

Sentiment Analysis and Demand Forecasting from Patient Reviews

Shayan Ahmed (Corresponding Author)

Department of Computer Science, IQRA University, Karachi, Pakistan

shayan.ahmed01@iqra.edu.pk

Muhammad Ahsan Hayat

Department of Computer Science, IQRA University, Karachi, Pakistan

<https://orcid.org/0009-0001-5063-7603>

muhammad.ahsan@iqra.edu.pk

Saira Riaz

Indus university

fsaira13@gmail.com

Muhammad Zamin Ali Khan

Department of Computer Science, IQRA University, Karachi, Pakistan

muhammad.zamin@iqra.edu.pk

ABSTRACT

The purpose of this paper is to offer pharmaceutical companies and physicians an opportunity to explore the experience of medicines. The challenge though is how to analyze the huge amount of unstructured text to be able to arrive at clear and useful insights. A solution in the present research is proposed which examines large volumes of reviews on medicines through natural language processing (NLP) and machine learning. A system has been designed to clean and process the text then extract features for analysis. In an attempt to quantify the sentiment of the patients, several classification models have been explored and the LinearSVC model is the most accurate and it succeeds in classifying sentiments with an accuracy of more than 94%.

Keywords: NLP, Sentiment Analysis, Machine Learning, Drug Reviews, Text Mining.

1. Introduction

The healthcare industry is rapidly shifting and there is an increased emphasis on how real-life patient experiences are used to inform all aspects of medical research, as well as the way pharmacies organize their shelves. Reviews that patients post online are among the most informative materials. These first-hand accounts are posted in medical

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

forums and drug websites, where they discuss the actual effectiveness of a drug, the side effects that they will experience and the general satisfaction with a particular treatment.

Although clinical trials offer important information in a controlled environment, reviewing patients can demonstrate the performance of a drug in real life among a large population. This real-world information can prove rather useful in predicting the future demand of drugs, suggesting drugs and providing recommendation about a supply chain.

The issue is the feedback from patient is crude and in text form hence difficult to analyze with the use of number crunching methods. This is a valuable but ignored in most of the regular forecasting. This is the reason a sentiment analysis is applied with the aid of NLP in classifying feedback as positive, negative or neutral. The study is being done by offering an approach of sentiment analysis onto patient review in predicting demand. A text is cleaned and processed by the use of NLP. For classification, methods to find useful features and machine learning models are being used. Furthermore, forecasting models are used on this sentiment pattern to aid in its trends. This two-component method is not only useful to get to know what the patient thinks, but results in a better pharmaceutical planning.

2. Related work/literature review

The possibility of how we can use blockchain, Internet of Things (IoT) or machine learning to transform pharmaceutical supply chains has prompted pharmacists to undertake a spirited effort towards finding a remedy. The high-profile initiatives such as K20 frame have demonstrated to us how blockchain together with predictive analytics can render drugs highly traceable, and the system in general is safer. With that said, the power of these approaches is that they offer transparency, and neither of these approaches generally disregards a potential source of information that is the raw sentiments of patients.[1]

A similar thing can be stated about the method of sentiment analysis, which evaluates the level of the sentiment of a text, whether it is a positive or a negative one. The method is commonly applied in other fields such as entertainment or marketing and in this context, one uses it to check the rating of a movie or the buzz on social media. Other health care uses include the use in health care to determine patient satisfaction or to create recommendation engines. What has lacked, though, is an important chance; little research has discovered effective methods of invigorating big-account examination of reviews delivered by patients with workable frameworks to forecast the demand.[2]

Majority of drug demand models have backward isoelectric episodes and sales on previous levels or medicated data. Each of these approaches lacks the comprehensive, up-to-date image of patient experiences and creates a fissure in our possession of ability to foresee the following demand by the market. Our research fills that gap. We introduce a mechanism of converting unstructured feedback obtained on patients into

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

objective sentiment indicators and putting them in predictions. We give a more dynamic, customer-oriented approach to foresee what would happen in terms of demand in the future, and facilitate wiser and adaptable supply chains.[3]

3. Proposed methodology/framework

A simple step is being used in converting crude review of patients to important sales forecasts. The system includes five large stages and they include data gathering, preprocessing, feature exaction, sentiment classification and thereafter prediction of the demand. A baseline population of more than 215,000 reviews of patients on the publicly available was selected initially.

