fit into the<br>candidate subset. Through grid fit anomaly subset. It the density coefficient is<br>higher than the density threshold, we define it as<br>a dense subset, which does not fit into the<br>candidate subset. Through grid filtering, the data<br>to set is initially screened t higher than the density threshold, we define it as<br>a dense subset, which does not fit into the<br>candidate subset. Through grid filtering, the data<br>set is initially screened to effectively remove the<br>most dense part of the d a dense subset, which does not fit into the<br>candidate subset. Through grid filtering, the data<br>set is initially screened to effectively remove the<br>most dense part of the data set and reduce the<br>data set to be used as smuch

execution of the algorithm, the impact of carry out removal proced<br>imension is eliminated. The data were method effectively red<br>normalized using the Min-Max Scaling method. sets. Because the local<br>for the data set with a dimension is eliminated. The data were method effectively reduces to<br>normalized using the Min-Max Scaling method<br>sets. Because the local density<br>for the data in the feature vector, Max and Min are feature of the cluster c normalized using the Min-Max Scaling method. sets. Because the local densit<br>For the data set with a feature, by traversing the sparse grid is small, whi<br>each data in the feature vector, Max and Min are feature of the clus candidate subset. Through grid filtering, the data<br>set is initially screened to effectively remove the<br>most dense part of the data set and reduce the<br>data set to be used as much as possible. Then,<br>through density-based det set is initially screened to effectively remove the<br>most dense part of the data set and reduce the<br>data set to be used as much as possible. Then,<br>through density-based detection method, noise<br>points are more accurately scr most dense part of the data set and reduce the<br>data set to be used as much as possible. Then,<br>through density-based detection method, noise<br>points are more accurately screened, which<br>improves the accuracy of the algorithm, data set to be used as much as possible. Then,<br>through density-based detection method, noise<br>points are more accurately screened, which<br>improves the running time and improves the<br>efficiency of the algorithm.<br>In the second through density-based detection method, noise<br>points are more accurately screened, which<br>improves the accuracy of the algorithm, greatly<br>reduces the running time and improves the<br>efficiency of the algorithm.<br>In the second points are more accurately screened, which<br>improves the accuracy of the algorithm, greatly<br>reduces the running time and improves the<br>efficiency of the algorithm.<br>In the second stage, based on the density-based<br>outlier dete mproves the accuracy of the algorithm, greatly<br>reduces the running time and improves the<br>efficiency of the algorithm.<br>In the second stage, based on the density-based<br>abnormal data set from the candidate subset and<br>carry ou reduces the running time and improves the efficiency of the algorithm.<br>In the second stage, based on the density-based outlier detection method, we finally calculate the abhormal data set from the candidate subset and carr efficiency of the algorithm.<br>In the second stage, based on the density-based<br>outlier detection method, we finally calculate the<br>abnormal data set from the candidate subset and<br>abnormal data set from the candidate subset a In the second stage, based on the density-based<br>outlier detection method, we finally calculate the<br>abnormal data set from the candidate subset and<br>carry out removal processing. The grid filtering<br>method effectively reduce outlier detection method, we finally calculate the<br>abnormal data set from the candidate subset and<br>carry out removal processing. The grid filtering<br>method effectively reduces the number of data<br>sets. Because the local dens results).



Journal of Intelligence and Knowledge Engineering (ISSN: 2959-C<br>
clusters the initial central point of the cluster, clustering results of<br>
improves the tightness within the cluster, clustering algorithm and<br>
eliminates the Iournal of Intelligence and Knowledge Engineering (ISSN: 2959-C<br>
clusters the initial central point of the cluster, clustering results of<br>
improves the tightness within the cluster, clustering algorithm and<br>
eliminates th 120 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-06<br>
clusters the initial central point of the cluster, clustering results of the improves the tightness within the cluster, clustering algorithm and<br>
elimi 120 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-<br>
clusters the initial central point of the cluster, clustering results of<br>
improves the tightness within the cluster, clustering algorithm and<br>
eliminates 120 Journal of Intelligence and Knowledge Engineering (ISSN:<br>
clusters the initial central point of the cluster, clustering results<br>
improves the tightness within the cluster, clustering algorithm<br>
eliminates the randomne 120 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620<br>
clusters the initial central point of the cluster, clustering results of the<br>
improves the tightness within the cluster, clustering algorithm and the 120 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-06<br>
clusters the initial central point of the cluster, clustering results of the<br>
improves the tightness within the cluster, clustering algorithm and the<br> 120<br>
algorithms, and the clustering effect can be eliminates the initial central point of the cluster, clustering results of the<br>
improves the tightness within the cluster, clustering algorithm and the<br>
eliminates the ran evaluated. The experimental results clearly indicate that, in the experimental results control in this paper and the selection, and makes the clustering also within the cluster, clustering algorithm and the election, and makes the c Figure 3, the BWP index, have all also entired the substrate the significantly better than K-Means also entired the Substrate of the substrate of the initial center and makes the cluster can be divided as shown in Figure 4 120 Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620)<br>
clusters the initial central point of the cluster, clustering results of the<br>
improves the tightness within the cluster, clustering algorithm and th Clusters the initial central point of the cluster, clustering results<br>improves the transmess of the initial center algorithms introduced<br>election, and makes the cluster can be divided as shown in Figure<br>in one stage. Figur

