

Written Report (2000 words)

Justify the use of data mining software to interpret data patterns in your own subject discipline - Marketing, Human Resources or Finance.

You will need to explain the main data mining approaches of prediction, association and clustering with specific examples from Marketing.

along with comparing and contrasting various types of software available on the market from open source to proprietary software as well as spreadsheet software such as Microsoft Excel.

PLAN:

INTRO: What data mining is IN BASIC. What you'll talk about.

P1: Explain first one and how they differ. How it can be used in marketing. Executive summary. Capturing and engaging

P2: explain the other one clustering. And association. Grab readers attention

P3: Why it's important and how its applied in business.

P4: Pros and cons of both

P5: cOMPARE AND contrast various types of software you can use. How it helps interpret data. OPEN SOURCE AND PROPRIETARY SOFTWARE

p6: which one is most advantageous to marketing with examples. How it was done in the past vs now

Open source to proprietary software vs microsoft excel

Conclusion: Round off all points making sure everything is covered and i cant lose many marks.

Start:

Data mining incorporates a variety of techniques used in examining large pre-existing databases in order to generate new relevant information. Businesses then interpret data patterns to gain competitiveness in the marketplace. Businesses will then analyze and use this data to manage and run their marketing programs. It is a new revolutionary asset for businesses to incorporate to increase their competitive edge over their competition.

There are multiple data mining methods businesses can use. One of which is predictive mining. Predictive mining analyze large amounts of data in regards to patterns and systematic relationships between variables. Businesses like data mining as it aims to predict and refine data sources in order to extract hidden value that businesses can use to their advantage. For example it can be used in sales forecasting; by analyzing when consumers bought products and in which location to try and predict when they will buy again. This is called E-commerce basket analysis. There is also database forecasting where businesses analyze purchasing patterns and look at their buyers demographics and psychographics to build a profile on them and what they like. This can then be used in marketing to specifically target the consumer by only showing them items that they are likely to like and want to buy. This then results in market segmentation which is a powerful marketing tool, as it allows you to split up the market into more niche groups, meaning each group is easier to target. This puts your business at a huge competitive advantage as your business is gaining data on customers buying habits, as well as their style and when they are likely to buy. This information is utilized by businesses to suggest and tailor how they market their products to the consumer.

Another type of mining is association mining. Association rule help to show the probability of relationships between data items within large data sets involved in various types of databases. This form of mining has multiple applications and is used to help discover sales correlations in transactional data. There are two parts to associative mining; an antecedent (if) and also a consequent (then). Antecedent is found within the data and a consequent is an item found in combination with the antecedent. Associative mining is created by searching data for frequent if-then patterns and it uses something called the criteria support and confidence systems to identify the most important connections. Support is an indication of how recurrent or frequently the item appears in the data and confidence is the number of times the if-then statements are found true. Association mining is an analytic tool that gathers information on what customers buy as well as providing insights into who they are and why they make certain purchases. This gives the marketing team the ability to monitor and study the buying habits of their customers, putting them at a much greater and more targeted starting point than their competition.

A third data mining approach used in marketing is clustering. Clustering refers to the grouping of items with similar characteristics and proportioning that data into different groups. This gives the marketing team the ability to examine both the internal and external pricing factors. Internal factors such as; pricing, product position in the market and staff skills. External factors include; competition and demographic of the customer base. This split is the mass market into a more niche one. Helping the business to identify groups within their customer base. The data has a big impact on sales and brand image, meaning its an

invaluable tool for the marketing team to boost sales and revenue. Applications of clustering include: for retail businesses; data clustering helps with monitoring customer shopping habits and sales campaigns as well as customer retention. In the insurance industry clustering is usually used in fraud detection, risk factor identification and consumer retention data. In banking clustering is used for credit scoring, customer segmentation and analyzing customer profitability. Therefore the application of this big data method is huge and widely used in businesses worldwide.

Within a business the product manager will work with the marketing manager to identify the segment of the population who is most likely to respond to your product. To do this you must have a good overall understanding of your market and then deploying the right technique to target them. For example a mining pattern may indicate that low average annual salary customers are more likely to incur a loan default. This information could help the marketing manager in devising a more practical and relevant loan strategy for similar future customers. There are several ways to interact with your customers using different communication channels. For example; direct marketing, telemarketing, radio, print advertisements, television and so on. It is only using data mining that an analyst would be able to conclude the most optimal channel to send the marketing through.

Data mining is a very powerful tool however as most things, it doesn't come without its downsides. Businesses have to be very careful that they ensure the data mining information that's been collected is very well protected. A lack of appropriate security system would put at risk the privacy information of users. Data mining can also require a business to invest more in staff training. Finally if the information collected is altered accidentally or is inaccurate in any way, it would hugely affect the decision making process of the marketing team as well as make their work irrelevant and inaccurate which wastes time and money. The pros of Data Mining greatly outway the cons however. Data mining contributes to strategic decision making by discovering key data that businesses then use to gain competitive edge and grow economies of scale further. It also allows huge databases to be analyzed at once, something that would take a human an extremely long time to do. Furthermore, it also improves your company's relationship with the customer as it allows you to offer your customers the products or services that they specifically want or similar things to what they have wanted in the past.

