UNIVERSITY OF LIVERPOOL

COMP516 - Research methods in computer science

LITERATURE REVIEW : Widely used Models of predictive analysis and its Applications in Healthcare, Manufacturing, Finance and Retail sector

INTRODUCTION

Why Predictive Analysis?

Data is the new gold and the ability to understand and analyse huge data is a key to business success.

Big data and predictive analysis is a huge and emerging field in computer science and a lot is still untapped.

Some member of our team upon graduation may choose to pursue a career in this field and this is a good step in understanding the field of predictive analysis

Project Objectives

- To review papers in this field and understand the concept and terminologies in Predictive analysis.
- To review the fundamental methodologies of predictive analysis.
- To discuss the application of these Methods in a few selected Industries and its benefits
- To highlight the most important developments in this field and across selected industries.
- To review the fundamental methodologies of predictive analysis .

Importance of Predictive analysis

PREDICTIVE ANALYSIS

- Informed decision making in business
- Customer Profiling and Segmentation
- Predicting patient patterns and disease diagnosis
- Cash flow prediction
- Fraud detection

Widely used Predictive models

Linear Regression	It is used to predict one variable's outcome which is dependent on another variable.
Logistic Regression	It is a classification model to predict categorical outcome using input variables. Its of two types, binary and multinomial.
Decision tree	This can be used to extract a suitable algorithm that predicts the best choice mathematically.
Time series analysis	Analysing a dataset collected over an interval of time.

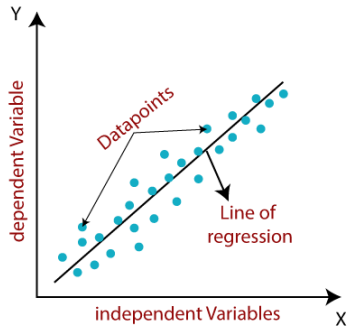

Figure 1: Linear Regression model graph[13]

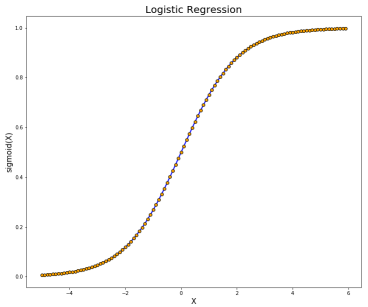

Figure 2: Logistic Regression model graph[14]

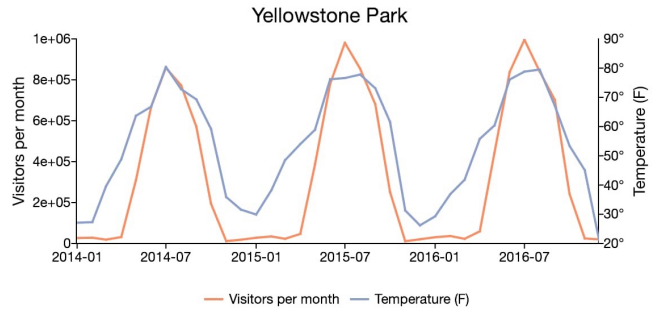

Figure 3: Time Series model graph[15]

RETAIL

"Exploring the determinants of and predicting the helpfulness of online user reviews using decision trees".[1]

- With large number of reviews available online, it's hard to decide the helpfulness of a product review.
- Positive reviews for a product can influence competition between retailers.

"Decision Tree Technique for Customer Retention in Retail Sector, Integrated Computing Technology, First International Conference, INTECH 2011".[2]

- Decision trees are more efficient in classifying customers.
- Based on retail dataset and millions of transactional records a decision tree model can be implemented to improve customer retention.

"Comparing the day temperature and holiday effects on retail sales of alcoholic beverages – a time-series analysis".[3]

- Time series analysis technique was applied to observe retail sales of alcoholic beverages over a period of time considering factors like temperature, holidays, seasonal/ non seasonal etc.

HEALTH CARE

"Predicting adoption of colorectal cancer screening among Korean Americans using a decision tree model".[4]

- Used Decision tree for recommending CRCS adoption based on self-efficacy, health status, risk of colorectal cancer for Korean Americans.
- The tree was designed based on mean scores calculated on each of these factors.

"Logistic regression was as good as machine learning for predicting major chronic diseases".[5]

- Variable importance mainly ethnicity, hb1ac, age, eGFR for major chronic diseases were calculated for logistic regression analysis.

FINANCE

"Measuring financial performance of Indian manufacturing firms: application of decision tree algorithms".[6]

- Return on equity, return on assets, return on capital employed performance are measured using financial ratios.

