
A Hybrid Approach Based Diet Recommendation System Using ML and Big Data Analytics

Muhib Anwar Lambay* and S. Pakkir Mohideen

*CSE Dept, B.S. Abdur Rahman Crescent Institute of Science and Technology,
Chennai, India*

E-mail: lambaymuhib@gmail.com, pakirmoitheen@crescent.education

**Corresponding Author*

Received 02 November 2021; Accepted 20 February 2022;

Publication 01 July 2022

Abstract

Recommendations are useful suggestions used by people from all walks of life. However, the usage of recommender systems plays a vital role in modern applications. They are found in different domains such as E-commerce. Concerning the health care industry, recommendations play a very crucial role. This industry has significance as it is linked to the lives of people and their well-being. Human health depends on the diet followed. Keeping this fact in mind, in this paper, we investigated healthy diet recommendations. The recommender systems that are existing in healthcare focused a little in this area. From the literature, it is understood that most of the frameworks on health recommendations are theoretical in nature. As food decides health, it is to be given paramount importance. In this paper, we proposed a hybrid mechanism based on Artificial Intelligence (AI) for big data analytics. Particularly we used Machine Learning (ML) for generating healthy diet recommendations. The proposed system is known as Hybrid Recommender System (HRS). It involves a hybrid approach with Natural Language Processing (NLP) and machine learning. An algorithm named Intelligent Recommender

Journal of Mobile Multimedia, Vol. 18_6, 1541–1560.

doi: 10.13052/jmm1550-4646.1864

© 2022 River Publishers

for Healthy Diet (IR-HD) is proposed to analyze data and provide healthy diet recommendations. IR-HD could generate recommendations on a healthy diet and outperform existing models. Python data science platform is used to implement the recommender system. The results of experiments showed that the system is capable of providing quality recommendations and it has performance improvement over the state of the art.

Keywords: Healthcare recommendations, recommender system, intelligent healthcare recommendation system, healthy diet recommender.

1 Introduction

Big data and cloud eco-system plays vital role in dealing with large amounts of data known as big data and perform analytics known as big data analytics. In the process of big data processing, there is need for Natural Language Processing (NLP) techniques as the pre-process to big data analytics. Big data processing is not effective without the usage of NLP as explored in [1]. Moreover, NLP works in multi-lingual context which is beneficial in cloud based applications. The linguistic analysis includes both translational routines and transformation routines. In fact, NLP is inter-disciplinary consisting of statistics, probability, machine learning (ML), information retrieval (IR) and linguistics. NLP is meant for understanding, analysing and interpreting written text in any human-spoken languages. With the emergence of big data which is voluminous in nature, NLP became more significant as the most of the big data is in the form of textual content. Data science is fast growing field that provides innovations in processing big data with the help of NLP. Various algorithms are being proposed for data analytics, visualization and interpretation [2].

There are many core tasks associated with NLP that are used in big data analytics as pre-processing. They include language modelling, word segmentation, Part-of-Speech (PoS) tagging, named entity recognition and parsing. In big data analytics, there is need for linguistic processing that involves aforementioned core tasks, syntactic parsing and co-reference resolution [3]. Though big data and NLP can be used in different applications, the applications in healthcare domain are highly significant as they influence health of humans across the globe. As studied in [13] humanitarian needs are identified with big data and data analytics. The promises of big data in healthcare industry is explored in [15] while real time health data analytics is the main focus in [16]. A framework for health and wellness applications is proposed in [17] while big data capabilities to fulfil healthcare needs is studied in [18].

Healthcare recommendations became an important application of late. Recommendations are useful in any field. It is more so in healthcare domain. From the literature, it is understood that there are many recommender systems found. However, in the area of healthcare domain, there is inadequate research. Moreover, most of the frameworks on health recommendations are theoretical in nature. Considering users, their preferences and basic nutrients of food of different categories, there is no sufficient research that exploits ML techniques. As the human health depends on the diet followed by them, in this paper we focused on developing a framework that exploits both big data and NLP for generating dietary recommendations. The importance of the proposed architecture is to provide intended healthy diet recommendations. The following are contributions of the paper.

1. We proposed a framework known as Hybrid Recommender System (HRS) which makes use of big data analytics and NLP for healthy diet recommendations.
2. We proposed a hybrid algorithm named Intelligent Recommender for Healthy Diet (IR-HD) for realizing the HRS framework.
3. We built a prototype application using Python data science platform to evaluate the performance of the proposed framework.

The remainder of the paper is structured as follows. Section 2 reviews literature on the big data analytics, NLP and healthcare recommendations. Section 3 provides the proposed framework and underlying algorithm. Section 4 presents experimental results. Section 5 covers results of experiments. Section 6 concludes the paper and provides directions for future work.

