International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I9P107

Volume 72 Issue 9, 44-47, September 2024

Case Study

Smart Health Companion: An App for Tracking Diagnoses and Recommending Next Steps

Kabita Paul

Principal Consultant, Genpact, NC, USA.

Corresponding Author : kabitapaul11@gmail.com

Received: 21 July 2024 Revised: 22 August 2024 Accepted: 13 September 2024 Published: 29 September 2024

Abstract *- With the overflow of health information online and misinformation in social media, there is a need for a reliable source of consumable health information along with a platform to keep track of medical records. This research project aims to bring a data science approach to the public health and health literacy domain to guide users with reliable health information and enable them to track their medical records. This project aims to build a web-based and mobile dashboard that enables users to enter their medical records like diagnosis values, create to-do lists and reminders, track health progress, and feed medical prescriptions. This project of building customized interactive tools is multi-disciplinary and collaborates with departments like Computer Science, Psychology, Healthcare, and Communication. The results of this research are expected to be a guide for further application development. Moreover, the analysis should be able to help and guide the preparation of user testing questionnaires and focus group/ survey questions for further data collection.*

Keywords *- Recommendation engine, Health informatics, Health tracker, Medical diagnosis, Health recommender.*

1. Introduction

With increasing digitization, people have access to a broad range of health-related data available [5]. Information that is related to one's state of health can come from different sources like online media or articles, electronic health records from the clinic, YouTube videos, and social media like Facebook, Twitter, and many more. Information is also available in different formats like activity tracker data stream, family history and genealogy, Electronic Medical Records (EMR) etc. People discuss their health condition with clinicians and family members, and they seek information online as well. As digitized information is constantly increasing, changing, and becoming enriched with new types of records, it has been challenging for people to decide which health information can be relied on. Health misinformation on social media and lack of capacity for health literacy are other constraints to consuming available health information with complete reliance [6]. It has become very important to define which health data is important to collect, store, manage, and share [10]. A person in a chronic condition [3] might want to monitor his/ her health progress between clinic visits. To bring the benefit of Big Data research, a Data Science approach is needed to meet the requirements of customers/ patients whose health literacy, education, and language skills affect the understandability of health information. The purpose of this research project is to undertake efforts towards building capacity for bringing artificial intelligence to health information by collecting and studying the pilot data and scientific literature in the field of digital health libraries.

2. The Solution

2.1. Product

This project broadly aims to build a tool that is envisioned as a browser-based dashboard tool, along with a smartphone app and personalized digital assistance that allows the patient (1) to annotate data for themselves (particularly important for lab values like glucose or physical activity); (2) create calendars and to-do lists, and reminder functions that are linked so that any entry allows for auto-population within the other sections; (3) generate push notifications for new electronic health record data; (4) allow to feed photos of medications taken via smartphone etc. The tool will emphasize customers with chronic pain and cancer patients more. The goal of this data analytics research is to collect and examine public health records and identify key health attributes using this ML model to be valuable for prototype building.

2.2. The Data Science Solution/ Methodologies

Keeping patients healthy and avoiding the worsening of disease is at the front of the priority list in the healthcare industry [7]. By collecting and analyzing aggregated data, key attributes that are most important to store and manage for specific medical conditions can be identified. By taking advantage of big data technologies collaborating with machine learning, recommendations can be made to patients with medical conditions or to people who want to achieve healthier lifestyles. The methodologies below will be used in this research.

3. Methods

3.1. Data Collection and Data Preparation

The main focus of this study is to gather and manage information on patients with chronic pain and cancer. We gathered and collected information online on cancer trial data. We found some useful datasets on Data World and Kaggle.com. We have also found some data on chronic back pain. However, combining various sources of data has been a challenge since those have different sets of attributes. We worked on overviewing the datasets and selecting the most relevant data for this project.

3.2. Exploratory Data Analysis

We have found some relevant data on the figshare.com website on Chronic pain in patients undergoing breast cancer surgery. This dataset looks very relevant to this analysis. Postsurgical pain is a relevant side effect following surgery that increases the risk of various complications and delays postoperative patient recovery. Lidocaine and magnesium are common anaesthetics used to reduce pain levels. This dataset contains pain level, and medication details for patients divided into three controlled study groups. The data has 116 patient records and 76 attributes. We wanted to see patterns in data and visualize the relationship between various factors and pain levels. At first glance, age, chemo level, and weight of the patient seem to be important factors. Therefore, we grouped the data by these factors and plotted them in the graph. We wanted to distinguish the behavior of various study groups. As height or weight alone did not provide much information, we created a new variable, "weight-height ratio", by combining them.

