Examination Of Free Software For Knowledge Discovery In Data

Section 1. summary of paper

1.1.References

Chen, X., Williams, G., Ye, Y., and Xu1, X. (2007). A Survey of open-source data mining systems*.

Website: pakdd07.pdf (togaware.com)

1.2.Purpose of study

The majority of companies are working to advance data mining as a contemporary technology. These open source data-mining platforms introduce a new developments in data-mining research, education, and industry. With these open-source solutions, companies don't have to stress about the cost of buying licenses to use the software, so they can focus more on training their staff about cutting-edge technology. Two examples of frequently used open-source software are MYSQL database and Weka opensource data mining platform. Open source data mining is efficient for small and medium-sized firms (SMEs) wanting to use business intelligence solutions for marketing, customer service, e-business, and risk management. Therefore, we must conduct a survey to understand more about the systems that are available and how they are used.

1.3 Research design and strategy

Data mining is the process of drawing valuable conclusions from massive amounts of data. The initial stage in the data mining process is to identify the business problem and comprehend the data called as **Data Understanding**. The second step **Data pre-processing** is critical phase will have an impact on the model's performance by cleaning, filtering, and converting data. At **Data modeling** stage, data-mining techniques like K-means clustering are put into practice. Classification, prediction,grouping,association rules, and interactive exploration techniques like link analysis are popular modeling procedures. In **Model evaluation** phase the model is assessed based on recall, accuracy, precision, and f1 score. This shows that the model is effective and that business problems can be solved using it. Following evaluation, **Deployment** phase the model is used to inform decisions, for as when employing a campaign model to rank consumers or a predictive model to anticipate prospective customers.

Based on the eight key qualities given below, the author of the article listed 12 open-source software applications.

• **Capacity to access a variety of data sources:** Data may be accessed via databases and data warehouses, and it can be in a variety of forms, including txt, csv, tables, and others. All these data must be easily accessible through a solid open-source system.

• **Data Pre-processing Capability:** Data Pre-processing is a crucial phase in the data mining process and is essential to finding a solution. A good system should be able to do numerous data-preprocessing tasks effectively.

• **Integration of many techniques:** No one algorithm can ever guarantee a problem-free outcome. Different data mining models should be able to combine with a good system. In order to choose the optimal model, the user will test out several ones.

• **The ability to work with enormous datasets:** Data mining issues call for lots of data to allow for better analysis. Good systems should therefore be able to handle a lot data.

• **Effective data and model visualization:** Users must explore the information and comprehend model. visualization feature is therefore essential for a better systems.

• **Extensibility:** Since data-mining is new technique, more of research is being done to develop new models. Thus, it is simple to incorporate new models into these open-source systems.

• **Interoperability with other systems:** Interoperability involves sharing of data and models. Canonical standards like CWM and PMML will be supported by a decent system.

• **Active development community:** These open-source programs need to be maintained by an active development community.

1.4 Conclusion

The main takeaways from this article include presentation of opensource data-mining systems and listing out important features in open sources systems. Author also explained both positive and negative aspects of opensource data-mining systems. Author has listed 12 important opensource data-mining systems that are useful for researchers, educational purposes but for commercial usage still we need to work on some feature to make them accessible.

1.5 Contribution

The author of this post provided a list of opensource data-mining systems and provided an explanation of the selection criteria utilizing the eight key attributes mentioned above. The author discussed data-mining process, several opensource data-mining programs, and aspects of their functionality and usability, as well as how they benefit various small-sized and medium-sized businesses.

SECTION 2.CRITICAL ANALYSIS

2.1 Overall Assessment

The author's detailed explanations made the material easy to read. In addition, the author did a great job of keeping a pleasant flow while offering different features for selection of opensource data-mining system. The paper's primary merit is its key features description and comparison based on 4 aspects and outcome analysis. I must commend writers' writing abilities.

2.2 Research methodology

Some significant aspects, such as general qualities, data source characteristics,functionality characteristics,usability characteristics,overall characteristics, were given weight by the author. Twelve pieces of open-source software are given scores based on these traits, and they are then contrasted. Based on the results of this study, we can choose the finest opensource data-mining solution.

Opensource data-mining systems are studied based on 4 aspects:
General Characteristics: some general characteristics like activity, licenses, programming-languages, and operating systems are used for comparison between open-source systems
Data Source Aspects: different data sources and data formats are used for comparison between open-source systems.
Functionality Aspects: Different functionality aspects available in open-source data mining systems are compared
Usability Aspects: It detiremes, how easily we can use open-source data mining systems.

2.3 New Knowledge Learned

I now know how data mining works, key characteristics of opensource data-mining systems, and the benefits and drawbacks in these systems.

2.4 Future Research

The majority of open-source solutions include downsides in addition to their benefits. Therefore, we must address these drawbacks going forward. Future open-source systems ought to be able to manage vast amounts of data as well as a variety of data sources. For these data mining technologies to be used effectively, proper documentation must be given.

Section 3.Question to discussed

Is it safe to utilize these open-source data mining platforms- There is high risk of vulnerabilities while using open-source software so we should consider security as an important factor while using them.