2 Related Work

This section reviews literature on big data analytics and NLP to solve real world problems. Monti et al. [1] proposed a methodology known as Cross-Language Information Retrieval (CLIR) that uses concepts of big data and NLP. The methodology is based on Lexicon-Grammar (LG) which is theoretical framework and big data analytics for empirical study. It supports ontology and SPARQL for information retrieval. The authors intended to apply in to different domains in future. Gudivada et al. [2] explored big data-driven NLP for different applications. They opined that for big data processing, NLP is essential. There are different NLP applications associated with big data. They include statistical language modelling, probability distribution, word segmentation, parsing, named entity recognition and Part-of-Speech (PoS)

tagging. Big data can be used to build language models and obtains hidden trends from the data. Big data enabled stochastic process is found essential. Agerri et al. [3] proposed a streaming based approach for big data and NLP. They used a distributed framework known as Apache Storm for the empirical study. They investigated and found that there is room for improving big data-driven NLP approaches. In future they wanted work on different level of granularity with respect to big data driven NLP.

Niakanlahiji et al. [4] focused on Advanced Persistent Threats (APTs) with respect to big data and NLP. They explored different APT techniques. They include initial access, exploitation, execution, defence evasion, information collection, activation logic and communication. These are the different categories of APT techniques that are commonly used. The technical documents are understood with APT techniques and NLP approaches. Thejaswini and Indupriya [5] employed NLP techniques to ascertain big data security issues. They used APTs for analysing threats in big data. Khader et al. [6] studied the influence of NLP on big data analytics. They used NLP techniques for big data analytics and found the utility of them. Machine learning techniques along with distributed programming frameworks like Hadoop are used. They found that accuracy of the methods is improved when NLP is used. Agogo and Hess [7] employed NLP methods like N-gram sequences, tokenization and PoS tagging. Social media data from Twitter is used to ascertain importance of big data analytics and NLP. Bertero et al. [8] employed NLP and big data for anomaly detection. They used NLP techniques and word2vec embedding in order to improve quality of analytics. Dessi et al. [9] explored big data for analytics and cognitive computing. They found that both NLP and cognitive analytics can help in bringing about semantic concepts. TF-IDF and machine learning process are combined to arrive at desired results. In future, they intend to work on cognitive-driven learning analytics.

Sun et al. [10] employed NLP techniques for mining opinions from social media. Different levels of granularity in big data analytics is achieved with NLP techniques. The granularity is investigated at document level, sentence level, fine-grained level, cross-domain level and cross-lingual level. They also employed deep learning models for improving performance. In future, they intend to continue the research on big data and NLP. Billal et al. [11] employed NLP techniques for pre-processing to big data analytics. They proposed hashtag segmentation algorithm and used different classification methods for empirical study. In future, they wanted to use multi-word expressions as pre-processing pipeline. Velupillai et al. [12] used NLP techniques for healthcare recommendations. They found that health related

research needs both big data analytics and also NLP to be used. Kreutzer et al. [13] employed NLP and Artificial Intelligence (AI) to improve humanitarian needs and reduce risks related to the humans. Xing et al. [14] employed NLP techniques to have content based recommendations of podcasts.

Huang et al. [15] explored on the challenges of big data usage in healthcare domain. They found different utilities of big data such as recommendations, epidemic surveillance, food safety monitoring, health monitoring, finding air quality, exploring nutrition and so on. Ta et al. [16] on the other hand focused on big data stream computing that is used in real time healthcare applications. They used different data sources such as clinical text, social media, Electronic Health Records (EHRs), biomedical images, sensing data, biomedical signals and genomic data. They proposed a framework for healthcare data analytics. Ahmad et al. [17] proposed a framework known as Health Fog for both wellness and health applications. They employed fog computing for efficient processing of data. In future they intend to integrate with social media data for health data analytics. Wang et al. [18] explored the technical perspective of big data and found its capabilities in healthcare industry. They found that big data architecture with NLP is suitable for healthcare data analytics. Suciu et al. [19] explored the usage of big data, Internet of Things (IoT), NLP and cloud computing for many secure e-Health applications. Banos et al. [20] proposed framework for mining using big data and NLP to know wellness details and mind digital health details. They explored different aspects of the framework for healthcare applications. Shakya et al. [21] explored on wearable devices being used in healthcare systems. They used it for computational enhancements. Manoharan and Samuel [22] used machine learning for diagnosing lung cancer early. Vijayakumar et al. [23] proposed a modified Particle Swarm Optimization (PSO) model for energy aware load distribution models. Joe et al. [24] proposed a recommender system for users that are context dependent with location based orientation. Haoxiang et al. [25] explored on data mining for perturbation and data analytics. From the literature, it is understood that most of the frameworks on health recommendations are theoretical in nature. There is need for empirical study that involves big data and NLP for bringing about healthcare recommendations.

3 Proposed Framework

Healthcare recommendations have attracted widespread research with the invent of big data and cloud ecosystem. With big data analytics, it is made possible to process complete data as a whole and bring about knowledge that

is hidden from the data. A framework by name Intelligent Healthcare Recommendation System (IHRS) is proposed and implemented. The framework is realized with NLP used for pre-processing and big data analytics algorithm for prediction and providing healthy diet recommendations.

3.1 Problem Statement

Provided the healthcare dataset, developing an intelligent framework that mines data using NLP and big data analytics (ML techniques) and provide healthy diet recommendations is the problem considered. Similar to various recommendations pertaining to weather, products, online services and healthcare services, food is an important component of human lives. With the pressures in modern era, people neglect eating healthy foods. People with preference over foods have links to many applications such as restaurants. With such information, in the long run, healthy diet recommender systems play vital role to enable people to have better eating habits. There is need for a framework to generate healthy diet recommendations. NLP is used for pre-processing of data while the data is analysed using ML algorithms like KNN. A hybrid algorithm is proposed to have quality healthy diet recommendations to public.