Open mining software is used for creating science applications and services. The data helps you to understand the collected data and to design data science workflows in relation to the results. It also allows you to provide easy integration with other technologies and many offer a marketplace to customize your mining experience. However most data today is unstructured and thus needs to be processed to extract useful information from it, to develop that into understandable and usable form. This is where data mining comes into it. There are many tools that use artificial intelligence, machine learning and other techniques to extract data. Rapidminer is one, its a program written in the Java language. This tool offers advanced analytics through template based frameworks. Its very easy to use and can be used for visualization, data preprocessing, predictive analytics, statistical modeling and many more. Being that it's so easy to use businesses will find this tool extremely useful. Because it

can portray all their data in a way that they can all understand it, no extra training or cost is needed to use this application.

WEKA is another open source analytic tool. This tool is very sophisticated and has uses in many different applications including visualization and algorithms for predictive modelling. It's a free application under the GNU license which puts it in a better position than Rapidminer because users can customize it however they want. WEKA supports clustering, classification, regression, feature selection and many more. R is a computing language for statistical computing and graphics, WEKA supports this too.

Another form it takes is proprietary software. Proprietary software remains the property of its owner or creator and is used by organizations as commercial software that can be bought, leased or licensed. Proprietary software can be licensed for a fee but copyrighting and distribution are prohibited. This contrasts free, open source software, that allows you to view, edit and redistribute the code that makes the entire program work. Microsoft office, DropBox, McAfee and Virus scan are all examples of proprietary software. Commercial proprietary software tends to have specific features that make it appeal to a specific audience as it aims to provide a better user experience. The main drawbacks are that you cannot modify the software, if additional features are required usually businesses will have to purchase them and the more advanced programs will result in the marketing team having to employ someone with the knowledge on how to use it or provide training to one of their members to learn how to use it. This is why open sourced programs are usually preferred because they are more established, having discussion boards and chat boxes for any questions or help. It also costs nothing as the systems are constantly being developed to remove bugs and improve the site meaning it has advanced functionality and better security to proprietary systems.

Microsoft Excel is a very popular software program used to create spreadsheets, execute calculations, produce charts and perform statistical analysis. Excel is used by professionals however it is deemed to be very easy to use and not as professional as other open sourced methods like R. R is a free, open sourcing programming language that's often used in big data analysis and statistical computing. R has many functions and abilities that are suited to large projects with multiple data sets. Excel is reported as being the starter program due to its wide use and easy formatting. R is more designed for larger data sets and to create a more detailed visualization. Deciding between the two is more based on your needs rather than choosing one because it's 'better'. If you're looking to create a quick presentation to demonstrate data for your coworkers, then excel is usually more applicable. For those planning on publishing large data sets with a lot of complicated information to stakeholders, then spending the time on R to create impressive professional work is advised. If you want to enter the financial field knowing R will put you at an advantage and those focusing on R also usually make a high salary, frequently starting off around £70,000 / year. In evaluation; in relation to marketing a sound understanding of excel is very advantageous and will be utilized frequently however unless you have a specialty to do with finance, it's not necessary to master R.

In conclusion, data mining is an enormous asset to add to a businesses arsenal, due to the advanced data collected that allows a business to specifically identify and tailor their marketing efforts to target the specific niche needs of each consumer. This in turn makes it

much more likely for the consumer to relate and to like a product that is suggested to them by the business, thus hopefully resulting in a sale. It increases profitability which is the core goal of most businesses. The link between market growth and economies of scale for a business through the use of data mining can be seen clearly here.

References:

<https://insidebigdata.com/2018/11/24/data-mining-predictive-analytics-things-care/> -7/11/19

<https://www.tutor2u.net/business/reference/data-mining> -7/11/19

<https://www.matillion.com/insights/5-real-life-applications-of-data-mining-and-business-intelligence/> -7/11/19

<https://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/> -7/11/19

https://www.sas.com/en_gb/insights/analytics/predictive-analytics.html -7/11/19

<https://bigdatadn.com/intelligence/why-use-clustering-in-data-mining/> -7/11/19

<https://opensourceforu.com/2017/03/top-10-open-source-data-mining-tools/> -7/11/19

https://link.springer.com/chapter/10.1007/978-3-319-20883-1_22 -7/11/19

<http://www.jmlr.org/papers/volume15/fournierviger14a/fournierviger14a.pdf>

<https://dl.acm.org/citation.cfm?id=1137997> -7/11/19

<https://www.businessinsider.com/yugabyte-open-source-polyform-project-amazon-2019-7?r=US&IR=T> -7/11/19

<https://www.businessinsider.com.au/category/mining> -7/11/19

<http://www.efficientsoftware.co.nz/microsofts-excel-vs-openoffices-calc-proprietary-or-open-source/> -10/11/19

Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner -

https://books.google.com/books?id=6WYV_ACVFuQC&dq=books+on+data+mining+and+business&lr=&source=gbs_navlinks_s - 4/12/19

<https://www.sciencedirect.com/science/article/pii/S0004370202003570> - 4/12/19

<https://www.businessinsider.com/r-how-a-data-mining-giant-got-me-wrong-2018-3> - 4/12/19

<https://www.businessinsider.com/sell-side-data-mining-comes-under-scrutiny-2019-3> - 4/12/19

<https://dl.acm.org/citation.cfm?id=2460327> - 4/12/19

https://www.researchgate.net/profile/Bing_Liu41/publication/2805722_Web_for_Data_Mining_Organizing_and_Interpreting_the_Discovered_Rules_Using_the_Web/links/5472bc000cf216f8cfae883d.pdf - 4/12/19

<https://ieeexplore.ieee.org/abstract/document/5386727> - 4/12/19

<https://www.tandfonline.com/doi/abs/10.2753/MIS0742-1222240107> - 4/12/19

<https://flossmole.org/sites/flosshub.org/files/dalle2.pdf> - 4/12/19

<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2575.2010.00357.x> - 4/12/19