3.3. Feature Selection and Feature Creation

As we have 76 variables, we performed feature importance analysis using the Boruta package in R to select important predictors. After that, we ran our initial linear regression model to predict patients' pain levels. Also, we have created new features like weight-height ratio as they alone cannot contribute much value to the analysis. Below is the result of the Boruta package with few selected variables as the full model becomes illegible.

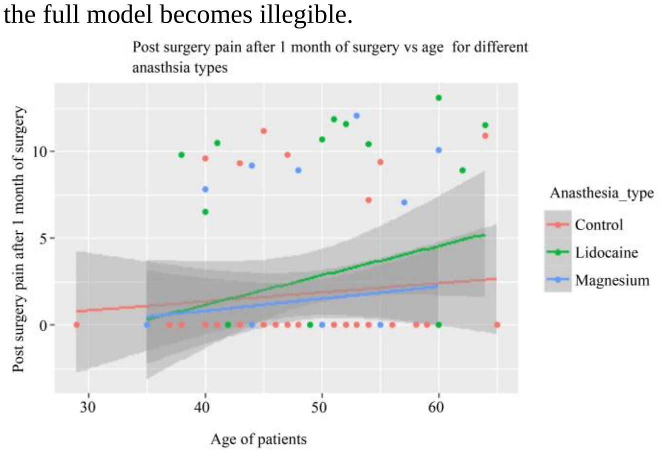

Fig. 1 Post-surgery pain after 1 month of surgery vs age

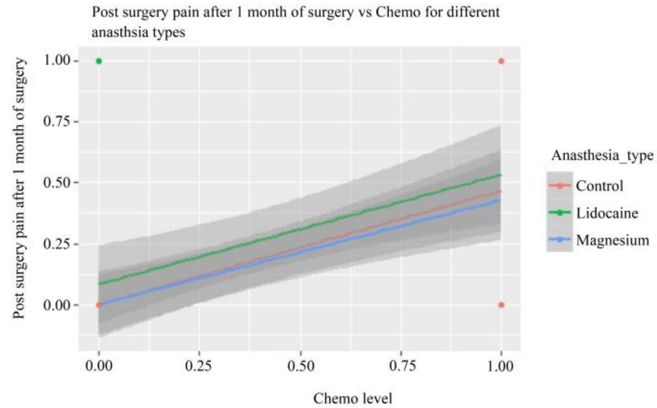

Fig. 2 Post-surgery pain after 1 month of surgery vs chemo

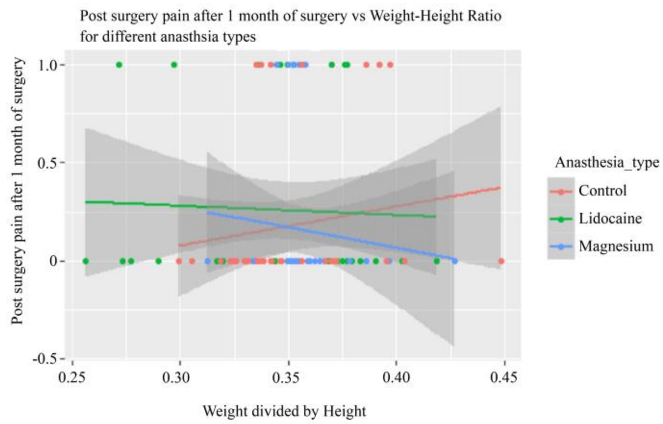

Fig. 3 Post-surgery pain after 1 month of surgery vs weight-height ratio

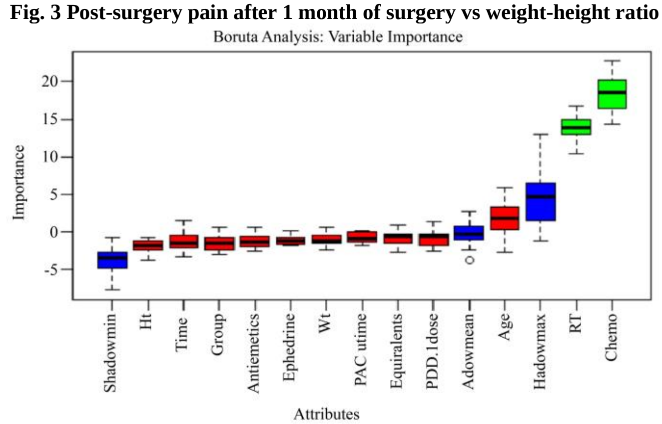

Fig. 4 Feature importance analysis result

3.4. Regression Analysis

The objective of the data analysis is to Analyze the effect of perioperative intravenous lidocaine infusion and magnesium infusion on functional recovery after general anesthesia in patients undergoing breast mastectomy [9]. 20 to 65-year-old female patients who were planning to undergo a mastectomy were included in this study and divided into three groups as per the type of anaesthesia provided. The target variable is chronic postsurgical pain (CPSP1 and CPSP3) after one month and three months of surgery. We performed Linear regression to determine how different factors contribute to pain level.

3.5. Recommendation System Design

We researched how to build a health article recommender system from scratch [1]. For existing systems, we have user preference data which helps to build a recommender system. There are three ways to recommend articles.

3.5.1. Popularity Model

In the article database, there will be some articles that many users like. We must decide the cutoff as per the number of entries in the database. Suppose there are 100 users and 50 articles. An article that is liked by more than five users (5%) can be considered popular.

3.5.2. Content-Based Filter

This filter is customized for users. If a user likes an article, we can recommend similar articles to the user. The question is how we measure similarity. A similarity matrix can be formed by analyzing the contents of articles. We need to decide the cutoff for the similarity value. The range of similarity values ranges from 0 to 1.

3.5.3. Collaborative filter

Users of similar preferences are recommended with similar articles. If users A and B like article C1 and user A likes article C2, then article C2 can be recommended to user B. All the above methods have a ‘cold start’ problem [8] which means new systems do not have enough data to implement the filters. Therefore, we need to first collect data by recommending random articles. The database structure should be something like below to start with.

Data table name: User Interaction