3.2 The Framework

The proposed framework is visualized in Figure 1. It has three important layers. They are known as cloud layer, middle layer and application layer. The cloud layer helps in data storage and processing. The middle layer contains underlying methods for healthy diet recommendations. The application layer is the application which is used by end users. The data (big data) is stored in cloud. It is used for generating user-food matrix. The food dataset contains details of various foods, their nutrition values and other dietary information. A recommender system is designed to have a hybrid approach which considers induced motivation model, users and foods. It is based on collaborative filtering approach.

As presented in Figure 1, the framework is to analyse big data that is from healthcare domain in order to generate healthy diet recommendations. The framework operates on the number of users denoted as N_u and number of foods denoted as N_f . Training data is used with attributes such as food id, user id and ratings. Two matrices are defined to achieve recommendations. They are known as user-food matrix and user-food ratings matrix.

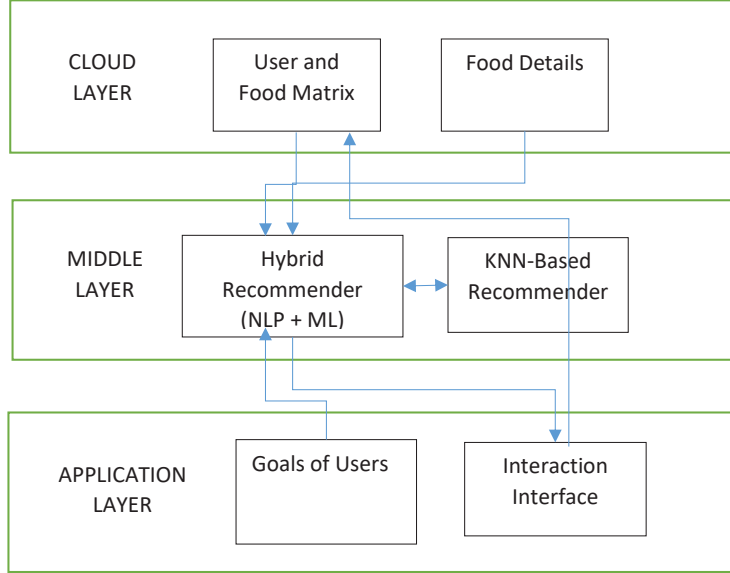


Figure 1 Proposed framework.

The two matrices are as in Equations (1) and (2).

$$A_{u,f} = [ID_u \ ID_f \ R_{u,f}] \quad (3, N_a) \text{ is the size of A such that } N_a \leq N_u \times N_f \quad (1)$$

$$B_{u,f} = \begin{pmatrix} R_{u1,f1} & \dots & R_{u1,fi} \\ \vdots & \ddots & \vdots \\ R_{ui,f1} & \dots & R_{ui,fi} \end{pmatrix} \text{ size of B is } (N_u, N_f) \quad (2)$$

Once users' ratings are predicted for various foods (where rating is given 1 to 10 scale), the aim is to reduce error rate known as Root Mean Square Error (RMSE) when tests are conducted. Both RMSE and training set error are defined as in Equations (3) and (4).

$$S_{Test} = ([ID_u \ ID_f], \dots), \text{ in which } ID_u, ID_f \in A_{u,f} \text{ and } R_{u,f} \notin A_{u,f} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{|S_{Test}|} \sum_{(u,f) \in S_{Test}} E_{u,f}^2}, \quad (4)$$

Where the true rating is denoted as $R_{u,f}$ while cardinality is denoted as $|S_{Test}|$. When RMSE is low, it is known as highly accurate prediction system. A health score is generated by comparing needs of users and nutrition of foods. The score is between 1 and 10. Based on the user needs the healthy diet recommendations are generated as in Equation (7). The result score is computed as in Equation (6) and the health score is computed as in Equation (5).

$$L_{u,P_f} = \begin{bmatrix} ID_f & P_{u,f} \\ \vdots & \vdots \end{bmatrix} (u, f) \in S_{Test} \text{ and} \quad (5)$$

A_{u,P_f} is a descending sorted list by $P_{u,f}$

$$H_{u,f} = \sum_{(u,f)} \in \frac{n_{f,i}}{h_{u,i}} \times a_i(u, f \in L_{u,P_f}) \quad (6)$$

$$L_{u,Result} = [P_{u,f} \times a_{u,t} + H_{u,f} \times a_{u,h}, \dots] (u, f) \in S_{Test}, \quad (7)$$

The descending list of records is denoted as L_u, P_f and $h_{f,i}$ denotes foods' nutritional value and $h_{u,i}$ denotes needs of users. The weight of nutritional components is denoted as $a_{u,i}$ and the weights are denoted as $(a_{u,t}, a_{u,h})$. The Euclidean distance function, cosine similarity function and Pearson similarity function are as provided in Equations (8), (9) and (10).

$$sim(x, y) = \frac{1}{1 + \sqrt{\sum(x - y)^2}} \quad (8)$$

$$sim(x, y) = \frac{(x \cdot y)}{\sqrt{(x \cdot x)(y \cdot y)}} \quad (9)$$

$$sim(x, y) = \frac{(x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{(x - \bar{x}) \cdot (x - \bar{x}) * (y - \bar{y}) \cdot (y - \bar{y})}} \quad (10)$$

These similarity functions are used to evaluate both KNN based and proposed hybrid recommender system.

