

A prediction model for benefitting e-commerce through usage of regional data: A new framework

Shefali Singhal, Poonam Tanwar

Department of Computer Science and Engineering, Manav Rachna International Institute of Research and Studies, Delhi, India

Article Info

Article history:

Received Feb 25, 2021

Revised Sep 6, 2021

Accepted Sep 28, 2021

Keywords:

Classification and clustering algorithms

Data extraction

E-commerce

ETL process

Predictive modelling

ABSTRACT

Today during 'Covid-20', people are more inclined towards online shopping. In general practice, analysis of browsing history and customer's micro behaviour against online shopping habits have been used for future suggestions. Due to this, the predictions made were suffering from over-similarity problem and the user was unable to find any novelty in the recommended items. Observing these issues, e-shopping quality can be enhanced by adding a factor other than similarity. The current research suggests and advertise those products which belongs to a person's region. For this research work the data has been collected on the basis of area-wise, like, country-based segregation. Here the considered dataset belongs to country, 'India', its culture, its handicraft and its citizens. Datasets and their combinations based on multiple attributes are input for the proposed predictive system. In this paper, existing data is also considered for collecting customers demographic details which is further mapped with the area-wise dataset. Also, a framework has been proposed which uses database and user query as input for its predictive system in order to generate default suggestions for the user other than the submitted query also.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Shefali Singhal

Department of Computer Science and Engineering

Manav Rachna International Institute of Research and Studies

Sector-43, Aravalli Hills, Delhi-Surajkund Road, Faridabad-121004, (Haryana), India

Email: shefalisinghal85@gmail.com

1. INTRODUCTION

The Consumer and their purchase style have been changed with time. As the web usage rate has increased, consumer is preferring online shopping which has many known reasons like, time-saving, brand visibility, offers, messaging, personalise recommendations and many more. The business personnel are also working parallelly to compete with the customer choices and the recommendations made by the system working in backend. It is the responsibility of businessmen to increase choices for consumer other than browsing history-based recommendations. Doan [1] has suggested some guidelines to improve online purchase intention like, value social influence, user-friendly website, and precise processing speed.

The business threats are for both the retailers and pure e-commerce dealers. There can be many reasons behind it. Dahiya [2] has briefed few reasons for e-commerce dealers like, customer trust factor regarding misuse of his personal information, customer feels uneasy initially while selecting any product without looking and touching product. Every website is not appealing for every customer due to type of his taste. For example, in case of clothing, there is variation in choices of different customers as in, what one prefers to carry. The same is the case with the person who is hosting a website and recommending products to the customers. It is quite tedious to come up with the right results as per the customer's choice. The sale

and business profit highly depend upon it. As per an article, Buckland [3], personalised recommendations hit the business by 43%. So, it is a considerable fact for which outcome-based measures are highly required. Customers sentiment and opinion plays a great role in revealing his lifestyle, choices and likings. Fan *et al.* [4] used sales and online review data of automobiles to extract sentiment information which is based on feedback or browsing history. With similar viewpoint, Hou *et al.* [5] examined the online reviews to predict product sales. Agarwal [6] found few Indian brands which promote the handicraft sale. But there is a problem of overspecialization when products are recommended as per customers past reviews repeatedly and customer is fed up with the monotonous pattern of product list displayed to him. Some more variables are required to add novelty and gain customers interest in one's business.

This research is for India which is well known for the diversity and regional skills of people defining the country's heritage. Since ancient times, the business was generally based on area-wise culture and agriculture, mostly in rural areas. Accordingly, there are various domains like agriculture and cultivation, weaving, pottery making business, handicraft, Kolhapuri footwear, furniture, grass-work, marble-work and many more. But after globalization and tech-advancement these rural businessmen suffered a lot due to low product sale. One of the reasons is advent of online shopping culture followed by the customers where he makes the choices as per the displayed products. In this research, location is considered as one of the ways to improve marketing strategy and also support retailers. For some businessmen location can be an issue but for some it is a boon also. Like, the businessmen near production site are privileged, location-wise. To accomplish it, the gap between customer and businessman must be bridged. For this customer location can be utilized to optimize the scenario which will simultaneously be beneficial for businessmen also.

Accordingly, data collection and data preparation tasks are the grounds to proceed for the outcome. Afterwards, data mapping has been done using the collected databases. Finally, a complete framework has been proposed which creates a predictive system and makes suitable predictions regarding products to the customer.

2. LITERATURE REVIEW

E-commerce industries are highly popular and they are proving to be a part of lifestyle. In metro cities, they are used by maximum people due to shortage of time and easy availability-based service. But in the current research, the concern is for rural areas business and people who are not handy with e-commerce industry. Kalambe [7], has discussed about challenges faced by urban and rural areas in India in context to e-commerce. Companies must try to reach rural areas also not only for product sales but also for utilizing the rural skills of businessmen. Mukherjee [8] has mentioned in an article about suicides by weavers of Andhra Pradesh due to less business income as the handicraft products are not reaching to the customers. Such scenarios are mandating the requirement of e-business by villagers. Kshetri [9] has analysed the factors behind rural-urban differences due to e-commerce. Wright [10] has discussed about Anglo-Indians who were left behind in rural areas and are facing problem in maintaining their culture and identity. So, this is true that developing countries face issues with e-commerce for rural areas people. Research in this direction may help people and industry both.