Fields: userId, content Id, user region, user Country, time stamp, Event Type, likes
Data table name: User Interaction:
Content Id, url, Title, text, language, topic

Step 1: Collect article links online to build the table below. We have to collect more relevant articles like the ones above to build the table. The initial target is 30 entries.
Step 2: Create dummy users and build a User_interaction table.
Step 3: Write business logic to implement in the system. Some of the important logics are:

1. Do not share the same article with one user. Check If the article is not already shared, follow other steps. If it was shared earlier and viewed by the user, then discard it.
2. Every time the user views and likes an article, save it to the user interaction database.
3. Filter relevant topics for the user. For back pain patients, we will filter out ‘back pain’ or ‘general’ articles and follow other steps as well.
4. If a user likes content, find other similar content using a similarity measure.

4. Results and Discussion

4.1. Prototype Building

We designed a web app and mobile dashboard prototype. Some of the important features are appointments, activity tracking, diagnosis reports, recommendations, diet, and articles.

4.2. Regression and Feature Importance Analysis

Only significant variables are shown in the table. Regression results are shown in Table 1. Aged patients experienced more severe pain (shown in Figure 1). Lidocaine anaesthesia caused more pain (shown in Figure 2). The level of Chemo and RT are very important deciding factors for pain (Figure 3). Magnesium caused less pain for patients with less Weight-height ratio. The effect is reversed in the control group.

4.3. Recommender System

We built a simple article recommender system using content-based filtering and personalized health data saved in users’ health libraries.

Table 1. Regression analysis R-Square = 0.5919 and C(p) = -11.1240

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.84112	0.62095	0.64856	8.79	0.0038
pre4	0.09330	0.02320	1.19347	16.18	0.0001
ephedrine	-0.17747	0.05598	0.74145	10.05	0.0020
Chemo	0.30374	0.09017	0.83718	11.35	0.0011
RT	0.20218	0.09918	0.30656	4.16	0.0441
MAP1	-0.00608	0.00298	0.30645	4.15	0.0441
MAP3	0.01128	0.00359	0.72676	9.85	0.0022
time2	-0.00853	0.00322	0.51747	7.01	0.0094
POD#1analg	0.16266	0.07576	0.34004	4.61	0.0342
POD#1ME	-0.02287	0.00823	0.57007	7.73	0.0065
PDD#1antie	0.47722	0.18446	0.49378	6.69	0.0111
PDD#1dose	-1.12935	0.55343	0.30721	4.16	0.0439
pod1	-0.07324	0.01584	1.57696	21.38	<.0001
pod5	-0.03672	0.01863	0.28647	3.88	0.0515

5. Key Concerns and Addressing Them

5.1. Data Privacy

As with some other industries, data privacy is a very big concern for the public health industry. We have read some research papers on data privacy in the Healthcare domain. It is essential to assure users that their data are safe and private and not shared with anyone else [4]. Let us take an example of how we addressed privacy concerns.