"This research used logistic regression model to predict the performance by accounting and financial variables of the nonfinancial firms listed in Pakistan stock exchange.".[7]

- Logistic Regression is used to predict the stock performance of non-financial firms by accounting and financial ratio variables.
- The model classified nonfinancial firms based on maximum likelihood of logistic method. The outcomes were of two types, good or poor performance based on the financial ratios and accounting variables.

MANUFACTURING

"Ontology-based decision tree model for prediction in a manufacturing network, Production & Manufacturing Research".[8]

- WEKA was used to make the predictive decision tree
- Attributes used for the decision tree – Product value, Supplier code, LOT value, Qty, Requested delivery days, On-time delivery

"Application of Logistic Regression for Production Machinery Efficiency Evaluation".[9]

- A model was built for the evaluation of machine efficiency based on available data.
- The variables selected were shift, device, occurrence of failure (yes or no) and no production order (yes or no).

FURTHER STEPS

- Further review of research papers related to predictive analysis (specifically reviewing more articles on the model application across these mentioned industries)
- Continuing on in-depth analysis of models with their applications.
- Trying to compare each model and attempt to extract insights about their efficiency over various applications.
- Identify challenges and future of predictive analysis in this industries.
- Create the structure of our paper using the top down approach.
- Review Final paper as a group and arrive at a conclusion.

REFERENCES

1. Zakir J, Seymour T,Berg K. Big Data Analytics. Issues in Information Systems, 16(2): 81-90, 2015.
2. Kasereka Henrys. Role of Predictive Analysis in Business. Master's thesis, Bugema University, Uganda, April 2022.
3. Mishra N, Silakari S. Predictive Analysis : A Survey, Trends, Applications, Opportunities & challenges. International Journal of Computer Science and Information Technologies, 3(3) : 4434 - 4438 , 2012.
4. Lee, S. and Choeh, J.Y. (2017), "Exploring the determinants of and predicting the helpfulness of online user reviews using decision trees", *Management Decision*, Vol. 55 No. 4, pp. 681-700. https://doi.org/10.1108/MD-06-2016-0398
5. Rose Tinabo, Decision Tree Technique for Customer Retention in Retail Sector, Integrated Computing Technology, First International Conference, INTECH 2011, pg 123.
6. Hirche, M., Haensch, J. and Lockshin, L. (2021), "Comparing the day temperature and holiday effects on retail sales of alcoholic beverages – a time-series analysis", *International Journal of Wine Business Research*, Vol. 33 No. 3, pp. 432-455. https://doi.org/10.1108/IJWBR-07-2020-0035.
7. Jin SW, Song CS. Predicting adoption of colorectal cancer screening among Korean Americans using a decision tree model. Ethn Health. 2022 Feb 9:1-15. doi: 10.1080/13557858.2022.2035693. Epub ahead of print. PMID: 35138199; PMCID: PMC9360189.
8. Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, Wong TY, Cheng CY. Logistic regression was as good as machine learning for predicting major chronic diseases. J Clin Epidemiol. 2020 Jun;122:56-69. doi: 10.1016/j.jclinepi.2020.03.002. Epub 2020 Mar 10. PMID: 32169597.
9. R.L., M. and Mishra, A.K. (2022), "Measuring financial performance of Indian manufacturing firms: application of decision tree algorithms", Measuring Business Excellence, Vol. 26 No. 3, pp. 288-307. https://doi.org/10.1108/MBE-05-2020-0073
10. Ali, S. S. ., Mubeen, M., Lal, I. ., & Hussain, A. . (2018). Prediction of Stock Performance by Using Logistic Regression Model: Evidence from Pakistan Stock Exchange (PSX). Asian Journal of Empirical Research, 8(7), 247–258. https://doi.org/10.18488/journal.1007/2018.8.7/1007.7.247.258
11. Zalan Mahmood Ayaz Khan, Salman Saeidlou & Mozafar Saadat (2019) Ontology-based decision tree model for prediction in a manufacturing network, Production & Manufacturing Research, 7:1, 335-349, DOI: 10.1080/21693277.2019.1621228
12. Anna Borucka & Małgorzata Grzelak (2019) Application of Logistic Regression for Production Machinery Efficiency Evaluation,9(22), 4770 DOI: 10.3390/app9224770
13. https://static.javatpoint.com/tutorial/machine-learning/images/linear-regression-in-machine-learning.png
14. https://miro.medium.com/max/1400/1*QY3CSyA4BzAU6sEPFwp9ZQ.webp
15. https://www.aptech.com/wp-content/uploads/2019/09/ts-pp-yellowstone-1.jpg