clusters the mittal central point of the cluster, clustering results of the eimproves the rightmass within the cluster, clustering algorithm and the eimental selection, and makes the other can be divided as shown in Figur mproves the tightness within the cluster, clustering algorithm and the eliminates the randomness of the initial center algorithms introduced previous in one stage. Figure 3 shows the experimental results of the other thre eliminates the randomness of the initial center<br>
selection, and makes the cluster can be divided<br>
in one stage. Figure 4 and Figure 4 and Figure 4 and Figure 3, the mproved clustering algorithm and<br>
the performance indica selection, and makes the cluster can be divided<br>in one stage. Figure 3 shows the experimental<br>in one stage. Figure 3 shows the experimental<br>the performance indicators of the other three<br>algorithms, and the clustering effec algorithm randomly selects the initial center of distance in the energiential results of the intervention of the other three  $\frac{1}{2}$  and index, accuracy rate and recall rate of clustering results obtained by the GPD-K-m results of the improved clustering algorithm and<br>algorithms, and the clustering effect can be<br>algorithms, and the clustering effect can be<br>evaluated.<br>The experimental results clearly indicate that, in<br>Figure 3, the BWP in factor of distance in the selection of the center point, the generated data with far distance in the selection of the center point, the generated data with far distance from the center point, the generated data with far di algorithms, and the clustering effect can be<br>
avalated.<br>
The experimental results clearly indicate that, in<br>
Figure 3, the BWP index, contour coefficient,<br>
Rand index, accuracy rate and recall rate of<br>
clustering results Experimental results clearly indicate that, in<br>
The experimental results clearly indicate that, in<br>
Figure 3, the BWP index, contour coefficient,<br>
Rand index, accuracy rate and recall rate of<br>
clustering results obtained The experimental results clearly indicate that, in<br>
Figure 3, the BWP index, contains and read red and read red of<br>
clustering results obtained by the GPD-K-means<br>
algorithm proposed in this paper are<br>
significantly bette Figure 3, the BWP index, contour coefficient,<br>
Rand index, accuracy rate and recall rate of<br>
algorithm proposed in this paper are<br>
significantly better than K-Means algorithm, K-<br>
Means  $++$  algorithm and Canopy  $+$  means<br> Rand mdex, accuracy rate and recell rate of<br>
algorithm proposed in this paper are<br>
algorithm proposed in this paper are<br>
algorithm randomly selects the initial center<br>
Means  $++$  algorithm and Canopy  $+$  means<br>
algorithm ra clustering results obtained by the GPD-K-means<br>
algorithm proposed in this paper are<br>
significantly better than K-Means algorithm, K-<br>
Means  $++$  algorithm and Canopy  $+$  means<br>
algorithm. The reason is that the traditiona algorithm proposed in this paper are significantly better than K-Means algorithm. The reason is that the traditional<br>algorithm and Canopy + means<br>algorithm The reason is that the traditional<br>algorithm randomly selects the Experimently better than K-Means algorithm, K-<br>
Means  $++$  algorithm and Canopy  $+$  means<br>
algorithm The reason is that the traditional<br>
algorithm randomly selects the initial center<br>
point. Although K-Means  $++$  incorporat Means  $++$  algorithm and Canopy  $+$  means<br>algorithm. The reason is that the traditional<br>algorithm randomly selects the initial center<br>point, Although K-Means  $++$  incorporates the<br>factor of distance in the selection of the algorithm. The reason is that the traditional<br>algorithm randomly selects the initial center<br>factor of distance in the selection of the center<br>point, the generated data with far distance from<br>the center point, the generate algorithm randomly selects the initial center<br>
fractor of distance the<br>
point, Although K-Means  $+$  incorporates the<br>
point, the generated data with far distance from<br>
the center point is more likely to be selected as<br>
th point. Although K-Means  $+$  incorporates the<br>factor of distance in the selection of the center<br>the center point, is more likely to be selected as<br>the enert point is more likely to be selected as<br>the next clustering center and F value are greatly improved, which is effective improvement of<br>clustering results of the influence in the effective improvement of the effective improvement of<br>and F value are greatly improvement of<br>the effective impr point, the generated data with far distance from<br>the center point is more likely to be selected as<br>still random. In the four data sets, the proposed<br>algorithm has the best performance in the<br>evaluation index, and the BWP i the center point is more likely to be selected as<br>
