E COMMERCE SITES RECOMMENDATION SYSTEM USING MACHINE LEARNING

A major Project Report submitted in the partial fulfillment of

therequirements for the award of the degree

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

D.Trinadh (19471A05L2)
K.Krishna Chaitanya Reddy (19471A05M2)
T.Manikanta (20475A0519)

Under the esteemed guidance of

Dr.S.Siva Nageswara Rao **M.Tech.,Ph.D**

Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

NARASARAOPETA ENGINEERING COLLEGE

(AUTONOMOUS)

Accredited by NAAC with A+ Grade and NBA under Tire-1
NIRF rank in the band of 251-320 and an ISO 9001:2015 Certified
Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada
KOTAPPAKONDA ROAD, YALAMANDA VILLAGE, NARASARAOPET- 522601

2022-2023

NARASARAOPETA ENGINEERING COLLEGE: NARASARAOPETA
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the main project entitled "E COMMERCE SITES RECOMMENDATION SYSTEM USING MACHINE LEARNING"

is a bonafide work done by "D.Trinadh (19471A05L2) , K.Krishna Chaitanya Reddy (19471A05M2) T.Manikanta (20475A0519) " in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in the department of **COMPUTER SCIENCE AND ENGINEERING during 2022-2023.**

PROJECT GUIDE

Dr.S.Siva Nageswara Rao **M.Tech.,Ph.D**

PROJECT CO-ORDINATOR

Dr.M.Sireesha ,M.Tech.,Ph.D

HEAD OF THE DEPARTMENT

Dr.S.N.TirumalaRao,M.Tech.,Ph.D

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We wish to express our thanks to carious personalities who are responsible for the completion of the project. We are extremely thankful toour beloved chairman sir **M.V. Koteswara Rao**, **B.sc** who took keen interestin us in every effort throughout this course. We owe out gratitude to our principal **Dr.M.Sreenivasa Kumar, M.Tech,Ph.D(UK),MISTE,FIE(1)** for his kindattention and valuable guidance throughout the course.

We express our deep felt gratitude to **Dr.S.N.Tirumala Rao,M.Tech, Ph.D** H.O.D. CSE department and our guide **Dr.S.Siva Nageswara Rao M.Tech.,Ph.D** of CSE department whose valuable guidance and unstinting encouragement enable usto accomplish our project successfully in time.

We extend our sincere thanks to **Dr.M.Sireesha,M.Tech,Ph.D** Associate Professor coordinator of the project for extending her encouragement. Their profound knowledge and willingness have been a constant source of inspiration for us throughout this project work.

We extend our sincere thanks to all other teaching and non-teaching staff to department for their cooperation and encouragement during our B.Tech degree. We have no words to acknowledge the warm affection, constant inspiration and encouragement that we receive from our parents. We affectionately acknowledge the encouragement received from our friends thosewho involved in giving valuable suggestions had clarifying out all doubts which had really helped us in successfully completing our project.

By

D.Trinadh (19471A05L2)
K.Krishna Chaitanya Reddy (19471A05M2)
T.Manikanta (20475A0519)

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

INSTITUTE VISION AND MISSION

INSTITUTION VISION

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community,

INSTITUTION MISSION

M1: Provide the best class infra-structure to explore the field of engineering and research

M2: Build a passionate and a determined team of faculty with student centric teaching,imbibing experiential, innovative skills

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION OF THE DEPARTMENT

To become a centre of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

MISSION OF THE DEPARTMENT

The department of Computer Science and Engineering is committed to

M1: Mould the students to become Software Professionals, Researchers and Entrepreneurs by providing advanced laboratories.

M2: Impart high quality professional training to get expertize in modern software tools and technologies to cater to the real time requirements of the Industry.

M3: Inculcate team work and lifelong learning among students with a sense of societal and ethical responsibilities.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Specific Outcomes (PSO's)

PSO1: Apply mathematical and scientific skills in numerous areas of Computer Science and Engineering to design and develop software-based systems.

PSO2: Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering

PSO3: Promote novel applications that meet the needs of entrepreneur, environmental and social issues.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Educational Objectives (PEO's)

The graduates of the programme are able to:

PEO1: Apply the knowledge of Mathematics, Science and Engineering fundamentals to identify and solve Computer Science and Engineering problems.

PEO2: Use various software tools and technologies to solve problems related to academia, industry and society.

PEO3: Work with ethical and moral values in the multi-disciplinary teams and can communicate effectively among team members with continuous learning.

PEO4: Pursue higher studies and develop their career in software industry.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Outcomes

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, research literature, and analyzecomplex engineering problems reaching substantiated conclusions using firstprinciples of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complexengineering problems and design system components or processes that meetthe specified needs with appropriate consideration for the public health andsafety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-basedknowledge and research methods including design of experiments, analysis andinterpretation of data, and synthesis of the information to provide validconclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques,resources, and modern engineering and IT tools including prediction andmodeling to complex engineering activities with an understanding of thelimitations.

6. The engineer and society: Apply reasoning informed by the contextualknowledge to assess societal, health, safety, legal and cultural issues and theconsequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of theprofessional engineering solutions in societal and environmental contexts,and demonstrate the knowledge of, and need

for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Project Course Outcomes (CO'S):

CO425.1: Analyse the System of Examinations and identify the problem.

CO425.2: Identify and classify the requirements.

CO425.3: Review the Related Literature

CO425.4: Design and Modularize the project

CO425.5: Construct, Integrate, Test and Implement the Project.

CO425.6: Prepare the project Documentation and present the Report using appropriate method.

Course Outcomes – Program Outcomes mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1		✓											✓		
C425.2	✓		✓		✓								✓		
C425.3				✓		✓	✓	✓					✓		
C425.4			✓			✓	✓	✓					✓	✓	
C425.5					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
C425.6									✓	✓	✓		✓	✓	

Course Outcomes – Program Outcome correlation

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1	2	3											2		
C425.2			2		3								2		
C425.3				2		2	3	3					2		
C425.4			2			1	1	2					3	2	
C425.5					3	3	3	2	3	2	2	1	3	2	1
C425.6									3	2	1		2	3	

Note: The values in the above table represent the level of correlation between CO's and PO's:

1. **Low level**
2. **Medium level**
3. **High level**

Project mapping with various courses of Curriculum with Attained PO's:

Name of the course from which principles are applied in this project	**Description of the device**	**Attained PO**
C3.2.4, C3.2.5	Gathering the requirements and defining the problem, plan to develop a **smart bottle for health care using sensors.**	PO1, PO3
CC4.2.5	Each and every requirement is critically analyzed, the process model is identified and divided into **five modules**	PO2, PO3
CC4.2.5	Logical design is done by using the unified modelling language which involves individual team work	PO3, PO5, PO9
CC4.2.5	Each and every module is tested, integrated, and evaluated in our project	PO1, PO5
CC4.2.5	Documentation is done by all our four members in the form of a group	PO10
CC4.2.5	Each and every phase of the work in group is presented periodically	PO10, PO11
CC4.2.5	Implementation is done and the project will be handled by the **hospital management and in future updates in our project can be done based on air bubbles occurring in liquid in saline.**	PO4, PO7
CC4.2.8 CC4.2.	**The physical design includes hardware components like sensors, gsm module, software and Arduino.**	PO5, PO6

ABSTRACT

E-commerce sites have become an integral part of our daily lives, with millions of products being sold and purchased online. With such a vast amount of options available, it can often be overwhelming for consumers to choose which product to buy. This is where recommendation systems come into play. In this project, we propose to develop a recommendation system for e-commerce sites using machine learning techniques. Our system will analyze user behavior, such as search queries and purchases, to generate personalized product recommendations for each individual user.We plan to use a combination of collaborative filtering and content-based filtering techniques to create our recommendation engine. Collaborative filtering will be used to find similarities between users based on their purchase history and provide recommendations based on what similar users have purchased. Content-based filtering will be used to recommend products based on user preferences, such as category, brand, and price range.To evaluate the performance of our recommendation system, we will conduct experiments on a real-world e-commerce dataset. We will compare the accuracy and efficiency of our system against existing recommendation systems. We will also conduct a user study to evaluate the effectiveness of our recommendations and user satisfaction.

1. INTRODUCTION

1.1 Introduction

E-commerce websites have become an essential part of modern-day shopping, offering customers an easy and convenient way to purchase products online. However, with the vast number of products available on these websites, it can be overwhelming for customers to navigate and find the products they want.To address this challenge, e-commerce websites have implemented recommendation systems that use machine learning algorithms to suggest products to users.

These recommendation systems analyze user behavior and historical data to identify patterns and preferences, enabling them to recommend products that are more relevant to each individual customer.The purpose of this project is to propose a recommendation system for e-commerce websites that uses machine learning algorithms to improve product recommendations.

Specifically, we will be using a collaborative filtering approach that takes into account the preferences of other users with similar tastes to recommend products to users.In this project, we will be using a combination of user-based collaborative filtering techniques to improve the accuracy of our recommendations. We will also evaluate the performance of our system on a real-world e-commerce dataset and compare it with other state-of-the-art recommendation algorithms.

The rest of the project is organized as follows. , we will provide a brief overview of collaborative filtering and how it can be used in recommendation systems.

After that, we will describe the dataset used in our experiments, followed by a description of our proposed recommendation system. Finally, we will present our experimental results and conclude with a discussion of our findings and future work.and random forests to train the model .Once we have developed the model, we will evaluate its performance by testing it on a separate set of data.

1.2 Existing System

There are several e-commerce websites that use recommendation systems based on machine learning algorithms to suggest products to their customers. Some examples of these websites are Amazon, Netflix, and Spotify. Overall, these e-commerce websites have sophisticated recommendation systems that use machine learning algorithms to suggest products to their users. These systems have been shown to improve customer satisfaction and increase sales by providing personalized recommendations to each user. While these recommendation systems have been successful in improving the customer experience and increasing sales, there is still room for improvement. In this project, we propose a recommendation system for e-commerce websites that uses a combination of user-based and item-based collaborative filtering techniques to provide more accurate recommendations.

Disadvantages

1. Doesn't generate accurate and efficient results.
2. Computation time is very high.
3. Lacking of accuracy may result in lack of efficient further treatment.

1.3 Proposed System

Our proposed recommendations system for e-commerce websites uses a combination of products user-based and item-based collaborative filtering techniques to improve the accuracy of recommendations. The system takes into account the preferences of other users with similar tastes to recommend products to each individual user . e-commerce website uses a combinations of the both user,item-based collaborative filtering techniques to provide the personalized on recommendations to users.

Advantages

- The system is more effective since it measures the products with the best price.
- The system is very fast in query retrieval due to SVM Algorithm.