3.3 The Proposed Hybrid Recommender System

The proposed hybrid recommender system is based on both matrix decompositions. The two matrices aforementioned are used in order to have better recommendations and minimize RMSE. The algorithm is known as Intelligent Recommender for Healthy Diet (IR-HD).

Algorithm 1 Healthy diet recommender

Algorithm: Intelligent Recommender for Healthy Diet (IR-HD)

Input: User preferences P, food dataset D

Output: Healthy diet recommendations R

- 1: Start
- 2: Initialize food vector F
- 3: Initialize user vector U
- 4: For each tuple d in D
- 5: Populate F
- 6: Populate U
- 7: End For
- 8: Construct user-food matrix using Equation (1)
- 9: Construct ratings matrix using Equation (2)
- 10: Compute health score using Equation (5)
- 11: Compute result score using Equation (6)
- 12: Compute recommendations using Equation (7) and populate R
- 13: Compute RMSE using Equation (4)
- 14: Return R
- 15: End

As presented in Algorithm 1, the pseudo code for the proposed recommender system is provided. Step 2 and Step 3 are used to initialize food and user vectors that are later used to construct matrices. The Step 4 through Step 7 is an iterative process to populate F and U from the food dataset. Afterwards, the user-food matrix is generated using Equation (1) and ratings matrix is created using Equation (2). Different computations are made before generating recommendations that are personalized. The results are the healthy diet recommendations that are stored in data structure R. The RMSE is computed for given results and the results are provided for the proposed algorithm and KNN in Section 4.

3.4 Dataset Details

A dataset known as food dataset is used for experiments. We also used user profiles that reflect their food preferences. The food dataset contains various attributes containing data related to basic nutritional values.

The food id is used in the user profiles data in the form of preferences so as to make the collaborative filtering with the proposed algorithm possible. Figure 1 gives an idea of filtering with the user-food matrix. With collaborative filtering the proposed system considers user specific recommendations.

Table 1 Details of food dataset

Sr. No.	Attribute Name	Description
1	Food ID	Unique ID of specific food
2	Name	Name of food
3	Type	Type or category of food which can hold a numeric value between 1 and 5. 1 – Dairy product 2 – Fruits 3 – Vegetables 4 – Poultry, lean meats, fish, eggs 5 – Cereal, legumes
4	Calories	Numeric value indicating calories of the food.
5	Fat	Numeric value reflecting fat value of food
6	Total carbs	Numeric value showing total carbohydrates associated with food.
7	Sodium	Numeric value showing sodium content in the food.
8	Protein	Numeric value indicating protein value of food.
9	Special	Takes two values such as “veg” or “0” where zero reflects non veg food.

4 Experimental Results

Empirical study is made with the Python data science platform based application. The results of KNN, Latent Factor Model (LFM) and the proposed hybrid algorithm are provided. RMSE is used as the metric to know the accuracy in prediction. The less in RMSE, the higher in accuracy.

As presented in Figure 2, the data analytics and NLP are employed to know the health score of food for different number of days. The day value is provided in horizontal axis and the percentage of weighting is provided in vertical axis. With respect to weight goals of people, the induced motivation model is used for empirical study.

As presented in Figure 3, the number of similar users is provided in horizontal axis and the vertical axis shows RMSE value. Less RMSE indicates higher in performance and the higher in RMSE indicates least performance. As the number of users is increased the RMSE is decreased indicating improvement in health score.

As presented in Figure 4, the steps used in the matrix factorization are shown horizontal axis. The values are taken from 3000 to 7000 incrementing by 500. RMSE on the other hand is shown in vertical axis. The results revealed that there is relationship between RMSE and steps. As the number of steps is increased, the RMSE came down showing improved performance.

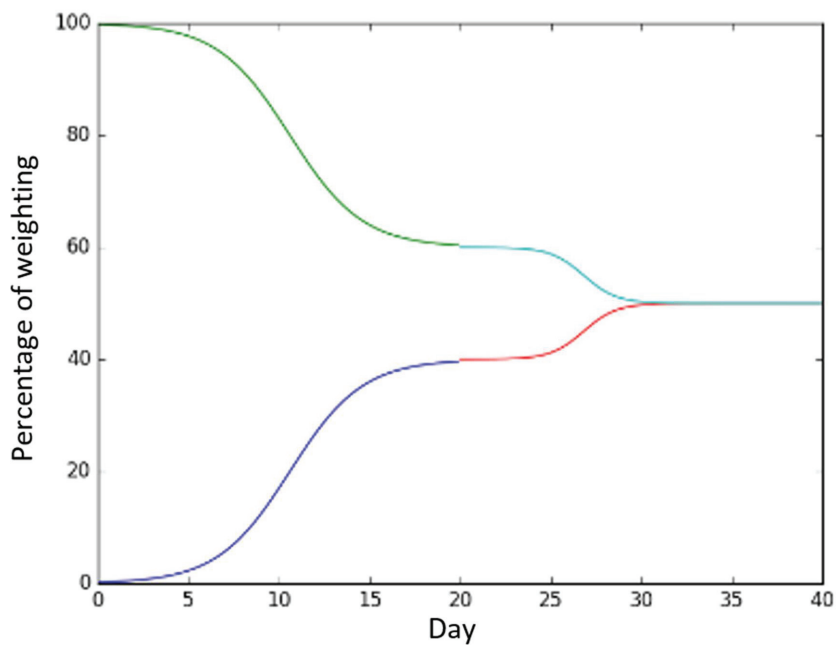


Figure 2 Results of induced motivation model.

