A Hybrid Smartphone Recommendation Model Using Different Machine Learning Algorithms

Kalash Chandak[1], Anuja Bokhare[2], Samaya Pillai[3] and Vanishree Pabalkar[4]

[1]Symbiosis Institute of Computer Studies and ResearchSymbiosis International (Deemed University) Pune, India
[2]Dr. Vishwanath Karad MIT World Peace University, Pune 411038, Maharashtra, India
[3]Symbiosis Institute of Digital and Telecom Management Symbiosis International (Deemed University) Pune, India
[4]Symbiosis Institute of Management Studies, Symbiosis International (Deemed University), Pune, India

E-mail address: kalashchandak@gmail.com, anuja.bokhare@gmail.com, samaya.pillai@sidtm.edu.in, vanishree.p@sims.edu
Correspondence: anuja.bokhare@gmail.com

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: In today's world, mobile phones are an essential part of many people's daily lives. With several smartphones to choose from, people are often confused about which smartphone will be best for their use. This study attempts to present an accurate and usable prediction model for real-life mobile phone reviews of individual users to improve the model's prediction accuracy. A review-based prediction model is built on individual users' behavior and choice using machine learning. This data has been collected using web scraping tools like UiPath and the Python Beautiful Soup library from different websites, followed by data pre-processing. In this study, different Collaborative Filtering-based machine learning algorithms have been used and compared. The algorithms used include KNN based on individual items or individual users and an unsupervised SVD-based model. This has been demonstrated using UiPath Studio and the Druid AI chatbot. The Druid chatbot provides smartphone recommendations and data based on user input. Upon entering a smartphone name, a UiPath process is triggered, sending results back to the chatbot. This UiPath-based chatbot delivers specifications and recommendations. Future enhancements include broader product recommendations, improved user understanding through advanced NLP training, and an overall better user experience. Additionally, there will be a focus on incorporating user feedback to continuously refine and enhance the prediction model, ensuring it remains relevant and highly accurate.

Keywords: Mobile Recommendation, Phones, Machine learning, KNN, collaborative filtering, SVD, NLP, chatbot, UiPath Studio, Druid AI

1. Introduction

The world is evolving rapidly in terms of technology nowadays. As far as today's times are concerned, every day is a busy day and the most crucial deciding factors are time, cost and efficiency. With technology evolving each day, it is becoming difficult to understand each and every minor detail of any website or web application that is developed or being developed.

Apart from understanding the complex user interface, the daily updates of every product also make it difficult to search for a particular product, analyze each and every specification of it and compare it on the website interface itself for final purchase. Plus, there is nothing over the e-commerce website or web application that recommends a particular product or a range of products that the customer should purchase at that very moment. In addition, it is also very difficult to analyze the fake products and filter them according to the current market situation. These are the features that the current e-commerce-based processes lack and can cause theft, cheating and misleading products trending over the web that the customer is unaware of. Manually taking care of all these activities is highly time consuming, tedious and can lead to inaccurate product recommendations and results.

For solving this problem, the required product data needs to be extracted at one place from single or multiple websites and then further analysed to provide recommendations in terms of trend, cost and user reviews. All of this can be achieved faster by automating the specific e-commerce process. So 'which technology can be used to automate the process?', 'which algorithm provides the best recommendations?' and 'How to demonstrate the process?'. All of these questions have been tried to answered in this study.

2. PREVIOUS STUDY

Author [1] have stated that recommender systems are computer programs that make product recommendations to customers based on a variety of factors. Recommender services are used by companies like Netflix, Amazon, and others to help their consumers locate the correct things or movies for them. TF-IDF and doc2vec are used to measure the film similarities using movie definition and keywords as inputs. For the first time, the film ranking experienced was used by users as input to the algorithm, and use of 'K nearest neighbors' as factorization of the matrix to estimate film scores for customers. Authors have discovered that collaboration works better than content. Predictive error and time filtering estimation. Authors [2] have stated that even as ages pass, the desire for new clothing persists. Everyone wishes to look and feel their best. As a result, authors have suggested an outfit recommender system that uses the KNN algorithm to recommend outfits based on user ratings. Here, authors attempting to decrease customer uncertainty about which outfit will go with their current clothing. The error occurs is low because the recommender system relies on the user's description.

Authors [3] have stated that individualized product recommendations are in high demand these days. The goal of this project is to develop a location-based agricultural product recommendation system that uses a unique KNN algorithm to ensure effective communication and transparency in agricultural trade marketing between buyers and sellers (farmers). The information provided by the farmers and buyers is dynamically recorded and updated in a database. Based on buyer desire, the recommender system suggests nearby suppliers and their agricultural products. The system's performance is analysed.

Authors [4] have focused on recommending books using the 'k-nearest neighbor' algorithm. A basic overview of various recommendation techniques like content based, collaborative and associative mining has been provided. The research explains each step systematically to its audience through a detailed process flow diagram. In the algorithm stage, after dataset factorization, the KNN is applied using Euclidean distance, cosine similarity and Pearson correlation methods to identify the item similarity. Then the association rules and frequent item set (Apriori algorithm) was used or identifying the frequently read books from the dataset. Finally, the research proposes that instead of the 'Apriori mining' algorithm, the Sparks ALS can also be used for more efficient data mining. Authors [5] have proposed the use of content-based filtering using KNN for recommending movies. The problems in content-based filtering are mentioned and its prevention methods are also stated. The cosine theta angle is used to find the movie similarity score. Finally using the KNN, the nearest neighbour value is calculated for the list of movie recommendations the user. Hence, it was stated that cosine similarity provides a greater accuracy that other distance metrics and gives a lower time and space complexity as well.

Authors [6] have proposed the restaurant recommender system based on cuisine search by the user. The recommender system has been developed using the KNN and MapReduce approach. The research addresses the issue of providing the restaurant recommendation based on user cuisines. Authors [7] have reviewed and provided a systematic analysis of different machine learning algorithms that can be used in recommender systems and provides a path for further research opportunities. It looks at ML algorithms that are not classified which results in difficult to choose an algorithm while implementing recommendation systems. So, this study provides a review on which algorithms are used in which applications and reviews the research in the same. Wro´blewska et al. (2018) have demonstrated the process of building an artificial intelligence system which performs formal document processing tasks in different kinds of business processes. Authors [8] have reviewed how recommender systems works, which is an application of Artificial Intelligence; can predict large scale data as per the requirements. However, in this paper artificial intelligence is not mentioned using which recommendation systems can be greatly improved. Also, the study looks at the comparison between various algorithms and methods that have been reviewed and included by using bar graphs, tables and demonstrations. Authors [9] have surveyed how RPA with machine learning and other Intelligent processes is disrupting the different business processes. It looks at how business processes are improving with RPA, Machine Learning and AI. It also looks at the performance measures in RPA and ML in Business processes and how effectively they work. However, this study doesn't assess whether RPA in business process provide the highest value. It also looks at the cost and time to implement RPA in a business process by analysing the tasks that can be automated and can be optimized with ML and AI. However there still need much work to be done in the field of AI with Robotic Process Automation [10] have described that AI and RPA will be Industrial Revolution 4.0 and this paper looks at and reviews RPA and AI techniques that can improve RPA processes in different organizational processes such as extraction, recognition, optimization and classification. The study describes AI techniques that are contributing to Industry 4.0 which are the most modern industries involved in revolutionizing today's industrial processes and techniques. In conclusion this research investigates the different tools and provides an in-depth comparison and how it is contributing to Industry 4.0.