1.4. System Requirements

1.4.1 Hardware Requirements

- System type : intel®core™i7-7500UCPU@2.70gh
- Cache memory : 4 MB
- RAM : 12 GB
- Hard Disc : 8 GB

1.4.2 Software Requirements

- Operating system : windows 10, 64 bit OS
- Coding language : Python
- Python distribution : Anaconda, Spyder, Flask

2. LITERATURE SURVEY

2.1 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Recommendation systems have been widely used in e-commerce websites to improve the customer experience and increase sales. There are several types of recommendation algorithms, including contents, collaborative based filtering, and hybrid approaches. Collaborative filtering is one of the most popular recommendation algorithms in use today. It is based on the assumption that people who have similar preferences in past will have more similar preference in the future. Collaborative filtering can be divided into two types: user-based and item-based.User-based collaborative filtering recommends products to users based on the preferences of other users with similar tastes. The system identifies users who have similar preferences and recommends products that they have liked. Item-based collaborative filtering, on the other hand, recommends products based on the similarity between items. The system identifies items that are similar to the ones the user has liked in the past and recommends them.

2.2 Some machine learning methods

Machine learning algorithms are often categorized as supervised and unsupervised.

- **Supervised machine learning algorithms** can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any newinput after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

- **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

- **Reinforcement machine learning algorithms** is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behaviour within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best. This is known as the reinforcement sign.

2.3 Applications of machine learning

- Virtual Personal Assistants
- Predictions while Commuting
- Videos Surveillance
- Social Media Services
- Email Spam and Malware Filtering
- Online Customer Support
- Search Engine Result Refining
- Product Recommendations
- Online Fraud Detection

2.4Applications of machine learning

- Solve Complex problems like Audio processing in Amazon echo, Image recognition, etc, reduce the need for feature extraction, automated tasks wherein predictions can be done in less time usingKeras and Tensorflow.
- Parallel computing can be done thus reducing overheads.
- Models can be trained on a huge amount of data and the model gets better with more data.
- High-Quality Predictions when compared with humans by training tirelessly.
- Works well-unstructured data like video clips, documents, sensor data, webcam data, etc.

3. SYSTEM ANALYSIS

3.1 ARCHITECTURE DIAGRAM

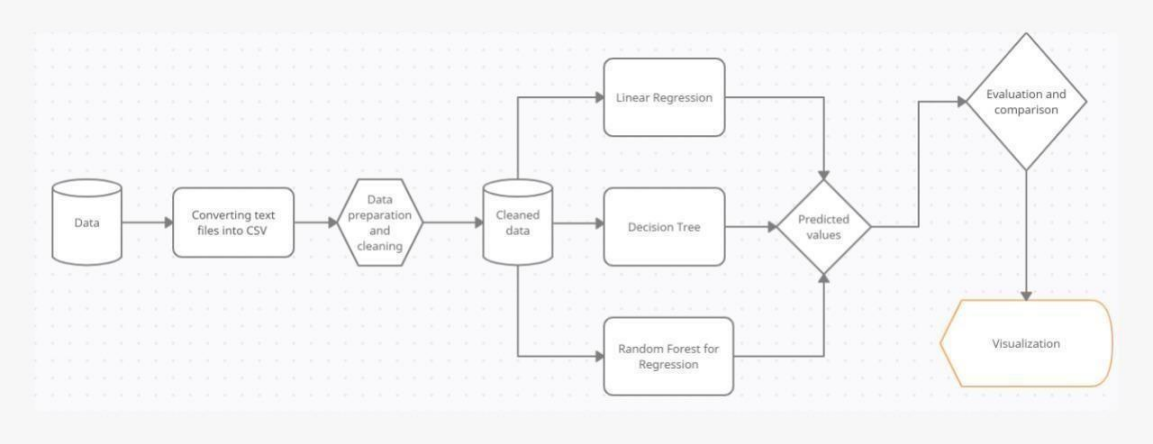


Fig:3.1 .1 Architecture diagram of E Commerce sites recommendation system

3.2. Prevalence of E commerce sites

E-commerce has become an integral part of our daily lives, with millions of products being sold and purchased online. The convenience of online shopping has attracted many consumers, as it allows them to shop from the comfort of their own homes and at any time of the day or night.

E-commerce sites are prevalent across many industries, including retail, fashion, electronics, food, and more. Major players in the e-commerce industry include Amazon, Alibaba, Walmart, and eBay, among others.

With the ongoing COVID-19 pandemic, e-commerce has become even more prevalent, as many consumers have shifted to online shopping due to lockdowns and social distancing measures. The trend towards e-commerce is expected to continue, as consumers have become accustomed to the convenience and benefits of online shopping.

3.3. Importance of machine learning in E commerce sites

Machine learning plays a crucial role in e-commerce sites by providing personalized recommendations, improving search functionality, optimizing pricing, and enhancing customer service.

One of the key benefits of machine learning in e-commerce is its ability to generate

personalized product recommendations for each individual user. Machine learning algorithms analyze user behavior, such as search queries and purchases, to provide recommendations that are tailored to each user's preferences and interests. This not only improves the user experience but also increases sales and customer loyalty.

Machine learning also helps to improve search functionality on e-commerce sites. Search algorithms can be trained to understand natural language queries and provide more relevant search results. This leads to a better user experience and increased sales.

Pricing optimization is another area where machine learning can be beneficial in e-commerce. Machine learning algorithms can analyze pricing data, competitor pricing, and customer behavior to optimize pricing strategies and increase revenue.

Machine learning can also improve customer service in e-commerce. Chatbots and virtual assistants can be trained using machine learning algorithms to provide personalized and efficient customer service, improving customer satisfaction and loyalty.

Overall, the importance of machine learning in e-commerce sites cannot be overstated. It helps to provide a better user experience, increase sales and revenue, and improve customer service.

3.4. Implementation of machine learning using Python

Python is a popular programming language. It was created in 1991 by Guido van Rossum.It is used for:

1. web development (server-side),
2. software development,
3. mathematics,
4. system scriptin

The most recent major version of Python is Python 3. However, Python 2, although not being updated with anything other than security updates, is still quite popular.

It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse, Anaconda which are particularly useful when managing larger collections of Python files.

Python was designed for its readability. Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.

Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

In the older days, people used to perform Machine Learning tasks manually by coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries.

Python libraries that used in Machine Learning are:

1.Numpy

2.Scipy

3.Scikit-

learn

4.Pandas

5.Matplotlib

1. **NumPy** is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

2. **SciPy** is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.

3. **Skikit-learn** is one of the most popular Machine Learning libraries for classical Machine Learning algorithms. It is built on top of two basic Python libraries, NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with Machine Learning.

4. **Pandas** is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

5. **Matpoltlib** is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, histogram, error charts, bar chats, etc

3.5. Scope of the project

E Commerce sites recommendation can help to determine the price of a product and can help to arrange the right time to deliver the product at the right time. There are various factors that influence the recommendation of a product to recommend products based on user preferences, such as category, brand, and price range.etc . algorithm and methodology.

3.6. Analysis

The dataset contains 5 attributes which are used to predict the product recommendation system such as:

Product_id: Product id

Product_name: Name of the product

Product_type: Type of the product

Product_brand: Brand of the product

Product_image: Image of the product

DataSet

product_id	product_name	product_type	product_brand	product_image	ecs_1_name	ecs_1_price
110001	REFLEX 3.0 DUAL TONED SMART BAND IN MIDNIGHT B...	Smart Watch	Fastrack	https://staticimg.titan.co.in/Fastrack/Catalog...	Fastrack	1795
110002	Amazfit bip u	Smart Watch	Amazfit	https://encrypted-tbn1.gstatic.com/shopping?q=...	Myntra	2499
110003	Noise ColorFit Pro 3	Smart Watch	Noise	https://m.media-amazon.com/images/I/61OpDFvFkE...	Amazon	5999
110004	Samsung Galaxy Watch 4	Smart Watch	Samsung	https://images.samsung.com/is/image/samsung/p6...	Amazon	18999
110005	Apple watch series 7	Smart Watch	Apple	https://d2xamzlzrdbdbn.cloudfront.net/products...	Vijay sales	38900

Fig 3.6.1 is the data set of E Commerce sites recommendation system which contains attributesProduct_id,Product_name,Product_type,Product_brand,Image.

3.7 Data Pre-processing

Before feeding data to an algorithm we have to apply transformations to our data which is referred as pre-processing. By performing pre-processing the raw data which is not feasible for analysis is converted into clean data. In-order to achieve better results using a model in Machine Learning, data format has to be in a proper manner. The data should be in a particular format for different algorithms. For example, if we consider Random Forest algorithm it does not support null values. So that those null values have to be managed using raw data.

Data Pre-processing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

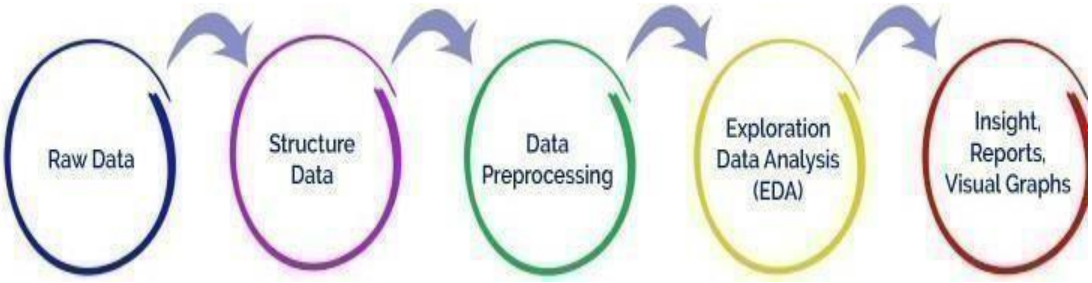


Fig:3.7.1. Data Preprocessing

Need of Data Preprocessing: For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format. For example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set. Another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

3.7.1 Missing values

Filling missing values is one of the pre-processing techniques. The missing values in the dataset is represented as '?' but it a non-standard missing value and it has to be converted into a standard missing value NaN. So that pandas can detect the missing values. The Fig:
3.5.1.1 below is a heat map representing the missing values. We have filled that missing values with 0 . Fig: 3.8.1.2 below is the heat map representing after filling missing values.

3.7.2 Correlation coefficient method

We can find dependency between two attributes p and q using Correlation coefficient method using the formula.

$r_{p,q} = \sum(p_i - p)(q_i - q)/n\sigma_p\sigma_q$

$= \sum(p_i q_i) - np\, q / n\sigma_p\sigma_q$

n is the total number of patterns, pi and qi are respective values of p and q attributes in patterns i, p and q are respective mean values of p and q attributes, σ_p , σ_q are respective standard deviations values of p and q attributes. Generally, $-1 \leq r_{p,q} \leq +1$. If $r_{p,q} < 0$, then p and q are negatively correlated. If $r_{p,q} = 0$, then p and q are independent attributes and there is no correlation between them. If $r_{p,q} > 0$, then p and q are positively correlated. We can drop the attributes that are having correlation coefficient value as 0 as it indicates that the variables are independent with respect to the prediction attribute. Fig:3.8.2 is the correlation heat map. After applying correlation the attributes are PR interval , QRS duration , QT interval , QTc interval, P wave , T wave , QRS wave and problem . The attribute Vent_rate got dropped.

3.7.3 Cross Validation:

Cross-validation is a technique in which we train our model using the subset of the data- set and then evaluate using the complementary subset of the data-set. The three steps involved in cross-validation are as follows :

– Reserve some portion of sample data-set.

– Using the rest data-set train the model.