the next clustering center, and the selection is<br>
algorithm has the best performance in the<br>
evaluation index, and the BWP index is<br>
obviously higher than other clusterin the next clustering center, and the selection is<br>
still random. In the four data sets, the proposed<br>
algorithm has the best performance in the<br>
evaluation index, and the BWP index is<br>
obviously higher than other clusterin Still random. In the four data sets, the proposed<br>algorithm has the best performance in the WP index is<br>obviously higher than other clustering<br>algorithms, which indicates that the intra-class<br>precision is stronger and the algorithm has the best performance in the<br>
evaluation index, and the BWP index is<br>
algorithms, which indicates that the intra-class<br>
algorithms, which indicates that the inter-class<br>
separation is stronger and the inter-c evaluation index, and the BWP index is<br>obviously higher than other clustering<br>algorithms, which indicates that the intra-class<br>generation is stronger and the inter-class<br>separation is better. Through performance<br>indicators by the detection is stronger and the interactions and the interaction of stronger and the interaction of th algorithms, which indicates that the intra-class<br>
separation is storter. Through performance<br>
indicators, it is found that accuracy, recall rate<br>
indicators, it is found that accuracy, recall rate<br>
and F value are greatly precision is stronger and the inter-class<br>
indicators, it is found hat accuracy, recall rate<br>
indicators, it is found that accuracy, recall rate<br>
and F value are greatly improved, which<br>
indicates that the effective impro separation is better. Through performance<br>
indicators, it is found that accuracy, recall rate<br>
and F value are greatly improved, which<br>
indicates that the effective improved, which<br>
indicates that the effective improvemen Indicators, it is found that accuracy, recall rate<br>
ind F value are greatly improved, which<br>
indicates that the effective improvement of<br>
clustering results depends on the effective From the observation of dat<br>
indicators, and F value are greatly improved, which<br>
indicates that the effective improvement of<br>
clustering results depends on the effective From the observation of dat<br>
selection of initial clustering enter by central<br>
indicators, r Indicates that the effective improvement of<br>
clustering results depends on the effective<br>
selection of initial clustering center by central<br>
indicators, reasonable selection of cluster<br>
indicators, reasonable selection of clustering results depends on the effective<br>
selection of diate<br>
selection of climal clustering centre by central<br>
indicators, reasonable selection of cluster<br>
number and removal of the influence of noise<br>
effect in this p selection of initial clustering center by central<br>
indicators, reasonable selection of cluster focusing and small span, are<br>
number and removal of the influence of noise<br>
points. It indicates that Canopy + K-Means<br>
still a mdicators, reasonable selection of cluster<br>
points. It indicates that Canopy + K-Means<br>
effect in this paper is good.  $\mu$ <br>
edustering algorithm can obtain better clustering<br>
edustering algorithm can obtain better clusteri number and removal of the influence of noise<br>
effect in this paper is<br>
colustreing algorithm can obtain better clustering<br>
clustering dignotine and that in better clustering<br>
effect. Secondly, the data is de-noised in the points. It indicates that Canopy + K-Means still a few singular values, clustering algorithm can obtain better clustering quality is high, and canceflect. Secondly, the data is de-noised in the observing Figure 5, we can c clustering algorithm can obtain better clustering<br>equality is high, and<br>effect. Secondly, the data is de-noised in the<br>obviously not applicate<br>the calculation process, which consumes a<br>discribution of<br>erration amount of ti effect. Secondly, the data is de-noised in the obviously not a<br>early stage, and the initial center is selected in observing Figure<br>the calculation process, which consumes a linear distributio<br>certain amount of time, but th early stage, and the initial center is selected in observing Figure 5, we can clear<br>the calculation process, which consumes a linear distribution of data in<br>certain amount of time, but the determination of small span of co

ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>clustering results of the GDD-K-Means<br>clustering algorithm and the three clustering<br>algorithms introduced previously are compared,<br>as shown in Figure 4 and Figure 5. ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>clustering results of the GDD-K-Means<br>clustering algorithm and the three clustering<br>algorithms introduced previously are compared,<br>as shown in Figure 4 and Figure 5. ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>clustering results of the GDD-K-Means<br>clustering algorithm and the three clustering<br>algorithms introduced previously are compared,<br>as shown in Figure 4 and Figure 5. ge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>clustering results of the GDD-K-Means<br>clustering algorithm and the three clustering<br>algorithms introduced previously are compared,<br>as shown in Figure 4 and Figure 5.



**Visualization**



**Visualization**

From the observation of data distribution in Figure 4, we can see that the data set has poor focusing and small span, and the cluster center of the intervalse of the paper is good. Although there are still a few singular strikted a few singular values, the cluster center quality is high, and canopy algorithm is observing Figure 5, we can see that the cluster center quality is high, and canopy algorithm is polyeuring the cluster center qua Figure 5. Wine Data Clustering Class Result<br>Figure 5. Wine Data Clustering Class Result<br>Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is good. Although th Figure 5. Wine Data Clustering Class Result<br>Figure 5. Wine Data Clustering Class Result<br>Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is good. Although th Figure 5. Wine Data Clustering Class Result<br>Figure 5. Wine Data Clustering Class Result<br>Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is good. Although th Figure 5. Wine Data Clustering Class Result<br>Figure 5. Wine Data Clustering Class Result<br>Visualization<br>From the observation of data distribution in<br>Figure 4, we can see that the data set has poor<br>focusing and small span, a Figure 5. Wine Data Clustering Class Result<br>Figure 5. Wine Data Clustering Class Result<br>Figure 5. Wine Data Clustering Class Result<br>From the observation of data distribution in<br>Figure 4, we can see that the data set has p Figure 5. Wine Data Clustering Class Result<br>Visualization<br>From the observation of data distribution in<br>Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is goo Figure 5. Wine Data Clustering Class Result<br>Visualization<br>From the observation of data distribution in<br>Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is goo **Visualization**<br> **From the observation of data distribution in**<br>
Figure 4, we can see that the data set has poor<br>
focusing and small span, and the clustering<br>
effect in this paper is good. Although there are<br>
still a few s From the observation of data distribution in<br>Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is good. Although there are<br>still a few singular values, the clu Figure 4, we can see that the data set has poor<br>focusing and small span, and the clustering<br>effect in this paper is good. Although there are<br>still a few singular values, the cluster center<br>still a few singular values, the focusing and small span, and the clustering<br>effect in this paper is good. Although there are<br>still a few singular values, the cluster center<br>quality is high, and canopy algorithm is<br>obviously not applicable to this data se effect in this paper is good. Although there are<br>still a few singular values, the cluster center<br>quality is high, and canopy algorithm is<br>obviously not applicable to this data set. By<br>observing Figure 5, we can clearly see still a few singular values, the cluster center<br>quality is high, and canopy algorithm is<br>obviously not applicable to this data set. By<br>observing Figure 5, we can clearly see that, the<br>linear distribution of data integratio quality is high, and canopy algorithm is<br>obviously not applicable to this data set. By<br>observing Figure 5, we can clearly see that, the<br>linear distribution of data integration and the<br>small span of coordinate interval can obviously not applicable to this data set. By<br>observing Figure 5, we can clearly see that, the<br>linear distribution of data integration and the<br>small span of coordinate interval can be<br>considered as concentrated distributio observing Figure 5, we can clearly see that, the linear distribution of data integration and the small span of coordinate interval can be considered as concentrated distribution. For linear distribution data, through the a

By examining the visualization chart of the Cluster Head selection fra<br>
above clustering results, it can be clearly also examined K-m<br>
observed that the improved GDD-K-Means<br>
clustering algorithm proposed in this paper and Fournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 202<br>
By examining the visualization chart of the Cluster Head selection<br>
above clustering results, it can be clearly using enhanced F<br>
obser Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 N<br>
By examining the visualization chart of the Cluster Head<br>
above clustering results, it can be clearly using enhar<br>
observed that the improved GD Cournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
By examining the visualization chart of the Cluster Head selection<br>
above clustering results, it can be clearly using enhanced K-<br>
obs Forms With the rapid increase of Internet data, the angle 2024, 13(18): 267-3677, and the rapid increase of Internet data clustering and has a cluster performs well in data clustering and has election fraction of Angle and Sournal of Intelligence and Knowledge Engineering (ISSN: 2959-0<br>
By examining the visualization chart of the<br>
above clustering results, it can be clearly<br>
observed that the improved GDD-K-Means<br>
methods of the method of th **5. Closing Remarks**<br> **5. Closing Remarks** and Knowledge Engineering (IS<br> **By** examining the visualization chart of the<br>
above clustering results, it can be clearly<br>
observed that the improved GDD-K-Means<br>
clustering algor Using the rapid increase of Internet data, the same of order the rapid increase of the same of the same of the same of the clustering results, it can be clearly above clustering results, it can be clearly using enhanced K-Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
By examining the visualization chart of the Cluster Head selection<br>
above clustering results, it can be clearly using enhanced K<br>
obse

