



The perspective of the opinion mining machine learning healthcare system

Sonal Sakalle^{1*}, Sakshi Rai²

1. Research Scholar, LNCT University, Bhopal, M.P., India

sonalsakalle@gmail.com ,

2. Professor, LNCT University, Bhopal, M.P., India

Abstract: This article presents opinion mining, which is based on the methods of mathematical statistics and machine learning, describes the features of applying regression analysis methods in the machine learning systems. The developed machine learning model includes the regression analysis modules based on the Bayesian linear, artificial neural network, decision tree, decision forest, and linear regressions. In the process of applying this machine learning model, using the mentioned algorithms, the corresponding regression models were constructed and their comparative analysis was performed, the results were analyzed. The results obtained indicate the feasibility of using opinion mining in the medical research using machine learning systems. The presented methods can serve as a basis for strategic development of a new directions of the medical data processing and decision-making in this field. We have identified the prospects for further research aimed at applying opinion mining methods to the healthcare system, namely, clustering, classification, anomaly detection

Keywords: Machine learning, Opinion mining, Clustering, Classification, Healthcare System

----- X -----

INTRODUCTION

The Healthcare industry collects huge amounts of healthcare data which, unfortunately, are not “mined” to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced opinion mining techniques can help remedy this situation. Opinion mining is often used during the knowledge discovery process and is one of the most important subfields in knowledge management. Opinion mining aims to analyze a set of given data or information in order to identify novel and potentially useful patterns (Fayyad et al., 1996). These techniques, such as Bayesian models, decision trees, artificial neural networks, associate rule mining, and genetic algorithms, are often used to discover patterns or knowledge that are previously unknown to the system and the users (Dunham, 2002; Chen and Chau, 2004). Opinion mining has been used in many applications such as health care, marketing, customer relationship management, engineering, medicine, crime analysis, expert prediction, Web mining, and mobile computing, among others [1,5,7]. The development of Information Technology has generated large number of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Opinion mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.

Opinion mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown. Once these

patterns are found they can further be used to make certain decisions for development of their businesses. Three steps involved are

1. Exploration
2. Pattern identification
3. Deployment

Medical knowledge advances as a result of advancements in information and digital technology used in healthcare. For instance, individuals with unique genetic profiles may now get highly focused medicines tailored to their needs. Utilising genetic data for therapy selection, however, requires access to diverse big data sets in order to provide statistically reliable recommendations [2]. By guaranteeing transparency of data transmission and audit of their usage, new methods of data gathering and exchange, like blockchain, might guarantee secrecy and security [2]. The development of automated decision-support tools is made feasible by advancements in artificial intelligence, namely machine learning, which allows for the creation of information and digital systems of machine learning.

The goal of opinion mining methods [1] is to discover associations in large datasets by use of computational algorithms. To aid in improved decision-making, they use heuristic approaches, which basically include looking for pertinent patterns and characteristics and then extracting more feature information [2]. Many other kinds of opinion mining algorithms exist [3]. The most popular ones are those using machine learning, decision trees, association rules, and neural networks.

Logistics, COX, lin-ear, random forest, support vector machines, NaiveBayes, and more recent advancements like KNN, GBDT, Histogram-based Gradient Boosting, XGBoost, LightGBM, and CatBoost are all part of machine learning algorithms. Particularly CatBoost, XGBoost, and LightGBM, these state-of-the-art algorithms have shown to be far superior to more conventional modelling approaches. In terms of performance, these three algorithms can compete with any sophisticated machine learning algorithm in the world. Practical data modelling and validation in real-world settings are necessary for finding the appropriate method for model creation. In my own experience, I have worked on several algorithmic modelling projects, including "Artificial Intelligence Learning of HUA Susceptible Gene Molecular Typing and Risk Prediction" and "Construction of Chronic Disease Prediction Models and Applications Based on Opinion mining." Over the course of my academic career, I have worked with tens of thousands of medical big data cases, using computer methods for modelling and validation. I have extensive experience with a wide range of tools, including statistical programs like SPSS and MedCalc as well as R, Python, ChatGPT 3.5, and others. As more and more statistical procedures are studied, statistical theory also grows. In addition to having some practical experience, I have written and published a number of scholarly articles on medical computer modelling and validation of chronic illnesses.