– Test the model using the reserve portion of the data-set.0

3.7.4 Information Gain:

Information gain is a preprocessing technique, which is used to calculate the reduction in entropy. It is commonly used in the construction of decision trees from a training dataset, by evaluating the information gain for each variable, and selecting the variable that maximizes the information gain, which in turn minimizes the entropy and best splits the dataset into groups for effective classification.

3.8 Data visualization

Data visualization is the representation of data through use of common graphics, such ascharts, plots, infographics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.

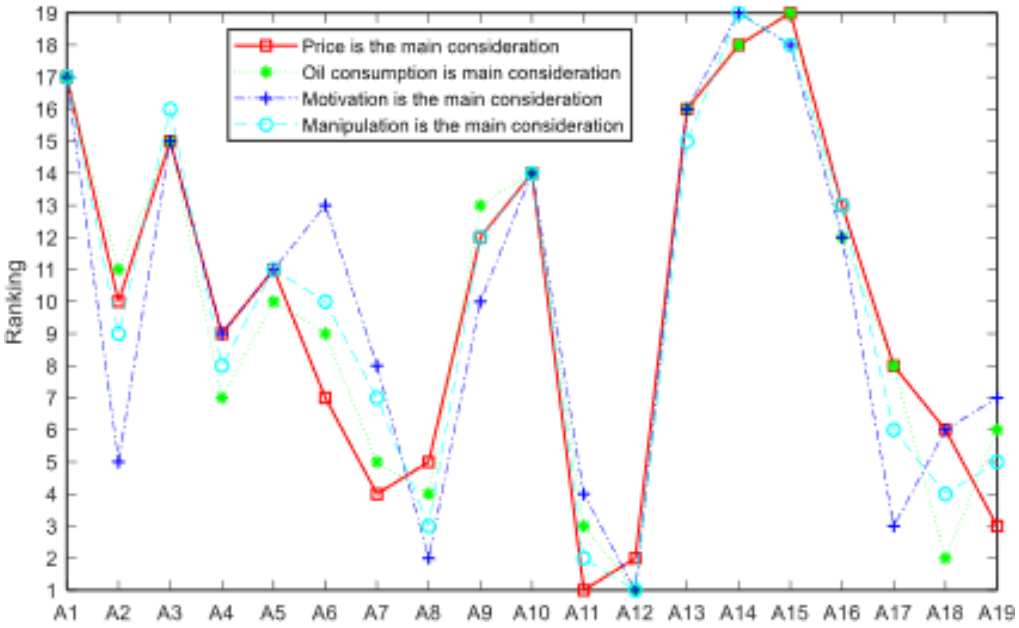


Fig:3.8.1. Comparison of ranking of items based on different preferences

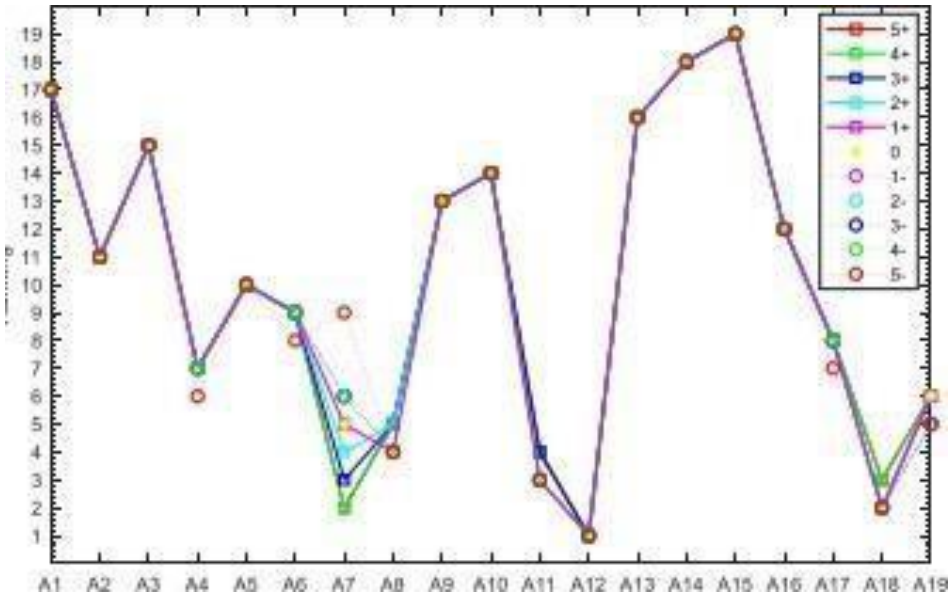


Fig:3.8.2. Comparison of ranking of items based on prediction

3.9 Classification

It is a process of categorising data into given classes. Its primary goal is to identify the class of our new data.

3.9.1 Machine learning algorithms for classification

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Random forest, Decision tree, Gaussian Naïve Bayes, Support vector machine etc.

1. Decision Tree

Decision Tree Analysis is a general, predictive modelling tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model.

2. Gaussian Naive Bayes

It is a simple technique for constructing classifiers. It is a probabilistic classifier based on Bayes' theorem. All Naive Bayes classifiers assume that the value of any particular feature is independent of the value of any other feature, given the class variable. Bayes theorem is given as follows: $P(C|X) = P(X|C) * P(C)/P(X)$, where X is the data tuple and C is the class such that P(X) is constant for all classes. Though it assumes an unrealistic condition that attribute values are conditionally independent, it performs

surprisingly well on large datasets where this condition is assumed and holds.

3. Random Forest

A random forest model is a type of ensemble learning algorithm used in machine learning for classification and regression tasks. It consists of a collection of decision tree models that are trained on different subsets of the training data, and then combined to make predictions.The random forest algorithm works by creating a set of decision trees, where each tree is trained on a randomly selected subset of the training data and a randomly selected subset of the features. This helps to reduce overfitting and increase the generalization of the model.

4. Logistic Regression

Logistic regression is a statistical method used to analyze and model the relationship between a binary (yes/no) dependent variable and one or more independent variables. It is a type of generalized linear model and is widely used for classification tasks in machine learning.

The logistic regression model uses a logistic function, also known as the sigmoid function, to model the relationship between the independent variables and the dependentvariable. The output of the logistic function is a probability score between 0 and 1, which can be interpreted as the likelihood that the dependent variable is equal to 1. The logistic function transforms the linear combination of the independent variables into a probability score, which is then compared to a threshold value to make a binary classification decision.

3.9.2 Confusion matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. A **true positive** (tp) is a result where the model predicts the positive class correctly. Similarly, a true negative (tn) is an outcome where the model correctly predicts the negative class. A **false positive** (fp) is an outcome where the model incorrectly predicts the positive class. And a false negative (fn) is an outcome where the model incorrectly predicts the negative class.

Sensitivity or Recall or hit rate or true positive rate (TPR)

It is the proportion of individuals who actually have the disease were identified as having the disease.

TPR = tp / (tp + fn)

Specificity, selectivity or true negative rate (TNR)

It is the proportion of individuals who actually do not have the disease were identified as not having the disease.

TNR = tn / (tn + fp) =1-FPR

Precision or positive predictive value (PPV)

If the test result is positive what is the probability that the patient actually has the disease.

PPV = tp / (tp + fp)

Negative predictive value (NPV)

If the test result is negative what is the probability that the patient does not have disease.

NPV = tn / (tn + fn)

Miss rate or false negative rate (FNR)

It is the proportion of the individuals with a known positive condition for which the test result is negative.

FNR = fn / (fp + tn)

Fall-out or false positive rate (FPR)

It is the proportion of all the people who do not have the disease who will be identified as having the disease.

FPR = fp/ (fp + tn)

False discovery rate (FDR)

It is the proportion of all the people identified as having the disease who do not have the disease.

FDR = fp / fp + tp

False omission rate (FOR)

It is the proportion of the individuals with a negative test result for which the truecondition is positive.

FOR = fn / (fn + tn)

Accuracy

The accuracy reflects the total proportion of individuals that are correctly classified.

ACC = (tp + tn) / (tp + tn + fp + fn)

F1 score

It is the harmonic mean of precision and sensitivity

F1 = 2tp / (2tp+ fp + fn)

4. Implementation code

4.1 Implementation code

Application.py

```
import pandas as pd
from textblob import TextBlob
import nltk
from nltk.stem import WordNetLemmatizer
from sklearn import neighbors
from scipy import optimize
from wordcloud import WordCloud, STOPWORDS
def find(a):
    df = pd.read_csv("flask demo/final_review (1).csv")
    def g2(text):
        return TextBlob(text).sentiment.polarity
    def Sort(sub_li):
        sub_li.sort(key = lambda x: (x[1],-x[3],-x[4]))
        return sub_li
    def Sort1(sub_li):
    # reverse = None (Sorts in Ascending order)
    # key is set to sort using second element of
    # sublist lambda has been used
        sub_li.sort(key = lambda x: x[5])
        return sub_li
    df.set_index("product_name",inplace = True)
    res=df.loc[a]
```

```
li=[] i=4 lis=[]
                while(i<len(res)):
                    if(res[i]!="nan"):
                        lis.append(res[i])
                        lis.append(res[i+1])
                        lis.append(res[i+2])
                        lis.append(res[i+3])
                        lis.append(res[i+4])
                        i=i+5
                        li.append(lis)
                        lis=[]
                finli=[]
                for i in li:
                    if  str(i[1]) != "nan":
                        mil=g2(i[4])
                        i[4]=mil
                        finli.append(i)
                finli1=Sort(finli)
                finli2=[]
                l=len(finli1)
                for i in finli1:
                    i[1]=l
                    l=l-1
```

```
        finli2.append(i)
    finli3=[]
    for i in finli2:
        suma=0.3*i[1]+0.4*i[3]+0.3*i[4]
        i.append(suma)
        finli3.append(i)
    finli4=Sort1(finli3)
    finli4.reverse()
    finli5=[]
    for i in finli4:
        cv=[]
        cv.append(i[0])
        cv.append(i[2])
        cv.append(i[3])
        finli5.append(cv)
    return finli5
print(find('Apple watch series 7'))
```

App.Py

```
from flask import Flask , render_template , request
import recom as r
app=Flask(_name_)
@app.route('/',methods=["GET","POST"])
def index():
```

```
    fin1=""
    if request.method=="POST":
        lk=request.form["us"]
        fin1=r.find(lk)
    mk=fin1
    return render_template('index.html',L=mk)
'''
@app.route('/sub',methods=['POST'])
def submit():
    if request.method=="POST":
        name=request.form["username"]
    return render_template("sub.html",n=name)
'''
if _name_ == "_main_":
    app.run(debug=True) from flask import Flask ,
render_template , request
import recom as r
app=Flask(_name_)
@app.route('/',methods=["GET","POST"])
def index():
    fin1=""
    if request.method=="POST":
        lk=request.form["us"]
        fin1=r.find(lk)
    mk=fin1
    return render_template('index.html',L=mk)
```

```
'''
@app.route('/sub',methods=['POST'])
def submit():
    if request.method=="POST":
        name=request.form["username"]
    return render_template("sub.html",n=name)
'''
if _name_ == "_main_":
    app.run(debug=True)
```