We have one feature in this app that pushes alerts about important medical information for that area. Suppose we would like to warn patients about some hospitals closing in Mecklenburg County due to COVID-19. We will not track the patient's location; rather, we will ask for the zip code, which is sufficient to implement the functionality.

5.2. User's Safety

We have gone through at least five research papers that are collected online. One of the papers that is found very helpful is "Safety first: Conversational Agent for Health Care" [3]. It explains what the potential dangers may lie in virtual health assistance when they do not perform as per expectation.

A natural language understanding system must have good models of all the off topics that users are likely to ask. Also, this paper emphasizes the privacy and safety aspects of health care virtual assistance.

5.3. App Engagement

App engagement starts from the time the user downloads the app from the app store and continues if the user uses the app. We studied the reasons for people giving up and not using the app. One of the important reasons might be that the app goes quiet for a small amount of time. When the user does not feel like using the app, and eventually deletes it. Push notification is identified as one of the important features that address the concern of app engagement. We designed this prototype to send users motivational messages and recommend relevant health articles depending on patients' medical conditions and health goals.

1. 6. Conclusion

The research has demonstrated that patients can track their health reports and diagnoses more efficiently using this recommender application. This study suggests that this health recommender application will bridge the gap between healthcare visits and report comprehension by enabling users with timely information and actionable insights. While this study shows great promise, future research and development are necessary to improve accuracy, enhance functionalities, and ensure proper integration with the healthcare system. It is also essential to expand the database to have sufficient use cases to improve the accuracy of results. This is a potential modern healthcare system promising to promote proactive health management and better health experience for its users.

References

[1] Andre Calero Valdez et al., *Recommender Systems for Health Informatics: State-of-the-Art and Future Perspectives*, Machine Learning for Health Informatics, Springer, Cham, pp. 391-141, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] Timothy Bickmore et al., *Safety First: Conversational Agents for Health Care*, Studies in Conversational UX Design, Springer, Cham, pp. 33-57, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ming Xia et al., "Effect of Intravenous Lidocaine on Chronic Postoperative Pain in Patients Undergoing Breast Cancer Surgery: A Prospective, Double-blind, Randomized, Placebo-controlled Clinical Trial," *Annals of Translational Medicine*, vol. 10, no. 14, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Kim Theodos, and Scott Sittig, "Health Information Privacy Laws in the Digital Age: HIPAA Doesn't Apply," *Online Research Journal Perspectives in Health Information Management*, vol. 18, 2020. [Google Scholar] [Publisher Link]

[5] Naresh Kasoju et al., "Digital Health: Trends, Opportunities and Challenges in Medical Devices, Pharma and Bio-Technology," *CSI Transactions on ICT*, vol. 11, pp. 11–30, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Salman Bin Naeem, and Maged N. Kamel Boulos, "COVID-19 Misinformation Online and Health Literacy: A Brief Overview," *International Journal of Environmental Research and Public Health*, vol. 18, no. 15, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] NEJM Catalyst, Healthcare Big Data and the Promise of Value-Based Care, NEJM Catalyst, 2018. [Online]. Available: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0290

[8] Kirk D. Borne, and Booz Allen, Data Scientist's Dilemma: The Cold Start Problem – Ten Machine Learning Examples, kdnuggets. [Online]. Available: https://www.kdnuggets.com/2019/01/data-scientist-dilemma-cold-start-machine-learning.html

[9] Effect of Perioperative Intravenous Lidocaine Infusion and Magnesium Infusion on the Functional Recovery After General Anesthesia in the Patients Undergoing Breast Mastectomy, Get well soon, 2022. [Online]. Available: http://getwellsoon.ai/trials/NCT02185859/

[10] Data Science Research: Personal Health Libraries for Consumers and Patients (R01 Clinical Trial Optional), National Institutes of Health Grants. [Online]. Available: https://grants.nih.gov/grants/guide/pa-files/PAR-19-072.html