SEGMENTATION AND COUPON ISSUANCE PREDICTION ON ONLINE SHOPPING USING MACHINE LEARNING TECHNIQUES

[1]Shaik Nowsheen Fathima, [2]Shaik Sumiya
[1,2]UG Student, [1,2]Department of Computer Science & Engineering, Geethanjali Institute of Science and Technology, Gangavaram, Andhra Pradesh, India

ABSTRACT
With the development of bigdata and deep learning technology, big data and deep learning technology have also been applied to the marketing field, which was a apart of business administration. Customer churn management is one of the important areas of marketing.
The objective of this project is to prevent customer churn and increase purchase conversion rate by issuing customized discount coupons to customers with high churn rate based on big data in real time. After segmenting customer segments with two-dimensional segment analysis, we predict coupons that suits to our customers. A two-dimensional cluster analysis-based churn rate estimation place an important role in order to easy to take proactive measures.Finally this will helps to find the conversion rate and sales growth.

INTRODUCTION
With the development of big data and deep learning technology, big data and deep learning technology have also been applied to the marketing field, which was a part of management. Also, growth in internet adoption has made digital coupons a popular promotional tool . Customized digital coupon issuance is a very important topic in online commerce. This is because maintaining existing customers is a more important business issue than acquiring new customers . Also, retaining existing customers is much more economically advantageous than acquiring new customers . In fact, the acquisition cost of new customers is known to be five to six times higher than the maintenance cost of existing customers. Companies that have effectively managed customer churn by improving customer retention are known to have a positive effect not only on the company's profitability but also on improving brand image by improving customer satisfaction. Customized coupon issuance research has traditionally been active in highly competitive and urgent sectors such as telecommunications, finance, distribution, and game industries, and has focused mainly on developing predictive models using machine learning and artificial intelligence technology . Also, recently, AI-based marketing using big data analysis and deep learning is emerging. Such AI-driven targeting can save huge amounts of marketing costs and raise online sales provided that the targeting model succeeds in estimating customer responsiveness accurately . In particular , in the case of online shopping malls, the average purchase conversion rate is around 2%.
Online shopping malls have the advantage of being easily accessed through the PC web or mobile web, but on the contrary, this advantage can be a disadvantage that it is easy to see and leave quickly. Therefore, even the slightest reduction of customer churn rate can lead to high conversions, which can lead to huge profits. Unlike offline shopping malls, online shopping malls are easy to collect data. All online behavioural characteristics of customers can be collected in real time in the shopping mall's 7924 This work is licensed under a Creative Commons Attribution-NonCommercial-No Derivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/ VOLUME 11, 2023 D. Seo, Y. Yoo: Improving Shopping Mall Revenue by Real-Time Customized Digital Coupon Issuance own DB. Therefore, it is possible to have a wealth of customer history data and to use it to understand customer tendencies.

In conclusion, if you use rich customer historical data to infer behaviours and tastes, you can increase customer conversion rates without special promotions. The easiest and most intuitive way is to issue personalized coupons to customers in real time. By selecting customers with a high risk of real-time churn and issuing real-time customized discount coupons, it is possible to increase sales by increasing the purchase

conversion rate without burdening special expenses such as promotional events. And to put these strategies into action, you need an AI-powered strategy. After AI automatically learns the histories of customers, it is possible to properly issue coupons by identifying the behaviours and tastes of individual customers. In summary, the integration of real-time customized digital coupon issuance presents a compelling opportunity for shopping malls to drive revenue growth, enhance customer engagement, and stay competitive in an increasingly digital marketplace. By leveraging the power of digital technology, data analytics, and personalized marketing, malls can deliver targeted promotions that resonate with consumers, foster loyalty, and ultimately drive sales.

MACHINE LEARNING

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data miningis a related field of study, focusing on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine developits own algorithm, rather than have human programmers specify every needed step.

The discipline of machine learning employs various approaches to help computerslearn to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the MNIST dataset has often been used.

MACHINE LEARNING APPROACHES

Early classifications for machine learning approaches sometimes divided theminto three broad categories, depending on the nature of the "signal" or "feedback" available to the learning system. The Fig 1.2 represents Types of Machine Learning.

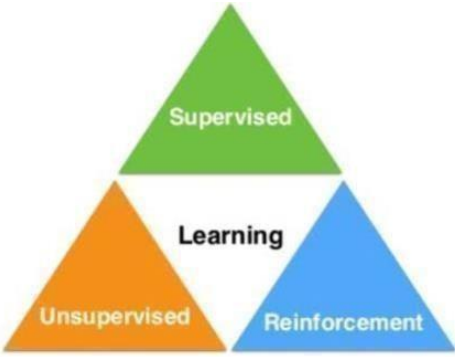

Fig:1. Types of Machine Learning

Supervised learning:

The computer is presented with example inputs andtheir desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.

