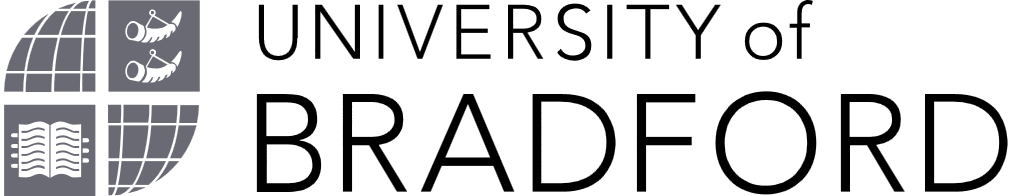
****

**DEPARTMENT OF BUSINESS ANALYTICS, CIRCULAR ECONOMY AND SUPPLY CHAIN**

**OIM7507 Artificial Intelligence and Data Science**

**OIM7507-B**

UoB Number: 20018888

I have read the University Regulations relating to plagiarism. In submitting this I certify that this assignment is all my own work and does not contain any unacknowledged work from any other sources.

Word Count: 2770

Introduction

Use of big data and data science techniques in organisations is fast becoming commonplace due to the relative abundance of data. Commercial companies and private companies, from all industries are benefiting from the implementation of data science techniques allowing them to gain insights on customers, staff, productivity and product performance amongst other uses (Cielen et al. 2016). Benefits appear to be very apparent with many examples of the use of data science techniques in modern organisations for example advertising which has observed success through personalised advertisements made possible by collation if data through Data Science. In this report, I will discuss the benefits and possible pitfalls of data science in organisations using named cases where possible and also use relevant worked examples where appropriate.

Business Intelligence

Early implementation of analytic techniques in organisations has often been coined as business intelligence in the early 1990’s. In essence, BI is the use of current and historical data by management in decision making (Sharda et al. 2021). Consisting of four major components which include a data warehouse for the source data, business analytic techniques, Business performance management and lastly a user interface such as a dashboard (Sharda et al. 2018). Organisations which are regarded as advanced in implementation of BI also appear to have more respect for measurement and evaluation of performance and proper implementation enables enhanced decision making which directly leads to improved firm performance (Kulkarni et al. 2017). The positives of BI are very evident, and it does appear that adding BI will improve performance when done correctly with a management team that is invested in participation. Literature suggests implementation is influenced significantly and has a higher success rate when top management “buys in” to its uses (Kulkarni et al. 2017). This is a positive if the culture of the organisation allocates resources, time and funding to allow management to embrace implementation however, this is a very big if as implementation may cause disruption and management may not want to embrace new technologies. Having a system that hinges its success largely on the willingness of management to implement may skew perception of its capabilities but fundamentally, BI can give organisations that implement it well a large advantage above competitors who do not.

Text Analytics/Text mining

In an increasing data-centric society, effective data acquisition, processing and interpretation is typically the difference between the most successful and progressive organisations. With a wide array of data available and an estimated 463 exabytes of data to be created daily in 2025 (Desjardins 2019), collating this data becomes even more important. Whilst structured data is readily processed and interpreted with relative ease, most data being created is in unstructured form. Daily, 500 million texts are sent, 294 billion emails sent and 4 petabytes of data created on Facebook (Desjardins 2019). Harnessing this information requires methods such as text analysis and sentiment analysis. Through a combination of language science and computer science with statistics, unstructured text data can be restricted into a more structured form for interpretation, for example Google performs a number of text mining techniques when presenting relevant query answers and spam filters via Gmail (Cielen et al. 2016). As an organisation, it has got to be understood that the most trusted form of information in advertising is that of the word of mouth (Adamopoulos et al. 2018), once understood it only makes sense to take advantage of the growing influence word of mouth is having on online consumer behaviour. Social media platforms can be effective tools in accumulating opinions of the masses and influencing the consumer, as stated in (Adamopoulos et al. 2018) word of mouth effectiveness and influence seemingly increases in those that have similar personality traits. During the 2016 presidential election in researchers were able to analyse and define six classes of twitter users and gauge who they would support solely based on their tweets and found discovered a link in homophily increase when there is reciprocal connections in terms of followers (Caetano et al. 2018). In the case of the 2016 election, it was surrounded with scandal due to the impact Cambridge Analytica was able to have on the election essentially creating propaganda based on the observed data collected on social media. Clearly, using text mining and analysis is powerful and has the potential to put any organisation at the top through understanding consumer behaviour. However, as highlighted by the election, there is a lot of potential for abuse and manipulation if the data is mishandled.