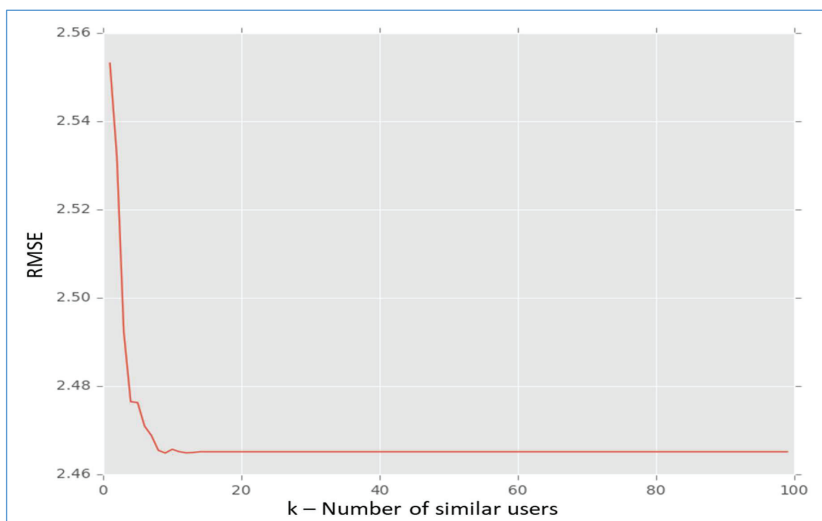


Figure 3 Cross-validation results of KNN.

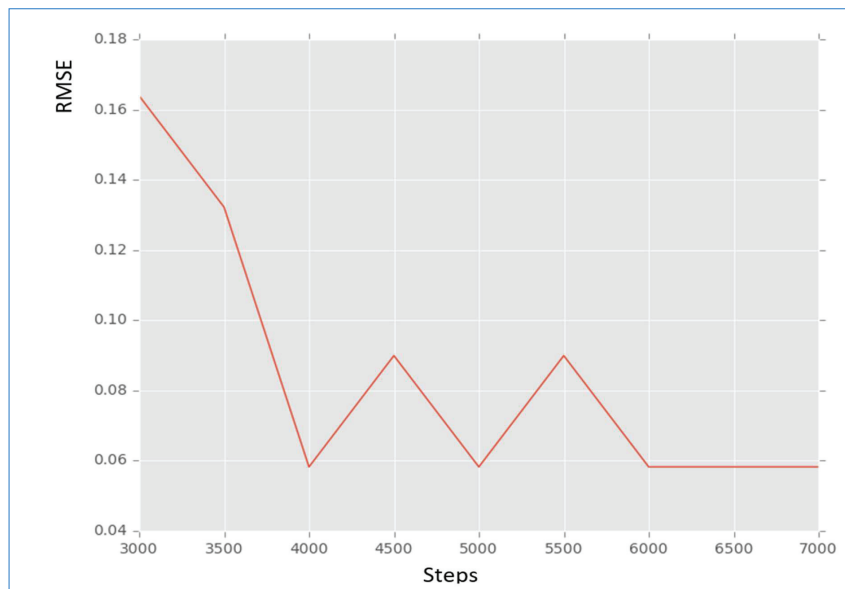


Figure 4 Cross validation results for the proposed hybrid algorithm.

When the proposed hybrid algorithm is used with different values in horizontal axis as shown in Figure 5, the RMSE is decreased gradually. The value provided in horizontal axis has its influence on the RMSE. The lesser in RMSE, the higher in performance.

As presented in Figure 6, the cross validation results are provided by changing a parameter as shown in horizontal axis. As it is changed, the RMSE shown in vertical axis is changing. Therefore, it is understood that the value given on the horizontal axis 0.0 to 1.0 is influencing RMSE value. When the value is less, the RMSE is less indicating highest accuracy.

As presented in Table 2, the RMSE metric is used for understanding the performance of different similarity measures like Euclidean distance, Cosine similarity and Pearson similarity measure.

As presented in Figure 7, the similarity functions are provided in horizontal axis and the RMSE value is shown in vertical axis. The RMSE values are observed with both training dataset and the whole dataset. The results revealed that with training dataset, Euclidean distance metric showed highest performance as it showed least RMSE. With respect to the whole dataset the least RMSE is exhibited by Pearson metric.

