A Cognitive Machine Learning Framework for Personalized Drug Suggestion

Dr. P. Velayutham[1], Chiruthala Shiva[2], Datti Naveen[3], Dunna Someswara rao[4], Edha Sarika[5]
[1]*Associate Professor, Department of Computer Science and Engineering Bharath Institute of Science and Technology, BIHER, Chennai, India*
[2,3,4,5]*Student, Department of Computer Science and Engineering Bharath Institute of Science and Technology, BIHER, Chennai, India*

Abstract— **A recommendation system can assist the user to compose an understanding of requirements and propose informed decisions from a lot of complicated knowledge. Recommendation from an analysis of sentiments seems to be a great challenge as user-generated content is represented using human language in several complicated ways. Many studies have focused on common fields such as reviews of electrical items, films, and restaurants, but not enough on health and medical issues. Sentiment analysis of healthcare in general and that of the drug experiences of individuals, in particular, may shed considerable light on how to focus on improving public health and reach the correct decision. In this paper, we design and implement a drug recommender system framework that applies sentiment analysis technologies on drug reviews. The objective of this research is to build a decision-making support platform to help patients to achieve more significant choices in drug selection. Firstly, we propose a sentimental measurement approach to drug reviews and generate ratings on drugs. Secondly, we take how much the drug reviews are useful to users, patient's conditions, and dictionary sentiment polarity of drug reviews into consideration. Then, we fuse those factors into the recommendation system to list appropriate medications. Experiments have been carried out using Decision Tree, K Nearest Neighbors, and Linear Support Vector Classifier algorithm in rating generation and Hybrid model in recommendation based on the given open dataset. The analysis is carried out to tune the parameters for each algorithm in order to achieve greater performance. Finally, Linear Support Vector Classifier is selected for rating generation to obtain a good trade-off among model accuracy, model efficiency, and model scalability.**

Index Terms— **Drug Recommendation, Sentiment Analysis, Drug Review, Machine Learning, Natural Language Processing, Rating Generation.**

I. INTRODUCTION

With the explosion of Web 2.0 platforms, there are vast quantities of content created by users, called social media. Therefore, too many scholars have been researching competent algorithms for sentiment analysis of content created by consumers over the past decade. The area of sentiment analysis, also known as opinion mining, analyzes the opinions, perceptions, beliefs, judgments, attitudes, and emotions of people, including products, services, organizations, personalities, events and topics. In recent years, these two domains of application have received great interest.

In sentimental research, the analyses are generally divided into two categories, positive and negative. But if all candidates' products reflect positive or negative feelings it is difficult for people to make choices. In order to make a decision, people need not only to know if the product is good but also how good it is. It is also accepted that various people have different preferences for sentimental expression. So, it is more important to offer numerical scores rather than binary decisions in many practical cases such as drug recommendation and build a system of decision support that assists people in selecting products. This new application field presents both challenges and research opportunities in medical health.

A recommendation framework aims to predict the preferences of users and make suggestions that would be of interest to users. Collaborative filtering (CF), content based (CB), knowledge-based (KB), and hybrid recommendation technologies, all of which have certain limitations, are concluded by traditional recommendation technology. CB has overspecialized

recommendations and CF has issues with sparsity, scalability, and cold-start problem. But some researchers focused on drug recommendation system from user reviews and have proved that the sentiment analysis of healthcare in general and that of user's drug experience, in particular, could shed significant light on the process to improve public health and make the right decisions, and this system combines with traditional recommendation system is more effective. In our research, we are focused on opinion mining in drug reviews, in which patients share their experiences and opinions about medicines and then classify the opinions into ratings, and even recommend a medication list that would be most appropriate for the patient. Implementing the proposed approach to sentiment analysis will not just be useful to patients but also to pharmacists and clinicians for valuable public opinion summaries.

II. RELATED WORK

Recommendations techniques aim to provide consumers with personalized goods and services to cope with the growing issue of overloading online information. Studies used numerous methods for sentiment analysis, and since the mid1990s, recommender model strategies have been recommended. Many early researches concentrate on document-level study and refer to e-business, government, e-learning, e-commerce/e-shopping, etourism, etc. However, the world of medicine contains rare recommending technologies.