Data Modelling

Like in the case of text analysis, processing and interpretation is key in data science and data mining process. Data science itself is referred to as a discipline focused on identifying useful patterns in data sets (Akerkar 2019). Just like in text analysis, there is potential for abuse but also potential for exponential growth from an organisational point of view. With a lot of organisations now having access to more evolved computer systems, cloud storage and data warehouses, it’s now more accessible for companies to move away from descriptive and diagnostic analytical techniques in favour of more advanced predictive and prescriptive tools (Sharda et al. 2021). The more advanced organisations are able to implement neural networks aiding response models which can indicate and predict consumer behaviour (Baesens et al. 2002). In this report, the experiment suggested implementation of Bayesian neural networks offers viable alternative to the favoured purchase incidence modelling techniques used and when used alongside customer profiling predictors other than recency, frequency and monetary, predictive powers of the models. Clearly, using advanced artificial intelligence appears to facilitate better decision making but also indicates customer retention which is very useful as retention of customers is thought to be more valuable than acquisition of new customers (Shaaban et al. 2012).

Data Quality

Regardless of the technique used, whether advanced or “primitive”, the biggest impede comes in the cleansing of data sets. This is the most time-consuming process in pre-processing of data. Data cleaning, consolidation and transformation are used as needed to create a set of data that can be interpreted most accurately (Abril 2011). 52% of data scientists stated poor data was the biggest daily obstacle whilst 66.7% of data scientist stated data organisation and cleansing was the most time consuming action (Biewald 2015). Given, sample size was relatively small, but the essence of the figures remains true. The stats indicate to a possible lack of data scientists in the industry. Organisations may look to hiring more scientists but unless commitment to proper implementation of data science and artificial intelligence, success may be allusive. The majority of issues surrounding data quality are derived from missing data, this is a given when taking into account data collection from sources such as IoT (Sharda et al. 2021). Missing data presents an issue as the decision then comes down to reduction of sample size of analysis by removing the whole set from sample or potentially doing nothing and risking incorrect generalisation based on the data as presented (Akerkar 2019).

New Insights

The scope of new technologies is one that elicits excitement based on the trajectory of sheer size of data and advancement in tech. It appears the more advanced machine gets, the closer to man it becomes (Turing 1950). Introduction of machine learning in the FinTech industry not only is improving hypothesis testing but also making it easier to remove human error and coincidences (Hendershott et al. 2021). For example, Jackwerth and Menner used machine learning to reject Ross recovery theorem demonstrating capabilities in testing of hypothesis and estimating probability of generating data by specific models (Hendershott et al. 2021). With further research and investment, it does appear machine learning will become increasingly involved in asset pricing through the use of return predictions and investor behaviour modelling. AI methods are slowly proving their worth in terms of more accurate and time saving methods in the financial industry for one. But they also are very useful in other organisations such as healthcare-based organisations and government organisations.

Conclusion

In conclusion, whilst living in such a data driven society, it seems almost impossible to escape the influence of artificial intelligence and data science. It does seem apparent that the most successful are those that embrace the change and adopt early. Based on the literature, it does seem to paint a very vivid picture of the future, one that’s very heavily influenced by a growing amount of data and technologies harnessing this data.

Part II

Tesco PLC is the name of one of the largest retail companies based in the UK. As part of this retail group, I will be analysing financial and industrial data, comparing with other retailers in hopes of recommending viable options to improve sales and profit. As of July 2021, our turnover for the year was £57.9bn, according to data obtained from FAME, cementing us as market leaders amongst retailers in the UK. This is of no surprise considering we are the largest retailer in the UK with over 3000 outlets (Retail Economics 2021). To remain at the top, its essential to evaluate how/why we are in the position before we can establish how we can better this in the near future.