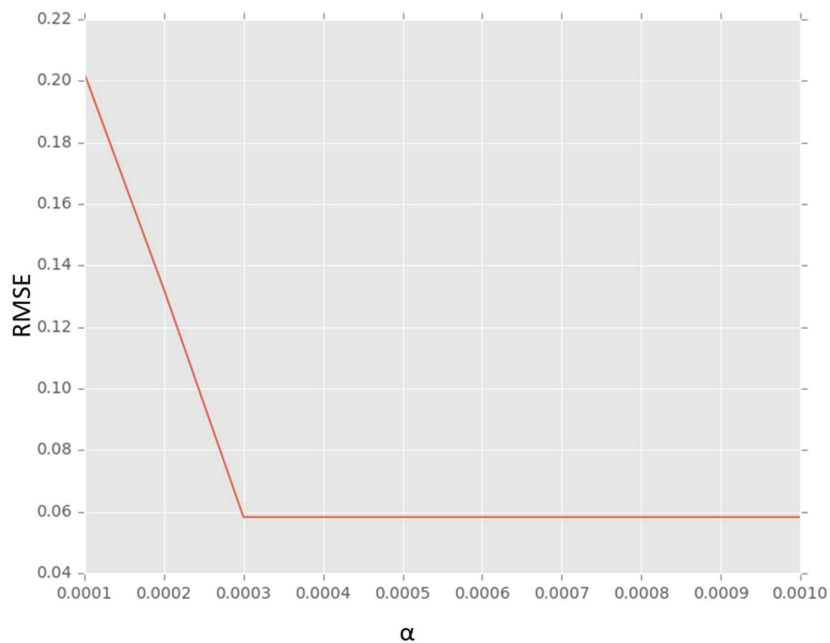


Figure 5 Cross validation results for the proposed hybrid algorithm.

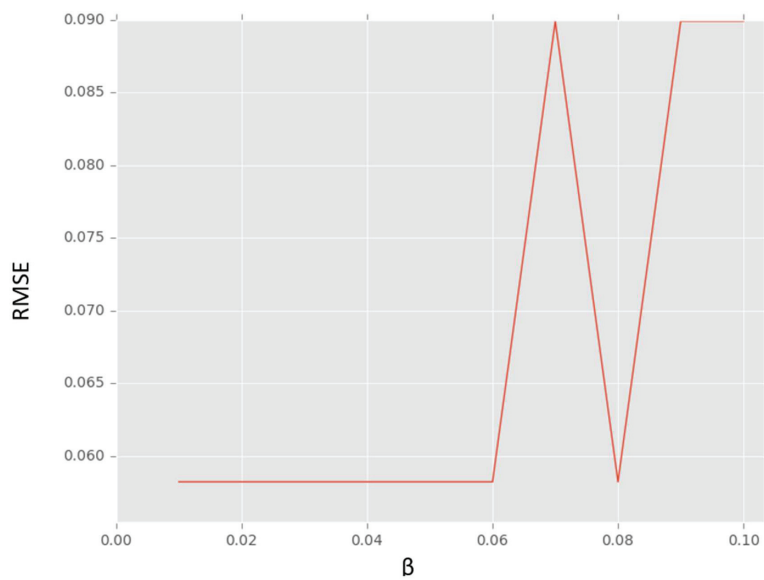
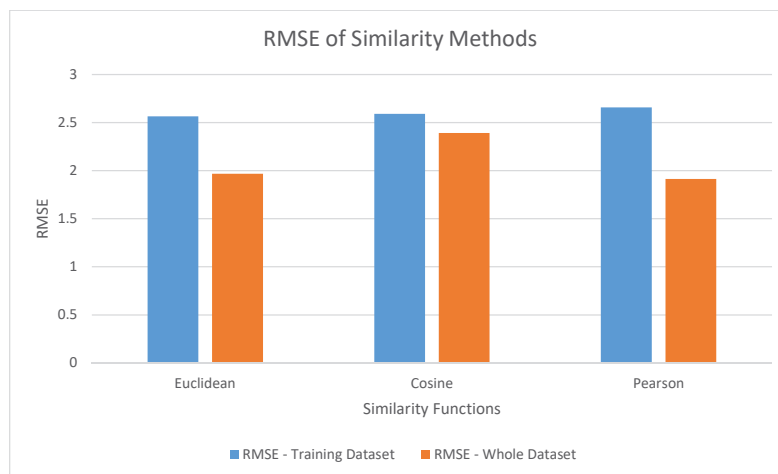


Figure 6 Cross validation results for the proposed hybrid algorithm.

Table 2 Performance comparison with different similarity metrics

Metrics	Euclidean	Cosine	Pearson
RMSE – Training Dataset	2.56594	2.59116	2.65868
RMSE – Whole Dataset	1.96741	2.39162	1.91329

**Figure 7** RMSE of similarity methods.**Table 3** Performance comparison among different algorithms

Metric	Naïve Approach	KNN	LFM	Hybrid Approach
RMSE	2.9742	2.57622	1.11279	0.18445

As presented in Table 3, the RMSE metric is used for understanding the performance of different algorithms like Naïve approach, KNN, LFM and the proposed hybrid approach.

As presented in Figure 8, the prediction algorithms are provided in horizontal axis and the RMSE value is shown in vertical axis. The RMSE values are observed for predicting taste rating. The results revealed that the hybrid approach proposed in this paper outperformed all other existing methods. It showed least RMSE indicating highest accuracy. The least performance is exhibited by naïve approach. LFM is found to be better than KNN while KNN showed better performance than naïve approach.