Authors [11] have reviewed different research that has been happening in RPA and AI over the years and how it is implemented in each industry. This looks at different sectors where different IA methods are used and what benefits it offers' gives a brief about where the research direction Is going and how it is affecting the industries. However, this doesn't take into account the costs for each process and application of this technology. In conclusion a historic and a holistic review of IA applications has been done on various aspects and look to support futures research for different applications

Authors [12] have investigated how RPA is being developed and with the onset of AI revolution RPA has become and ever important field for research and development. It looks at how demand for RPA tools have increased and vendors and organizations look for more and more RPA solutions. In conclusion it looks at how in future AI and RPA can be developed using Data driven approach toeplace repetitive tasks that can be done only by humans. Auth et al. (2019) have attempted a study about RPA and enterprise architectures affect each other and what needs to be done in future to improve this relation. It looks as how enterprises look at RPA with high expectations in regards to automation but with this enterprise's architectures need to be changed depending on complexity of the RPA tasks. Finally, the paper looks at how RPA influences and impacts EA design and implementation and where further research need to be done for its optimization and many questions that arise from this need to be answered.

Authors [13] have compared how RPA with AI applied to education in regards to create a new kind of robot teacher. The study demonstrates this by developing a robot teacher by using RPA, AI and computer vision. In this paper a prototype robotic teacher has been developed and demonstrated to support the student teach education process. Finally, the paper demonstrates this using a video demonstration and using images from students and their experience. However due to limited literature in RPA in this field the potential for further research is high and this paper provides a baseline and feasibility of such a technology in education.

Authors [15] has done a comparative study in the different RPA tool that are available in the market such as UiPath, Automation Anywhere and Blue Prism. It looks at how RPA is being develop and how it will lead to industry 4.0 revolution.In order to implement RPA various tools are being develop and this paper does a comparative study of the same. In conclusion this paper looks at where each tool can be helpful in the industry and how these tools are revolutionizing RPA and giving users the power to automate with very little coding. Authors [14] have studied a user preference/voting based smartphone recommendation system. They have used a dataset with user preference was used and model with user preferences and actual weights of different phones was trained to group the mobile phone users according to interests. The phones were clustered using type of phones based on weights obtained from attributes linked to the voting by users. They did a controlled experiment with 1000 users and gather the information online. In conclusion of the paper the proposed systems gave better recommendation to users that the equal weight-based benchmarks systems.

Authors [16] have determined that the noisy data in current mobile phone datasets due to different factors can give bad accuracy when trained due to over fitting problems. To address this issue, they have proposed a robust model for real life data of users in order to improve accuracy of the model. They have handled the data by removing noisy instances by using a dynamic noise threshold using na¨ıve bias classifier. They use decision tree machine learning algorithm to build the prediction model. Their experiment results are very accurate on real life mobile datasets of different user and have a good accuracy and precision. Many businesses offer a variety of items in a highly competitive high-end product market to vie for market share in various sectors. Consumers struggle to compare and select the best items because of the abundance of competitive product alternatives and the wealth of available information. Most of a product's features must satisfy customer needs for it to be successful. Consumer perception ratings of product features are difficult to quantify and forecast [17]

The recommender systems can be broadly classified as memory-based and model-based. The memory based accesses the database directly. On the other hand, the model-based approach leverages the transaction data to build a model that can produce recommendations. Several implications can be drawn from the memory-based approach and model-based collaborative filtering experiment: A study was carried out by [18] to compare memory-based versus model-based collaborative filtering on software libraries. The findings of the study demonstrate that the memory-based strategy excels in two areas: precision and recall.

The advent of social commerce has been rendered obsolete by the introduction and rapid development of digital computing and cell phones (S-commerce). The authors [19] revealed that using the built application, the social commerce construct, which includes ratings and reviews, forums and communities, as well as suggestions and referrals, had various effects on customer purchase intentions. Authors from [20] frequently consider the advantages and disadvantages of a choice while consulting close friends and family for guidance. Finding the proper product, however, has become a difficult task in today's fast-paced modern world. Various algorithms, and their hybrid combinations have been discussed by the authors to make the recommendation system strong. Like Feature selection, various clustering techniques, neural networks, principal component analysis etc.

Author [21] describe that in order to give users more options, different cellphones have varying specifications. Budget, brand, camera, storage, and many other factors are typically taken into account by smartphone consumers. But because all of these criteria must be taken into account, purchasers of smartphones could find it difficult to express their preferences precisely and to compare their preferences for different smartphone features. A progressive web application (PWA) was the project goal. The project's functionality testing demonstrated that the system successfully recommends three cell phones with an accuracy rate of 85% based on user preferences, achieving the project's goal. A subclass of machine learning systems called recommender systems uses complex information filtering techniques to speed up user searches and offer the most pertinent content [22].

Consumers typically look for products based on their categories and visit stores that sell such products, such as supermarkets for food or stationary stores for pencils. In the study, the authors [23] rated the Where2Buy shopbot app system for smartphones, which can find and filter local stores that provide the needed goods. Our system supports voice or text searches to streamline the input process, and fuzzy matching is supported to broaden the search parameters. Compared to the other Hong Kong shopbots, Where2Buy is preferred by more than half of the participants. [24] The goal of the context-aware idea is to close the communication gap between users and information systems so that the latter actively comprehend user context and requirements. In order to create context-aware recommendation systems, this study merges the notion of context-awareness with association algorithms (CARS). According to the user's desired product functional criteria, the study [25] uses the product characteristics and product reviews to recommend products. Ontologies map out the relationships among functional needs, product characteristics (specifications), and user perceptions of those features. Recommendations and justifications discovered by exploring ontology's semantic relations.

Authors [26] in their paper, the authors came to the conclusion that standard recommendations used in browsing activity records performed poorly. DPRMBDMU, a novel recommendation model that can analyse a mobile user's browsing data to determine their interests while also satisfying differential privacy, was proposed based on the conclusion and differential privacy. Lastly, the authors demonstrated that DPRMBDMU ensures -differential privacy by using real datasets to confirm the model's effectiveness. Based on the findings, we suggested the DPRMBDMU recommendation model, which effectively provides personalised recommendations based on mobile users' browsing habits while also ensuring the security of their data. The outcome of the trial confirms that DPRMBDMU is more effective than the conventional recommendation algorithm based on mobile users' browsing patterns. As per authors from [27] the smartphone business is expanding in prominence and is seen as a lucrative sector in Malaysia as a result of the country's growing smartphone usage trend. It is obvious that the current mobile phone models on the market today offer more than simply call-and-text capabilities; they also have a variety of practical features like photo-taking, voice-recording, file-organizing, and similar capabilities, making them popular among regular customers to own. Because of this, it is crucial for academics and marketers to comprehend the underlying causes and effects of this smartphone dependency issue in order to comprehend customer behavior [28].

In order to maintain smartphone sales and its market share, it is crucial for marketers to comprehend how dependent customers are on smartphone usage. This study was conducted to better understand the critical variables that can affect smartphone usage dependence and the potential consequences of dependence on purchase behaviour. This was accomplished by using the Theory of Uses and Gratification to justify and explain the causes that drive people's dependence on smartphones and the Theory of Media Dependence to clarify dependence and purchase patterns. The motivation for using smartphones was suggested [29] to cause dependency on them based on the pleasure theory. It was then suggested that entertainment, social needs, social influence, and convenience were the precursors to smartphone dependence. This is consistent with [30] assertion that the Malaysian smartphone market remained in its developing stage and that customers were still primarily using their cellphones for social interaction and search. Nonetheless, [31] have claimed that Malaysians still favour using a laptop or personal computer.

According to [32], dependence on smartphones would undoubtedly rise when the benefits of smartphones were highly perceived and valued. As the market develops, it is anticipated that consumers will become increasingly dependent on smartphone usage, which will affect their purchasing decisions in the future. This could aid local smartphone marketers in promoting, advertising, and selling their product in the region. Authors claim that [33][25] when a customer plans to purchase a high-tech item with complicated functions, such as a smartphone, notebook, camera, PC, car, server, etc., they frequently find it challenging to express what they need. This is due to the fact that the majority of people are unfamiliar with the technical aspects of these product categories [34]. Nonetheless, while collecting user requirements, the majority of recommender systems that have been built still make direct reference to product characteristics. In order to create more accurate recommendations, this study [35] suggests a novel method for product recommendations based on functional requirements of the product. The suggested strategy makes use of a mapping between functional requirements, component products, and component supporting features [36].