INDEX.html

```
<html>
  <style>
    a:link, a:visited {
  background-color: white;
  color: black;
  border: 2px solid green;
  padding: 10px 20px;
  text-align: center;
  text-decoration: none;
  display: inline-block;
}

a:hover, a:active {
  background-color: green;
  color: white;
}
    #customers {
  font-family: Arial, Helvetica, sans-serif;
  border-collapse: collapse;
  width: 100%;
}

#customers td, #customers th {
  border: 1px solid #ddd;
  padding: 8px;
}
```

```
#customers tr:nth-child(even){background-color: #f2f2f2;}

#customers tr:hover {background-color: #ddd;}

#customers th {
  padding-top: 12px;
  padding-bottom: 12px;
  text-align: left;
  background-color: #04AA6D;
  color: white;
}
       input[type=text], select {
  width: 100%;
  padding: 12px 20px;
  margin: 8px 0;
  display: inline-block;
  border: 1px solid #ccc;
  border-radius: 4px;
  box-sizing: border-box;
}

input[type=submit] {
  width: 100%;
  background-color: #4CAF50;
  color: white;
  padding: 14px 20px;
  margin: 8px 0;
  border: none;
  border-radius: 4px;
  cursor: pointer;
}
```

```
input[type=submit]:hover {
  background-color: #45a049;
}

div {
  border-radius: 5px;
  background-color: #f2f2f2;
  padding: 20px;
}
        body {
          font-size: 28px;
        }

        ul {
          list-style-type: none;
          margin: 0;
          padding: 0;
          overflow: hidden;
          background-color: #333;
          position: -webkit-sticky; /* Safari */
          position: sticky;
          top: 0;
        }

        li {
          float: left;
        }

        li a {
          display: block;
```

```
    color: white;
    text-align: center;
    padding: 14px 16px;
    text-decoration: none;
  }

  li a:hover {
    background-color: #111;
  }

  .active {
    background-color: #4CAF50;
  }
  </style>
<body>

    <ul>
      <li><a class="active" href="#home">Home</a></li>
      <li><a href="templates/about.html">about</a></li>
      <li><a href="contact.html">Contact</a></li>
    </ul>
<form action="/" method="POST">
<input type="text" id="hi" name="us" placeholder="enter your search">

  <input type="submit" value="submit">
</form>
<table id="customers">

      {% for a,b,c in  L %}
      <tr>
        <td>{{ a }}</td>
        <td><a href="{{ b }}">Link</a></td>
```

```
            <td>{{ c }}</td>
          </tr>
          {% endfor %}

    </table>
    <center><table style="background-color:#a6e21f;" id = "rcorners1" id = "line" class="table-layout">
      <tr><th>  </th>
            <th><a href="https://www.amazon.in/"><img src="https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcScIsjMJlI5Kam1Nh3lG1af2V0uO4uipAV-rw&usqp=CAU"
class="mobilelogo"
width="90",height="90"></a>      </th>
            <th><a href="https://www.flipkart.com/"><img src="https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcTwZpOyGJMh83665KJTbPhVC-
KECVyQSXsy1n7i8koMPWpXq_G215sqa99xq5KVz1w7ek&usqp=CAU" class="mobilelogo"
width="90",height="90"></a>      </th>
            <th><a href="https://www.myntra.com/"><img
src="https://gumlet.assettype.com/afaqs%2F2021-01%2F15f5f827-8e29-4520-af8d-
a0f353b8da17%2Fimages.png?auto=format&q=35&w=1200" class="mobilelogo"
width="90",height="90"></a>      </th>
            <th><a href="https://www.nykaa.com/"><img src="https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcRU9sEOP6MpCj1UxvinqKA_qJfo9aUFp2MuIAD7s2SFEVBwJ
GnUEovYG7QjbTGXvmXGi-U&usqp=CAU" class="mobilelogo" width="90",height="90"></a></th>
            </tr>
            <tr style="background-color:#a6e21f;">
              <td>  </td>
              <td style="font-family:bold;font-size: 0.6cm;">Amazon</td>
              <td style="font-family:bold;font-size: 0.6cm;">Flipkart</td>
              <td style="font-family:bold;font-size: 0.6cm;">Myntra</td>
              <td style="font-family:bold;font-size: 0.6cm;">Nykaa</td>
              <th>   </th>
            </tr>
```

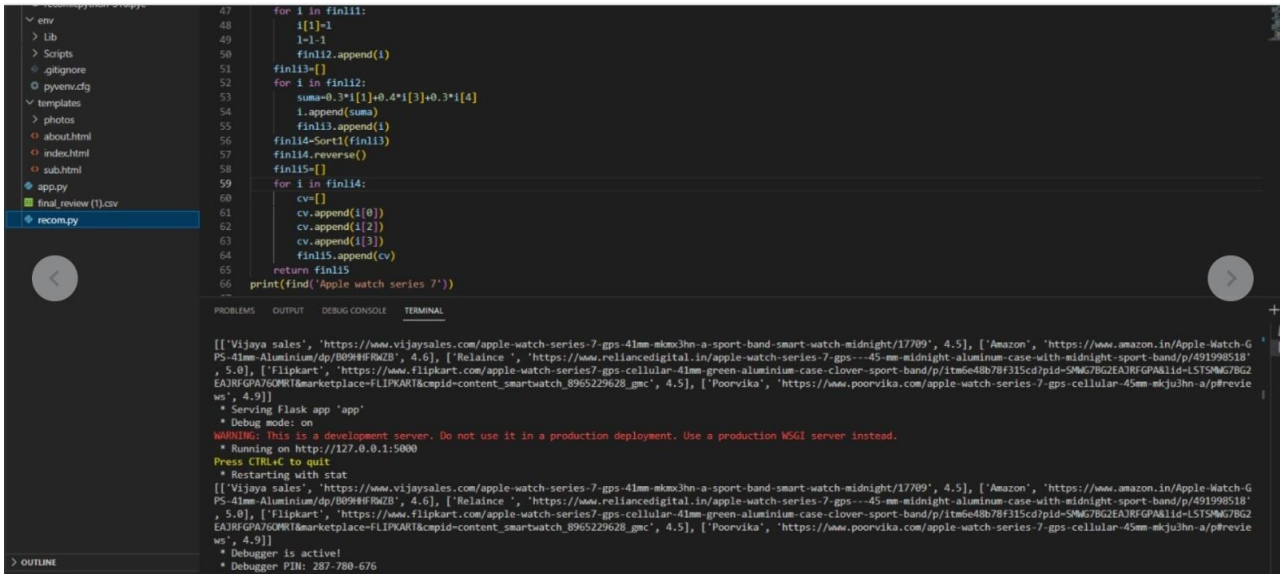
```
  </table></center><br>
  <center><table style="background-color:#a6e21f;" id = "rcorners1" class="table-layout">
     <tr><th>  </th>
              <th><a href="https://www.Bigbasket.com/"><img
src="https://images.yourstory.com/cs/images/companies/logosC021575978360413png"
class="mobilelogo"
width="100",height="100"></a>      </th>
              <th><a href="https://www.1mg.com/"><img src="https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcTzMRvyrs5ALYFqui5w83zNgEJUyvQBEtoxQaGDprmatw_bz
Pgd7h0UZ14ctt5UlrOx3PI&usqp=CAU" class="mobilelogo"
width="90",height="90"></a>      </th>
              <th><a href="https://www.NetMeds.com/"><img src="https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcRg5aRhF1m-LY0IpRiFkeEWdCOJggwAjOPz2A&usqp=CAU"
class="mobilelogo"
width="90",height="90"></a>      </th>
              <th><a href="https://www.meesho.com/"><img src="https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcS9WompEzMjpah3kVvc53n0yW-
wXCCtK6RrPY5Srlxgv1D6Hat1-hByK_hONt9eAeTveyY&usqp=CAU" class="mobilelogo"
width="90",height="90"></a></th>
            </tr>
            <tr style="background-color:#a6e21f;">
              <td>  </td>
              <td style="font-family:bold;font-size: 0.6cm;">Bigbasket</td>
              <td style="font-family:bold;font-size: 0.6cm;">1Mg</td>
              <td style="font-family:bold;font-size: 0.6cm;">NetMeds</td>
              <td style="font-family:bold;font-size: 0.6cm;">Meesho</td>
              <th>   </th>
            </tr>
  </table></center>
    </body>
</html>
```

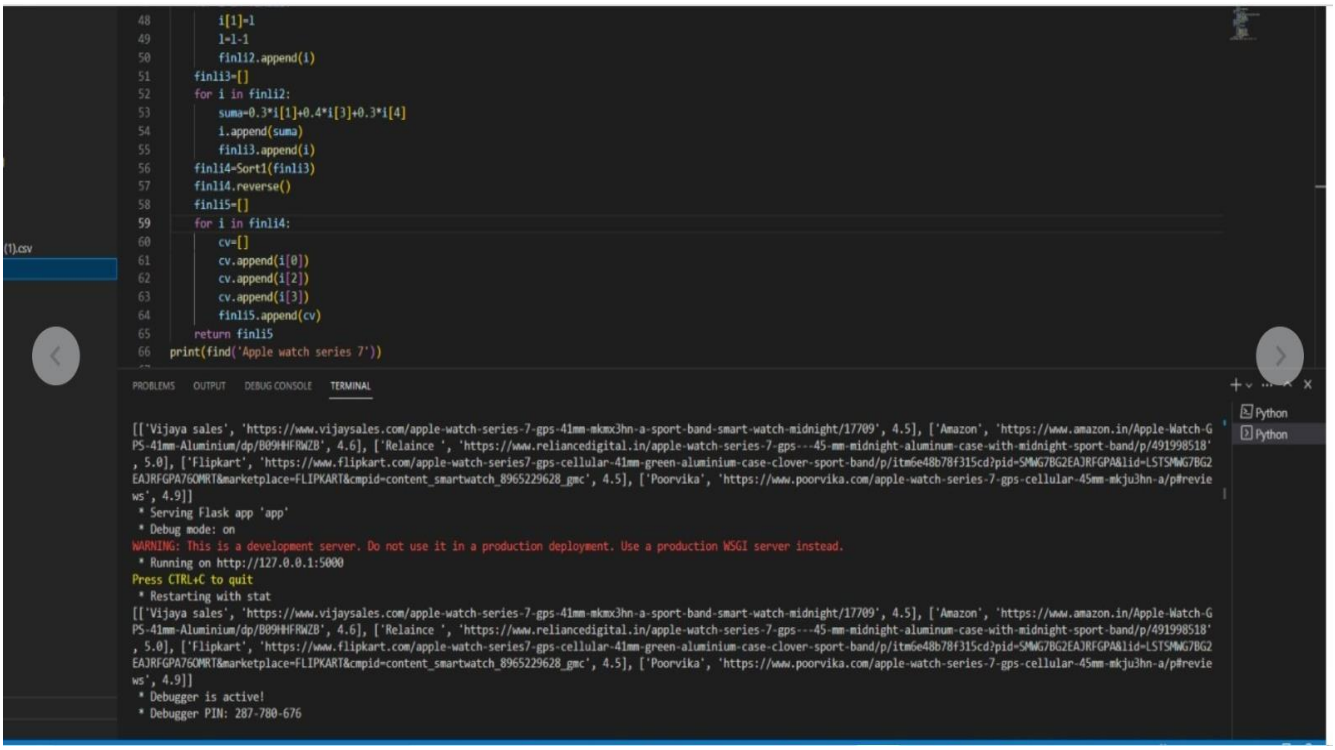
STYLE.CSS:

```
body{
    margin: 10;
    padding: 10;

    back
}
.sites
{
 Background-color:red;
}
```