Journal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 2024<br>
By examining the visualization chart of the Cluster Head selection fra<br>
above clustering results, it can be clearly using enhanced K-m Dournal of Intelligence and Knowledge Engineering (ISSN: 2959-0620) Vol. 2 No. 3, 20<br>
By examining the visualization chart of the Cluster Head selection<br>
above clustering results, it can be clearly using enhanced<br>
observed By examine the visualization chart of the Cluster Head selection fractions<br>above clustering results, it can be clearly using enhanced K-m<br>observed that the improved GDD-K-Means mechanism. Journal of Am<br>clustering algorithm By examining the visualization chart of the Cluster Head selection fra<br>
observed dat the improved GDD-K-Means<br>
clustering algorithm proposed in this paper<br>
eclustering algorithm proposed in this paper<br>
performs well in dat above clustering results, it can be clearly<br>
observed that the improved GDD-K-Means<br>
clustering algorithm proposed in this paper<br>
performs well in data clustering and has<br>
significant clustering effects.<br>
[4] Sabbagh A A, observed that the improved GDD-K-Means mechanism. Journal clustering algorithm proposed in this paper and Smart Environm performs well in data clustering and has significant clustering effects. [4] Sabbagh A A, Hamzs Enhan clustering algorithm proposed in this paper<br>performs well in data clustering and has 335.<br>significant clustering effects. [4] Sabbagh A A, I<br>Enhanced K-M<br>for Phishing At A, I<br>manced K-M<br>for Phishing At A, Bonco<br>noise point performs well in data clustering and has 335.<br>
significant clustering effects. [4] Sabbagh A A, I<br>
Enhanced K-M<br>
for Phishing At A, I<br>
with the rapid increase of Internet data, the<br>
vith the rapid increase of Internet data significant clustering effects.<br>
5. Closing Remarks<br>
With the rapid increase of Internet data, the<br>
19 Sabbagh A A, Hamze K,<br>
Enhanced K-Means Clu<br>
19 Sabbagh A A, Hamze K,<br>
will also increase of Internet data, the<br>
2024, 5. Closing Remarks<br>
With the rapid increase of Internet data, the<br>
With the rapid increase of Internet data, the<br>
2024, 13(18): 3677-3677.<br>
noise points of data will also increase, which [5] Klen M A, Bonduà S, Kasn<br>
will **S. Closing Remarks** for Phishing Attack D<br>With the rapid increase of Internet data, the 2024, 13(18): 3677-36'<br>will affect the clustering effect of data. In this<br>will affect the clustering effect of data. In this A fuzzy With the rapid increase of Internet data, the  $2024$ ,  $13(18)$ :  $367/366$ <br>noise points of data will also increase, which [5] Klen M A, Bonduà S, K<br>will affect the clustering effect of data. In this A fuzzy K-Means algor<br>n noise points of data will also increase, which<br>
will affect the clustering effect of data. In this A fuzzy K-Means algorithm<br>
paper, the method based on grid filtering outliers<br>
and so provides a guarantee for the accuracy will attect the clustering eftect of data. In this<br>
napper, the method based on grid filtering outliers<br>
not only ensures the purity of sample data, but<br>
also provides a guarantee for the accuracy and<br>
timeliness of k-mean paper, the method based on grid filtering outliers<br>
and the other discontinuity sets. Internalses provides a guarantee for the accuracy and the discontinuity sets. Internalses of k-means clustering algorithm in 182105879-1 not only ensures the purity of sample data, but<br>
discontinuity sets. In<br>
discontinuity sets. It<br>
dischains a glorithm in<br>
selecting the initialization center. Through [6] Ahmad W, Singh<br>
dynamic outlier detection, the comp also provides a guarantee for the accuracy and<br>
the constrained and Mining<br>
selecting the initialization center. Through [6] Ahmad W, Singh A, Ki<br>
dynamic outlier detection, the computational Optimizing Energy Efficier<br>
co tmeliness of K-means clustering algorithm in  $1821058/9-1058/9$ .<br>
selecting the initialization center. Through [6] Ahmad W, Singh A, K<br>
selecting to direct the computational optimizing Energy Efficie<br>
complexity is reduce selecting the initialization center. Through [6] Ahmad W, Singt<br>dynamic outlier detection, the computational Sensor Networks u<br>compensity is reduced and the computational Sensor Networks u<br>efficiency is improved to some ex dynamic outlier detection, the computational<br>
complexity is reduced and the computational<br>
effeciency is improved to some extent. The<br>
effeciency is improved to some extent. The<br>
computations<br>
relationship between distance complexity is reduced and the computational<br>
efficiency is improved to some extent. The Cluster Head Selection<br>
GPD-K-Means algorithm weighs the of Communication<br>
relationship between distance and density Information Secur efficiency is improved to some extent. The Cluster Head Selecti<br>
GPD-K-Means algorithm weighs the of Communicatie<br>
relationship between distance and density Information Security<br>
through the weight and the central point in GPD-K-Means algorithm weighs the of Communication I<br>
relationship between distance and density Information Security, 2024, The<br>
intough the weight and the central point index, [7] Zeng B, Li S, Gao X. Thr<br>
and solves the relationship between distance and density<br>
thromation Security, 2024,<br>
through the veight and solves the problem that randomly select the<br>
mans sector clustering<br>
number of initial central points due to the<br>
wireless senso through the weight and the central point index,<br>
and solves the problem that randomly select the<br>
number of initial central points due to the wireless sensor ne<br>
influence of the initial center selection, so that<br>
influenc miluence of the initial center selection, so that<br>
the improved clustering effect is better. Through<br>
the results of different evaluation indicators, it<br>
can be found that the proposed algorithm<br>
achieves better results th the improved clustering effect is better. Through<br>
the results of different evaluation indicators, it<br>
can be found that the proposed algorithm<br>
achieves better results than the traditional images using K-means<br>
algorithm algorithm in the processing of different data<br>
improvement and improvement in the screening<br>
improvement in the screening<br>
of noisy data. The next step of this paper will be<br>
of noisy data. The next step of this paper will unts, but there is still room for further<br>
overnent and improvement in the screening<br>
organization of Retension<br>
organization of Patterns in<br>
e processing and analysis of text data.<br> **Example 19** Preciado J L A, Aké C S<br>
e

## **Acknowledgments**

KYLK030).