Customer reviews are very significant in analysing the online business. These ratings help in further analysing the products to be purchased by the customer. Jain and Kumar [11] identified the characteristics of customer reviews to help businessmen in improving marketing strategies. He applied analytics on social media data to identify demographic, psychographic and lifestyle information about customer. Obal and Lv [12] promoted banner ads with high visual complexity and calculated effective cost per activity using predictive modelling methods. Zhang *et al.* [13] discussed about online product recommendations (OPR) where a model has been proposed to find out positive and negative factors of OPR. Xu [14] analysed the effect of online customer reviews and the parameters which justify satisfaction. Huang *et al.* [15] proposed a graph model for e-commerce recommender systems. All such researches clarify that predictions can be done by using customer's information and their liking or disliking for the product. This is one of the research gaps where it is observed that product recommendation is done only using user feedback which creates over similarity in recommendations due to which no variety is seen by the user in the recommended items.

Singhal *et al.* [16] discussed Machine learning techniques which deal with various applications today belonging to different domains. These techniques help in data extraction. Castelli [17] discussed about knowledge extraction from customer's reviews using amazon.com database. He has discussed about prediction of successfully available products on amazon. Eslami *et al.* [18] analysed a fact that lengthy reviews give most helpful idea about products and reviews. Chou and Chuang [19] did feature extraction in online reservation system for restaurant where predictive modelling has been applied. Liébana-Cabanillas and Lara-Rubio [20] explored mobile payment system from merchant's point of view which indicated that m-payment behaviour is not applicable globally. Ge [21] developed a distributed prediction modelling

framework for plant-wide process of industry where reasons for low performance were diagnosed. Also, TE benchmark process was used to evaluate the proposed model performance.

Chin *et al.* [22] compared traditional partial least squares structural equation modelling (PLS-SEM) with other methods like PLSpredict, CVPAT and model selection criteria (i.e. Bayesian information criterion (BIC), BIC weight, geweke-meese criterion (GM), GM weight, HQ and HQC) where low performance of traditional method has been found. Min *et al.* [23] used predictive modelling in a case study on chronic obstructive pulmonary disease (COPD) where prediction performance was analysed with respect to parameters like knowledge driven features, data driven features and one-year history of patient before discharge. Kim *et al.* [24] suggested benefits and challenges in multi-omics predictive analytics. Also, classification of omics was done. Singhal *et al.* [25] analysed classification algorithms which are suitable for e-commerce domain. Malhotra *et al.* [26] reviewed search-based techniques for predictions and effort estimations while using predictive modelling. Wazurkar *et al.* [27] suggested a decision-making process to handle large size dataset with minimum errors and failures using predictive analysis. Tax *et al.* [28] found in his research that long-short term memory (LSTM) neural network performs better while predicting about next event of current activity along with its time duration. Metzger *et al.* [29] used ensemble prediction techniques and applied on an industry dataset. It was found that there was a positive effect on cost estimation for proactive business process adaptation. Kang *et al.* [30] developed a predictive model for environmentally sustainable product purchases. This model showed a better prediction as compared to the theories applied earlier. For current purchases, he found that personal norms are better predictors for those who take risks. Mehdizadeh *et al.* [31] reviewed descriptive and predictive modelling techniques and their multiple applications with respect to road safety.

Busari *et al.* [32] developed a predictive model for the trip pattern for low density area. The study found that a car ownership highly increases the number of trips on daily basis where transportation planning and proper infrastructure may add on to daily trips. Yang *et al.* [33] did a survey on social networks using predictive analytics where he identified few basic tasks which can be performed by using social networks. Like, link prediction, group formation, decision support system, risk analysis and planning. They have also studied about applications based on social network which predicts those persons who are likely to follow some other persons. Francescomarino *et al.* [34] developed combination of techniques, recurrent neural network and long-short term memory cells which resulted into a better performance while predicting about future activities of ongoing current activity. Harl *et al.* [35] used gated graph neural network for decision making. He analysed that different activities which are included in the process, effected the prediction. Singhal *et al.* [36] proposed a logical structure to improve presentation of a web page, making it more user-friendly to promote e-business. Musso *et al.* [37] used machine learning predictive models to predict low and high levels of performances in subjects in educational systems. Márquez-Chamorro *et al.* [38] summarises about basic concepts of predictive monitoring area in business processes.

Overall, it has been observed that predictive modelling plays a vital role in business-oriented applications and also in decision support systems. It is a way to overcome unwanted delays and flaws in system along with cost optimization. At the same time, one more research gap is observed that the till now predictive systems were developed where clustering of users were made but in current research clustering of products is done to make recommendations.