The authors have [37][38][39] gather details about the objects' descriptions and customer evaluations from open websites, then apply sentiment analysis techniques to model the similarities at the user- and item-levels, respectively. Authors [40] have conducted a study on One hundred university students. They participated in a poll, and the results showed that price, Random Access Memory (RAM), CPU, internal memory, and camera were the most important factors in choosing a smartphone. Authors have discussed that [41] information retrieval in mobile devices depends on a number of variables, including the device's location, screen size, and CPU speed. In addition to providing a more thorough explanation of the difficulties encountered, this paper [42] provides an overview of the technologies connected to mobile recommender systems.

Authors have illustrated [43][44][45] a genetic algorithm (GA)-based strategy; for improving the smartphone's recommendation system was put forth. To ensure that the consumer may utilise it across various platforms, it is implemented using a progressive web application (PWA) platform [46][47].

3. PROBLEM STATEMENT

In today's time phones have become an important part of everyone's daily life where everyone uses them for communication, entertainment, work and education. Smartphones is the literal definition of the world in pockets. Those who own a smartphone can no longer imaging living without one. India is also the second ranked country in terms of average time spent on mobile websites among users in Asia pacific. There are around 74.8 crore smartphone users in India in 2020 and will increase to 93.10 crore by 2022. Also, about 13.4 crore smartphones were sold across India in the year 2017 and is estimated to increase to about 44.2 crores in 2022 [48] as shown in Figure 2. Also, over time the combination of very high sales volumes and smartphone consumer behaviour in India has made it a very attractive market for foreign vendors. Also, worldwide smartphone users are increasing and is estimated to reach 7.516 billion by 2026 [49] as shown in Figure 1. As per Consumer behaviour, 97% of consumers turn to a search engine when they are buying a product vs. 15% who turn to social media. If a seller succeeds to publish smartphones based on user's behaviour/choice at the right place, there are 90% chances that user will enquire for the same.

Although mobile phones in today's time have many features in common in terms of hardware and software, manufacturers are still trying to bring uniqueness to their products by adding some more new features to the existing features. This has made the mobile development a challenge and manufacturers welcome the challenge with a great set of innovative designs. The growing number of brands and models created the fierce market competition. Therefore, it is inevitable to research and implement new innovations and design and at the same time it is also desirable to know the trending thoughts of potential customers. Therefore, the aim of this research paper is to compare different machine learning algorithms and build an accurate smartphone recommendation system based on individual consumer's behaviour or choice.

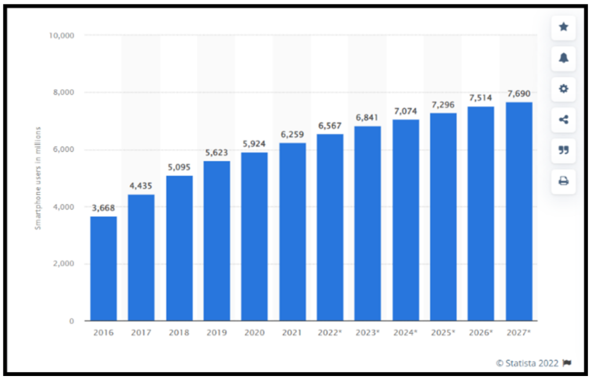

Figure 1. Number of smartphone users from 2016 to 2027 (in millions) (S. O'Dea, 2022)

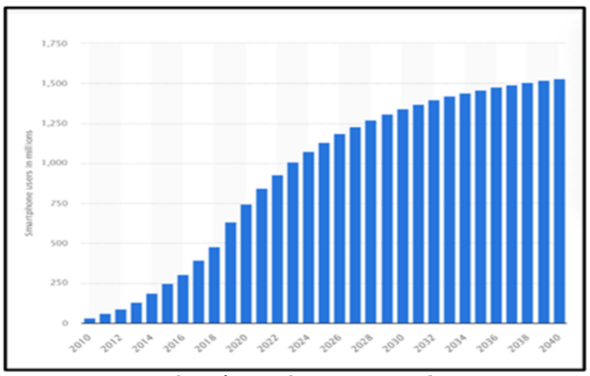

Figure 2. Number of smartphone users in India in 2010 to 2020, with estimates until 2040(in millions) (Sun, 2021)

4. Research Methodology

In this research various methods and techniques were used to compare different machine learning algorithms and the steps in involved in the research are shown in figure 3. These steps involve data collection, data pre-processing, model building and implementation using different methods and processes.

Various tools such as Jupiter notebook, python, UiPath Studio and Druid AI chat bot were used for these processes. The figure 3 shows the various steps involved in conduction the necessary steps to procure smartphone data and model building and training on the processed data. Then the steps for integrating the smartphone recommendation model in UiPath and Druid AI chatbot for demonstrating the results in a user-friendly format

5. Research Steps

A. Data Mining

In this step latest smartphone data was collected using python script from “91mobile.com” website using python web scraping methods such as beautiful soup and requests. The data about smartphones such as name, price, reviews, no of reviews and URL was collected and stored in a csv file.

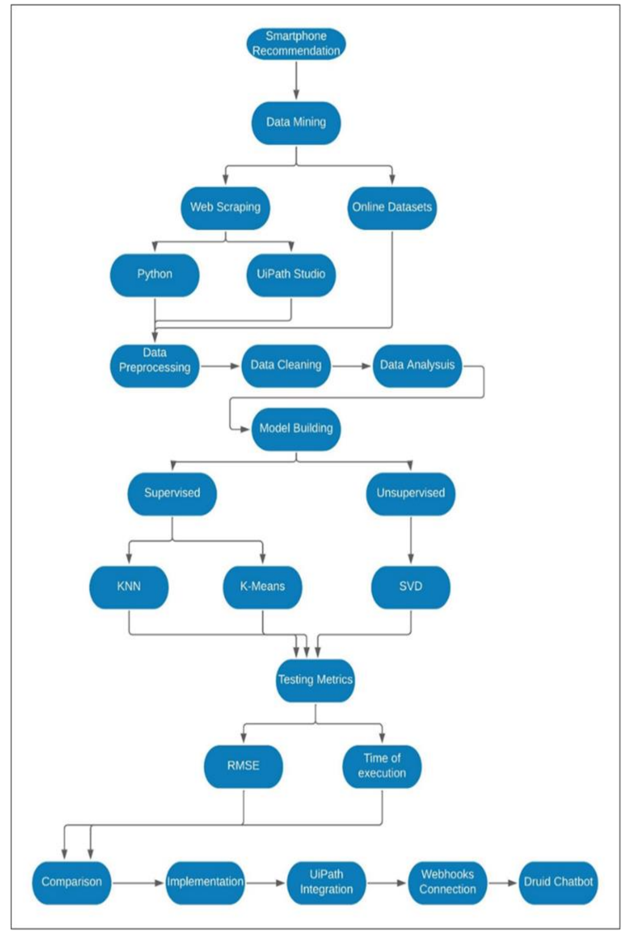

Figure 3. Proposed Smartphone Recommendation Model

B. Data Cleaning and Analysis

In this step the mined data was cleaned and cleaning was done on the data such as removing unknown and missing values. Then the smartphone data was analysed using pandas and NumPy. Based on the data the best attributes were selected to train the model. Attributes such as price, review count and ratings were used to train the model.

C. Model Building and Testing

In this step based on the data analysis attributes were selected for training and models were built using different algorithms such as KNN user based, KNN item based and SVD. And once the models were built, the models were compared on different parameters such as RMSE score and wait time for getting the smartphone recommendations for different models. And based on that a new hybrid model was implemented to get the smartphone recommendations and was demonstrated using a chatbot in next step.