## **References**

- or and South The next step of this paper will be<br>
e processing and analysis of text data.<br>
In the processing and analysis of text data.<br>
In the processing and analysis of text data.<br>
In the Processing and analysis of text
- to remove outliers more accurately and apply it<br>
thentification of Pattern<br>
to the processing and analysis of text data.<br>
Combined with PCA am<br>
Combined Wisualization. Mathems<br>
This paper is supported by Changii College<br>
V e processing and analysis of text data. among 208 Countries: K-N<br> **nowledgments**<br>
spaper is supported by Changji College 2591-2591.<br>
3 Campus level Research Project (No. [10] Jahandoost A, Torghabeh<br>
LK030). S, et al. Crud mowledgments<br>
spaper is supported by Changii College<br>
3 Campus level Research Project (No. [10] Jahandoost A,<br>
LK030).<br>
S. et al. Crude one<br>
means cluster<br>
emhanced by command of Big L<br>
Performance of Cluster Heads Selecti nowledgments<br>
Subsetted by Changii College<br>
Subsetted by Changii College<br>
Subsetted and the communisation. N<br>
S. et al. Crude of<br>
means cluster Heads Selection in [11] Nowak A B, C<br>
DEC Protocol Using K-Means Algorithm<br>
Su This paper is supported by Changii College 2591-2591.<br>
2023 Campus level Research Project (No. [10] Jahandoost A, Torg<br>
KYLK030). S. et al. Crude oil pr<br>
means clustering<br>
enhanced by dense<br>
[1] Juwaied A, Strumillo J L. I Fuzzy-based States and Benderick Communication<br>
Examples level Research Project (No. [10] Jahandoost A,<br>
LK030).<br>
S, et al. Crude<br>
means cluster<br>
means cluster<br>
means cluster<br>
Performance of Cluster Heads Selection in [11] LK030).<br>
S, et al. Crude oil primes<br>
emhanced by dense.<br>
Juwaied A, Strumillo J L. Improving<br>
plannal of Big Data, 2<br>
Performance of Cluster Heads Selection in [11] Nowak A B, Czesła<br>
DEC Protocol Using K-Means Algorithm<br>
- 

9-0620) Vol. 2 No. 3, 2024<br>Cluster Head selection framework in WSN<br>using enhanced K-means clustering<br>mechanism. Journal of Ambient Intelligence<br>and Smart Environments, 2024, 16(3): 309-<br>335. 9-0620) Vol. 2 No. 3, 2024<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A. Hamze K. Kha 9-0620) Vol. 2 No. 3, 2024<br>
121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K 9-0620) Vol. 2 No. 3, 2024<br>
121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K 335. 2959-0620) Vol. 2 No. 3, 2024 121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
[4] Sabbagh A A, H 9-0620) Vol. 2 No. 3, 2024<br>
121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Ehanced K-Means Clus 9-0620) Vol. 2 No. 3, 2024<br>
121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K 9-0620) Vol. 2 No. 3, 2024<br>
121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K