3. RESEARCH METHODOLOGY

3.1. Data-set preparation phase

In earlier researches, datasets based on online customer reviews has been already prepared and collected by researchers before like Hou *et al.* [5] analysed online review data collected from Amazon.com which is ideally a large volume of data. But here two things are different. Firstly, customer ratings are not considered which has been collected as per browsing history or any other feedback mode. More precisely, sentiment analysis is not the part of research as it has no novelty. Secondly, the dataset is specifically for India and Indian citizens.

3.1.1. Data model

The proposed data model is inspired by demographic details and other social details of a citizen. There are two different databases which has been considered here: i) region-wise data, ii) existing user data, and iii) user-product mapping

– Region-wise data

India is a culture rich country where every different region is identified by its culture and practices. As discussed in introduction, there are unlimited domains which define skillset of Indians belonging to different regions. These domains can be varied in the sense that they are promoting different business sectors

like weaving, printing, furniture work, handicraft and many others. This richness in variety of skillsets in every region is the key which defines Indian heritage.

Every village, city and town belonging to a state covers a different form of products. These products are developed by the people who are belonging to that place and they are trained for making those items. It is an art of such skilled persons which they have learned from their ancestors. But unfortunately, more often these skilled persons are not that much educated overtime. Due to low level of literacy, they could not match themselves with the fast pace of online business applications and their practice [37]. Hence, they are lagging behind financially. In rural areas, there are skilled persons but they face unemployment and poverty. One could not reach them to help them financially but yes usage of their skill-based products can be improved by some rate gradually. Online customers are willing to get such products which they have seen in their houses since childhood, but they could not find them easily during online shopping. These observations can be formulated in the form of a record which can be used to create a product and region-based database.

Considering the above scenario, the region-wise database has been prepared. The data has been collected from various authorised resources. Indian maps have been also considered while data was collected [39]. These maps indicated the products in symbolic form for all the states. There are 28 states in India as per the year 2021. The database has been created for all the states. While preparing state-wise database, many categories of products like wood work, glass beads, zari, metal work and many more were observed. So, domains have been categorised and then included in the database. In Figure 1, a snapshot of the created database has been displayed. Here data for few states have been mentioned to give a glimpse. This database will help once we identify the region of a user.

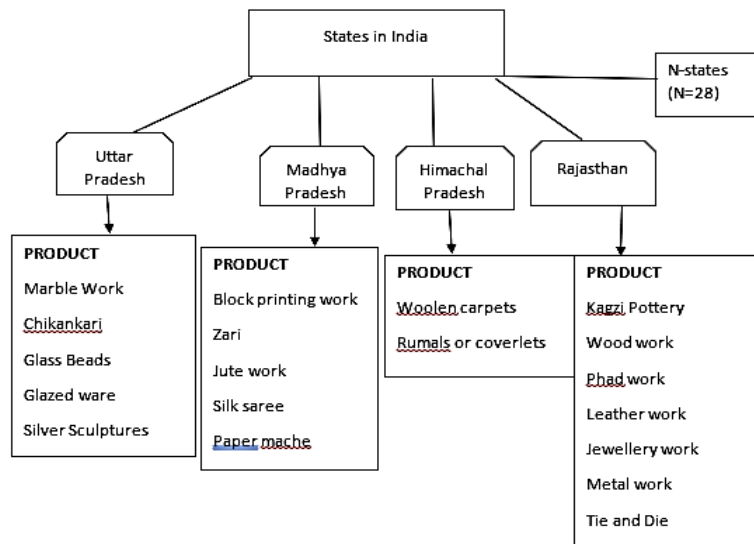


Figure 1. Region-wise database

– Existing user data

On the globe, everywhere there is a system of record keeping about people. This record can have different basis. Few records are containing a person's personal details, few may contain his official details, few may contain his citizenship details and many other different types of details. Such type of data is considered as raw data. It is a general practice that raw data is refined before using anywhere like, in any application. Here comes the usage of data mining. Data mining is used to extract data from an existing data as per requirement. One can say that by using data mining, an informed data can be created.

In DBMS, there is a concept of relational database. For a particular requirement there can be multiple tables, containing different attributes among which few may be common. There is also a command of joins which helps in joining multiple tables to develop a single table containing combination of maximum required attributes. Using all these techniques one can create a database of user's details which may support in online business. Existing user database has been created for the current research. The citizen details which are already existing for a user (belonging to a country) can be utilized here for the current model. This database has been created from various datasets which are collected from the authorized sources. Multiple datasets mean here that for example, Amazon online shopping site customers dataset is available in many

forms for the same Indian customer. The difference may be in number of attributes, type of attributes or data values.

In Table 1 a sample table has been shown for the existing user data. This table contains User id, Gender, Age, State, District, Tehsil, Occupation and Pincode as attributes. User id is an attribute which is generally used in databases for identifying a customer uniquely. While joining multiple tables containing same user data this user id is used as primary key. State, District, Tehsil and Pincode are attributes which compile address details through which region in India can be matched. The attribute ‘occupation’ helps in creating further combinations of customers for clustering.