To begin with, figure one is a simple overview of some analysis presented in a dashboard format, this can allow the manager to pick the most relevant and meaningful illustration with ease.

Chart, line chart

Description automatically generated

Figure 1) Dashboard presenting different graphs.

Before delving into the more ‘complex’ data, it was necessary to have an overview of the summary of the data. Upon first glance at the data, it appears the general trend is an increase in every variable until 2019, from 2020 onwards it appears there is a drop in every variable. The reason for this drop is not clear upon initial glance, however, the data could be applied to draw hypothesis that may be applicable. Initial analysis of the turnover was indicative of the general trend. Figure 2 depicts the turnover across the 5-year period, It appears to suggest a positive correlation overall thus leading to a belief that there may be positive relationship between the period in time and the amount of turnover produced. However, the data also does appear to show a drop off in turnover I the years of 2020-2021. Without having a complete picture of all variables that may have played a part in this it is difficult to suggest any valid reason for this occurrence, with the current situation in the world (Covid-19 Pandemic) its likely that the pandemic did play a roll in this time period potentially even being the main reason for this drastic reduction in turnover. The R-Square value suggests a weak relationship between the two variables thus suggesting no link. Due to this revelation, it would be unwise to suggest that this is a dip based on the pandemic and is likely to pass post lockdown. It does appear to suggest further investigation required. The graphs mentioned in the dashboard were used throughout the report and other graphs that were not mentioned in the dashboard but still are featured in my own time.

Chart, line chart

Description automatically generated

Figure 2) Graph showing the change in turnover across the 5 years. R-squared value.

Similar to the trend in turnover, it appears that profit increased across the 5-year period overall, but the amount of profit fell considerably post 2019. Figure 3 shows the trend in profit, it also shows the change in profit margin which also follows the same trend suggesting there was something that occurred post 2019 that affected both the profit and the profit margin. The data would seem to suggest turnover decreasing would cause a decrease in the profit and profit margin as it is likely products are not selling as much. The trend lines are pretty much mirror images with the profit margin and the profit before taxation. Due to no indication of what caused this decline, It is difficult to suggest alterations that could cause a uptorn in the profits. At this stage it was still very difficult to draw conclusions however it was ideal and helpful in terms of forecasting patterns that would inform us at a later point.

Chart, line chart

Description automatically generated

Figure 3) Profit before taxation and profit margin over the 5 years

To understand what could have happened in 2019 onwards, further analysis could be undertaken to develop connections. Figure 4 begins to potentially identify a link that explains the decrease in profit. Figure 4 shows a graph of the number of employees in the 5-year period from 2017-2021 and also the profit before taxation in the same time period. The number of employees generally fell from 2017 to 2021. The year of 2019 had the greatest number of employees, this also appeared to have the year with the highest profit prior to taxation, although it is not proven that these variables are linked it could be said that increasing employees increases profit potentially due to increased productivity or increased number of stores. It could be argued that we need to continue hiring staff as it suggests more staff may lead to more profit. It may also be a suggestion to increase the number of stores to stimulate the sales of products as discussed above. It may also be an indication that we need to begin thinking of new promotional or advertisements to begin gaining the attention of customers. This will drive the sales up thus increasing the profit as long as we keep running costs constant over the next year.

Chart, line chart

Description automatically generated

Figure 4) Graph of Profit before taxation and the number of employees.