Mohammad Ehsan Basiri at el. suggest two deep fusion models for evaluating drug reviews focused on 3-way decision theory. The first fusion model, a 3-way fusion of a single deep model with a traditional training algorithm (3W1DT), was developed as a primary classifier using a deep learning method and a traditional learning process as a secondary method employed when there is weak trust in the deep method during the labelling of test samples. The 3-way fusion of three deep models including a traditional model (3W3DT), three deep and a traditional model is learned over the whole training data in the second suggested deep fusion model and each classifies the test sample uniquely. Then to categorize the test drug review, the most assured classifier is chosen.

Sairam Vinay Vijayaraghavan at el. decided to evaluate reviews of different medications that were analyzed in the form of texts and were also given a ranking on a scale of 1-10. In general, they divided the drug number rating into three classes: negative (1-4), neutral (4-7) or positive (7- 10). For the medications that belong to the comparable condition, there are many reports and they wanted to examine how various terms are used by the reviews for different conditions to influence the drug ratings.

Their goal was primarily to introduce supervised machine learning classification algorithms that use textual review to determine the class of the ranking. Various features, such as Term Frequency Inverse Document Frequency (TFIDF) as well as Count Vectors (CV), were primarily performed. In the data collection, they had learned models on the most common factors, such as 'Depression', 'Birth Control' and 'Pain'.

Vikrant Doma at el. introduced a drug recommendation assistant, constructed using machine learning methods as well as natural language processing, which generates its accuracy with several major datasets. The introduced method allows the contrasting effects, feedback, rankings, to be manifested and then recommends the most "effective" medication for a given person. The findings of the predictive study found that the death rates of the top deadliest diseases in the U.S. all rose significantly between 2005 and 2015, between the ages of 55 and 80. The next top five medical conditions (depression, birth control, pain, acne, anxiety), which will be widespread in the near future and the five medications used to treat them, can be predicted with some degree of precision based on the existing trends.

Ahmed Abdeen Hamed at el. are launching their initial work to build a recommendation scheme for social networks called T-Recs. The system is a time aware alternative medicine recommendation system based on Twitter. For three consecutive weeks, they compiled a collection of tweets containing unique hashtags (#throwingup, #headache, #itching, etc.) (about 500,000 tweets). The individual tweets were manually reviewed by a domain expert (a medical doctor) who examined the feelings of the tweet along with the timestamp of the tweet. The domain expert assigned the tweet/group of tweets a preliminary label using this knowledge. Authors then trained a classifier (Decision Tree algorithm) using the hashtags and their labels, taking into account other variables (i.e. age, gender, conditions of comorbidity) that must be given by the

Tweeter by taking a questionnaire. The recommender framework makes suggestions based on the contents of the tweet and the questionnaire factors for hereinbefore tweets after the questions are answered.

Chun Chenatel. use actual medical online cases and carry out large-scale data processing and build an Implicit Drug Recommendation Feedback and Crossing Recommendation (IFCR) framework then support doctors in drug choice. The new platform is intended to examine epileptics' psychiatric records in order to determine the relationship between the syndromes and the medications. The authors studied two recommendation algorithms and developed an ensemble model that merged IFCR with ANN to maximize their respective advantages in drug recommendation.

Rahul Majethia at el. describe People Save: a drug recommendation and feedback system for doctors based on internet-based contextual patient reviews. Authors filter information sources to check for the feasibility of crowdsourcing and then assess the effectiveness of the drug based on its reported detrimental effect on a patient. This helps to eliminate certain drugs, which would almost certainly have an adverse effect on the health of the patient and thus obtain a set of recommended drugs.

Chenetal. Provide a context-based, personalized antihypertensive drug recommendation system, and develop a contextual support framework. They suggest a guidance framework focused on domain ontology for diabetes medication. The software builds ontological awareness of each patient's condition regarding the essence of the medication as well as the nature of distribution and health risks, and also apply Semantic Web Rules Language (SWRL) and the Java Expert System Shell (JESS) to construct a possible patient preface. Alternatively, Gottlieb et al. suggest a clinical recommendation technique based on similarity. Shimada et al. replace the medical recommendation method with a decision tree classifier algorithm that encourages decision making by doctors and advises first-line (over prescription) care. With the exception of the earlier study, Zhang et al. define a hybrid recommendation structure that connects artificial neural networks with case-based reasoning.

Diana Cabral and Ricardo B. C. Prudêncio developed a system for separating aspects and classifying them in drug reviews. The produced solution consists of two key phases. In the aspect extraction, a system focused on syntactic pathways of attachment is designed to extract opinion pairs in drug comments, composed of an aspect word paired with an emotion modifier. In order to classify the opinion pairs by aspect category (e.g., condition, adverse effect, dose, and efficacy), a supervised classification is established in aspect classification based on domain and linguistic tools. The authors performed screening studies on sample sets of three different diseases: ADHD, AIDS, and Anxiety.