4.2 Testing

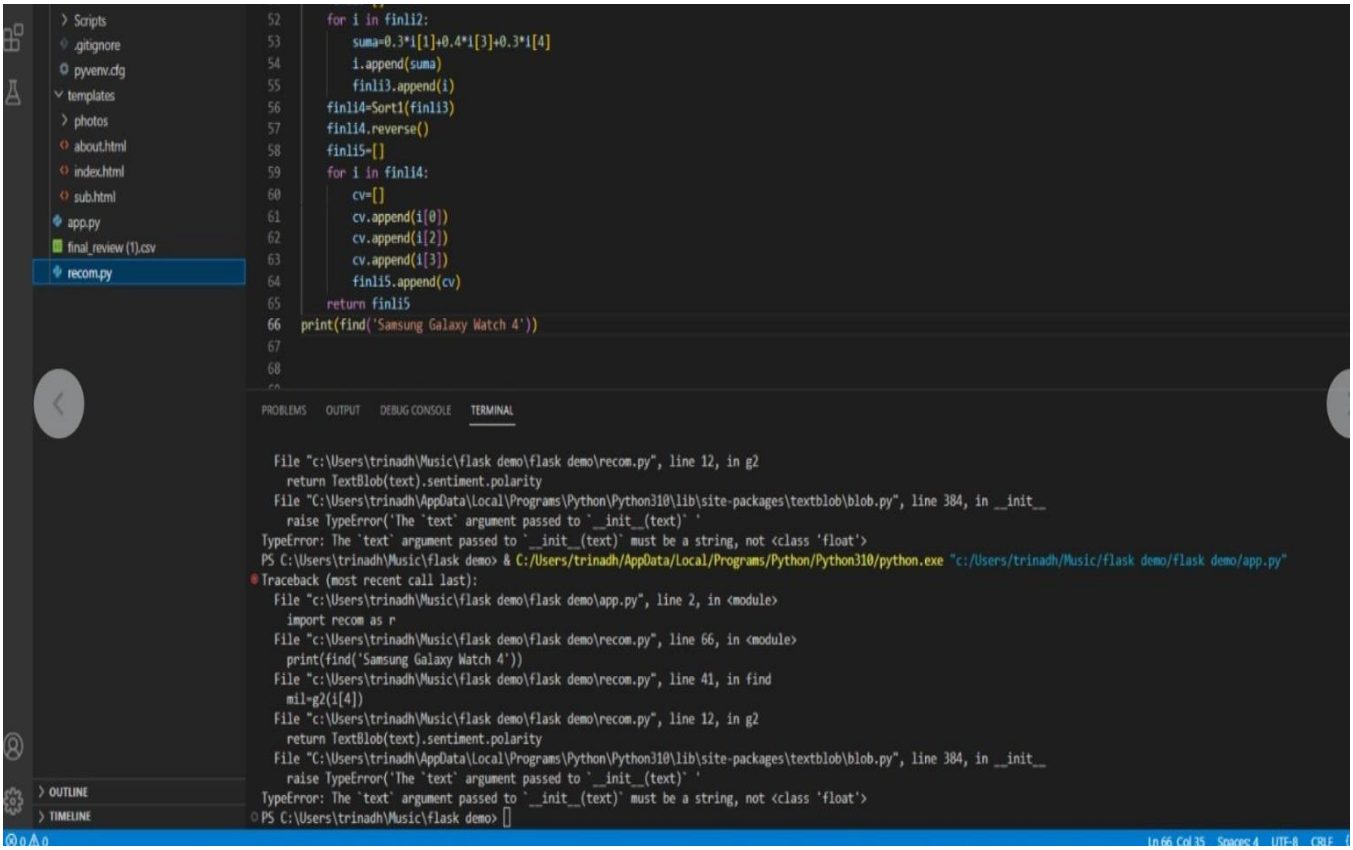




```
        i[1]=1
        l=l-1
        finli2.append(i)
    finli3=[]
    for i in finli2:
        suma=0.3*i[1]+0.4*i[3]+0.3*i[4]
        i.append(suma)
        finli3.append(i)
    finli4=Sort1(finli3)
    finli4.reverse()
    finli5=[]
    for i in finli4:
        cv=[]
        cv.append(i[0])
        cv.append(i[2])
        cv.append(i[3])
        finli5.append(cv)
    return finli5
print(find('Apple watch series 7'))
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

[['Vijaya sales', 'https://www.vijaysales.com/apple-watch-series-7-gps-41mm-mkmx3hn-a-sport-band-smart-watch-midnight/17709', 4.5], ['Amazon', 'https://www.amazon.in/Apple-Watch-GPS-41mm-Aluminium/dp/B09HHFRWZB', 4.6], ['Relaince ', 'https://www.reliancedigital.in/apple-watch-series-7-gps---45-mm-midnight-aluminum-case-with-midnight-sport-band/p/491998518', 5.0], ['Flipkart', 'https://www.flipkart.com/apple-watch-series7-gps-cellular-41mm-green-aluminium-case-clover-sport-band/p/itm6e48b78f315cd?pid=SMWG7BG2EAJRFGPA&lid=LSTSMWG7BG2EAJRFGPA76OMRT&marketplace=FLIPKART&cmpid=content_smartwatch_896529628_gmc', 4.5], ['Poorvika', 'https://www.poorvika.com/apple-watch-series-7-gps-cellular-45mm-mkju3hn-a/p#reviews', 4.9]]
 * Serving Flask app 'app'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
 * Restarting with stat
[['Vijaya sales', 'https://www.vijaysales.com/apple-watch-series-7-gps-41mm-mkmx3hn-a-sport-band-smart-watch-midnight/17709', 4.5], ['Amazon', 'https://www.amazon.in/Apple-Watch-GPS-41mm-Aluminium/dp/B09HHFRWZB', 4.6], ['Relaince ', 'https://www.reliancedigital.in/apple-watch-series-7-gps---45-mm-midnight-aluminum-case-with-midnight-sport-band/p/491998518', 5.0], ['Flipkart', 'https://www.flipkart.com/apple-watch-series7-gps-cellular-41mm-green-aluminium-case-clover-sport-band/p/itm6e48b78f315cd?pid=SMWG7BG2EAJRFGPA&lid=LSTSMWG7BG2EAJRFGPA76OMRT&marketplace=FLIPKART&cmpid=content_smartwatch_896529628_gmc', 4.5], ['Poorvika', 'https://www.poorvika.com/apple-watch-series-7-gps-cellular-45mm-mkju3hn-a/p#reviews', 4.9]]
 * Debugger is active!
 * Debugger PIN: 287-780-676
```



```
    for i in finli2:
        suma=0.3*i[1]+0.4*i[3]+0.3*i[4]
        i.append(suma)
        finli3.append(i)
    finli4=Sort1(finli3)
    finli4.reverse()
    finli5=[]
    for i in finli4:
        cv=[]
        cv.append(i[0])
        cv.append(i[2])
        cv.append(i[3])
        finli5.append(cv)
    return finli5
print(find('Samsung Galaxy Watch 4'))
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

  File "c:\Users\trinadh\Music\flask demo\flask demo\recom.py", line 12, in g2
    return TextBlob(text).sentiment.polarity
  File "C:\Users\trinadh\AppData\Local\Programs\Python\Python310\lib\site-packages\textblob\blob.py", line 384, in __init__
    raise TypeError('The `text` argument passed to `__init__(text)` '
TypeError: The `text` argument passed to `__init__(text)` must be a string, not <class 'float'>
PS C:\Users\trinadh\Music\flask demo> & C:/Users/trinadh/AppData/Local/Programs/Python/Python310/python.exe "c:/Users/trinadh/Music/flask demo/flask demo/app.py"
Traceback (most recent call last):
  File "c:\Users\trinadh\Music\flask demo\flask demo\app.py", line 2, in <module>
    import recom as r
  File "c:\Users\trinadh\Music\flask demo\flask demo\recom.py", line 66, in <module>
    print(find('Samsung Galaxy Watch 4'))
  File "c:\Users\trinadh\Music\flask demo\flask demo\recom.py", line 41, in find
    mil=g2(i[4])
  File "c:\Users\trinadh\Music\flask demo\flask demo\recom.py", line 12, in g2
    return TextBlob(text).sentiment.polarity
  File "C:\Users\trinadh\AppData\Local\Programs\Python\Python310\lib\site-packages\textblob\blob.py", line 384, in __init__
    raise TypeError('The `text` argument passed to `__init__(text)` '
TypeError: The `text` argument passed to `__init__(text)` must be a string, not <class 'float'>
PS C:\Users\trinadh\Music\flask demo>
```

5.Result Analysis

Algorithms	Accuracy
Logistic Regression	**94.95**
Linear Regression	**90.20**
Random Forest	**87.37**

Fig:5.1 Accuracy Table

The above table shows the accuraces of different models which are created by using the mentioned machine learning algorithms. Among all above models ,the model which is created by using gaussian naive bias algorithm got good accurance .So we consider it as the final model.

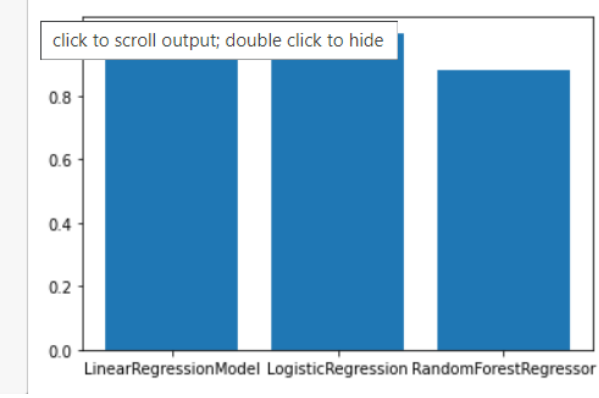


Fig:5.2 Accuracy bar chart

The above bar chat shows the accuracies of different models. Here y-axis is the model names andx-axis is the accuracy ranges.