D. UiPath and Druid Ai chatbot Integration

Once the model is trained and tested the model is saved in a “sav” file, which is a statistics data file format for saving the trained model. This is done using ‘Joblib’ which is a part of the SciPy ecosystem and provides utilities for pipelining Python jobs. It provides utilities for saving and loading Python objects that make use of NumPy data structures efficiently. This proves to be useful for some machine learning algorithms that require a lot of parameters or store the entire dataset like 'K-Nearest Neighbor' algorithms. The saved model is then loaded into a python file which is used in UiPath to which takes smartphone name as input and gives recommendations as output.

Then we integrate the UiPath process with druid chatbot so that when the user interacts with the chatbot the UiPath process gets executed in the backend and sends the results back to the Druid AI chatbot. The chatbot then shows the output to the user and the output can be characterized as smartphone recommendations or smartphone specification based on the user input. The flow of steps in Druid AI chatbot and UiPath Studio are:

- Druid Workflow Steps:-

I: When the user accesses the chatbot and enters something that the chatbot understands like "get phone data", "phone info" etc. The chatbot process is triggered and the chatbot asks the user to select one of the options from the two options i.e., "Phone Recommendations" and "Phone data". And once the user selects the desired option above the next step is triggered.

II: If the user selects phone recommendations or enters input such as give me "phone recommendations" this step is triggered and the chatbot asks the user for phone name and sends the phone name to the UiPath process using webhooks and the appropriate UiPath process gets executed and sends the phone recommendations back to the chatbot and it is displayed to the user.

III: And if the user selects phone data in 1st step or enters input such as give me "phone specifications" this step is triggered and the chatbot asks the user for phone name and sends the phone name to the UiPath process and then the escarping method gets executed and collects the phone specifications. Once the UiPath process ends it sends back the phone specifications to the chatbot and the phone specifications are displayed to the user.

- UiPath Workflow Steps:-

I: Firstly, if the user input from chatbot is to get Phone Recommendations then the saved smartphone recommendation machine learning model in the python script is executed in UiPath with phone name as input else if the user input is Phone specifications, then the web scraping feature of UiPath is used to collect smartphone specifications from "91mobiles.com". And phone specifications such as price, ram, processors, battery, camera and display are collected and saved if the phone exists.

II: Then the output of the python script for smartphone recommendations or the collected phone specifications form web scraping is sent back to the chatbot using webhooks to display the information back to the user in the Druid AI chatbot.

6. MATERIALS AND METHODS

A. KNN Algorithm

The KNN algorithm has been used and performs a similarity check of the various smartphones based on the provided dataset and is expected to give back the smartphone recommendations. For the recommendation within KNN, the collaborative filtering technique has been used in the research project. Using the available data points from the connected dataset, the algorithm is used to classify a new data point based on the similarity which is categorized in two techniques i.e., user-based recommendations and item-based recommendations.

B. Content-based filtering

These methods use a set of distinct qualities of an item to suggest other items with comparable features. Also, Recommender systems prove to be a good alternative to search algorithms since they assist people in find things that might not have discovered otherwise. It's worth noting that recommender systems are frequently deployed with search engines that index non-traditional data. There are several recommender systems that have been employed in the creation of machines, including content-based, collaborative, and hybrid techniques (combining the above two approaches).

C. Collaborative filtering

The collaborative filtering approach collects and analyses data based on user behaviours, preferences, or activities, and then predicts what users would like based on their commonalities with other users. It is based on the notion that people who have previously agreed will have similar tastes in the future. This technique is based on the user item interaction. So, in this it assumes the past product ratings and predicts the similar kind of products for the future. The system generates recommendations using only information about the ratings for different smartphones in this case. Collaborative filtering is one of the base methods in the field of information retrieval and filtering.

D. UIPath Studio

For the development of the RPA process workflow the tool which has been used is the UiPath Studio (2021) [50] here. The community version of the tool has been used for the process development. The user interface of UiPath is simple to understand as compared to its counterparts making complex processes easier to automate in less time. Apart from basic programming knowledge, no unique skill set is required to use UiPath studio. The tool allows creation of blank projects using the blank process function. Inside the blank process the user can create a sequence or flowchart of steps for creating an automation process. This tool along with python has been used for data mining using web scraping. The python script has been integrated within the UiPath process to get the additional functionalities of automation. This tool has also been used to integrate with Druid AI for backend process to create predictions based on input and sendthe output to the chatbot.

E. UiPath Cloud Orchestrator

It allows us to collect, manage, analyse, operate, scale and deploy various kinds of automation processes in just a matter of seconds. The automation cloud orchestrator proves to be an integral component while dealing with its integration with the druid environment. The UiPath python process has been uploaded on the automation cloud as a package. Further, the default tenant API Access credentials have been passed to the druid chatbot using the orchestrator tenant settings. The UiPath cloud orchestrator (2021) [51] is connected to Druid Ai chatbot [52] using webhooks for API calls. Webhooks also allow external systems to subscribe and listen to different types of Orchestrator events in UiPath. This is used to send and retrieve data and notifications to and from the cloud platform.

F. Druid Ai Oxygen platform

Druid AI (2021) is an NLP based chatbot platform that has been used to deploy chatbots. In this research project it is used to demonstrate the smartphone recommendation model. Various types of druid customizations have been performed in the research project like creating a new druid bot, importing various types of solutions required for the recommendation chatbot and training the bot using NLP. Conversation flows have also been created using drag and drop into a single diagram through the imported solutions. Finally, the Orchestrator API Access credentials have been passed in the druid chatbot settings which acts as the druid connector.

7. DATASET ANALYSIS

In this study, the user-reviews dataset that was collected using automated data collection using tools like python beau- tiful soup and UiPath web scraping. It contains 14,15,133 rows which contain ratings data and 11 columns. In this data the ratings are on a scale from 1 to 10. In this the dataset there are six files which contain part of the 14,15,133 rows in each file. In the dataset there is a list of ratings that users have given to different smartphones on different websites. In this dataset there are a range of examples where users have given ratings to different smartphones and multiple users have given multiple reviews on various websites.

The vast majority of the ratings are 10 and it has an un- even distribution. The most ratings are from users that are anonymous and. This dataset has 11 columns, Data Type and Value range/format as shown in Table 1. Firstly, since there’s a lot of unwanted data, data cleaning and analysis is done and then the task is to visualize and combine the dataset before training based on KNN algorithm and testing. Further different attributes will be applied on KNN and also applied on collaborative filtering using SVD and compare the outcomes of each mode. Firstly, the values and data types for each attribute has been checked in the dataset. And then checked for Missing and Null Values and Check for Unique Values.

Looking at the Data in Figure 4 and 5 following steps are applied for data cleaning:

1. Remove irrelevant features
2. Round-off score feature to nearest integer
3. Impute missing values in score feature with median
4. Remove samples with missing values in 'Product' and 'author' feature and also 'Anonymous'
5. Remove duplicates, if any

Table 1 Attribute Description Table

Sr.no	Attribute	Description	Data Type	Value-Range/ Format
1.	Phone_url	Link to website	String	-
2.	Date	Date of review	Date	-
3.	Language	Language of review	String	In ISO 639
4.	Country	Country of review	String	In ISO 639
5.	Source	Website source name	String	-
6.	Domain	Main website link	String	-
7.	Score	Rating by user	Float64	0.2-10
8.	Score_max	Max rating	Float64	0-10
9.	Extract	Review description	String	-
10.	Author	Name of person who gave the review	String	-
11.	Product	Name of the product	String	-

```
Missing count and percentages for each column are:
 phone_url          0 (0.0%)
date                0 (0.0%)
lang                0 (0.0%)
country             0 (0.0%)
source              0 (0.0%)
domain              0 (0.0%)
score           63489 (4.49%)
score_max       63489 (4.49%)
extract         19361 (1.37%)
author          63202 (4.47%)
product             1 (0.0%)
dtype: object
```