Wen-Hao Chiang at el. built a joint model with a recommendation component and an Adverse Drug Reactions (ADR) label prediction component to recommend a collection of to-avoid / secure drugs that will induce / will not induce ADRs when taken along with the prescription. The authors also developed real datasets on drug-drug interaction and the corresponding evaluation protocols.

III. PROPOSED MODEL

Our drug rating generation and recommender system framework primarily contains five modules, can be seen in Figure 1, which are data preprocessing module (including feature extraction), rating generation module, model evaluation module, dictionary sentiment analysis module, and recommendation model module.

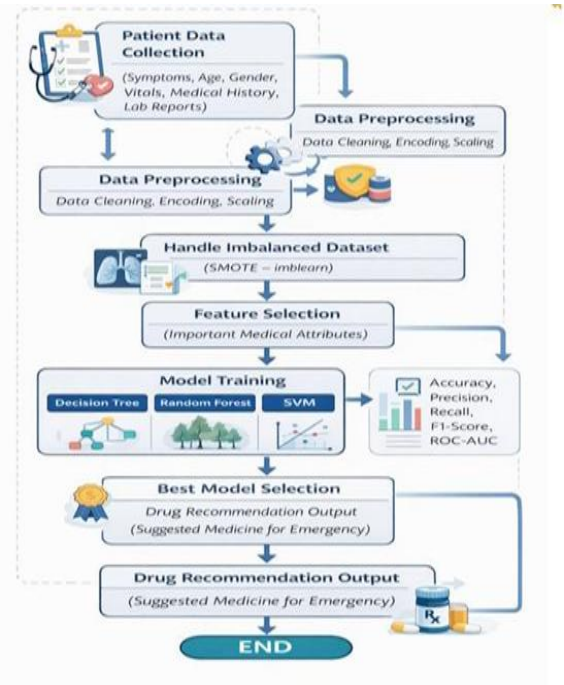

Figure 1: Rating generation and recommender system framework

A. Data preprocessing

Data cleaning is the method of finding and fixing (or removing) damaged or defective information from a record set, which refers to discovering missing, incorrect, defective, or irrelevant sections of the data and then adding, changing, or deleting dirty or coarse data. Proper data preparation is a required step, not only for a valid experiment but in the first place to allow the mining of a dataset using the means of machine learning. A collection of preprocessing steps needed to allow the machine learning system and algorithms to read and analyze the data, as well as to reduce the dataset to contain the necessary data points and attributes for the analysis. Similarly, the production or measurement of additional attributes from the data could also be important if such derived attributes might assist the analysis and thus allow better predictions. When we've used social media data, the data sets need to be cleaned wisely.

Essentially, social media data cannot be processed in a single way. Therefore, we used our techniques for properly analyzing sentiments to clean up those data.

These are the following tools we used for preprocessing our drug dataset:

- Tokenization: The process by which the running text is segmented into terms and phrases. The key task is to divide a text into tokens while throwing off other characters like dots.
- Stop word: This approach filters out and omits some very common words that seem to offer little to no meaning to the NLP target. This excludes the generic terms which are not insightful about the relevant material.
- Handling Negative Adjectives: The thing is to avoid the elimination of words in a given sentence will wipe out relevant details and change the meaning. Under these conditions, we can pick a minimum to stop word list or tests of this form (not good / not so good, not bad / not so poor) in our data under certain conditions and delete them from stop words.
- Stemming: Refers to the slicing process of the end or beginning of words to remove affixes (lexical additions to the word's root).
- Lemmatization: Has the purpose of reducing a word to its base form and of grouping various forms of the same word together.

B. Feature Selection

Unigrams: Within the unigrams model every single word is considered as a feature in the preprocessed review corpus. Thus, it is straightforward to create a feature vector for a review. First, it creates a lexicon of every word that occurs in the analysis corpus. A word-opinion matrix is then built, where the arrival (i, j) is the quantity of word i occurrence in the j^{th} opinion.