Unsupervised learning:

No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

Reinforcement learning:

A computer program interacts with a dynamicenvironment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent) as it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.

other approaches or processes have since developed that don't fit neatly into this three-fold categorization, and sometimes more than one is used by the same machine learning algorithms. For example, topic modeling, dimensionality reduction or meta learning.[8] As of 2020 deep learning has become the dominant approach for much ongoing work in the field of machine learning.

OBJECTIVE

The objective of this paper is to design an efficient system that can detects of issuing the digital coupons to customer with more accuracy and predicting the customer churn towards improving the shopping mall revenue by using Machine Learning models.

LITERATURE SURVEY

In this chapter we review some papers to get knowledge and understanding on the techniques had been proposed. All those techniques have the same aim which is used to issue the digital coupons and clustering the data. As Archimedes once said, "Man has always learned from the past. After all, you can't learn history in reverse!" it is essential for man to learn from history. Thus, considering all past researches, the most relevant research glimpses have been picked to
be explained in detail. The overview shall discuss relevant aspects contributing to our research.

What Makes a Consumer Redeem Digital Coupons?Behavioral Insights from Grounded Theory Approach

P. Naval and N. Pandey

The growth in internet adoption has made digital coupons a popular promotional tool.However, the extant literature on digital coupons is at an embryonic stage and requires theory building. This study adopts an inductive grounded theory approach to explore the new horizons for digital Coupon redemption. The results showed that the intention to search and subscribe is a key driver of redemption, while coupon proneness and coupon acquisition value are itsdimensions. Social media reviews, past experience, privacy risk, and customization are more relevant today for coupon redemption. The study also proposes segmentation of coupon users into deal buyers, planners, and convenience seekers based on consumer behavior.

LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning.

M. Sarkar and A. D. Bruyn In predictive modeling, firms often deal with high-dimensional data that span multiple websites, demographics, purchase types, and product categories. websites, demographics, purchase type and product categories. Traditional customer response models rely heavily on feature engineering, and their performance depends on the analyst's domain knowledge and expertise to craft relevant predictor.

As the complexity of data increases, however, traditional models grow exponentially complicated. In this paper, we demonstrate that long-short term memory (LSTM) neural networks, which rely exclusively on raw data as input, can predict customer behaviors with great accuracy. In our first application, a model outperforms standard benchmarks. In a second, more realistic application, an LSTM model competes against 271 hand-crafted models that use a wide variety of features and modeling approaches. It beats 269 of them, most by a wide margin. LSTM neural networks are excellent candidates for modeling customer behavior using panel data in complex environments (e.g., direct marketing, brand choices, clickstream data, churn prediction).

An intelligent framework for online product recommendation using collaborative filtering
G. Chandrasekaran and D. J. Hemanth

Recommendation systems have become a vital area of research in recent times. These recommendation systems are very much needed for e-commerce applications to identify the products liked by a customer which helps the companies to promote product sales and improve their product quality. It also helps the users to arrive at the purchasing decision without reading the online reviews about the product. The key idea behind the proposed work is to analyze the user preference for the products from the online data by employing the collaborative filtering-based recommendation framework. The concept of collaborative filtering is best suited for recommendation systems involving a large set of product users. It generates a user-item matrix and finds the list of products liked by the individual users. It gives prediction regarding the product that a user could buy in the future and also recommends the products which are liked by the customers who have similar interests. It gives a comparative analysis in terms of performance metrics and accuracy of different collaborative filtering techniques.

Time-Aware Smart Object Recommendation in Social Internet of Things
Y. Chen, M. Zhou, Z. Zheng, and D. Chen

With a large number of possible smart objects in Social Internet of Things (SIoT), a recommendation system is of great necessity to help users find smart objects they need. However, traditional recommendation techniques usually exploit user's rating or feedback information, which are impractical as such kind of user preference information is difficult to collect in the SIoT environment. In addition, temporal context plays an important role in smart object recommendation since most users tend to utilize different objects at different time slots in a day, e.g., making coffee at morning and playing games on weekends. In this article, we propose a time-aware smart object recommendation model by jointly considering user's preference over time and smart object's social similarity. We first learn user's preference over time from his/her object usage events with a latent probabilistic model. Then, we estimate the smart object's social similarity by embedding their heterogeneous social relationships into a shared lower dimensional space. Finally, we generate the recommendation list with an item-based collaborative filtering. We conduct a comprehensive experimental study based on two real-world data sets, and the experimental results show our method outperforms all baselines significantly in terms of recommendation effectiveness .