Table 1. Sample table for customer details

S.No.	Gender	Age	State	District	Tehsil	Occupation	Pincode
1	M	24	Gujrat	Valsad	Valsad	Technician	396001
2	F	53	Rajasthan	Sirohi	Abu Road	Other	307026
3	M	35	Uttar Pradesh	Gonda	Utraula	Executive	271604
4	M	44	Tamilnadu	Madurai	Uttamapalayam	Administrator	625533
5	M	32	West Bengal	Darjiling	Naxalbari	Artist	734429
6	F	40	West Bengal	Jalpaiguri	Kumargram	Educator	736203

– User-product mapping

Whenever E-commerce based databases are observed, always there is some type of mapping. Similarly, in the current research mapping has been done. Mapping means using different attributes of different tables and making a matrix structure. Here users and products have been considered as the attributes which are cross verified and thus values are generated. Table 2 is a sample table where it is observed that values are either 0 or 1. So, if a user has selected or purchased that product earlier then the value is 1 otherwise it is 0. From here similarity can be calculated among users and thus similar users can be assigned to the specific group or cluster of users. Thus, the mapping plays a very significant role in providing a knowledgeable input to the proposed predictive system framework.

Table 2. User-product mapping

Products User	P1	P2	P3	P4
User 1	1	1	0	0
User 2	0	1	1	0
User 3	1	0	1	1

3.2. Proposed framework for the predictive system

Earlier, when business was totally based on manual procedures and calculations then it was very typical to analyse and keep the updates of one’s business state. The major drawback was that even after much hard work, it was very tough to do analysis and make any assured decision related to business for future. Estimating any business performance is another issue which is not very easy to resolve because there are various parameters for performance analysis. Based on a particular factor, the performance may vary for the same business if checked with respect to another parameter. All such issues can be resolved with the usage of technologies. As per the state of art, in an online business, or more precisely, E-commerce sector, maximum tasks are sequenced and executed on the basis of some technologies like predictive system, business analytics, supply chain management, big data analytics and many more. The current research involves the concept of predictive modelling.

As per Heo *et al.* [40], Predictive modelling is a technique which uses current and historical data to make predictions about the most probable future results. This clarifies that data input is required for a predictive system. Proceeding in the same direction, Data sets have been created in the current research. The details have been discussed above. In Figure 2, at the top of the framework, there is a block named “User”. From ‘user’ block, in right direction, there is a database connected with name ‘Existing user data’. This is the database which already exists for every citizen of a country. How and from where this database has been gathered, has already been discussed above. Moving further, there are blocks which represents Data Pre-processing phase in sequence. Let’s first discuss what is actually data pre-processing. Alexandropoulos *et al.* [41] has discussed about data pre-processing in detail. However, in current research, few required phases have been considered. For solving any prediction issue, it is required to find out the parameters and attributes from the databases which are mandatory inputs. This is done through data extraction.

Vyas *et al.* [42] discussed ETL process. On the existing data set, first of all data extraction has been performed. Data extraction is generally the first step towards ETL, that is, extraction-transformation-loading. Data extraction simply means to fetch out or segregate those attributes along with their values which are actually needed for the data analysis part. Most of the times, data is very poorly organised. Also, data is collected from various sources due to which there can be different types of data, like, continuous data, batch data, and many more.