As presented in Table 4, the performance of the recommender models is provided in terms of accuracy, precision, recall and F-score.

As presented in Figure 9, the performance metrics are provided in horizontal axis while the performance is shown in vertical axis. Different

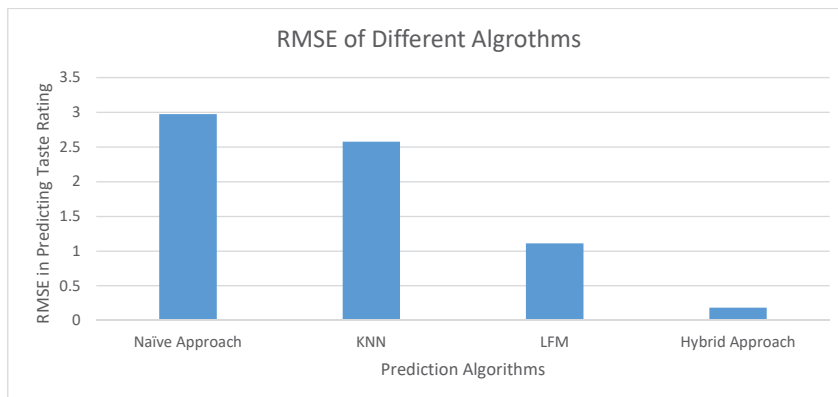


Figure 8 RMSE of different algorithms in predicting taste rating.

Table 4 Performance comparison among different recommender models

Recommender Models	Performance			
	Accuracy	Precision	Recall	F1-score
Naive Approach	0.639394	0.8756	0.6368	0.737347
k-NN	0.665263	0.88555	0.66665	0.760665
LFM	0.952527	0.9552	0.9552	0.9552
Hybrid Approach	0.957547	0.96515	0.9552	0.960149

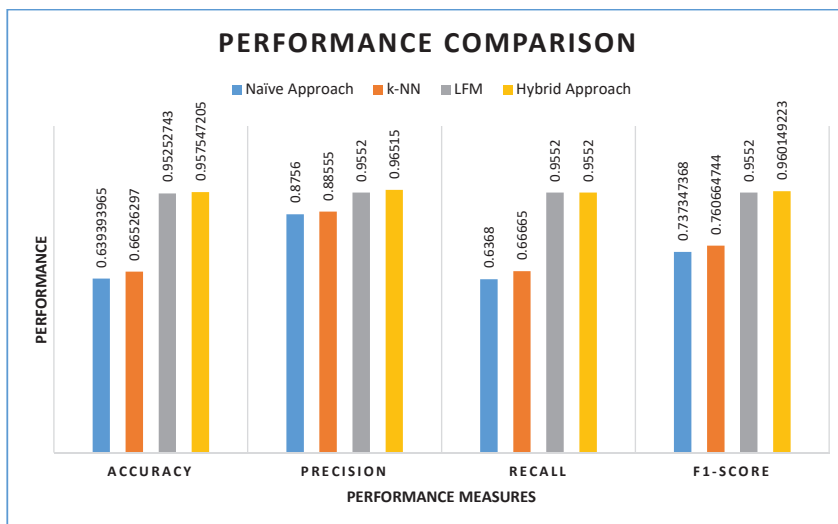


Figure 9 Performance of recommender models.

algorithms have shown varied performance. The highest values for accuracy, precision, recall and F1-score are exhibited by the proposed method. It shows that the hybrid approach outperforms the existing methods.

5 Conclusions and Future Work

In the contemporary era, big data and cloud computing infrastructure play vital role in solving real world problems. In addition to these big data analytics needs the help of NLP techniques as most of the data is in textual format. Healthcare industry is rich in such data that needs to be processed and used in different applications. Many healthcare applications came into existence. However, with the big data analytics and NLP there is room for more in the industry. In this paper, we proposed a framework with hybrid approach based on ML for big data analytics and NLP for pre-processing of data for generating diet recommendations. It leverages health diet in the form of recommendations to people. A framework named Healthcare Recommendation System (HRS) is proposed with a hybrid algorithm known as Intelligent Recommender for Healthy Diet (IR-HD). Empirical study is made with food dataset using Python data science platform. The experimental results revealed that the proposed framework is able to provide healthy diet recommendations with high accuracy. An important problem in this paper is that it does not work properly when training samples are very less. This problem is known as cold start problem. In our future endeavours, we enhance the framework to deal with cold start problem and also incorporate deep learning methods to leverage its performance further.

Declarations

Funding

Authors declare that there is no funding support for this research work.

Conflicts of Interest

Authors declare that there are no conflicts of interest among themselves.