Unigrams & Bigrams: The unigrams model is a method of extraction of features commonly used in the processing of natural languages. Implementation is very simple and, in many situations, yields remarkably good results. An underlying downside, however, is its lack of ability to identify relationships between pairs (e.g., a word and its suffix, a word, and its negation, etc.), since it detaches each term. We are adding bigrams to the model of unigrams to reflect the influence of phrases such as 'good medicine and 'not bad. In addition, the dictionary also includes all the 2-tuples of terms (i.e. all pairs of consecutive words) that occur in the corpus of the study The matrix is calculated the way it used to; it has more rows now.

Unigrams, Bigrams & Trigrams: We are now adding trigrams (i.e. all triples of consecutive words) to the unigrams & bigrams model to capture the effect of sentences like 'not very good. Remember, however, that the same trigram will seldom occur across different articles because it is unlikely that two different people would use the same 3-word term in their analysis. The results of this model are therefore not supposed to be very different from the model of unigrams & bigrams.

IV. METHOD

In this section, we discussed some supervised machine learning algorithms which are used to generate rating from drug review and recommendation model that recommend an appropriate medication to remove the specific condition.

Decision Tree (DT): One of the most widely used hierarchical models for supervised learning that identifies local regions as series of recursive separations via decision nodes in the test function. The intuition behind the algorithm of the decision tree is simple, but still quite powerful. It divides data into two subsections to keep the information in each section quite homogeneous (all information in the section is of the same target class) than the prior/alternate

subsections; the two subsections can then be separated again before the homogeneity or any other time-based pausing thresholds are met. In increasing the decision tree, the same predictor parameter can be applied to many places. The ultimate aim of separation is to evaluate the correct variable correlated with the appropriate threshold to increase subgroup/branch homogeneity.

K-Nearest Neighbors (KNN): KNN is extremely easy to use and yet carries out complex classification tasks in its most basic form. The algorithm is lazy as the learning phase is non experienced. Rather, all data is used for training when a new data point or instance is classified. KNN is an algorithm of nonparametric learning, which means the underlying data are unassuming. This is a very useful feature because most real- world data don't follow theoretical assumptions, such as linear separateness, uniform distribution, etc. This measures precisely the difference from a single sample to every other training samples. Then it chooses K-nearest data samples and allocates the data sample to the category that most K data samples exist.

Support Vector Machine (SVM): The SVM concept is based on the Structural Risk Minimization principle of computational learning theory and one of the most reliable and effective techniques used in machine learning. In this theory, data is evaluated and the boundaries of decisions are described by having hyperplanes. In the case of input data that cannot be easily separated, it utilizes 4 kernel structures for classification tasks including linear, polynomial, radial based, and sigmoid functions by mapping the input data into high dimensional feature space to allow the data conveniently separable. The hyperplane divides the text vectors of each class in a way that the distinction is held as large as possible.

Linear SVC is analogous to SVC with parameter kernel=`linear'. Learning the hyperplane in linear SVM takes place by using linear algebra to transform the problem. Clear insight is that, rather than using observations themselves, the linear SVM is mostly rephrased employing inner product with any two data points. A sum of the multiplication of the input values for each pair is the inner product between two vectors. The equation for making a prediction for a new input using the dot product between the input (x) and each support vector () is calculated as follows:
= +, (1)

The equation no. 1 involves the calculating of the inner products of a new input vector (x) with all support vectors in training data. From the training data on the learning algorithm, coefficients and (for each input) must be calculated. The dotproduct is called the kernel and can be re-written as: = ∗ (2)
The kernel determines the similarity or distance of new data from support vectors. The dot product is a measure of similarity used for linear SVM or a linear kernel since the distance between the inputs is linear convention.

Recommendation Model: We are used to predicting what a consumer will give the "rate" or "preference." Recommendation engines are tools for data filtering that use algorithms and data to inform a single user of the most important products. Or they are simply an automated form of a "shop counterman". For a product, you ask him, he displays not only the drug but also the items you may purchase. They are well trained in cross selling and upselling.
As there is increasing data on the Internet and the number of users has increased dramatically, searching, mapping, and providing businesses with the information they need, in accordance with their preferences and tastes, is important.

Algorithm 1: Drug recommendation model 1 for all drug Reviews do: Mean of Review Score= mean (Dictionary Sentiment Polarity*Useful Count+ Rating).
2 for all condition in Patient Condition do: sort (group by (condition & drug name with Mean of Review Score, descending order)

V. RESULT AND DISCUSSION

Rating Generation: Throughout our research, we used the drug dataset from UCI Machine Learning Repository obtained by Felix Gräßer and Surya Kallumadi which is gathered from Drug.com and Druglib.com, the largest and widely visited pharmaceutical database portals that offer information to both customers and healthcare professionals. The dataset is presented with the users assigned a ranking score from 1 to 10, which can be adjusted in scale 1 to

5 and used for the purpose of predicting the review rating.