UCI Drugs.com dataset 2008-2022. The individual records in this dataset have hard details such as name of the drug, the reason of the drug is administered, and so on and the number of stars rated and the written text constituting the written comment and a mark of how useful other people rated the comment.

Id	drugName	condition	review	rating	date	usefulCount
206461	Valsartan	Left Ventricular Dysfunction	"It has no side effect, I take it in combinati...	9	20-May-12	27
95260	Guanfacine	ADHD	"My son is halfway through his fourth week of ...	8	27-Apr-10	192
92703	Lybrel	Birth Control	"I used to take another oral contraceptive, wh...	5	14-Dec-09	17
138000	Ortho Evra	Birth Control	"This is my first time using any form of birth...	8	03-Nov-15	10
35696	Buprenorphine / naloxone	Opiate Dependence	"Suboxone has completely turned my life around...	9	27-Nov-16	37

FIGURE 1: PATIENT'S REVIEW DATASET

After that the text is cleaned to be analyzed which includes the removal of duplicates, punctuations, and stop words or so trivial common words. The text was then the term frequency abridged and the word tokenized the words in their radical scent forms (lemmatization) and systematized the writing to where it was even.

The words had to be decipherable to the machine learning algorithms to enable the machine learning consider the text experienced as a number. We sought to do well to extract this feature under two applications strategy warfare algorithms: TF-IDF and Count Vectorizer. These are algorithms that are relatively efficient map subjective text language into quantitative data value in the devotion of capturing the frequency and the number of words in the humungous set of reviews.

These examples of processed data were then trained to make a number of machine learning models- including

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

LinearSVC, Logistic Regression, Random Forest, Multinomial Naive Bayes and LightGBM to carry out sentiment analysis. Their work was to read every review and categorize its general positiveness, negativity, or even neutrality.

Thirdly, we determined the sentiment score of a drug as the number of such classification in the pre-case set periods of time and creating an index of sentiment. Associating the index and actual indicators of drugs demand assisted us in creating forecasting models relying on the changes in the grasp of common people in creating forecasts on the close future trends in popularity and demand of medication.

ratingSentiment	ratingSentimentLabel	vaderReviewScore	vaderSentimentLabel	vaderSentiment
2	positive	0.0000	neutral	0
2	positive	0.9070	positive	2
0	neutral	0.7096	positive	2
2	positive	0.7184	positive	2
2	positive	0.9403	positive	2

FIGURE 2: MODIFIED DATASET

4. Result and Analysis

The results of the analysis are quite promising, and feedback of patients can be taken as a very potent sentiment indicator. When applying the machine learning algorithms to test, the strongest contender was the LinearSVC algorithm, one of the highest cross validation scores of over 94% when provided with TF-IDF values. The next better algorithm was the stable and strong Logistic Regression, and its cross-validation scores was 90%. Multinomial Naive Bayes model produced good results with accuracy of 83% and LightGBM was at 83%. Interestingly, the Random Forest model did not suit well to perform the task as its accuracy decreased to less than 50%. When examining the errors of the models closer, confusion matrices were utilized and depicted that they succeeded in making most errors during the misclassifications of the neutral and negative reviews. This can be explained by the fact that the language in these classes may be quite subtle and ambiguous. Nevertheless, the models were highly successful in detecting both positive and negative sentiments with their accuracy and recall rates of positive and negative sentiments ranging above 90%. This assures that the forecasts were very precise and reliable.

But above all, we discovered a direct and significant connection between patient sentiment and the real-world demand. Medication with predominantly positive feedbacks always had good demand in the long-run. On the other hand, drugs with an increase in negative feedback also got appreciable reduction in their usage. An example is that the most common proportions were usually constant and high with medicines

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

tendentially characterized by words that described their effect such as effective, or life-changing. But on the other hand, the terms 'side effects' or, 'ineffective' which are used to classify characters were far more likely to experience reduced proportions used.

These quantitative findings we supplemented with qualitative spectacles such as word clouds. Such pictures provided an immediate example glance of the most frequent phrases in the highest-level sentiment categories and supported the readings of the statistics by revealing an actual voice of the patient experience.