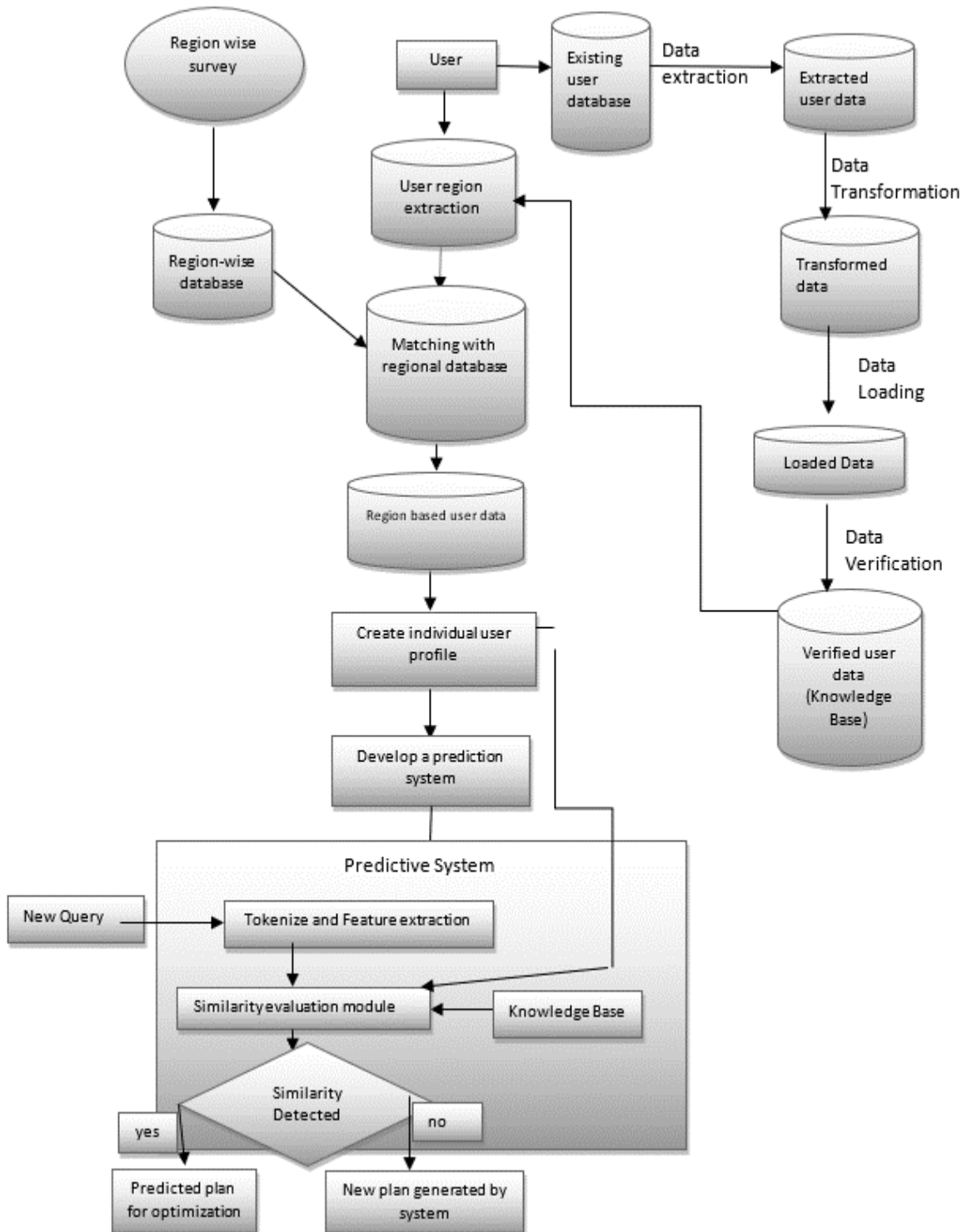


Figure 2. Block diagram for the proposed framework

All such shortcomings are considered during data extraction as, it combines the heterogeneous data. Also, it identifies the relevant data out of bulk data and removes duplicate data also. Overall, it refines the data for business intelligence. Hence, the resulting data is a mined data. In Figure 2, after data extraction, the resulting data is named as 'extracted user data'. This data is a relevant database on which further processes can be applied completely. After extracting data, transformation is required. Vyas *et al.* [42] and Runtuwene *et al.* [43] has discussed about transformation process. Transformation cleanse and aggregate data to give it a form such that it can be used for analysis. Cleaning means applying various operations which are customized as per the condition of extracted data. Operations may apply things like:

- Providing values to blank places, like, replacing NULL values by zero.
- Converting required text with respect to data type, units, date, time and many more.
- Apply data integrity rules to remove issues like same data with different names and vice-versa.
- Joining or splitting tables or table columns using key attributes.
- Apply business-based functions or rules to generate new calculated values from the extracted data.

In Figure 2, data transformation is applied on extracted user data, resulting into 'transformed data'. Transformed data is organized and formatted in improved way as compared to extracted data. It is also compatible with the application and system. Hence analytics can be applied on transformed data. This data needs to be loaded afterwards.

Data loading' is done after transforming data. Loading the transformed data into a warehouse, cloud or some other storage system is done in this phase. This helps in streamlining the data for future requirements. Higher efficiency and flexibility play a very important role as data is growing day by day. This is one of the major requirements of many business companies which are based on E-commerce. In Figure 2, 'Loaded data' is received after applying data loading operation. Last but not the least, loaded data passes through the final phase of verification.

Data verification' is an operation which is applied on the loaded data. Already data transformation stage has been crossed but still data verification is required for validating the data. Data validation ensures high accuracy in results. During data verification few things are cross checked like, data content, format of data, data duplicity and many other such type of database issues. Hence, after data verification, a verified user data, i.e., 'Knowledge base' is received as represented in Figure 2.

As discussed earlier in database section, there is a 'region-wise database' which has been collected after performing 'region-wise survey' as shown in Figure 2. Both the region-wise database and existing database are used to perform matching operation as seen in block, 'Matching with regional database' in Figure 2. Matching is done on the basis of user's region contained in knowledge base. When the region values collected from both the regional database and knowledgebase matches then the user profile can be prepared.

Region based user data' block is developed by receiving region values from both types of databases. Join operation on tables is applied to perform matching and segregating a single database per user. This part is very important for the current research as it is the key element which adds to the novelty regarding E-commerce sector where region-based products concept has been proposed. Using this data, 'Individual user profile' is created as shown in Figure 2. This user profile helps in identifying user's on internet. For any E-commerce website, user is as important as products in context to sale. Hence tracking user is a mandatory task in online business. Once user profile has been created, its required to use it as input for further processing. In the proposed work a predictive system has been framed. Here the user details can be used as input values. A predictive system always provides some probable outcomes which can relate and effect future decisions. In this research a predictive system has been proposed as shown in Figure 2.