Numerous classification algorithms were used to evaluate the reliability of the approaches for classifying sentiments, including Decision Tree, KNN, and LinearSVC. We split the data on the medications into two sections, consisting of 70% training data and 30% evaluation data. Except in this model, 5 groups of experiments were carried out to prevent accidental error. For this study, the outcomes obtained for the test sample are analyzed using different model consistency parameters on each single class label. Positive, negative, and neutral comments are equally important when deciding on medication use in this specific drug review area. The overall accuracy rates, as well as per class accuracy rates for each model constructed from the test data samples are shown in Table 1, 2, and 3. From tables 1, 2, and 3, it is found that among the three methods, DT and Linear SVC classifier models with accuracy 76.79% and 83.08% respectively better than KNN classifier models with accuracy 55.41%.

Comparing accuracy, precision, recall, and f1-score of Linear SVC classifier model with DT and KNN classifier models, Linear SVC based prediction models perform well in all aspects. LinearSVC serves as an outstanding interface in keeping the process of learning since it has standardized approximations in high dimensional space. The findings of this section indicate that user experience contributes well to estimate the accuracy of the rating.

Table 1: Classification result of Decision Tree

Class	Precision	Recall	F1-score
1	0.85	0.88	0.87
2	0.60	0.52	0.55
3	0.63	0.59	0.61
4	0.61	0.57	0.59
5	0.72	0.73	0.72

Table 2: Classification result of KNN

Class	Precision	Recall	F1-score
1	0.55	1.00	0.71
2	0.77	0.01	0.01
3	0.77	0.02	0.04
4	0.86	0.02	0.04
5	0.73	0.00	0.00

Table 3: Classification result of Linear SVC

Class	Precision	Recall	F1-score
1	0.84	0.98	0.90
2	0.89	0.47	0.61
3	0.83	0.52	0.64
4	0.83	0.52	0.64

Dictionary Sentiment Analysis: In the analysis of dictionary sentiment, we performed emotional analysis using an emotional word dictionary to resolve the limitations of the package built with the data from movies. To make up for this, we used the Harvard emotional dictionary (http:/www.wjh.harvard.edu/~inquirer/spreadsheet _ guide.ht m) to perform additional emotional assessments.

First, we count the number of words included in the dictionary and specified positive ratio in preprocessed data = the number of positive words/ (the number of positive words + the number of negative words). If the ratio is less than 0.5, we have classified it as negative and if it is greater than 0.5, we have classified it as positive. We graded it as neutral with remainders, which includes the sentence without any positive or negative terms.

Recommendation: Recommendation Model (Algorithm 1) filters drug name based on condition and descriptions of the drug and recommend an appropriate medication to remove the condition. Table 4 shows the output of our recommendation model.

Table 4: Drug list with the specific condition

Condition	Drug Name
Attention	Deficit
Hyperactivity	Bupropion
	Wellbutrin Desoxyn
	Modafinil
	Urecholine
Abdominal Distension	Voltaren
Zen Shoulde	Diclofenac
	Naproxen
	Indocin

VI. CONCLUSION

In this paper, a recommendation model is proposed by mining sentiment information from social users' reviews. The drugs are recommended for the patient automatically after comprehensively considering each

symptom. We use user drug reviews into Decision Tree, KNN, and Linear SVC machine learning algorithms to achieve the rating generation task. Tuning the various parameters, we have tried to increase the accuracy of the rating generation. We compared the algorithms and all of them have different performances concerning the prediction accuracy. Finally, the Linear SVC model is selected for rating generation to obtain a good tradeoff among model accuracy (83.0%), model efficiency, and model scalability. Then take this result in Hybrid Recommendation Model to list appropriate medications. Besides, we conducted the emotional analysis using an emotional word dictionary to overcome limitations of a package formed with movie data. The results of this study show that the sentimental attributes contribute greatly to the prediction of drug rating, as well as recommendations. It also demonstrates significant improvements on a realworld dataset compared to current strategies. In our future work, when evaluating the context, we can find more linguistic rules, and to incorporate phrase-level sentiment analysis, we may adapt or build hybrid factorization models such as tensor factorization, or deep learning techniques. We also plan to enhance the accuracy and reliability of the recommendation model further.