Figure 4. Count of Missing Values

```
print('Number of unique values in each feature
Number of unique values in each feature:
 phone_url        5556
date             7728
lang               22
country            42
source            331
domain            384
score              86
score_max           1
extract       1321353
author         801103
product         61313
dtype: int64
```

Figure 5. Unique Value Count

In this study the distribution of data in terms of number of reviews per phone and number of reviews per customer distribution is shown in Figure 6,7 and 8. Figure 8 and 9 shows the score distribution given by users for different smartphones before and after cleaning. Names like 'Anonymous', 'unknown', 'Amazon Customer', 'Cliente Amazon', 'e-bit', 'Client d'Amazon', 'Amazon Kunde', 'Anonymous ', 'einer Kundin', 'einem Kunden' can be interpreted in the same way i.e., an 'unknown customer' as shown in Figure 8. After replacing Anonymous data, the distribution is shown in Figure 8

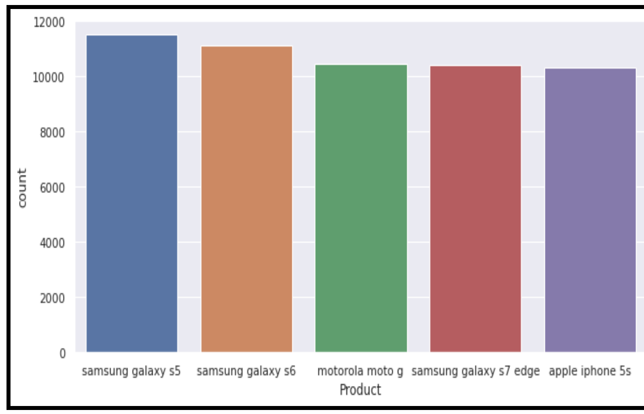

Figure 6. Most Rated Phones

The Figures 9 and 10 describe the number of reviews per score mark. There is a very uneven distribution in each table and in this study the data is handled by find outliers and cleaning the data before training and testing.

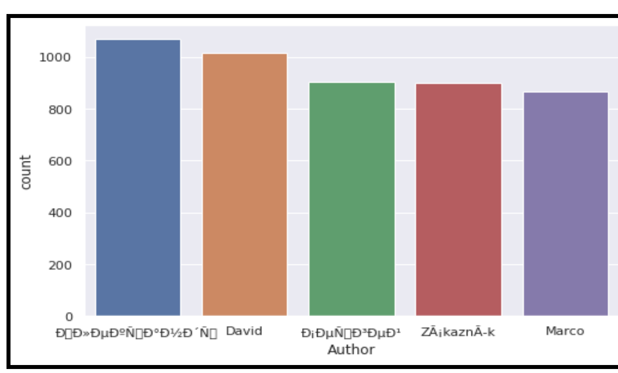

Figure 7. Distribution of number of ratings/users before Cleaning

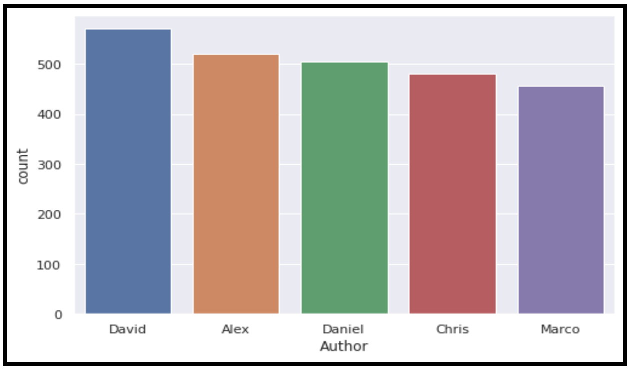

Figure 9. Distribution of number of ratings/users After Cleaning

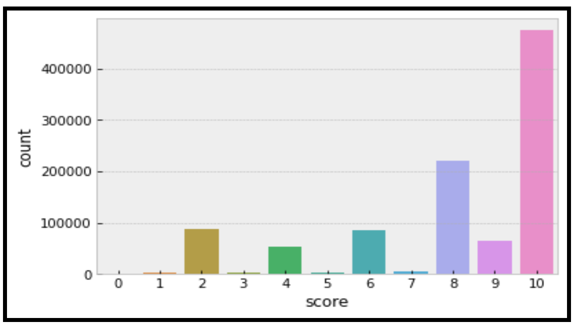

Figure 9. Score Distribution

```
revs1['score'].value_counts()

10     542758
8      253001
2       99720
6       98263
9       73406
4       60066
7        7297
5        4140
1        2691
3        2211
0          42
Name: score, dtype: Int64
```

Figure 10. Score Distribution

8. Dataset Dscription

After data analysis and cleaning, data splitting and training is done on the data using the different algorithms. The data was split into 25% Test set and 75% Training set. The data was trained on three different algorithms that is KNN with means for item-based reviews, KNN with means for user-based reviews and collaborative filtering using SVD. For KNN with means user-based k value of 50 was used and sim options were "pearson-baseline" and "user-based" for training. And for KNN with means item-based k value of 50 and sim options were "pearson-baseline" and "item-based" for training. And default values were used for SVD using Collaborative filtering. After training, the model was saved using the 'Joblib' library in Python for later usage in UiPath and Druid chatbot for running the code and displaying the output in a user-friendly format.

9. Model Testing

In this study we are testing the RMSE and wait time for the smartphone recommendation models based on different algorithms. The Figures 11,12 and 13 show the true ratings vs predicted ratings for the different models for the test users. Table 2 shows the predicted ratings by different algorithms and Actual ratings given by the test users to the different phones and shows the prediction error for the different algorithms.

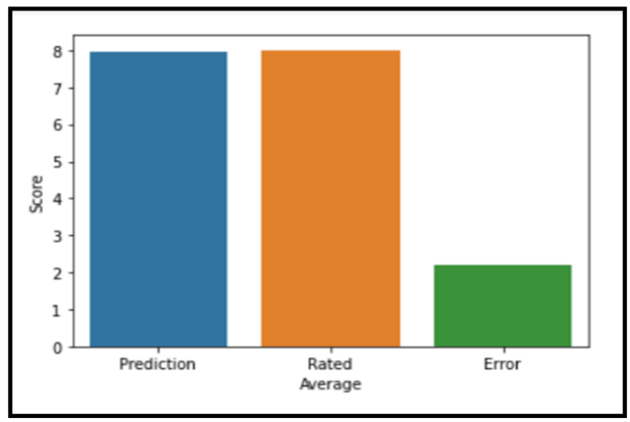

Figure 11. Average Rating for SVD Based Model

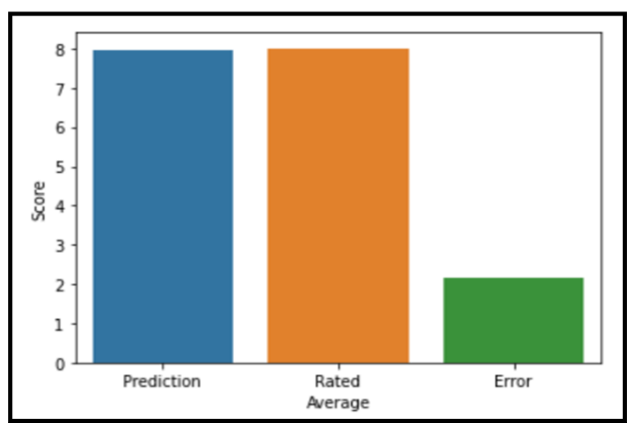

Figure 12 Average Rating for KNN (Item Based) Model

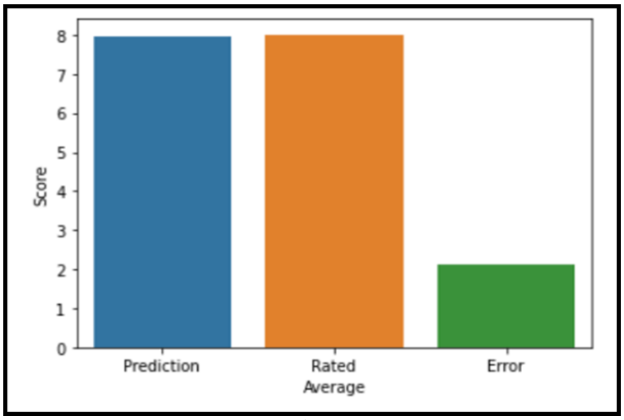

Figure 13. Average Rating for KNN (User Based) Model

Table 2 shows the predicted ratings by different algorithms and Actual ratings given by the test users to the different phones and shows the prediction error for the different algorithms.

