Name : PRATHAM HARAM

Roll no : A012

Class : TY IT

Subject : Business Intelligence

TOPIC :

DATA MINNING

INTRODUCTION:

Data mining is the process of discovering patterns and relationships in large data sets by using various statistical, algorithmic, and visualization techniques. It is a way of finding useful insights and knowledge from large and complex datasets.

Data mining involves analyzing data from different perspectives and summarizing it into useful information that can be used to make informed decisions. This information can be used to improve the efficiency and effectiveness of various business processes, such as marketing, finance, and operations.

There are several steps involved in the data mining process, including:

- ✓ Gathering and collecting data from various sources
- ✓ Preprocessing and cleaning the data to remove any noise or inconsistencies
- ✓ Selecting the relevant data for analysis
- ✓ Applying data mining algorithms to discover patterns and relationships in the data
- ✓ Validating the results of the data mining analysis
- ✓ Visualizing and presenting the results in a meaningful way.
- ✓ Data mining is an important tool for businesses, as it allows them to make better decisions based

on data-driven insights rather than relying on gut instincts or past experience.

Review of 1st Research Paper :

<u>Survey on Anomaly Detection using Data Mining Techniques</u>

Anomaly detection, also known as outlier detection, is the process of identifying unusual or unexpected patterns in a dataset that do not conform to the expected behavior. It is a key task in data mining and machine learning, and it has many practical applications, such as fraud detection, network intrusion detection, and fault diagnosis.

There are several data mining techniques that can be used for anomaly detection, including:

Statistical methods: These methods involve calculating statistical measures, such as mean, median, and standard deviation, and identifying data points that are significantly different from the rest of the data.

Clustering: Clustering algorithms, such as k-means and DBSCAN, can be used to identify data points that do not belong to any cluster, or that are significantly different from the other points in the same cluster.

Classification: Supervised learning algorithms, such as decision trees and support vector machines,

can be used to classify data points as normal or anomalous based on a labeled training dataset.

Neural networks: Artificial neural networks, such as autoencoders and deep belief networks, can be trained to detect anomalies by learning the normal patterns in the data and flagging any deviations from these patterns.

To effectively detect anomalies, it is important to have a good understanding of the data and the expected behavior, as well as to consider the costs of false positives and false negatives. Anomaly detection can be a challenging task, but it is an important tool for identifying unusual patterns and events that may be of interest or require further investigation.

Review of 2nd Research Paper :

Data Mining: Web Data Mining Techniques, Tools and Algorithms: An Overview

Web data mining is the process of extracting, processing, and analyzing data from the World Wide Web for a variety of purposes, such as market research, customer relationship management, and competitive intelligence. Web data mining involves using a variety of techniques, tools, and algorithms to extract and process data from web sources, such as web pages, social media, and online forums.

Some examples of web data mining techniques include:

Web scraping: Web scraping involves using specialized software or programming scripts to extract data from web pages and other online sources.

Web crawling: Web crawling involves using automated software to scan and index web pages and other online resources.

Natural language processing: Natural language processing (NLP) techniques are used to extract

and analyze text data from web sources, such as social media posts, online reviews, and blogs.

Some examples of web data mining tools include:

Python: Python is a popular programming language for web data mining, with libraries such as Beautiful Soup and Scrapy for web scraping, and NLTK for natural language processing.

R: R is a programming language and software environment for statistical computing and graphics, with packages such as rvest and httr for web scraping, and tm and qdap for natural language processing.

RapidMiner: RapidMiner is a data science platform that provides a range of tools and algorithms for web data mining, including web scraping, text mining, and machine learning.

Some examples of algorithms used in web data mining include:

Decision trees: Decision trees are a type of machine learning algorithm that can be used to classify and predict outcomes based on data from web sources.

Clustering: Clustering algorithms, such as k-means and DBSCAN, can be used to group data from web sources into clusters based on similarities.

Association rules: Association rule mining algorithms, such as Apriori, can be used to identify patterns and relationships in data from web sources, such as co-occurring products or services.

Review of 3rd Research Paper :

A review of data mining techniques

Having the right information at the right time is crucial for making the right decision. The problem of collecting data, which used to be a major concern for most organizations, is almost resolved. In the millennium, organizations will be competing in generating information from data and not in collecting data. Industry surveys indicated that over 80 percent of Fortune 500 companies believe that data mining would be a critical factor for business success by the year 2000 (Baker and Baker, 1998). Obviously, DM will be one of the main competitive focuses of organizations. Although progresses are continuously been made in the DM field, many issues remain to be resolved and much research has to be done.