Predictive system takes user query as input where the query may contain information about product or type of product. The internal structure of proposed predictive system contains mainly three modules: 'tokenize and feature extraction', 'similarity evaluation' and 'similarity detection'. First of all, 'tokenize and feature extraction' module takes the user query as input. This module performs tokenization. Tokenization means bifurcating the input text into set of meaningful words. These partitioned words are called as tokens. These tokens are further user used for feature extraction. Feature extraction means reducing the existing dataset or word content in such a way that the resulting text set is able to summarize the maximum information. Thus, after feature extraction the data is in a state using which further predictions are possible. For example, using feature extraction one can predict whether the fabric of a cloth is cotton or not. Hence, this module is helpful in adding meaning to the user query.

After feature extraction, similarity evaluation is done. For finding similarity, similarity evaluation module considers user profile, knowledge-base and output of previous module (tokenize and feature extraction) as input values as shown in Figure 2. Similarity check operation is performed here to find the similarity in product choices by user as compared to other users. Also, region-wise similarity is checked to match current user with the other similar users. This module plays a significant role in research as prediction

results are based on it. Finally, the results of similarity evaluation module are used. In similarity detection module, condition is checked whether similarity exists or not. The percentage of similarity also matters in this case as its never an ideal state as in when one can predict with hundred percent yes or no. The percentage decides about similarity existence. For example, if a person asks about chicken kurta fabric then there is high probability for liking towards other chicken fabric-based suits. Thus, the prediction system may predict some optimized results if similarity exists. But if similarity percentage is very low then in that case the model will itself suggest some other plan on the basis of similar user clusters as shown in Figure 2.

4. CONCLUSION

Currently world is facing problem of COVID-19 to its maximum extent. As people are not in a situation to visit crowded places, they are meant to remain in their houses. In this situation they are trying to purchase things online. The current research work emphasizes on rural area based skilled businessmen who are having very little finance. They are not much literate and due to E-commerce industry; they are facing issues of low sale and low income. In this situation they can join hands with the E-commerce business holders and supply their products or they can themselves start online sale. To support these people a framework has been proposed here which emphasises on increase in regional product sale. This model considers region of a person or customer as input and then predict the products to be advertised on a person's screen. This will promote the regional business to some extent and help the poor regional businessmen to sell their product. Also, customer satisfaction improves as the customer is able to find and select those products which belongs to their region. This is however going to be very fruitful among Indians as every religion is full of various festivals during which they pursue the ancestral practice of using things which are produced in their area.

5. FUTURE SCOPE

The current research work generates many parameters and factors for further research. Researchers can work in several other directions making this research work as baseline. It has been observed that few skills are common among more than one state in India, like, silk weaving is done in Telangana, Tamilnadu, Maharashtra, Gujarat, and Assam. This combination can be used in future for further clustering. Another factor is that as the proposed database is created for India, however, such type of database can be collected for any other country also and promote E-commerce industries. One more direction is that using this data, 'Individual user profile' is created as shown in Figure 2. This user profile is playing role in identifying user's on internet. So, this data can be used further for different types of customer-based clustering. Another suggestion is that many other parameters can be added to find similarity from the databases. Seeing so many suggestions, researchers can perform well in this area.

REFERENCES

- [1] T. Doan, "Factors affecting online purchase intention: A study of Vietnam online customers," *Management Science Letters*, vol. 10, no. 10, pp. 2337-2342, 2020, doi: 10.5267/j.msl.2020.3.001.
- [2] M. Dahiya, "Study on e-commerce and it's impacts on market and retailers in India," *Advances in Computational Sciences and Technology*, vol. 10, no. 5, pp. 1495-1500, 2017.
- [3] J. Buckland, "Best E-commerce articles of 2018 with 10 lessons to guide you into the new year," *shopify.in*, 2019.
- [4] Z. P Fan, Y. J Che, and Z. Y Chen., "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis," *Journal of Business Research*, vol. 74, pp. 90-100, 2017, doi: 10.1016/j.jbusres.2017.01.010.
- [5] F. Hou, B. Li, A. Y. L. Chong, N. Yannopoulou and M. J. Liu., "Understanding and predicting what influence online product sales? A neural network approach," *Production Planning and Control*, vol. 28, no. 11-12, pp. 964-975, 2017, doi: 10.1080/09537287.2017.1336791.
- [6] P. Agarwal, "These 5 brands are reviving Indian handicrafts and crafting a livelihood for rural artisans," *yourstory.com*, India, 2019.
- [7] D. Kalambe, "Challenges and oppourtunities of ecommerce in urban and rural sector in india," *Paripex-Indian Journal of Research*, vol. 8, no. 9, 2019.
- [8] B. Mukherjee, "How Globalization and Tech-advancement Are Benefitting Rural Craft Communities," *entrepreneur.com*, India, 2018.
- [9] N. Kshetri, "Rural e-commerce in developing countries," *IT Professional*, vol. 20, no. 2, pp. 91-95, 2018, doi: 10.1109/mitp.2018.021921657.
- [10] R. D. Wright, "The shattering of cultural identity: the Anglo-Indian community in rural India," *International Journal of Anglo-Indian Studies*, vol. 2, no. 1, 2017.