REFERENCES

[1] Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2012.

[2] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1910–1921, Sep. 2016, doi: 10.1109/TMM.2016.2575738.

[3] Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," in *Proc. IEEE 11th Conf. Industrial Electronics and Applications (ICIEA)*, 2016, pp. 1383–1388, doi: 10.1109/ICIEA.2016.7603801.

[4] R. Majethia, V. Mishra, A. Singhal, K. L. Manasa, K. Sahiti, and V. Nandwani, "People Save: Recommending effective drugs through web crowdsourcing," in *Proc. 8th Int. Conf. Communication Systems and Networks (COMSNETS)*, 2016, doi: 10.1109/COMSNETS.2016.7440000.

[5] R. C. Chen, Y. H. Huang, C. T. Bau, and S. M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," *Expert Systems with Applications*, vol. 39, no. 4, pp. 3995–4006, Mar. 2012, doi: 10.1016/j.eswa.2011.09.061.

[6] J.-C. Na and W. Y. M. Kyaing, "Sentiment analysis of user-generated content on drug review websites," *Journal of Information Science Theory and Practice*, vol. 3, no. 1, pp. 6–23, Mar. 2015, doi: 10.1633/jistap.2015.3.1.1.

[7] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, and U. R. Acharya, "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques," *Knowledge-Based Systems*, vol. 198, p. 105949, Jun. 2020, doi: 10.1016/j.knosys.2020.105949.

[8] S. Vijayaraghavan and D. Basu, "Sentiment analysis in drug reviews using supervised machine learning algorithms," *arXiv preprint arXiv:2003.11643*, 2020.

[9] V. Doma *et al*., "Automated drug suggestion using machine learning," in *Advances in Intelligent Systems and Computing*, vol. 1130, 2020, pp. 571–589, doi: 10.1007/978-3-030-39442-4_42.

[10] Hamed, R. Roose, M. Branicki, and A. Rubin, "TRecs: Time-aware Twitter-based drug recommender system," in *Proc. IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining (ASONAM)*, 2012, pp. 1027–1031, doi: 10.1109/ASONAM.2012.178.

[11] Chen, D. Jin, T. T. Goh, N. Li, and L. Wei, "Context awareness-based personalized recommendation of anti-hypertension drugs," *Journal of Medical Systems*, vol. 40, no. 9, pp. 1–10, Sep. 2016, doi: 10.1007/s10916-016-0560-z.

[12] Gottlieb, G. Y. Stein, E. Ruppin, R. B. Altman, and R. Sharan, "A method for inferring medical diagnoses from patient similarities," *BMC Medicine*, vol. 11, no. 1, p. 194, Sep. 2013, doi: 10.1186/1741-7015-11-194.

[13] K. Shimada, K. Fujikawa, K. Yahara, and T. Nakamura, "Antioxidative properties of xanthan on the autoxidation of soybean oil in

cyclodextrin emulsion," 1992. [Online]. Available: https://pubs.acs.org/sharingguidelines

[14] Q. Zhang, G. Zhang, J. Lu, and D. Wu, "A framework of hybrid recommender system for personalized clinical prescription," in *Proc. 10th Int. Conf. Intelligent Systems and Knowledge Engineering (ISKE)*, 2015, pp. 189–195, doi: 10.1109/ISKE.2015.98.

[15] Cavalcanti and R. Prudêncio, "Aspect-based opinion mining in drug reviews," in *Lecture Notes in Computer Science*, vol. 10423, 2017, pp. 815–827, doi: 10.1007/978-3-319-65340-2_66.

[16] W.-H. Chiang, L. Shen, L. Li, and X. Ning, "Drug recommendation toward safe polypharmacy," *arXiv preprint arXiv:1803.03185*, 2018.

[17] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," in *Proc. 2nd Int. Conf. Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2017, pp. 294–298, doi: 10.1109/ICITISEE.2017.8285514.

[18] N. Asghar, "Yelp Dataset Challenge: Review rating prediction," 2016.

[19] S. Fletcher and M. Z. Islam, "Decision tree classification with differential privacy: A survey," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–33, Aug. 2019, doi: 10.1145/3337064.

[20] Basti, C. Kuzey, and D. Delen, "Analyzing initial public offerings' short-term performance using decision trees and SVMs," *Decision Support Systems*, vol. 73, pp. 15–27, May 2015, doi: 10.1016/j.dss.2015.02.011.