LITERATURE REVIEW

Nilashi, M., Ibrahim, O., Samad, S., Ahmadi, H., Shahmoradi, L., & Akbari, E. (2019) Ischemic stroke therapy success rates were calculated using a number of machine learning methods. The authors take into account DNN, RF, and LR methods for the same purpose. Two thousand forty-three patients are used to assess the efficacy of a machine learning method; the results show that deep learning provides the most

precise prediction of Ischemic Stroke outcomes.

Patient records and real-time data were merged and analyzed by Ali, L., Rahman, A., Khan, A., Zhou, M., Javeed, A., and Khan, J. A. (2019) to provide a diagnosis recommendation. In addition, the author explores the difficulties associated with the quality of services for healthcare applications.

In this study Kanimozhi, U., Ganapathy, S., Manjula, D., & Kannan, A. (2019), we create a u-healthcare monitoring system that utilizes the Internet of Things and Cloud to Fog(C2F) computing. Enhanced communication between healthcare facilities and smart homes is made possible by the suggested approach. When compared to cloud-based systems, it processes data quickly and with fewer delays, and it can meet all the needs of developing models.

Gokulnath, C. B., & Shantharajah, S. P. (2019), is an application of the support vector machine for predicting cardiovascular disease. In the context of illness diagnosis, it is a dependable and adaptable healthcare monitoring system. The goal of this system is to keep tabs on patients who are at a different location using their mobile phones and a mix of wearable devices and embedded technologies. The suggested monitoring system may also interact with first aid software to facilitate rapid patient rescue. In addition, patient information is gathered by using an ontology based on user models. In order to make the health monitoring system and first aid software compatible for users in their homes, physicians and first aid software utilize this information. The diagnostic system also has certain custom-made fuzzy rules. Overall, the purpose to suggested monitoring system is to recommend the appropriate therapy in case of serious health problems.

Expert systems and fuzzy logic were used to create the Ebola fuzzy informatics system by Kirk, R. A., and D. A. Kirk (2017). The suggested system is intended for Ebola virus disease diagnosis and advice. There was also a survey taken on the topic online. It has been shown that 31% of the general public is unaware of the therapeutic options available for Ebola. There were others, at 28%, who thought it could be healed. In a recent survey, 41% of respondents believed that Ebola is spread by the air and water, while 34% disagreed. It is also observed that 24% of individuals are unaware of how the Ebola virus is spread. While 23% are aware that it is spread by insects like mosquitoes, another 30% are unaware that this is even a possibility. Forty-five persons are used for a more in-depth exam. In addition, therapy for Ebola virus disease is recommended by 77% of the population. While 84% of respondents agree with the statement in question, 16% strongly disagree. In addition, 82% of respondents believed the suggested system to be simple to use, compared to 13% who disagreed and 20% who were ambivalent.

OPINION MINING

Opinion mining is a technique for assessing the positive, negative, or neutral polarity of a piece of text. Without any human input, machine learning technologies may learn to understand emotions based on examples gleaned from texts. The text categorization aids in categorizing the polarity of the viewpoints expressed in the text. Positivity, negativity, apathy, ambivalence, favorability, unfavorability, etc., Managing enormous volumes of data requires theoretical expertise in this field of preparation and analysis. Discovery data, carriage data, stock transaction data, grid data, internet data, web-index data, business-data, health-data, and many more are all examples of legitimate sources of a plethora of information. Businesses may benefit from real-time data from a number of sources, including consumer behavior and

forecasts.

Optimizing opinion analysis that recognizes workers' attitudes based on the data obtained, carefully examining the material to arrive at illuminating findings, and unearthing buried feelings are all ways to achieve this goal. By obtaining preliminary findings and predicting outcomes via fruitful study, intelligent analysis of the views of big datasets created the groundwork for in-depth investigation. In many fields, including business, stock trading, recommendation models, smart outcomes, legislation, politics, service sectors, etc., it has become a new and inventive template.