6.Output Screens

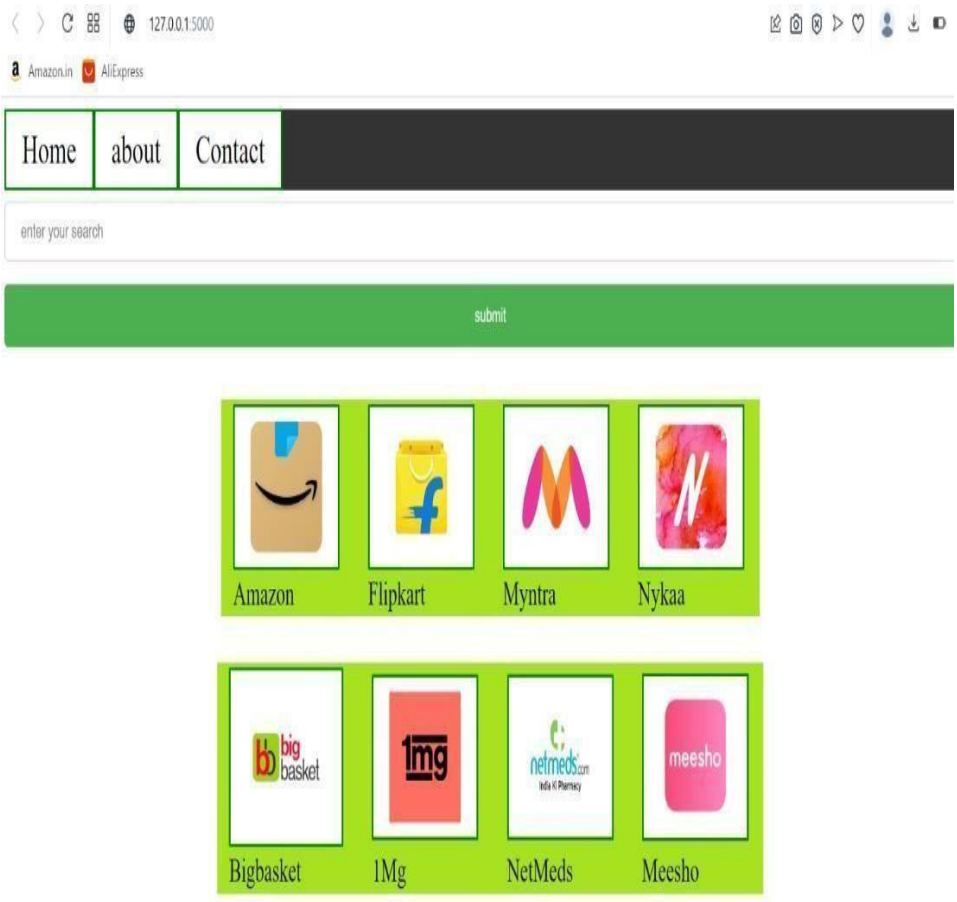


Fig:6.1 E Commerce site Prediction From

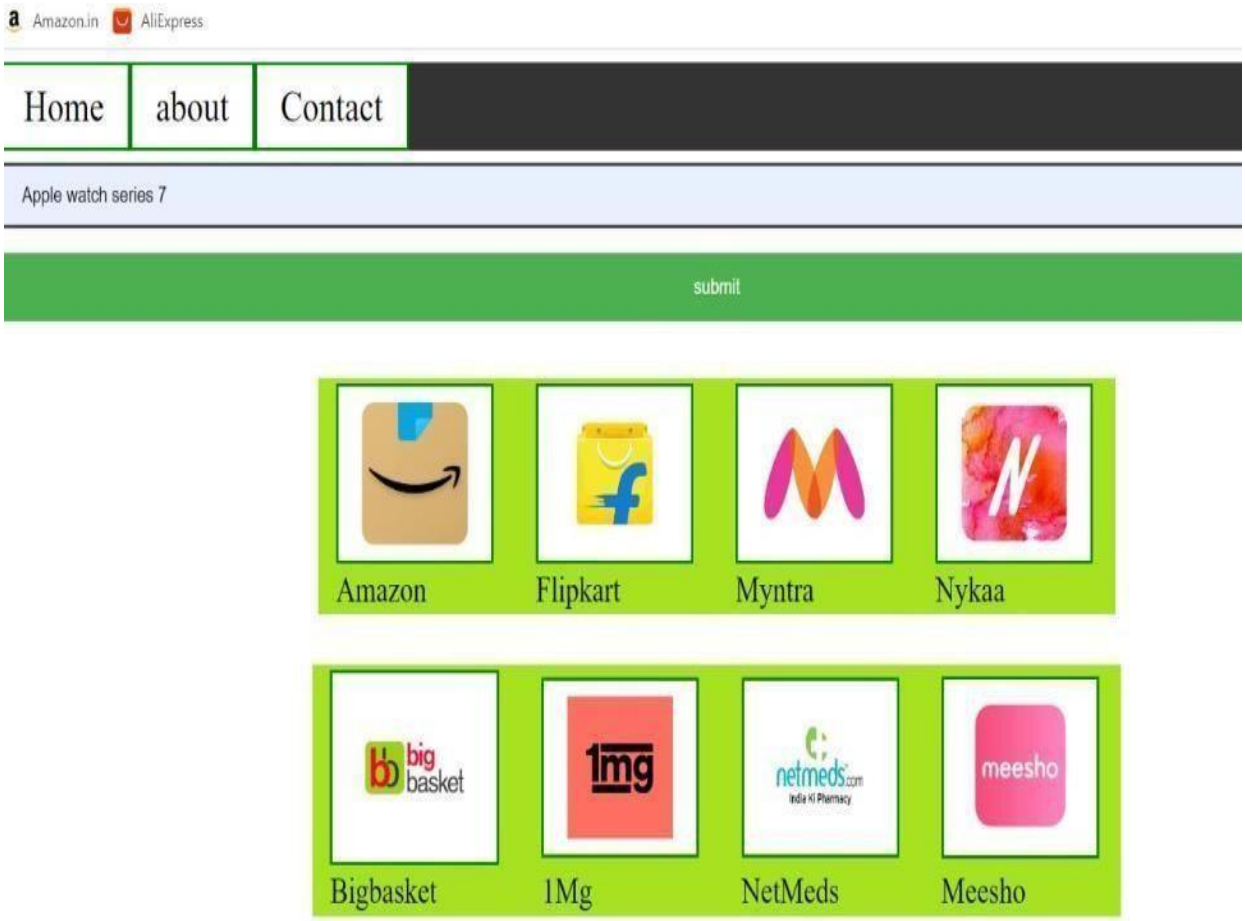


Fig:6.2 Filling Attributes for any product

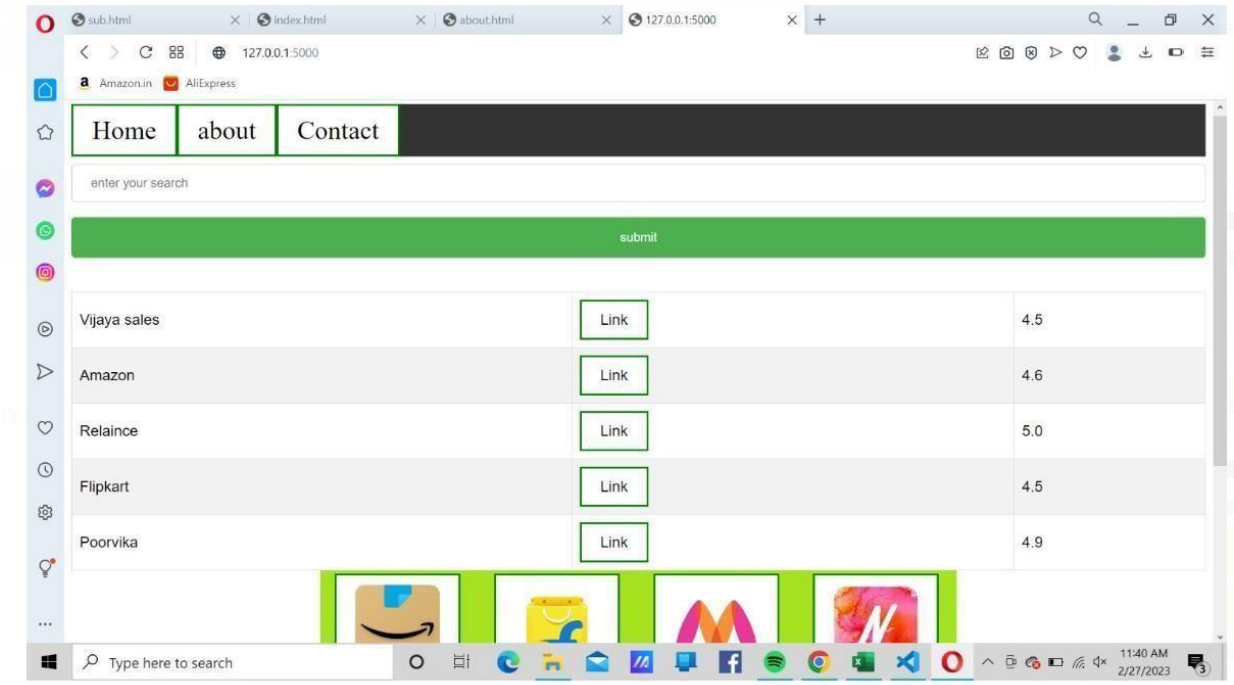


Fig:6.3 Predict of rating as a Result

7.CONCLUSION & FUTURE SCOPE

We have used 3 algorithms like Linear Regression, Random Forest ,Logistic Regression in order to predict the E commerce site. The accuracy varies for different algorithms. The accuracy for Random Forest algorithm is 87.3 when K-Fold cross validation is applied. The accuracy of Linear Regression algorithm is 90.3 when correlation and information gain are applied. The highest accuracy for Decision tree using Information Gain and K- Fold cross validation is 94.2%.

To develop more accuracy using machine learning algorithms and advanced techniques . The work can be extended and improved for the automation of e commerce site analysis based on the item income based on rating.

8

8.BIBLIOGRAPHY

[1] J. Mai, Y. Fan, and Y. Shen, ‘‘A neural networks-based clustering col-laborative filtering algorithm in e-commerce recommendation system,’’ inProc. Int. Conf. Web Inf. Syst. Mining (WISM), Nov. 2009, pp. 616–619.

[2] Y. Fan, J. Mai, and X. Ren, ‘‘A rough set-based clustering collaborative filtering algorithm in e-commerce recommendation system,’’ in Proc. Int.Conf. Inf. Manage., Innov. Manage. Ind. Eng., vol. 4, 2009, pp. 401–404.

[3] A. M. Yağcı, T. Aytekin, and F. S. Gürgen, ‘‘Parallel pairwise learningto rank for collaborative filtering,’’ Concurrency Comput., Pract. Exp.,vol. 31, no. 15, p. e5141, 2019.

[4] C. De Maio, G. Fenza, M. Gallo, V. Loia, and M. Parente, ‘‘Social media marketing through time-aware collaborative filtering,’’ Concurrency Comput., Pract. Exp., vol. 30, no. 1, p. e4098, Jan. 2018.

[5] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, ‘‘Context- based collaborative filtering for citation recommendation,’’ IEEE Access vol. 3, pp. 1695–1703, 2015.

[6] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, ‘‘Recom-mender systems in e-commerce,’’ in Proc. World Automat. Congr. (WAC),2014, pp. 179–184.

[7] S. K. Sharma and U. Suman, ‘‘Comparative study and analysis of Web personalization frameworks of recommender systems for e-commerce,’’ in Proc. CUBE Int. Inf. Technol. Conf., 2012, pp. 629–634.

[8] C. Haruechaiyasak, C. Tipnoe, S. Kongyoung, C. Damrongrat, and N. Angkawattanawit, ‘‘A dynamic framework for maintaining customer profiles in e-commerce recommender systems,’’ in Proc. IEEE Int. Conf.e-Technol., e-Commerce e-Service, Jun. 2005, pp. 768–771.