	Count Vectorizer	TF-IDF
LinearSVC	0.939	0.945
LogisticRegression	0.943	0.903
MultinomialNB	0.853	0.832
RandomForestClassifier	0.494	0.494
LGBMClassifier	0.823	0.837

TABLE 1: CROSS VALIDATION SCORE

	Count Vectorizer	TF-IDF
LinearSVC	0.928	0.938
LogisticRegression	0.878	-
MultinomialNB	-	-
RandomForestClassifier	-	-
LGBMClassifier	0.76	0.834

TABLE 2: PREDICTION SCORE

5. Discussion

In this study, we explored an approach to perceive patient sentiment as a predicting demand for drugs, which was a relatively underused approach. We have set the possibility of making use of the rich qualitative data of the patient reviews within both the natural language and machine learning models in such a manner that predictions will be more accurate than those models based on number ratings. And with the world of pharmaceuticals looking on, it is not theory in this instance, it is reality. By adding this methodology, more discerning supply lines can be created where shortages and unneeded overstocking would be reduced sharply and inventory would be much more tightly and aptly aligned with actual patient needs.

The restrictions of the method should also be mentioned. Like all user-generated content, patient reviews are biased and subjective. We also had relatively low percentages of reviews who left unbiased comments in our corpus so that classification was an issue. The models were also formulated on based on historical up to 2022 and failed to factor in occurrences of sudden shocks such as regulations or grand campaigns that found their way to the demand.

Just despite such limitations, this study is a scaled template that is applicable to integrate

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

patient feedback in forecasting. The described system is not unique to the demand of drugs, and can be generalized to and applied to other areas of healthcare such as the real-time monitoring of patient satisfaction rates up to predicting adverse terms and driving the new era of customized care.

6. Conclusion and future work

The current literature can be seen as an expedient avenue to exploit patient comments as an instrument towards anticipating the need of medication. A sentiment we have seen to be of more or less degree. Non-probability of patients can elicit the reliability of talent, and the accuracy can be regarded as high and as a vertically suitable predictor of subsequent demand, through natural language processing and machine learning. LinearSVC had the best performance scores of 94 percent and higher, hence justifying the fact that sentiment analysis is an outstanding predictor in health professions. This additional work highly qualifies as something to be looked into in the future. Such frameworks as BERT could be used as a deep-learning tool and extract more fine-grain emotion cues through the text with its help. More dynamic and comprehensive in its predictions, the study should have incorporated reviews in a variety of languages and real time data retrieved in the social media. Also, predictions can also be further combined with external forces such as drug pricing, routing availability and current prescribing trends. In essence, this move will hinder the promotion of more patient-centered healthcare paradigm. Implementing an efficient and data-driven system by adjusting the supply chain to the actual voices of the patient can make the entire transformation into a realized organization that met the needs of the people that it serves.