Table 2 Model testing Comparison

Algorithm/ Average Rating for Test Users	**SVD**	**KNN (item Based)**	**KNN (User Based)**
Predicted	7.9726237 75829122	7.9506137 669598775	7.9441834 91508518
Actual	8.0088602 8725984	8.0088602 8725984	8.0088602 8725984
Prediction Error	2.1812320 082481746	2.1396308 23272711	2.1288921 12093161

10. Result And Discussion

The experimental and result analysis was done using python with the python scikit-surprise library to implement the proposed algorithm. For this, phone review records from people with real names was only used and anonymous data was removed. Also, products having ¿50 ratings and users who have given ¿ 50 ratings been only considered. The total number of rows and columns in the final dataset were 85773 x 3. Input data was first transformed before being used as an input to the training of the different machine learning models. Figure 16 shows the comparison between mean ratings vs rating count for highest rated 50 phones in the dataset.

Firstly, CF SVD based model was tested. The accuracy measure of the model was tested using Root Mean Square Error (RMSE). The input data was then trained on KNN algorithm using item-based recommendation model and then tested using RMSE. Lastly, the input data was trained on KNN algorithm using user-based recommendation model and then tested using RMSE.

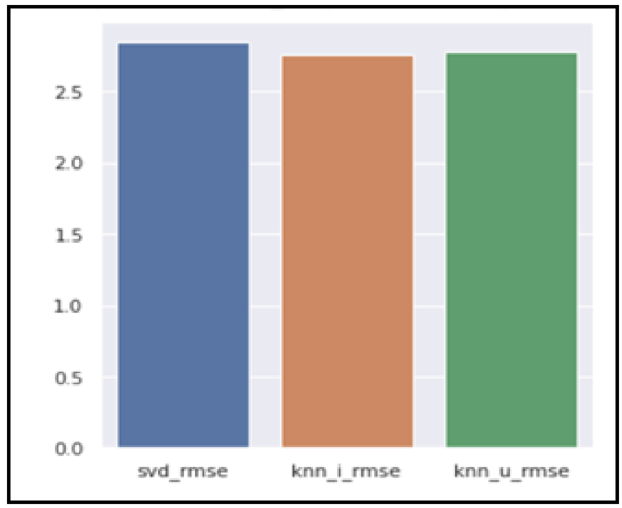

Figure 14. RMSE score comparison between the models

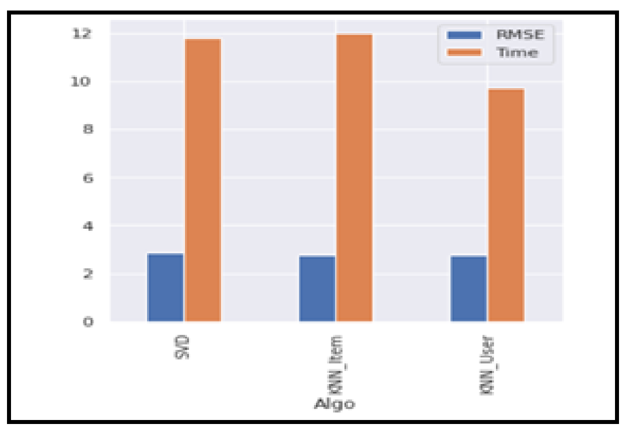

Figure 15. Wait Time and RMSE Comparison

The wait time and RMSE score graphs for different models is shown in Figure 14 and 15 and values in Table 3. Measures in Table 3 shows that the best algorithm with the least waiting time and good RMSE is KNN (User based model)

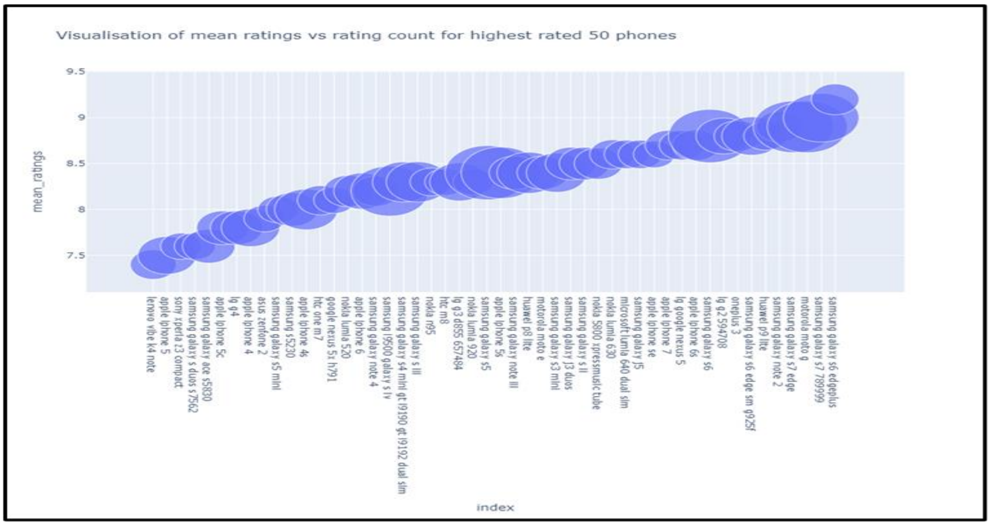

Figure 16. Comparison between mean ratings vs rating count for highest 50 phone

Table 3 Model Comparison

Score/Algorithm	SVD	KNN (item based)	KNN (user based)
RMSE	2.87	2.75	2.77
Wait Time (in seconds)	11.8	12	9.73

11. CHATBOAT OUTPUT

The Figures 17 and 18 illustrates the output of the chatbot when the user provides an input and enters the smartphone name and based on the smartphone input recommendations and related phones details are shown back to the user. In the chatbot a hybrid model was implemented using KNN user and item-based models.

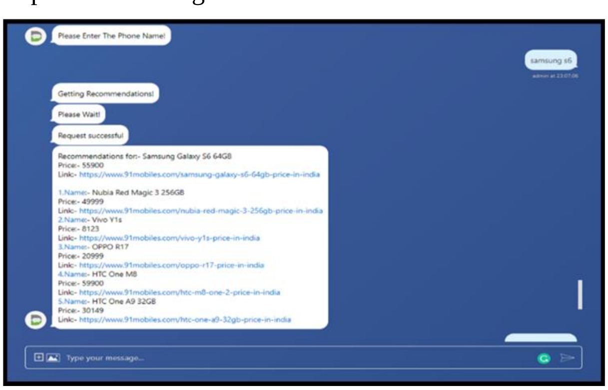

Figure 17. Final output recommendations

Figure 18. Final output Phone Specs

12. CONCLUSION

To summarize the most popular phone (rated 10 by highest number of people) was Overall Verykool t742 and amongst top users it was Doogee y6. Overall data is highly skewed towards 'Amazon customers' from different countries. This may also be because 'Amazon' is the biggest trader for phones in the world. Although correct 'user' names from 'Amazon' should have used. Most of the authors have given the rating of '10' or '8'. Both KNN (item-based) and KNN (user-based) and SVD have roughly similar RMSE. However, KNN using user-based ratings has less wait time and overall better performance and accuracy compared to all other dataset.