Overall, data mining techniques are useful for finding insights and knowledge in large and complex datasets, and they have a wide range of applications in various fields such as business, finance and healthcare.

Review of 4th Research Paper :

Application of Data Mining Techniques to Healthcare Data

Data mining techniques have the potential to transform healthcare by providing data-driven insights and decision-making tools for a variety of applications, including disease diagnosis and treatment, population health management, and quality improvement.

Here are some examples of how data mining techniques can be applied to healthcare data:

1. Disease diagnosis and treatment: Data mining techniques can be used to identify patterns and relationships in patient data that can help to improve disease diagnosis and treatment. For example, machine learning algorithms can be used to predict the likelihood of a patient developing a certain disease based on their medical history, lifestyle, and other risk factors.

2. Population health management: Data mining techniques can be used to analyze data from large populations to identify trends and patterns that can inform public health policies and interventions. For example, clustering algorithms can be used to

group individuals into distinct subpopulations based on their risk factors and needs.

3. Quality improvement: Data mining techniques can be used to identify areas of opportunity for quality improvement in healthcare organizations. For example, association rule mining can be used to identify patterns in patient data that may indicate opportunities for preventative care or more efficient resource utilization.

Overall, data mining techniques have the potential to provide valuable insights and support decision-making in healthcare, leading to improved patient outcomes and reduced costs.

Review of 5th Research Paper :

10 CHALLENGING PROBLEMS IN DATA MINING RESEARCH

There are many challenging problems in data mining research, including:

1. Big data: One of the main challenges in data mining is dealing with very large datasets, known as big data. These datasets often have high dimensionality and require specialized algorithms and infrastructure to process and analyze efficiently.

2. Data quality: Ensuring the quality and integrity of the data is a major challenge in data mining. This includes dealing with missing or incorrect data, as well as handling data that is noisy, redundant, or biased.

3. Complex data: Data mining algorithms often struggle with complex data, such as data with multiple levels of abstraction or data that is structured in a non-linear fashion.

4. High-dimensional data: Data with a large number of features or dimensions can be difficult for data mining algorithms to handle, as it can lead to the curse of dimensionality, where the number of samples required to accurately model the data

increases exponentially with the number of dimensions.

5. Privacy and security: Protecting the privacy and security of data is a major concern in data mining, as it often involves handling sensitive personal or confidential information.

6. Interpreting and explaining results: One of the challenges in data mining is interpreting and explaining the results of the analysis in a way that is meaningful and actionable for the intended audience.

Overall, these challenges highlight the need for ongoing research and development in data mining to address the growing complexity and diversity of data, and to enable data-driven decision-making in a wide range of applications.

Dupli Checker

PLAGIARISM SCAN REPORT

10% Plagiarised	90% Unique	

Date	2023-01-04
Words	268
Characters	1888

Content Checked For Plagiarism

Review of 1st Research Paper :

Survey on Anomaly Detection using Data Mining Techniques

Anomaly detection, also known as outlier detection, is the process of identifying unusual or unexpected patterns in a dataset that do not conform to the expected behavior. It is a key task in data mining and machine learning, and it has many practical applications, such as fraud detection, network intrusion detection, and fault diagnosis.

There are several data mining techniques that can be used for anomaly detection, including:

Statistical methods: These methods involve calculating statistical measures, such as mean, median, and standard deviation, and identifying data points that are significantly different from the rest of the data.

Clustering: Clustering algorithms, such as k-means and DBSCAN, can be used to identify data points that do not belong to any cluster, or that are significantly different from the other points in the same cluster.

Classification: Supervised learning algorithms, such as decision trees and support vector machines, can be used to classify data points as normal or anomalous based on a labeled training dataset.

Neural networks: Artificial neural networks, such as autoencoders and deep belief networks, can be trained to detect anomalies by learning the normal patterns in the data and flagging any deviations from these patterns.

To effectively detect anomalies, it is important to have a good understanding of the data and the expected behavior, as well as to consider the costs of false positives and false negatives. Anomaly detection can be a challenging task, but it is an important tool for identifying unusual patterns and events that may be of interest or require further investigation.

Matched Source

Similarity 25%
Title:
Survey on Anomaly Detection using Data Mining TechniquesSurvey on Anomaly Detection using Data ... - ScienceDirect.com

https://www.sciencedirect.com/science/article/pii/S1877050915023479