As a field of study, opinion analysis wasn't there until a Web application, together with product marketing, administrative research, advertising, legislative concerns, proposals, and business were among his many enlargements by 2006. It has become clearer that people of all ages, from youngsters to seniors, are actively sharing their thoughts and ideas in various online forums. In order to glean information that may be used in analytics based on people's feelings, an opinion analysis method has arisen. Although there are few evaluations that touch on the topic of big data and conventional approaches such task categorization, development, and perspective analysis, these studies seldom discuss their combined potential.

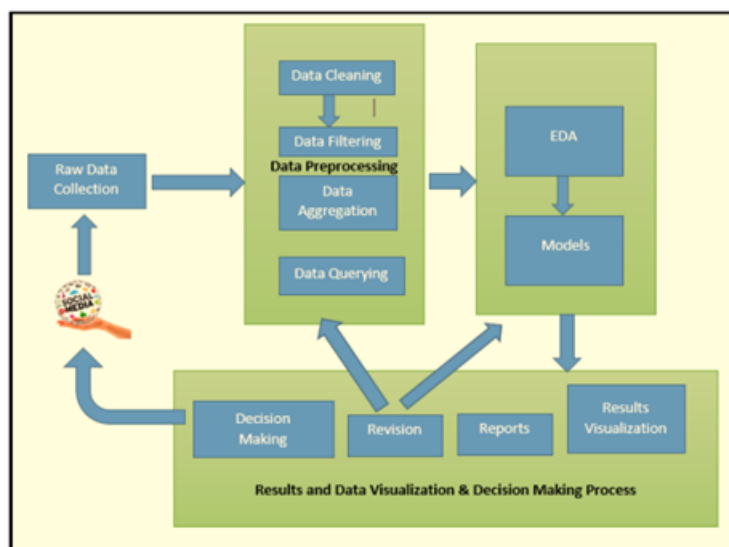


Figure 1: Process of Data

More specifically, taking into account opinion mining and its application, and filtering meaningful information from enormous volumes of data, may enhance the quality of healthcare decision-making.

MACHINE LEARNING

Machine learning (ML) is a subfield of artificial intelligence (AI) that focuses on generalizing inferences from a data sample to solve a problem. Generally speaking, learning progression has to predict the unknown requirements required by the system for a given dataset. The initial outputs of the structure may be predicted by computing the dependencies later. By using machine learning and psycholinguistic elements, is able to assess the polarity of drug users' attitudes in the year 2020. Machine learning (ML) is an emerging and rapidly expanding field in the study of health, with a variety of applications in which generalization is done by various algorithms and approaches that explore an n-dimensional space for a given collection of samples.

Machine Learning Techniques

Machine learning (ML) has emerged as a stimulating field with several applications in biomedical research, and ML-based approaches are becoming more prominent. By analyzing large data sets and making predictions about different kinds of cancer, ML can help uncover their patterns and linkages. A strategy that combines several types of clinical data is used in the suggested study. However, all models have the same issue: insufficient external validation, accuracy, and forecast of the models' predictive performance.

By examining the data and their connections, machine learning algorithms are able to glean useful insights from massive datasets. This study employs supervised and unsupervised learning paradigms from the realm of machine learning. Supervised learning methods include those used for classification, regression prediction, and time series analysis. It helps construct models that are used to categorize novel and previously unknown data. Unsupervised learning relies on clusters to identify commonalities and differences in a dataset.

Supervised Learning

Opinion inquiry may be treated as an order job, where machine learning approaches are a plausible computational instrument for this issue. One branch of machine learning techniques analyzes previously collected data to provide an emotional context for a given piece of text. Subsequently, it encourages the layout of grouping borders. It incorporates a model that may be used to describe secret data in accordance with the judgment they convey.