- [11] V. K. Jain, and S. Kumar., "Improving customer experience using sentiment analysis in e-commerce," In *Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics*, IGI Global, pp. 216-224, 2017, doi: 10.4018/978-1-5225-0997-4.ch012.
- [12] M. W. Obal and W. Lv., "Improving banner ad strategies through predictive modelling," *Journal of Research in Interactive Marketing*, vol. 11, no. 2, pp. 198-212, 2017, doi: 10.1108/JRIM-08-2016-0092.
- [13] H. Zhang L. Zhao, and S. Gupta, "The role of online product recommendations on customer decision making and loyalty in social shopping communities," *International Journal of Information Management*, vol. 38, no. 1, pp. 150-166, 2018, doi: 10.1016/j.ijinfomgt.2017.07.006.
- [14] X. Xu, "Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants," *Journal of Business Research*, vol. 106, pp. 196-210, 2020, doi: 10.1016/j.jbusres.2018.07.022.
- [15] Z. Huang, W. Chung, and H. Chen., "A graph model for E-commerce recommender systems," *Journal of the American Society for information science and technology*, vol. 55, no. 3, pp. 259-274, 2004, doi: 10.1002/asi.10372.
- [16] S. Singhal and P. Tanwar, "Machine learning techniques and their problem-specific domain based excellence," *Journal of Critical Reviews*, vol. 7, no. 6, pp. 1841-1850, 2020, doi: 10.31838/jcr.07.13.339.
- [17] M. Castelli, L. Manzoni, L. Vanneschi, and A. Popovic, "An expert system for extracting knowledge from customers' reviews: The case of Amazon. Com Inc," *Expert Systems with Applications*, vol. 84, pp. 117-126, 2017, doi: 10.1016/j.eswa.2017.05.008.
- [18] S. P. Eslami, M. Ghasemaghaei, and K. Hassanein., "Which online reviews do consumers find most helpful? A multi-method investigation," *Decision Support Systems*, vol. 113, pp. 32-42, 2018, doi: 10.1016/j.dss.2018.06.012.
- [19] Y.C. Chou and H. H. C. Chuang, "A predictive investigation of first-time customer retention in online reservation services," *Service Business*, vol. 12, no. 4, pp. 685-699, 2018, doi: 10.1007/s11628-018-0371-z.
- [20] F. Liébana-Cabanillas and J. Lara-Rubio., "Predictive and explanatory modeling regarding adoption of mobile payment systems," *Technological Forecasting and Social Change*, vol. 120, pp. 32-40, 2017, doi: 10.1016/j.techfore.2017.04.002.
- [21] Z. Ge, "Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes," *Journal of Process Control*, vol. 65, pp. 107-117, 2018, doi: 10.1016/j.jprocont.2017.08.010.
- [22] W. Chin, J. H. Cheah, Y. Liu, H. Ting, X. J. Lim, and T. H. Cham., "Demystifying the role of causal-predictive modeling using partial least squares structural equation modeling in information systems research," *Industrial Management and Data Systems*, vol. 120, no. 12, pp. 2161-2209, 2020, doi: 10.1108/IMDS-10-2019-0529.
- [23] X. Min and F. Wang, "Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: a case study on COPD," *Scientific reports*, vol. 9, no. 1, pp. 1-10, 2019, doi: 10.1038/s41598-019-39071-y.
- [24] M. Kim and I. Tagkopoulos, "Data integration and predictive modeling methods for multi-omics datasets," *Molecular omics*, vol. 14, no. 1, pp. 8-25, 2018, doi: 10.1039/c7mo00051k.
- [25] S. Singhal and P. Tanwar, "Analysis of classification algorithms compatible with e-commerce technology," *International Journal of Advanced Science and Technology*, vol. 29, no. 3, pp. 14329-14336, 2020.
- [26] R R. Malhotra, M. Khanna, R. R. Raje, "On the application of search-based techniques for software engineering predictive modeling: A systematic review and future directions," *Swarm and Evolutionary Computation*, vol. 32, pp. 85-109, 2017, doi: 10.1016/j.swevo.2016.10.002.
- [27] P. Wazurkar, R. S. Bhadoria, and D. Bajpai, "Predictive analytics in data science for business intelligence solutions," In *Proc of 7th International Conf on Communication Systems and Network Technologies (CSNT)*, 2017, pp. 367-370, doi: 10.1109/CSNT.2017.8418568.
- [28] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," In *Proc of International Conf on Advanced Information Systems Engineering*, Springer, Cham, 2017, pp. 477-492, doi: 10.1007/978-3-319-59536-8_30.
- [29] A. Metzger and F. Föcker, "Predictive business process monitoring considering reliability estimates," In *Proc of International Conf on Advanced Information Systems Engineering*, Springer, Cham, 2017, pp. 445-460, doi: 10.1007/978-3-319-59536-8_28.
- [30] J. Kang and F. Moreno, "Driving value to action: Predictive modeling for environmentally sustainable product purchases," *Sustainable Production and Consumption*, vol. 23, pp. 224-235, 2020, doi: 10.1016/j.spc.2020.06.002.
- [31] A. Mehdizadeh *et al.*, "A review of data analytic applications in road traffic safety. Part 1: descriptive and predictive modelling," *Sensors*, vol. 20, no. 4, p.1107, 2020, doi: 10.3390/s20041107.
- [32] A. Busari *et al.*, "Trip pattern of low-density residential area in semi urban industrial cluster: Predictive modelling," *Int. J. Hum. Capital Urban Manage*, vol. 2, no. 3, pp. 211-218, 2017, doi: 10.22034/ijhcum.2017.02.03.005.
- [33] M. Yang, W. H. Hsu and S. T. Kallumadi, "Predictive analytics of social networks: A survey of tasks and techniques," In *Social Media Marketing: Breakthroughs in Research and Practice*, IGI Global, pp. 823-862, 2018, doi: 10.4018/978-1-5225-5637-4.ch042.
- [34] C. Di Francescomarino, C. Ghidini, F. M. Maggi, G. Petrucci, and A. Yeshchenko., "An eye into the future: leveraging a-priori knowledge in predictive business process monitoring," In *Proc of International Conf on Business Process Management*, Springer, Cham, 2017, pp. 252-268, doi: 10.1007/978-3-319-65000-5_15.
- [35] M. Harl, S. Weinzierl, M. Stierle, and M. Matzner, "Explainable predictive business process monitoring using gated graph neural networks," *Journal of Decision Systems*, pp. 1-16, 2020, doi: 10.1080/12460125.2020.1780780.
- [36] S. Singhal and P. Tanwar, "Logical structure for user friendly dynamic web page visualization for small screen terminals promoting E-Business," *Recent Patents on Engineering*, (Bentham Science Publishers), 2020.-online version awaited, doi: 10.2174/1872212114999201109204536.