- 
- 2959-0620) Vol. 2 No. 3, 2024<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
[4] Sabbagh A A, Hamze 9-0620) Vol. 2 No. 3, 2024<br>
121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K 9-0620) Vol. 2 No. 3, 2024 121<br>
Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K Cluster Head selection framework in WSN<br>
using enhanced K-means clustering<br>
mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K, Khan S, et al. An<br>
Enhanced K-M Cluster Head selection framework in WSN<br>using enhanced K-means clustering<br>mechanism. Journal of Ambient Intelligence<br>and Smart Environments, 2024, 16(3): 309-<br>335.<br>Sabbagh A A, Hamze K, Khan S, et al. An<br>Enhanced K-Means C 182105879-105879. mechanism. Journal of Ambient Intelligence<br>
and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
[4] Sabbagh A A, Hamze K, Khan S, et al. An<br>
Enhanced K-Means Clustering Algorithm<br>
for Phishing Attack Detections. Electronics,<br> and Smart Environments, 2024, 16(3): 309-<br>
335.<br>
Sabbagh A A, Hamze K, Khan S, et al. An<br>
Enhanced K-Means Clustering Algorithm<br>
for Phishing Attack Detections. Electronics,<br>
2024, 13(18): 3677-3677.<br>
Klen M A, Bonduà S, K 335.<br>
Sabbagh A A, Hamze K, Khan S, et al. An<br>
Enhanced K-Means Clustering Algorithm<br>
for Phishing Attack Detections. Electronics,<br>
2024, 13(18): 3677-3677.<br>
Klen M A, Bonduà S, Kasmaeeyazdi S, et al.<br>
A fuzzy K-Means algo Sabbagh A A, Hamze K, Khan S, et al. An<br>Enhanced K-Means Clustering Algorithm<br>for Phishing Attack Detections. Electronics,<br>2024, 13(18): 3677-3677.<br>Klen M A, Bonduà S, Kasmaeeyazdi S, et al.<br>A fuzzy K-Means algorithm based Enhanced K-Means Clustering Algorithm<br>for Phishing Attack Detections. Electronics,<br>2024, 13(18): 3677-3677.<br>Klen M A, Bonduà S, Kasmaeeyazdi S, et al.<br>A fuzzy K-Means algorithm based on Fisher<br>distribution for the identifi for Phishing Attack Detections. Electronics,<br>2024, 13(18): 3677-3677.<br>Klen M A, Bonduà S, Kasmaeeyazdi S, et al.<br>A fuzzy K-Means algorithm based on Fisher<br>distribution for the identification of rock<br>discontinuity sets. Int 2024, 13(18): 36//-36//.<br>
[5] Klen M A, Bonduà S, Kasmaeeyazdi S, et al.<br>
A fuzzy K-Means algorithm based on Fisher<br>
distribution for the identification of rock<br>
discontinuity sets. International Journal of<br>
Rock Mechanic
- Klen M A, Bondua S, Kasmaeeyazdı S, et al.<br>
A fuzzy K-Means algorithm based on Fisher<br>
distribution for the identification of rock<br>
discontinuity sets. International Journal of<br>
Rock Mechanics and Mining Sciences, 2024,<br>
1 A fuzzy K-Means algorithm based on Fisher<br>distribution for the identification of rock<br>discontinuity sets. International Journal of<br>Rock Mechanics and Mining Sciences, 2024,<br>182105879-105879.<br>Ahmad W, Singh A, Kumar S, et a distribution for the identification of rock<br>discontinuity sets. International Journal of<br>Rock Mechanics and Mining Sciences, 2024,<br>182105879-105879.<br>Ahmad W, Singh A, Kumar S, et al.<br>Optimizing Energy Efficiency in Wireles discontinuity sets. International Journal of<br>
Rock Mechanics and Mining Sciences, 2024,<br>
182105879-105879.<br>
Ahmad W, Singh A, Kumar S, et al.<br>
Optimizing Energy Efficiency in Wireless<br>
Sensor Networks using Enhanced K-Mean Rock Mechanics and Mining Sciences, 2024,<br>182105879-105879.<br>
[6] Ahmad W, Singh A, Kumar S, et al.<br>
Optimizing Energy Efficiency in Wireless<br>
Sensor Networks using Enhanced K-Means<br>
Cluster Head Selection. International Jo 1821058/9-1058/9.<br>
Ahmad W, Singh A, Kumar S, et al.<br>
Optimizing Energy Efficiency in Wireless<br>
Sensor Networks using Enhanced K-Means<br>
Cluster Head Selection. International Journal<br>
of Communication Networks and<br>
Informat
- 
- Ahmad W, Singh A, Kumar S, et al.<br>
Optimizing Energy Efficiency in Wireless<br>
Sensor Networks using Enhanced K-Means<br>
Cluster Head Selection. International Journal<br>
of Communication Networks and<br>
Information Security, 2024, Optimizing Energy Efficiency in Wireless<br>
Sensor Networks using Enhanced K-Means<br>
Cluster Head Selection. International Journal<br>
of Communication Networks and<br>
Information Security, 2024, 16(3): 565-573.<br>
Zeng B, Li S, Gao Sensor Networks using Enhanced K-Means<br>Cluster Head Selection. International Journal<br>of Communication Networks and<br>Information Security, 2024, 16(3): 565-573.<br>Zeng B, Li S, Gao X. Threshold-driven K-<br>means sector clusterin 930-953. of Communication Networks and<br>
Information Security, 2024, 16(3): 565-573.<br>
[7] Zeng B, Li S, Gao X. Threshold-driven K-<br>