References

- [1] Nguyen, A., Pellerin, R., Lamouri, S. and Lekens, B., 23. Dealing with the dynamics of demand peaks and declines of the pharmaceutical products in periods of disruption using the sentiment analysis of the news. *International Journal of Production Research*, 61(9) pp.2829-2840.
- [2] Shah, A., Yan, X.B. and Shah, S.A., 2021. Monitoring the healthcare experience of patients in the times of the COVID-19 pandemic: Sentiment analysis and topic modeling of patient ratings of doctors. *Diplomatic Journal of Engineering Research*, 9(3A).
- [3] Abirami, A.M. and Askarunisa, A., 2017. Sentiment analysis model to highlight the power of the online review within the healthcare industry. *Online Information Review*, 41 (4), pp.471-486.
- [4] Uddin, M., Salah, K., Jayaraman, R., Pesic, S., & Ellahham, S. (2021). Blockchain for drug traceability: Architectures and open challenges. *Health Informatics Journal*. DOI:10.1177/14604582211011228.
- [5] Haq, I., & Esuka, O. M. (2018). Blockchain technology in pharmaceutical industry to prevent counterfeit drugs. *International Journal of Computer Applications*. DOI:10.5120/ijca2018916579.
- [6] Bryatov, S.R., & Borodinov, A.A. (2019). Blockchain technology in the pharmaceutical supply chain. Researching a business model based on hyperledger fabric. URL: <https://ceur-ws.org/Vol-2416/paper18.pdf>.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- [7] Shah, Y., Verma, Y., Sharma, U., & Sampat, A. (2023). Supply Chain for Safe & Timely Distribution of Medicines using Blockchain & Machine Learning. DOI: 10.1109/ICSSIT55814.2023.10061049.
- [8] Meghla, T.I., Rahman, M.M., Biswas, A.A., Hossain, J.T., & Khatun, T. (2021). Supply Chain Management with Demand Forecasting of Covid-19 Vaccine using Blockchain and Machine Learning. DOI: 10.1109/ICCCNT51525.2021.9580006.
- [9] Tirkolaee, E.B., Sadeghi, S., Mooseloo, F.M., Vandchali, H.R., Amini, S. (2021). Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas. DOI: 10.1155/2021/1476043.
- [10] Muhammad Ahsan Hayat, Imran Ali Channa, Ubaidullah Khan, Nazia Alfred Fernandes, Urooj Tariq, & Khan Ikram Uddin. (2025). BRIDGING CLASSICAL STATISTICS AND MODERN AI TOWARD INTERPRETABLE DATA-SCIENCE MODELS. *Spectrum of Engineering Sciences*, 3(9), 525–537. Retrieved from <https://thesesjournal.com/index.php/1/article/view/1031>
- [11] Salim, M.M., Park, L., & Park, J.H. (2023). A Machine Learning based Scalable Blockchain architecture for a secure Healthcare system. DOI: 10.1109/ICTC55196.2022.9952962.
- [12] Roozkhosh, P., Pooya, A., & Agarwal, R. (2022). Blockchain acceptance rate prediction in the resilient supply chain with hybrid system dynamics and machine learning approach. DOI: 10.1007/s12063-022-00336-x.
- [13] G. Yang, Z. Pang, M. J. Deen, M. Dong, Y. Zhang, N. Lovell, and A. M. Rahmani. (2020). Homecare Robotic Systems for Healthcare 4.0: Visions and Enabling Technologies. DOI: 10.1109/JBHI.2020.2990529.
- [14] A. E. Azzaoui, T. W. Kim, Y. Pan, and J. H. Park. (2021). A Quantum Approximate Optimization Algorithm based on Blockchain Heuristic Approach for Scalable and Secure Smart Logistics Systems. DOI: <https://doi.org/10.22967/HCIS.2021.11.046>.
- [15] Y. Wang, D. Kim, and D. Jeong. (2020). A Survey of the Application of Blockchain in Multiple Fields of Financial Services. DOI: <https://doi.org/10.3745/JIPS.04.0185>.
- [16] M. A. Hayat, J. Baig, S. Ahmed, and A. F. Ayubi, "Towards Early and Accurate Disease Detection Through Multimodal Predictive Modeling: Fusion of Electronic Health Records, Medical Imaging, and Omics Data Using Interpretable Machine Learning," *Policy Research Journal*, vol. 3, no. 11, pp. 1–29, Nov. 2025.
- [17] M. T. Sarfraz, M. A. Hayat, S. Ahmed, M. Ali, and A. Murtaza, "Machine Learning-Based Intrusion and Anomaly Detection Models for Securing IoT Networks Against Emerging Cyber Threats," *Spectrum of Engineering Sciences*, vol. 3, no. 9, pp. 1445–1457, Sep. 2025.
- [18] M. A. Hayat, S. Ahmed, M. R. Khan, M. Zaka, F. Irfan, and R. Zaka, "Blockchain-Secured IoT Framework for Smart Waste Management in Urban Environments," *The Critical Review of Social Sciences Studies*, vol. 3, no. 3, pp. 1462–1467, Aug. 2025.
- [19] A.A.Mazlan, S.M. Duad, S. M. Sam, H. Abas, S.Z. A. Rasid, and M. F. Yusof. (2021). Scalability Challenges in Healthcare Blockchain System— A Systematic Review. DOI: 10.1109/ACCESS.2020.2969230.
- [20] Q. Zhou, H. Huang, Z. Zheng, and J. Bian. (2020). Solutions to Scalability of Blockchain: A Survey. DOI: 10.1109/ACCESS.2020.2967218.
- [21] UCI. DrugReviewDataset. Available online: URL: <https://archive.ics.uci.edu/dataset/462/drug+review+dataset+drugs+com>.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- [22] Musadique Hussain, Ahmed Faraz Ayubi, Satyadhar Joshi, Sumayya bibi, Nazia Alfred Fernandes, Muhammad Ahsan Hayat, & Huzaifa Hussain. (2025). META-LEARNING FOR FEW-SHOT AND ZERO-SHOT LEARNING: ENABLING RAPID ADAPTATION TO NEW TASKS WITH LIMITED DATA. Policy Research Journal, 3(11), 653–669. Retrieved from <https://policyrj.com/1/article/view/1311>