References

- [1] Huang, T., Lan, L., Fang, X., An, P., Min, J., and Wang, F. (2015). Promises and Challenges of Big Data Computing in Health Sciences. *Big Data Research*, 2(1), pp. 1–29.
- [2] Van-Dai Ta, Chuan-Ming Liu, and Nkabinde, G. W. (2016). Big data stream computing in healthcare real-time analytics. 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). pp. 1–6.
- [3] Wang, X., Yang, L. T., Kuang, L., Liu, X., Zhang, Q., and Deen, M. J. (2018). A Tensor-Based Big-Data-Driven Routing Recommendation Approach for Heterogeneous Networks. *IEEE Network*, 33(1), pp. 64–69.
- [4] . Manogaran, G., Varatharajan, R., and Priyan, M. K. (2017). Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System. *Multimedia Tools and Applications*, 77(4), pp. 4379–4399.
- [5] Abbas, A., Bilal, K., Zhang, L., and Khan, S. U. (2015). A cloud based health insurance plan recommendation system: A user centered approach. *Future Generation Computer Systems*, pp. 99–109.
- [6] Lv, Z., Song, H., Basanta-Val, P., Steed, A., and Jo, M. (2017). Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. *IEEE Transactions on Industrial Informatics*, 13(4), pp. 1891–1899.
- [7] Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), pp. 137–144.
- [8] Zhang, Y., Qiu, M., Tsai, C.-W., Hassan, M. M., and Alamri, A. (2017). Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data. *IEEE Systems Journal*, 11(1), pp. 88–95.
- [9] Fang, R., Pouyanfar, S., Yang, Y., Chen, S.-C., and Iyengar, S. S. (2016). Computational Health Informatics in the Big Data Age. *ACM Computing Surveys*, 49(1), pp. 1–36.
- [10] Ahmad, M., Amin, M. B., Hussain, S., Kang, B. H., Cheong, T., and Lee, S. (2016). Health Fog: a novel framework for health and wellness applications. *The Journal of Supercomputing*, 72(10), pp. 3677–3695.
- [11] Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. (2017). *IEEE Access*, 5, pp. 5247–5261.

- [12] Thaduri, A., Galar, D., and Kumar, U. (2015). Railway Assets: A Potential Domain for Big Data Analytics. *Procedia Computer Science*, 53, pp. 457–467.
- [13] Wang, Y., Kung, L., Ting, C., and Byrd, T. A. (2015). Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care. 2015 48th Hawaii International Conference on System, pp. 1–10.
- [14] Saggi, M. K., and Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, 54(5), pp. 758–790.
- [15] Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A. Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, 36(5), pp. 748–758.
- [16] Grover, P., and Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Global Journal of Flexible Systems Management*, 18(3), pp. 203–229.
- [17] Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., . . . Fratu, O. (2015). Big Data, Internet of Things and Cloud Convergence – An Architecture for Secure E-Health Applications. *Journal of Medical Systems*, 39(11). pp. 1–8.
- [18] Zhong, R. Y., Newman, S. T., Huang, G. Q., and Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*, 101, pp. 572–591.
- [19] Jin, X., Wah, B. W., Cheng, X., and Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), pp. 59–64.
- [20] Banos, O., Bilal Amin, M., Ali Khan, W., Afzal, M., Hussain, M., Kang, B. H., and Lee, S. (2016). The Mining Minds digital health and wellness framework. *BioMedical Engineering OnLine*, 15(S1) pp. 1–22.
- [21] Shakya, Subarna, and Lalitpur Nepal. “Computational Enhancements of Wearable Healthcare Devices on Pervasive Computing System.” *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 2, no. 02 (2020): 98–108.
- [22] Manoharan, Samuel. “Early diagnosis of Lung Cancer with Probability of Malignancy Calculation and Automatic Segmentation of Lung CT scan Images.” *Journal of Innovative Image Processing (JIIP)* 2, no. 04 (2020): 175–186.
- [23] Vijayakumar, T., and Mr R. Vinothkanna. “Efficient Energy Load Distribution Model using Modified Particle Swarm Optimization Algorithm.” *Journal of Artificial Intelligence* 2, no. 04 (2020): 226–231.

- [24] Joe, Mr C. Vijesh, and Jennifer S. Raj. “Location-based Orientation Context Dependent Recommender System for Users.” *Journal of trends in Computer Science and Smart technology (TCSST)* 3, no. 01 (2021): 14–23.
- [25] Haoxiang, Wang, and S. Smys. “Big Data Analysis and Perturbation using Data Mining Algorithm.” *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 19–28.

Biographies



Muhib Anwar Lambay has 12+ years of experience in teaching field, received his Master of Technology (M.Tech) in Computer Science & Engineering from Jawaharlal Nehru Technological University, Hyderabad, India in year 2014 and Bachelor of Engineering (B.E.) in Computer Engineering from University of Mumbai, India in year 2008. Currently pursuing his Ph.D in Computer Science & Engineering from B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai. He presented various academic as well as research-based papers at several National and International Conferences. He is holding many publications in his area and many more papers are in pipeline. Presently working as an Assistant Professor in the department of Computer Engineering at Theem College of Engineering, Affiliated to University of Mumbai. His main research work focuses on Big Data Analytics and Machine Learning.



S. Pakkir Mohideen received his Ph.D in the field of Personalized Ontology based Adaptive Learning System from Anna University, Chennai in 2017 and M.E. in Computer Science & Engineering from Anna University, Chennai in 2007. He is currently working as an Associate Professor & Head of the Computer Application department at B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai. He participated in several high profile conferences, has many publications and is presently working on many more papers. In addition to his academic career, He received a “Certificate of Appreciation” award for having produced excellent academic results. His areas of research include Big data analytics, Data Mining, Information Retrieval System.