means sector clustering algorithm for<br>
wireless ensor networks. EURASIP Journal<br>
on Wireless Communi Information Security, 2024, 16(3): 565-573.<br>Zeng B, Li S, Gao X. Threshold-driven K-<br>means sector clustering algorithm for<br>wireless sensor networks. EURASIP Journal<br>on Wireless Communications and<br>Networking, 2024, 2024(1):
- rowement and improvement in the screening and the server of this paper will be [9] Preciado J L A, Aké C S.<br>
emove outliers more accurately and apply it allentification of Patterns in<br>
emoves outliers more accurately and a Zeng B, L1 S, Gao X. Threshold-driven K-<br>means sector clustering algorithm for<br>wireless sensor networks. EURASIP Journal<br>on Wireless Communications and<br>Networking, 2024, 2024(1): 68-68.<br>Kaizheng W, Yitong F, Shunzhen Z, et means sector clustering algorithm for<br>wireless sensor networks. EURASIP Journal<br>on Wireless Communications and<br>Networking, 2024, 2024(1): 68-68.<br>Kaizheng W, Yitong F, Shunzhen Z, et al.<br>Cloud detection from Himawari-8 spec wireless sensor networks. EURASIP Journal<br>
on Wireless Communications and<br>
Networking, 2024, 2024(1): 68-68.<br>
Kaizheng W, Yitong F, Shunzhen Z, et al.<br>
Cloud detection from Himawari-8 spectral<br>
images using K-means ++ clus 2591-2591. Networking, 2024, 2024(1): 68-68.<br>
[8] Kaizheng W, Yitong F, Shunzhen Z, et al.<br>
Cloud detection from Himawari-8 spectral<br>
images using K-means  $+$  clustering with<br>
the convolutional module. International<br>
Journal of Remo Kaizheng W, Yitong F, Shunzhen Z, et al.<br>Cloud detection from Himawari-8 spectral<br>images using K-means ++ clustering with<br>the convolutional module. International<br>Journal of Remote Sensing, 2024, 45(3):<br>930-953.<br>Preciado J Cloud detection from Himawari-8 spectral<br>images using K-means ++ clustering with<br>the convolutional module. International<br>Journal of Remote Sensing, 2024, 45(3):<br>930-953.<br>Preciado J L A, Aké C S, Martínez V F.<br>Identificatio images using K-means ++ clustering with<br>the convolutional module. International<br>Journal of Remote Sensing, 2024, 45(3):<br>930-953.<br>Preciado J L A, Aké C S, Martínez V F.<br>Identification of Patterns in CO2 Emissions<br>among 208 the convolutional module. International<br>Journal of Remote Sensing, 2024, 45(3):<br>930-953.<br>Preciado J L A, Aké C S, Martínez V F.<br>Identification of Patterns in CO2 Emissions<br>among 208 Countries: K-Means Clustering<br>Combined w Journal of Remote Sensing, 2024, 45(3):<br>
930-953.<br>
[9] Preciado J L A, Aké C S, Martínez V F.<br>
Identification of Patterns in CO2 Emissions<br>
among 208 Countries: K-Means Clustering<br>
Combined with PCA and Non-Linear t-SNE<br>
V 930-953.<br>
Preciado J L A, Aké C S, Martínez V F.<br>
Identification of Patterns in CO2 Emissions<br>
among 208 Countries: K-Means Clustering<br>
Combined with PCA and Non-Linear t-SNE<br>
Visualization. Mathematics, 2024, 12(16):<br>
25 Preciado J L A, Akè C S, Martinez V F.<br>Identification of Patterns in CO2 Emissions<br>among 208 Countries: K-Means Clustering<br>Combined with PCA and Non-Linear t-SNE<br>Visualization. Mathematics, 2024, 12(16):<br>2591-2591.<br>Jahando
	- Identification of Patterns in CO2 Emissions<br>
	among 208 Countries: K-Means Clustering<br>
	Combined with PCA and Non-Linear t-SNE<br>
	Visualization. Mathematics, 2024, 12(16):<br>
	2591-2591.<br>
	Jahandoost A, Torghabeh A F, Hosseini A<br> among 208 Countries: K-Means Clustering<br>
	Combined with PCA and Non-Linear t-SNE<br>
	Visualization. Mathematics, 2024, 12(16):<br>
	2591-2591.<br>
	[10] Jahandoost A, Torghabeh A F, Hosseini A<br>
	S, et al. Crude oil price forecasting us Combined with PCA and Non-Linear t-SNE<br>Visualization. Mathematics, 2024, 12(16):<br>2591-2591.<br>Jahandoost A, Torghabeh A F, Hosseini A<br>S, et al. Crude oil price forecasting using K-<br>means clustering and LSTM model<br>enhanced by Visualization. Mathematics, 2024, 12(16):<br>2591-2591.<br>Jahandoost A, Torghabeh A F, Hosseini A<br>S, et al. Crude oil price forecasting using K-<br>means clustering and LSTM model<br>enhanced by dense-sparse-dense strategy.<br>Journal o 2591-2591.<br>
	Jahandoost A, Torghabeh A F, Hosseini A<br>
	S, et al. Crude oil price forecasting using K-<br>
	means clustering and LSTM model<br>
	enhanced by dense-sparse-dense strategy.<br>
	Journal of Big Data, 2024, 11(1): 117-117.<br>
	No
	-
	- Jahandoost A, Torghabeh A F, Hossemi A<br>S, et al. Crude oil price forecasting using K-<br>means clustering and LSTM model<br>enhanced by dense-sparse-dense strategy.<br>Journal of Big Data, 2024, 11(1): 117-117.<br>Nowak A B, Czesław H