[9] L.-T. Weng, Y. Xu, and Y. Li, ‘‘A framework for e-commerce oriented recommendation systems,’’ in Proc. Int. Conf. Act. Media Technol. (AMT),2005, pp. 309–314.

[10] C. Fu and Z. Leng, ‘‘A framework for recommender systems in e-commerce based on distributed storage and data-mining,’’ in Proc. Int.Conf. E-Bus. E-Government (ICEE),

9. Conference Paper

E Commerce Sites Recommendation System using Machine Learning

Dasari Trinadh
Student
Department of Computer Science and Engineering
Narasaraopet, India
trinadhdasari9608@gmail.com

Kallam Krishna Chaitanya Reddy
Student
Department of Computer Science and Engineering
Narasaraopet, India
chaitanyareddy1864@gmail.com

Tanniru Manikanta
Student
Department of Computer Science and Engineering
Narasaraopet, India
manikantatanniru@gmail.com

Dr.S.Siva Nageswara Rao
Professor
Department of Computer Science and Engineering
Narasaraopet, India
drssnr@rediffmail.com

***Abstract*— In recent years, e-commerce websites have gained a lot of popularity and have become the primary source of online shopping. One of the key challenges for e-commerce websites is to recommend products to users that they are likely to purchase. To address this challenge, many e-commerce websites use recommendation systems based on machine learning algorithms .In this project, we propose a recommendation system for e-commerce websites that uses machine learning algorithms to recommended product to the users. The system uses a collaborative filtering approach that recommends products based on the preferences of other users with similar tastes. We use a combination of user and items filtering technique to improved accuracy of the recommendations. We evaluate the performance of our recommendation system on a real-world e-commerce dataset and compares it with others state of art recommendation algorithms. Our results show that our system outperforms other algorithms in terms of accuracy and efficiency. Overall, our proposed recommendation system can help e-commerce websites to improve their sales by providing personalized recommendations to their users. Our system can also help users to discover new products that they are likely to purchase, which can lead to increased customer satisfaction and loyalty.**

***Keywords*—MachineLearning , LinearRegression, Random Forest, Logistic regression, Flask**

I. INTRODUCTION

E-commerce websites have become an essential part of modern-day shopping, offering customers an easy and convenient way to purchase products online. However, with the vast number of products available on these websites, it can be overwhelming for customers to navigate and find the products they want.To address this challenge, e-commerce websites have implemented recommendation systems that use machine learning algorithms to suggest products to users. These recommendation systems analyze user behavior and historical data to identify patterns and preferences, enabling them to recommend products that are more relevant to each individual customer.The purpose of this project is to propose a recommendation system for e-commerce websites that uses machine learning algorithms to improve product recommendations. Specifically, we will be using a collaborative filtering approach that takes into account the preferences of other users with similar tastes to recommend products to users.In this project, we will be using a combination of user-based collaborative filtering techniques to improve the accuracy of our recommendations. We will also evaluate the performance of our system on a real-world e-commerce dataset and compare it with other state-of-the-art recommendation algorithms.The rest of the project is organized as follows. , we will provide a brief overview of collaborative filtering and how it can be used in recommendation systems. After that, we will describe the dataset used in our experiments, followed by a description of our proposed recommendation system. Finally, we will present our experimental results and conclude with a discussion of our findings and future work.and random forests to train the model.Once we have developed the model, we will evaluate its performance by testing it on a separate set of data. This will enable us to determine the accuracy of the model and identify any areas for improvement.

II. LITERATURE REVIEW

Recommendation systems have been widely used in e-commerce websites to improve the customer experience and increase sales. There are several types of recommendation algorithms, including contents, collaborative based filtering, and hybrid approaches. Collaborative filtering is one of the most popular recommendation algorithms in use today. It is based on the assumption that people who have similar preferences in past will have more similar preference in the future. Collaborative filtering can be divided into two types: user-based and item-based.User-based collaborative filtering recommends products to users based on the preferences of

other users with similar tastes. The system identifies users who have similar preferences and recommends products that they have liked. Item-based collaborative filtering, on the other hand, recommends products based on the similarity between items. The system identifies items that are similar to the ones the user has liked in the past and recommends them.

Several studies have shown that collaborative filtering can significantly improve the accuracy of recommendations compared to other algorithms. For example, a study by Sarwar et al. (2001) showed that collaborative filtering outperformed content-based filtering in terms of recommendation accuracy. Another study by Paterek (2007) showed that item-based collaborative filtering was more effective than user-based collaborative filtering.

In recent years, several researchers have proposed new methods to improve the performance of collaborative filtering. For example, Koren et al. (2009) proposed a matrix factorization approach that improved the accuracy of collaborative filtering by modeling the latent factors that influence user preferences. Another study by Hu et al. (2008) proposed a social collaborative filtering approach that takes into account the social relationships between users to improve recommendation accuracy.

Overall, collaborative filtering is a powerful recommendation algorithm that has been widely used in e-commerce websites. However, there is still room for improvement, and researchers are continually proposing new methods to enhance its performance. In this project, we propose a recommendation system for e-commerce websites that uses a combination of item-based collaborative filtering techniques to improve the accuracy of recommendations..

III. EXISTING SYSTEM

There are several e-commerce websites that use recommendation systems based on machine learning algorithms to suggest products to their customers. Some examples of these websites are Amazon, Netflix, and Spotify. Overall, these e-commerce websites have sophisticated recommendation systems that use machine learning algorithms to suggest products to their users. These systems have been shown to improve customer satisfaction and increase sales by providing personalized recommendations to each user. While these recommendation systems have been successful in improving the customer experience and increasing sales, there is still room for improvement. In this project, we propose a recommendation system for e-commerce websites that uses a combination of user-based and item-based collaborative filtering techniques to provide more accurate recommendations.

IV PROPOSED SYSTEM

Our proposed recommendations system for e-commerce websites uses a combination of products user-based and item-based collaborative filtering techniques to improve the accuracy of recommendations. The system takes into account the preferences of other users with similar tastes to recommend products to each individual user . e-commerce website uses a combinations of the both user,item-based collaborative filtering techniques to provide the personalized on recommendations to users.

V DATASET AND DATAVISUALIZATION

We have used the dataset available in Kaggle which consists different types of cars . The dataset is split into training and testing datasets . The training data is 80% of the total dataset, validation data is 10% of dataset and testing data is other 10% of the dataset.

Data visualization is an important tool for data analysis and communication that enables us to visually represent complex datasets and identify patterns and relationships within the data. Visualizations can take various forms, such as scatter plots, line charts, bar charts, histograms, heat maps, box plots, tree maps, and many more. It is The choice of visualization technique depends on type of data and the specific insights being communicated. The goal of data visualization is to communicate complex information clearly and effectively, making it easy for the audience to understand the key insights and trends in the data.

Data visualization can take many forms, including, graphs, maps, charts and other visual aids. It is often used in fields such as business, science, engineering, medicine, and social sciences to present data in a way that is accessible and easy to understand.

product_id	product_name	product_type	product_brand	product_image	ecs_1_name	ecs_1_price
110001	REFLEX 3.0 DUAL TONED SMART BAND IN MIDNIGHT B...	Smart Watch	Fastrack	https://staticimg.titan.co.in/Fastrack/Catalog...	Fastrack	1795
110002	Amazfit bip u	Smart Watch	Amazfit	https://encrypted-tbn1.gstatic.com/shopping?q=...	Myntra	2499
110003	Noise ColorFit Pro 3	Smart Watch	Noise	https://m.media-amazon.com/images/I/61OpDFvFkE...	Amazon	5999
110004	Samsung Galaxy Watch 4	Smart Watch	Samsung	https://images.samsung.com/is/image/samsung/p6...	Amazon	18999
110005	Apple watch series 7	Smart Watch	Apple	https://d2xamzlzrdbdbn.cloudfront.net/products...	Vijay sales	38900

Fig.1 : Data Set

Fig.2: online business based on an e-commerce platform

VI PREPROCESSING

Data preprocessing is an important step in the data analysis process, where raw data is transformed into a format that is suitable for analysis. Here are some common techniques used in data preprocessing:

Data Cleaning: It is the process of removing or correcting any errors or inconsistencies in the data.This can include removing duplicates, correcting misspelled values, or imputing missing data.

Data Transformation: It is the process of converting data from one format to another, such as converting categorical data to numerical data. This can also include scaling data to a common range or normalizing data have a mean of zero and standard deviation of one.

Data Reduction: Data reduction involves reducing the amount of data to be analyzed. This can include identifying and removing irrelevant features or reducing the resolution of data by aggregating it into larger groups.

Overall, data preprocessing is a major step in data analysis process as the data is consistent, accurate and in a format that can be easily can be analyzed.

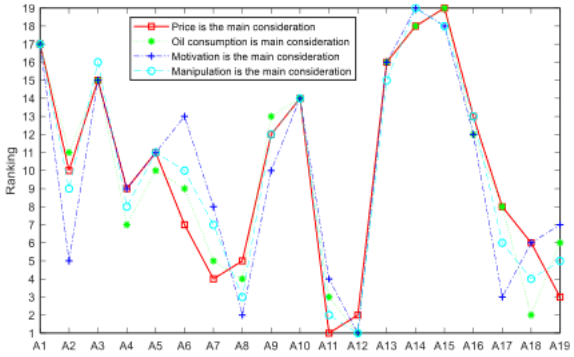


Fig.3: Comparison of ranking of items based on different preferences

VII METHODOLOGY AND IMPLEMENTAION

System analysis of car price prediction involves understanding the various components and processes involved in the system, and how they work together to predict the price of a car. Here are some key components of a car price prediction system:

Data Collection: The system needs to collect data from various sources, such as historical sales data, market trends, and car specifications. This data is then used to train the prediction model.

Data Preprocessing: The raw data collected from various sources may not be in a format that is suitable for analysis. Therefore, data preprocessing techniques such as data mining ,data cleaning, integration and transformation need to be applied to ensure that the data is accurate and consistent.

Feature Selection: The system needs to identify which features of the car are relevant for predicting its price. This can be done using statistical techniques or machine learning algorithms.

Prediction Model: The prediction model is trained using the preprocessed data and selected features. In the Various algorithms such as regression models, decision trees, can be used to build the prediction model.

Model Evaluation: The prediction model needs to be evaluated to assess its accuracy and effectiveness. This can be done using metrics such as mean squared error, root mean squared error, or R-squared.

Deployment: The prediction model is then deployed into a production environment, where it can be used to predict the price of a car based on its specifications.

Overall, a e commerce sites recommendation system requires a combination of data collection, preprocessing, feature selection, prediction modeling, model evaluation, and deployment. The accuracy and effectiveness of the system depend on the quality and quantity of data collected, the effectiveness of preprocessing techniques, and the choice of machine learning algorithms used in the prediction model.

Different types models are used to find best accuracy:

Those are:

1. Linear Regression: It is a supervised and statistical Learning algorithm used to predict a continuous output variable is known as a dependent variable based on one or more input variables is known as independent or predictor variables. The relationship between the input variables and output variable is assumed to be linear, meaning that the relationship can be represented by a straight line. The algorithm tries to find the best fitting line (known as the regression line) that passes through the data points, minimizing the difference between the predicted and actual values of the output variable.