On the other hand, looking at the bigger picture it appears there may be other reasons which explain the general trend in the amount of profit. After analysing the Gearing (%), there could be a link between the gearing and the profit before taxation. Gearing generally decreased across the 5-year period as expressed in figure 5. It was at its lowest point in 2019, this is likely to be due to repayment of non-current liabilities or an increase in equity and investment by shareholders. Depending on the reason for the fall in Gearing, it may indicate that the fall in profit is not necessarily due to a lack of sales or a lack of customers but instead just due to the company reducing liabilities. Although Tesco is one of the largest retailers in the UK, there has been a rising interest in the online grocery market mostly due to the rising influence of the Corona Virus which left the opportunity to “improve” business methods. Comparisons were made with a leading online retailer “Ocado”. Figure 6 shows the Gearing (%) of Ocado over a 5-year period, in comparison it does not on face value appear to have a general trend like the Tesco Gearing (%) does with periods of increasing and decreasing percentage. Profits/losses before taxation of Ocado paint a bleak picture in comparison to that of Tesco, only in the first year did the profit stay above zero with all the other years falling into a loss. It appears, based on this data, online sales does not surpass the business model in place currently at Tesco.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Figure 5 and 6) Graph on Gearing from 2017 – 2021 (Left w/Tesco) and Right 2021 – TBC (Right w/Ryan

In conclusion, it appears that analysis of financial data does give interesting insights and also allows for hypothesis and conclusions to be drawn. It is essential for a business so large to implement analytic techniques as we can begin to allow for the business to run more efficiently as expressed in the report. Suggestions can be made as required given the right data which in turn should benefit the business. On the other hand, the data given appears to be enough to make suggestions based on trends it does not have enough data to accurately gauge if the trends have a causation or if they have any link at all between variables thus making it difficult to gauge what suggestions would have the highest success rate if implemented. I would suggest on a grand scheme increasing the number of stores and employees, this may affect the overall budget on salaries but data does suggest it may also increase the profit and turnover thus making it worthwhile. In addition to this, I would suggest creating a marketing campaign geared at increasing the number of people going out to the retail stores as this also appeared to be one way to promote the stores rather than the Dr. I would also suggest greater focus on instore shopping due to online shopping platforms losing money overall it may seem.

Reference List

Abril, R. M. (2011) The Quality Attribution in Data, Information and Knowledge. Encyclopedia of Knowledge Management.

Adamopoulos, P., Ghose, A. and Todri, V. (2018) The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms. Information Systems Research 29 (3), 612-640.

Akerkar, R. (2019) Introduction to Artificial Intelligence. Artificial Intelligence for Business. Cham: Springer International Publishing. 1-18.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J. and Dedene, G. (2002) Bayesian neural network learning for repeat purchase modelling in direct marketing. European Journal of Operational Research 138 (1), 191-211.

Biewald, L. (2015) CrowdFlower 2015: Data Scientist Report. CrowdFlower:

Caetano, J. A., Lima, H. S., Santos, M. F. and Marques-Neto, H. T. (2018) Using sentiment analysis to define twitter political users’ classes and their homophily during the 2016 American presidential election. Journal of Internet Services and Applications 9 (1), 18.

Cielen, D., Meysman, A. D. B. and Ali, M. (2016) Introducing Data Science : Big Data, Machine Learning, and More, using Python Tools. Place of publication not identified: Manning Publications.

Desjardins, J. (2019) How much data is generated each day? Visual Capitalist, World Economic Forum.

Hendershott, T., Zhang, X., Zhao, J. L. and Zheng, Z. (2021) FinTech as a Game Changer: Overview of Research Frontiers. Information Systems Research 32 (1), 1-17.

Kulkarni, U., Robles-Flores, J. and Popovič, A. (2017) Business Intelligence Capability: The Effect of Top Management and the Mediating Roles of User Participation and Analytical Decision Making Orientation. Journal of the Association of Information Systems 18, 516-541.

Shaaban, E., Helmy, Y., Khder, A. E. and Nasr, M. M. (2012) A Proposed Churn Prediction Model.

Sharda, R., Delen, D. and Turban, E. (2018) Business intelligence, analytics, and data science : a managerial perspective. Fourth edition.

Global edition. edition. Harlow, England: Pearson.

Sharda, R., Delen, D. and Turban, E. (2021) Analytics, data science, & artificial intelligence : systems for decision support. Eleventh edition.

Global edition. edition. Harlow, England: Pearson.

Turing, A. M. (1950) Computing machinery and intelligence. MIND LIX (236), 433-460.

Top 10 UK Retailers. (2021) Retail Economics: Retail Economics  2021.