The proposed recommendation system acts as a perfect decision-making system since current research uses a robust smartphone dataset and performed various analysis and data pre-processing on it. Then it recommends a list of smartphones for each particular user that have given reviews to different phones. The comparison between algorithms was done and base on priority these algorithms can be deployed on factors such as wait time and accuracy. Different machine learning concepts like RMSE are used to test the data and provides an accuracy test. Hence, KNN proves to be an efficient Collaborative filtering algorithm in recommending Smartphones. Finally, this model can be further improved by using the sentiment analysis and Natural language processing by analysing the description of the reviews written by users and not just user rating. Further tuning can be done based on different requirement and scenarios and the smartphone recommendation tool can be game changer in the future of smartphone marketing.

Here the Druid chatbot is giving recommendations and data regarding smartphones that is entered by the user. Once the user enters the smartphone name, the UiPath process is getting triggered and the output is sent back to the chatbot for the user to see. Finally, the UiPath process based chatbot has been used to get not only the smartphone specifications but also recommendations based on the smartphone model typed by the user in the chatbot. In future more integrations can be added to give recommendations about a variety of different products. And with more NLP training, the chatbot will be able to understand the user better regarding his/her requirements.

REFERENCES

[1] Vinay Kumar, A., Neeraj, M., Akash Reddy, P. and Chavan, A., 2022. Machine Learning-Based Human Activity Recognition Using Smartphones. In Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021 (pp. 573-582). Singapore: Springer Nature Singapore.

[2] Kamali, P., Sudha, P. and Akshaya, S.O., (2018) 'Outfit Recommender System using KNN Algorithm', International Journal of Engineering Research Technology, Vol. 6, No. 7.

[3] Sachin, J., Geethatharani, P., Surya, M.K. and Kavin, K.V., (2019) 'Location Based Agricultural Product Recommendation System Using Novel KNN Algorithm', International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 5, No.2, pp. 945-950.

[4] Bhagirathi, Y.N. and Kiran, P., (2019) 'Book Recommendation System using KNN Algorithm', International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 2, No.6, pp. 336-339.

[5] Singh, R.H., Maurya, S., Tripathi, T., Narula, T. and Srivastav, G., (2020) 'Movie recommendation system using cosine similarity and KNN', International Journal of Engineering and Advanced Technology, Vol.9, No. 5, pp. 556-559.

[6] KesavaDasu, K., Kamal, K.R. and Pranith, G., (2021) 'A Cuisine Based Recommender System Using k-NN and Mapreduce Approach', International Journal of Recent Trends in Multidisciplinary Research, Vol. 1, No. 1, pp. 09-17.

[7] Portugal, I., Alencar, P. and Cowan, D., (2018) 'The use of machine learning algorithms in recommender systems: A systematic review', Expert Systems with Applications, Vol. 97, pp. 205-227.

[8] Chate, P.J., (2019) 'THE USE OF MACHINE LEARNING ALGO- RITHMS IN RECOMMENDER SYSTEMS: A SYSTEMATIC RE- VIEW', International Journal of Research and Analytical Reviews, Vol. 6, No. 2, pp. 671-681.

[9] Chakraborti, T., Isahagian, V., Khalaf, R., Khazaeni, Y., Muthusamy, V., Rizk, Y. and Unuvar, M., (2020) 'From robotic process automation to intelligent process automation', International Conference on Business Process Management, Vol. 393, pp. 215-228.

[10] Ribeiro, J., Lima, R., Eckhardt, T. and Paiva, S., (2021) 'Robotic process automation and artificial intelligence in industry 4.0–a literature review', Procedia Computer Science, Vol. 181, No. 2, pp. 51-58.

[11] Ng, K.K., Chen, C.H., Lee, C.K., Jiao, J.R. and Yang, Z.X., (2021) 'A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives', Advanced Engineering Informatics, Vol. 47, p.101246.

[12] Van der Aalst, W.M., Bichler, M. and Heinzl, A., (2018) 'Robotic process automation', Business Information Systems Engineering, Vol. 60, No. 4, pp. 269-272.

[13] Lasso-Rodr´ıguez, G. and Gil-Herrera, R. (2019), 'Robotic Process Automation Applied to Education: A New Kind of Robot Teacher?' In ICERI2019: 12th annual International Conference of Education, Research and Innovation, Seville, Spain, pp. 2531-2540.

[14] Khan, S., (2020) 'Comparative analysis of rpa tools-uipath, automation anywhere and blueprism', International Journal of Computer Science and Mobile Applications, Vol. 8, No. 11, pp. 1-6.

[15] Reddy, Y.S. and Govindarajulu, P., (2018) 'A mobile phone recommen- dation system with user centric voting approach', International Journal of Computer Science and Network Security, Vol. 18, No. 2, pp. 91-102.

[16] Sarker, I.H., (2019) 'A machine learning based robust prediction model for real-life mobile phone data', Internet of Things, Vol. 5, pp. 180-193.

[17] Yuen, K. K. F. (2017). The fuzzy cognitive pairwise compar- isons for ranking and grade clustering to build a recommender system: An application of smartphone recommendation. Engineer- ing Applications of Artificial Intelligence, 61(April 2016), 136–151. https://doi.org/10.1016/j.engappai.2017.02.001

[18] F. McCarey, M. Cinneide, N. Kushmerick, A Recommender Agent for Software Libraries: An Evaluation of Memory-Based and Model- Based Collaborative Filtering Proceedings of the IEEE/WIC/ACM international conference on Intelligent Agent Technology, 2006, pp. 154-162.

[19] Ashoer, M., Murdifin, I., Basalamah, J., As'ad, A., Pramukti, A. (2021). Integrating Social Commerce Constructs into Mobile Application Ser- vice; A Structural Equation Model. Journal of Physics: Conference Series, 1783(1). https://doi.org/10.1088/1742-6596/1783/1/012043

[20] Balasubramanian, J., Koppaka, S., Rane, C., Raul, N. (2020). Parameter Based Survey of Recommendation Systems. SSRN Electronic Journal, 1–7. https://doi.org/10.2139/ssrn.3569579

[21] Binti, N., Azam, A. (2020). Universiti Teknologi MARA SRcS: Smartphone Recommendation System Using Genetic Algorithm. July.

[22] Biswas, P. K., Liu, S. (2022). A hybrid recommender system for recommending smartphones to prospective customers. Expert Systems with Applications, 208. https://doi.org/10.1016/j.eswa.2022.118058

[23] Chan, K. C., Cheung, T. L., Lai, S. H., Kwan, K. C., Yue, H., Pang, W. M. (2017). Where2Buy: A Location-Based Shopping App with Products-wise Searching. Proceedings - 2017 IEEE Interna- tional Symposium on Multimedia, ISM 2017, 2017-Janua, 438–443. https://doi.org/10.1109/ISM.2017.87

[24] Chen, M.Y., Wu, M.N., Chen, C.C., Chen, Y.L. and Lin, H.E., 2014. Recommendation-aware smartphone sensing system. Journal of applied research and technology, 12(6), pp.1040-1050.