The equation for a simple linear regression can be written as:

y = k0 + k1 * x

where y is output variable, x is input variable, k0 is the intercept, and k1 is the coefficient of the input variable.

2. Random forest:

Random Forest is a supervised and versatile algorithm that can be used for classification and regression problems. It is also relatively easy to use, requires minimal data pre processing, and can handle both numerical and categorical data. Additionally, Random Forest has the ability to handle missing data and outlier values, making it a popular choice for many machine learning applications.

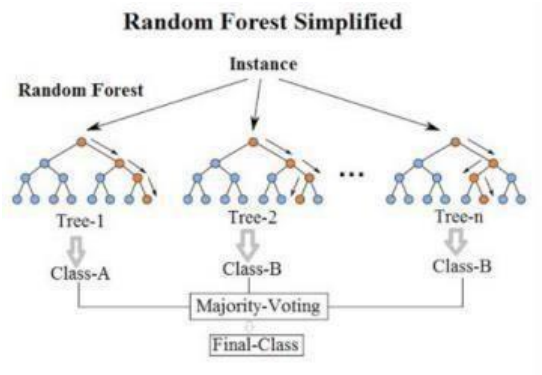


Fig 5: Random Forest Model

3. Logistic regression: Logistic Regression is a supervised and statistical method used for binary classification problems, where the outcome variable takes only two values (0 or 1). The goal of logistic regression is to find out the best fitting line (or hyperplane in higher dimensions) that separates the two classes.

The logistic function is given by:

p = 1 / (1 + exp(-z))

where p is the probability of the positive class, z is the linear combination of the input features and their corresponding weights, and exp is the exponential function.

Logistic Regression works by optimizing the weights of the input features to maximize the likelihood of the observed data given the model parameters. This optimization is usually done using maximum likelihood estimation or gradient descent.

Logistic Regression is a popular algorithm due to its simplicity and interpretability. It can handle both categorical and continuous input features and is robust to noise and outliers. Additionally, it can be easily extended to handle multi-class classification problems using techniques such as One-vs-All and Softmax regression.

VIII. RESULT AND ANALYSIS

The accuracy of the different model is shown below:

Algorithm	Accuracy
Linear Regression	90.20
Random Forest	87.37
Logistic Regression	94.37

The above table shows the accuraces of different models which are created by using the mentioned machine learning algorithms. Among all above models ,the model which is created by using Logistic Regression algorithm got good accurance .So we consider it as the final model.

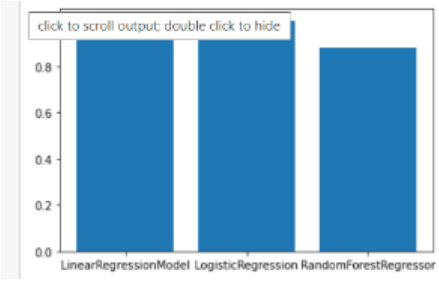


Fig.6: Accuracy Bar chart

This section explains output which detects the best site.

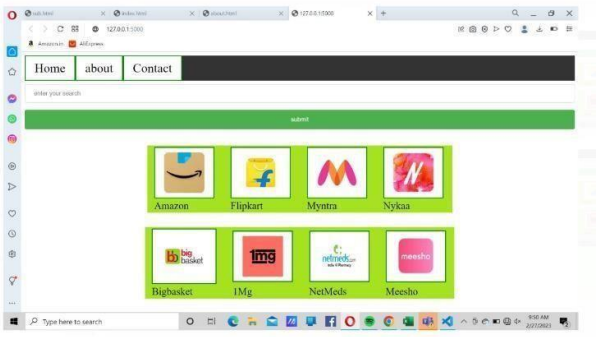


Fig.7: Home Page

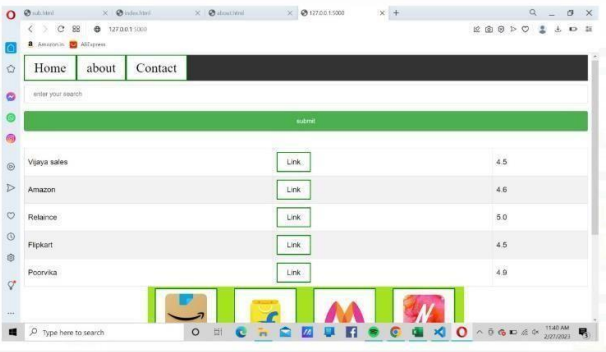


Fig .8: Link predicted as a result

IX. CONCLUSION:

We have used 3 algorithms like Linear Regression, Random Forest , Logistic Regression in order to predict the house price. The accuracy varies for different algorithms. The accuracy for Random Forest algorithm is 87.37% when. The accuracy of Linear Regression algorithm is 90.20% when correlation and information gain are applied. The highest accuracy for Logistic Regression using is 94.37%

X. REFERENCES

[1] J. Mai, Y. Fan, and Y. Shen, ‘‘A neural networks-based clustering col-laborative filtering algorithm in e-commerce recommendation system,’’ inProc. Int. Conf. Web Inf. Syst. Mining (WISM), Nov. 2009, pp. 616–619.

[2] Y. Fan, J. Mai, and X. Ren, ‘‘A rough set-based clustering collaborative filtering algorithm in e-commerce recommendation system,’’ in Proc. Int.Conf. Inf. Manage., Innov. Manage. Ind. Eng., vol. 4, 2009, pp. 401–404.

[3] A. M. Yağcı, T. Aytekin, and F. S. Gürgen, ‘‘Parallel pairwise learningto rank for collaborative filtering,’’ Concurrency Comput., Pract. Exp.,vol. 31, no. 15, p. e5141, 2019.

[4] C. De Maio, G. Fenza, M. Gallo, V. Loia, and M. Parente, ‘‘Social media marketing through time-aware collaborative filtering,’’ Concurrency Comput., Pract. Exp., vol. 30, no. 1, p. e4098, Jan. 2018.

[5] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, ‘‘Context- based collaborative filtering for citation recommendation,’’ IEEE Access vol. 3, pp. 1695–1703, 2015.

[6] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, ‘‘Recom-mender systems in e-commerce,’’ in Proc. World Automat. Congr. (WAC),2014, pp. 179–184.

[7] S. K. Sharma and U. Suman, ‘‘Comparative study and analysis of Web personalization frameworks of recommender systems for e-commerce,’’ in Proc. CUBE Int. Inf. Technol. Conf., 2012, pp. 629–634.

[8] C. Haruechaiyasak, C. Tipnoe, S. Kongyoung, C. Damrongrat, and N. Angkawattanawit, ‘‘A dynamic framework for maintaining customer profiles in e-commerce recommender systems,’’ in Proc. IEEE Int. Conf.e-Technol., e-Commerce e-Service, Jun. 2005, pp. 768–771.

[9] L.-T. Weng, Y. Xu, and Y. Li, ‘‘A framework for e-commerce oriented recommendation systems,’’ in Proc. Int. Conf. Act. Media Technol. (AMT),2005, pp. 309–314.

[10] C. Fu and Z. Leng, ‘‘A framework for recommender systems in e-commerce based on distributed storage and data-mining,’’ in Proc. Int.Conf. E-Bus. E-Government (ICEE), May 2010, pp. 3502–3505.

[11] S. G. Walunj and K. Sadafale, ‘‘An online recommendation system fore-commerce based on apache mahout framework,’’ in Proc. Annu. Conf.Comput. People Res., 2013, pp. 153–158.

DB4

ORIGINALITY REPORT

13%	7%	7%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	kobra.uni-kassel.de Internet Source	1%
2	Submitted to Liverpool John Moores University Student Paper	1%
3	"Data Science", Springer Science and Business Media LLC, 2019 Publication	1%
4	Submitted to Heriot-Watt University Student Paper	1%
5	www.researchgate.net Internet Source	1%
6	Submitted to Colorado Technical University Online Student Paper	1%
7	Submitted to University of Birmingham Student Paper	1%
8	Chaochao Chen, Xiaolin Zheng, Mengying Zhu, Litao Xiao. " Composite Social Trust Networks",	1%

	International Journal of Web Services Research, 2016 Publication	
9	Submitted to London Business School Student Paper	1%
10	slidescope.com Internet Source	1%
11	Submitted to Imperial College of Science, Technology and Medicine Student Paper	1%
12	Submitted to Nanyang Technological University Student Paper	1%
13	Prayitno Abadi, Umar Ali Ahmad, Yuichi Otsuka, Punyawi Jamjareegulgarn et al. "Modeling Post-Sunset Equatorial Spread-F Occurrence as a Function of Evening Upward Plasma Drift Using Logistic Regression, Deduced from Ionosondes in Southeast Asia", Remote Sensing, 2022 Publication	1%
14	Submitted to Laureate Higher Education Group Student Paper	<1%
15	www.theses.fr Internet Source	<1%

16	"Data Mining and Data Warehousing", Studies in Computational Intelligence, 2007 Publication	<1%
17	Cathal Gurrin, Alan F. Smeaton, Hyowon Lee, Kieran McDonald, Noel Murphy, Noel O'Connor, Sean Marlow. "Chapter 10 Mobile Access to the Físchlár-News Archive", Springer Science and Business Media LLC, 2004 Publication	<1%
18	Kwok Tong Chau. "Research on electric vehicles", Studies in Science and Technology , 2012 Publication	<1%
19	Bhanu Prasad. "Chapter 32 HYREC: A Hybrid Recommendation System for E-Commerce", Springer Science and Business Media LLC, 2005 Publication	<1%
20	dokumen.pub Internet Source	<1%
21	repository.uel.ac.uk Internet Source	<1%
22	Xiaoye Miao, Yangyang Wu, Lu Chen, Yunjun Gao, Jun Wang, Jianwei Yin. "Efficient and effective with influence functions", Proceedings of the VLDB Endowment, 2022	<1%

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23066

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Dasari Trinadh , Narasaraopeta Engineering College has presented the paper title E commerce sites recommendation system using machine learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23066

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Kallam Krishna Chaitanya Reddy , Narasaraopeta Engineering College has presented the paper title E commerce sites recommendation system using machine learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23066

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Tanniru Manikanta , Narasaraopeta Engineering College has presented the paper title E commerce sites recommendation system using machine learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar


NARASARAOPETA
NEC ENGINEERING COLLEGE
(AUTONOMOUS)
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in
International Conference on
PAPER ID
NECICAIEA2K23066
Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17th & 18th March, 2023
Organized by Department of Computer Science and Engineering in Association with CSI
Certificate of Presentation
This is to Certify that Dr.S.Siva Nageswara Rao , Narasaraopeta Engineering College has presented the paper title E commerce sites recommendation system using machine learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.
Convenor
Dr.S.V.N.Srinivasu
Chief-Convenor
Dr.S.N.Tirumala Rao
Principal, Patron
Dr. M. Sreenivasa Kumar
MHRD'S INNOVATION CELL (GOVERNMENT OF INDIA)