To produce a classifier by learning from training samples was stated as the objective of a supervised learning algorithm by Elkan and Charles in 2010. Predictions may be made using test examples and a classifier. The term "supervised approach" is used to describe this strategy. The data is provided as a labeled dataset from which a model may learn to predict the result of the problem. The two main categories of problems that may be tackled using supervised learning are classification difficulties and regression problems. The expected worth of an independent variable may be calculated using a categorization code. The input data might be seen as belonging to a certain category. Think of a collection of animal photos with labels like "tiger," "lion," etc. The algorithm then has to decide which of these groups the newly-discovered photos belong to. Some examples of classifying algorithms are as follows:

- Logistic-Regression
- Naïve-Bayes-Classifier
- Support-Vector-Machines

HealthCare Perspective

The majority of the health field is rapidly developing for the better. It improves machine intelligence via higher-resolution diagnostic imaging and other forms of enlightenment technology. For a short time, it will swap out all of your broken organs. Even if money and resources are restricted by technology, it often seems like there are endless opportunities. Commonly used systems exist to provide efficient administration of healthcare resources, an area known as (CDSS). A CDSS is implemented to facilitate the

clinician's decision-making and to collect the clinician's prior knowledge and expertise. CDSS aims to improve the ability to predict the occurrence of future health problems. The ability to accurately forecast medical outcomes is critical for medical care systems to allocate their resources (such as doctors, hospital beds, and other personnel) and plan future actions. Unfortunately, despite the efficacy of many of the established therapeutic procedures, progress in this field remains slow. All the established approaches rely on biochemical markers as illness indicators. In most cases, there is no direct correlation between an indicator and a medical outcome; an indicator's presence alone does not strongly predict that a patient would acquire a disease. This indicates that the interplay between these factors is not as simple as with traditional physician prediction procedures based on productive signs, which contributes to the mediocre effectiveness of sickness prediction. By fusing together health perspectives and cutting-edge mathematical principles, a potent help tool is created to advance the healthcare sector's designated toughest outcomes.

CONCLUSION

Helpful tools for healthcare organisations and clinicians to use when making decisions. To fully use ML in healthcare, it is important to acknowledge the limits of current datasets and understand how a system models the intricate healthcare ecosystem. When discussing healthcare data from a systemic viewpoint, we mean an all-encompassing database that records all financial, operational, and clinical details of care procedures at the person, group, and population levels. The Institute of Medicine suggests including social and behaviour measurements into electronic health records (e.g., degrees of isolation, physical activity, geocoding, stress, etc.) in order to better understand the many elements that contribute to complex illnesses and impact the health of both individuals and populations [26]. Interoperability across various EHR systems is a challenging but not impossible problem to solve [19, 27]. Creating scalable and secure data warehouses to store sensitive medical information and provide responsible access to researchers when needed will simultaneously need fresh research problems in the fields of data security and differential privacy as we rejigger EHRs. Due to the fact that data scientists and administrators are gaining access to EHRs alongside medical practitioners, the reorientation of ethical and legal regulations is necessary to accommodate this dual-purpose usage of EHRs.

References

1. Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., & Akbari, E. (2019). A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. *Journal of infection and public health*, 12(1), 13-20.
2. Zhao, D., Liu, H., Zheng, Y., He, Y., Lu, D., & Lyu, C. (2019). A reliable method for colorectal cancer prediction based on feature selection and support vector machine. *Medical & biological engineering & computing*, 57(4), 901-912.
3. Kanimozhi, U., Ganapathy, S., Manjula, D., & Kannan, A. (2019). An intelligent risk prediction system for breast cancer using fuzzy temporal rules. *National Academy Science Letters*, 42(3), 227-232.
4. Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, 22(6), 14777-14787.

5. Kirk, R. A., & Kirk, D. A. (2017, July). Introducing a Decision-Making Framework to Help Users Detect, Evaluate, Assess, and Recommend (DEAR) Action Within Complex Sociotechnical Environments. In International Conference on Human Interface and the Management of Information (pp. 223-239). Springer, Cham.