- [37] M. F. Musso, E. C. Cascallar, N. Bostani and M. Crawford, "Identifying reliable predictors of educational outcomes through machine-learning predictive modeling," In *Frontiers in Education Frontiers*, vol. 5, pp. 104, 2020, doi: 10.3389/feduc.2020.00104.
- [38] A. E. Márquez-Chamorro, M. Resinas, and A. Ruiz-Contes, "Predictive monitoring of business processes: a survey," *IEEE Transactions on Services Computing*, vol. 11, no. 6, pp. 962-977, 2017.
- [39] Usha, "Handicrafts in North India," India, 2013.
- [40] J. Heo, H. Lim, S. B. Yun, S. Ju, S. Park, and R. Lee, "Descriptive and predictive modeling of student achievement, satisfaction, and mental health for data-driven smart connected campus life service," In *Proc. of the 9th International Conf. On Learning Analytics and Knowledge*, Tempe AZ USA, 2019, pp. 531-538, doi: 10.1145/3303772.3303792.
- [41] S. A. N. Alexandropoulos, S. B. Kotsiantis and M. N. Vrahatis, "Data preprocessing in predictive data mining," *The Knowledge Engineering Review*, vol. 34, 2019, doi: 10.1017/S026988891800036X.
- [42] S. Vyas, and P. Vaishnav, "A comparative study of various ETL process and their testing techniques in data warehouse," *Journal of Statistics and Management Systems*, vol. 20, no. 4, pp. 753-763, 2017, doi: 10.1080/09720510.2017.1395194.
- [43] J. Runtuwene, I. R. Tangkawang, C. T. M. Manoppo, and R. J. Salaki, "Comparative analysis of extract, transformation and loading (ETL) process," In *Proc. Of 2nd International Conf. on Innovation in Engineering and Vocational Education, Manado, Indonesia*, vol. 306, 2018, doi: 10.1088/1757-899X/306/1/012066.

BIOGRAPHIES OF AUTHORS



Shefali Singhal received her Btech. Degree in Computer Engineering from Bundelkhand University, Jhansi and Mtech. Degree from YMCA University, Faridabad. She is pursuing her PhD. in Computer Science From MRIIRS, Faridabad. Currently she is working as Assistant Professor in MRIIRS with a teaching experience of more than 10 years. Her main research interests are predictive modelling, machine learning, big data analytics, algorithms and computer networks. She authored a book, 'Analysis and Design of Algorithms', BpB publications in 2018. Also, she has published papers in Scopus Indexed Journals. She has also chaired session in Rajasthan (India).



Poonam Tanwar has 17 years of Teaching Experience working as associate prof. in Manav Rachna International Institute of Research & Studies, Faridabad. She has filled 4 patents. She was Guest Editor for Special issue of "Advancement in machine learning (ML) and Knowledge Mining (KM)" for International Journal of Recent Patents in Engineering (UAE). She has organized various Science & Technology awareness program for rural development. Beside this she has even published more than 40 research papers in various International Journals and Conferences. She is Technical program committee member for various International Conferences.