[25] Ken Arnett, D. S., Baizal, Z. K. A., Adiwijaya. (2015). Rec- ommender system based on user functional requirements using Euclidean fuzzy. 2015 3rd International Conference on

Infor- mation and Communication Technology, ICoICT 2015, 455–460. https://doi.org/10.1109/ICoICT.2015.7231467

[26] Ding, Z., Qin, Z., Qin, Z. G. (2017). Differential privacy for recom- mendation model based on web browsing behaviors of mobile users. 2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2017, 195–199. https://doi.org/10.1109/ICCWAMTIP.2016.8079836

[27] Harun, A., Soon, L. T., Mohd Kassim, A. W., Sulong, R. S. (2015). Smartphone dependency and its impact on purchase behavior. Asian So- cial Science, 11(26), 196–211. https://doi.org/10.5539/ass.v11n26p196

[28] A.R. Pratama, "Exploring personal computing devices ownership among university students in indonesia",International Federation for Informa- tion Processing, pp. 835–841, 2017

[29] S. Belbag, A. Gungordu, T. Yumusak, and K. G. Yilmaz, "The Eval- uation of Smartphone Brand Choice: an Application with The Fuzzy Electre I Method," Int. J. Bus. Manag. Invent., vol. 5, no. 3, pp. 2319–8028, 2016

[30] Karimuddin,A.(2012).Nielson's Founds on Asia's Smartphones Utilization. Retreived July 17,2012,from http://dailysocial.net/en/2012/06/25/nielsen%E2%80%99s-found-on-smartphone-utilization-in-asia/

[31] Loh, J. (2011). Tablets, smartphones ownership on the rise. RetrievedJuly 16, 2023

[32] Alcaniz, B. E., Mafe, C. R., Manzano, J. A., Blas, S. S. (2008). Influence of online shopping information dependency and innovativeness on internet shopping adoption. Online Information Review, 32(5), 648- 667.

http://dx.doi.org/10.1108/14684520810914025

[33] D.H. Widyantoro, Z.K.A. Baizal, "A framework of conversational rec- ommender system based on user functional requirements," In: 2nd IEEE International Conference on Information and Communication Technology (ICoICT), IEEE, 2014, pp. 160 – 165. 978-1-4799-7752- 9/15/$31.00

[34] Kuhlmeier, D., Knight, G. (2003). Antecedents to internet-based pur- chasing: a multinational study. International Marketing Review,22(4),460-473. http://dx.doi.org/10.1108/02651330510608460

[35] P. Kazienko, P. Kołodziejski, "Personalized integration of recommenda- tion methods for E-commerce, " Int. J. Comput. Sci. Appl., vol. 3, no.3, , Aug. 2006, pp. 12–26

[36] A. Felfernig, R. Burke, "Constraint-based recommender systems:technologies and research issues," Proceedings of the 10th international conference on Electronic commerce, ACM, August 2008

[37] Li, G., Chen, Q., Zheng, B., Hung, N. Q. V., Zhou, P., Liu, G. (2020). Time-aspect-sentiment recommendation models based on novel similarity measure methods. ACM Transactions on the Web (TWEB), 14(2), 1-26

[38] A. Oulasvirta, T. Ratternbury and E. Raita, "Habits make smartphone use more pervasive". Personal and Ubiquitous Computing vol. 16, issue 1, pp. 105-114, January 2012

[39] C. Ling, W. Hwang and G. Salvendy,"Diversified users, satisfaction with advanced mobile phone features", Universal Access in the Information Society, vol. 5, no. 2, pp. 239- 249, 2006. DOI: 10.1007/s10209-006- 0028-x

[40] Okfalisa, O., Rusnedy, H., Iswavigra, D. U., Pranggono, B., Haerani, E., Saktioto, T. (2020). Decision Support System For Smartphone Recommendation: The Comparison Of Fuzzy AHP and Fuzzy ANP In Multi-Attribute Decision Making. Sinergi, 25(1), 101-110

[41] Polatidis, N., Georgiadis, C. K. (2013, September). Mobile recom- mender systems: An overview of technologies and challenges. In 2013 Second International Conference on Informatics Applications (ICIA) (pp. 282-287). IEEE

[42] A. B. Mohammed, —Selling smartphones to generation Z: Under- standing factors influencing the purchasing intention of smartphone,‖ International Journal of Applied Engineering Research, vol. 13, no. 6, pp. 3220–3227, 2018

[43] F. Ricci, L. Rokach, B. Shapira and P. Kantor, "Recommender systems handbook", Springer, 2011

[44] F. Ricci, "Mobile Recommender Systems". International Journal of Information Technology and Tourism vol. 12, issue 3, pp. 205-231, 2011

[45] Samah, K. A. F. A., Azam, N. A., Hamzah, R., Chew, C. S., Riza, L.S. (2021). Optimizing Smartphone Recommendation System through Adaptation of Genetic Algorithm and Progressive Web Application. International Journal of Advanced Computer Science and Applications,12

[46] D. Kamalapurkar, N. Bagwe, R. Harikrishnan, S. Shahane, and G. Manisha, —Phone recommender: sentiment analysis of phone reviews,‖ International Journal of Engineering Sciences Research Technology, vol. 6, no. 5, pp. 212–217, 2017

[47] S. Jain and B. Singh, —Consumer behavior toward mobile phone handsets,‖ in International Conference on Innovative Computing and Communications, 2019, pp. 61–69

[48] Sun, Shangliao. (2021) Smartphone users in India 2010-2040. [online] Technology Telecommunications, Telecommunications, Statista.

https://www.statista.com/statistics/467163/forecast-of-smartphone-users-in-india/ (Accessed 15 January 2022)

[49] O'Dea, S. (2022) Smartphone subscriptions worldwide 2016-2027.[online]Technology Telecommunications,

Telecommunications, Statista.

https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ (Accessed 03 March 2022)

[50] UiPath studio. [online] https://www.uipath.com/ (Accessed 15 October 2021).

[51] UiPath cloud orchestrator. [online] https://cloud.uipath.com/ (Accessed 15 October 2021).

[52] Druid AI. [online] https://www.druidai.com/ (Accessed 15 October 2021).

Kalash Chandak: Student of MSC(CA) from SICSR, Pune. A Tech Enthusiast Studying Masters in Computer Science at USC. Experienced in Web Development, Operating Systems Development and Data Analysis.

Dr. Anuja Bokhare is working as Assistant Professor in School of Computer Science and Engineering, Dr.Vishwanath Karad MIT World Peace University,Pune. She worked at SICSR,SIU,Pune for 12years. She has achieved M.Phil(C.S) from YCMOU, Nasik and PhD (CS) from Symbiosis International (deemed University), Pune India. She has 22 years of experience in the field of academics. Her research interests includes Artificial Intelligence, Machine Learning, Deep Learning, Graph Theory, Software engineering and fuzzy logic.. She had published 39 research papers in international journal and conferences indexed by Scopus (Q2,Q3) along

with one book in her account. She served as reviewer for international conferences and journals from IEEE and Elsevier. She is a senior member of IEEE, Life menber of ISTE, member of IAENG and Soft Computing Research Society.

Samaya Pillai Iyengar is a full-time faculty in Symbiosis Institute of Digital and Telecom Management, International University, Pune (India). She holds MCA, MCM and MPhil degree under the Faculty of Management. She is Perusing her PhD with (Savitribai Phule Pune University) SPPU, Pune University under Computer-Management. She has 18 years of teaching experience with specialisation in databases. Her research interests focus around computer sciences and big data domains. She has taught subjects in programming -Java, Python, C. Databases –Oracle, MySQL, MongoDB, Operating Systems, MIS, Basic Statistics, and Networking etc. She has more than 30 publications in Scopus. And more than 15 papers in Conference proceedings.

Dr Vanishree Pabalkar
Associate Professor

Dr Vanishree Pabalkar is associated with Symbiosis Institute of Management Studies. She has corporate experience of 4 years in Service sectors with organizations like ITC, and VSNL. After this, she moved to Academics. She is a Gold medallist in Post-Graduation from SPPU. She has done her B Sc, MBA, MBS in the field of Marketing. Her PhD was in the area of Telecom Sector. Her areas of interest include Marketing Management, Sales and Distribution, Business to Business Marketing, Services Marketing and Cases in Marketing. She has published several research papers in National and International Conferences. She has also published research papers in repute journals. She is the author of a book on Marketing Research. She has 7 Patents to her credit.

She has also been part of Research Projects. She has completed a Research Project on Prisoners Rehabilitation in Pune and Nasik. And the next project that she worked was based on Women Sanitation. Currently, she is working on the next Project on the lines of Civic engagement of Youth in rural areas in Maharashtra. She has also been part of Management Development Programs and Trainings in the field of Marketing.