EXISTING SYSTEM

Customer Relationship Management (CRM) is all about finding the customers. Collecting their information along the way and using that information to enhance their experience and foster long term relationships.

Some of the machine learning algorithms can be used in CRM are Decision trees, SVM.

The Drawbacks of CRM are Limited control and flexibility, It may also result in duplication of tasks.

User Registration and Profiling:

Shoppers need to register with the shopping mall's mobile app or website.

During registration, users provide basic information such as demographics, preferences, and shopping behaviour.

Data Collection and Analysis:

Collect and analyze data from various sources, including purchase history, browsing behavior, and location data.

Use machine learning algorithms to gain insights into individual shopper preferences and trends.

Real-Time Analytics Engine:

Implement a real-time analytics engine that processes incoming data and identifies relevant patterns.

Utilize machine learning models to predict user behaviour and preferences based on historical and real-time data.

Coupon Generation System:

Develop a coupon generation system that creates personalized coupons for individual shoppers.

Take into account factors such as past purchases, preferences, and real-time location.

Integration with POS Systems:
Integrate the coupon system with point-of-sale (POS) systems to ensure seamless redemption at checkout.
Enable real-time validation of coupons to prevent fraud and misuse.
Mobile App/Website Interface:
Provide a user-friendly interface on the mobile app or website where shoppers can easily view and access their personalized coupons.
Implement push notifications to alert users about new, relevant coupons when they are in proximity to the shopping mall.

PROPOSED SYSTEM
In this project we proposed Machine Learning technology to detect customer behaviour,identifying the customer spending and nature loyalty metrics. The System consists of the following steps :two dimensional clustering analysis, Suggesting the coupon issuance. One of the best and simplest ways to identify and extract items from the dataset. Our proposed system aims to provide real-time customized digital coupon recommendations to online shopping customers to prevent churn and increase conversion rate.
User-Friendly Mobile App:
Develop a user-friendly mobile application for both iOS and Android platforms. Include a seamless and intuitive interface for easy navigation.
Smart User Registration:
Implement a quick and easy registration process. Allow users to sign up using social media accounts for a faster onboarding experience.
User Profiling and Preferences:
Collect user data, including demographics, shopping preferences, and behaviour. Encourage users to voluntarily provide information through surveys to enhance personalization.
Real-Time Data Processing:
Utilize a cloud-based infrastructure for efficient real-time data processing. Implement big data analytics to process and analyze user data promptly.
Machine Learning Algorithms:
Employ machine learning algorithms to predict user preferences based on historical and real-time data.Continuously update and refine algorithms for better accuracy.
Personalized Coupon Generation:
Design a coupon generation engine that tailors discounts based on user profiles . Take into account factors like purchase history, preferences, and real-time location.
METHODOLOGY
SYSTEM ARCHITECTURE
Machine Learning algorithms are totally subject to data since it is the most vital perspective that makes model training possible. On the other hand, if won't be able to makesense out of that data, before feeding it to ML algorithms, a machine will be useless. In straight forward words, we generally need to take care of the right data for example the datain the right scale, group, and containing important features, for the problem we need a machine to solve.
This makes data preparation the most important step in the ML process. Data preparationdefined as the procedure that makes our dataset more appropriate to work with in the ML process.

Customer Segmentation

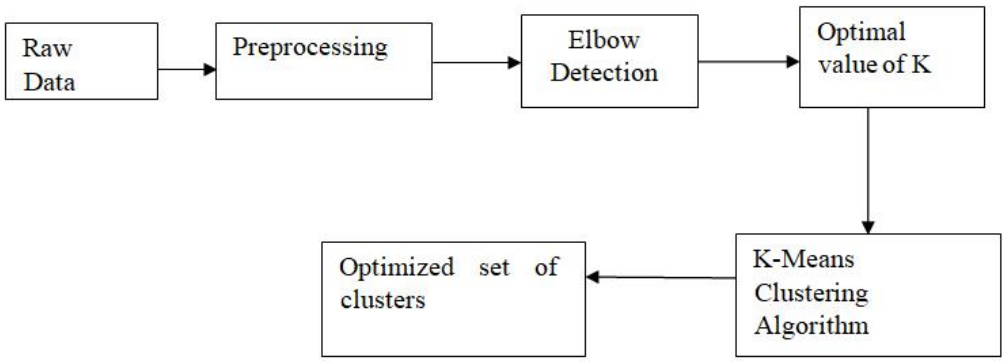

Fig.2. Customer Segmentation Model

Coupon Recommendation

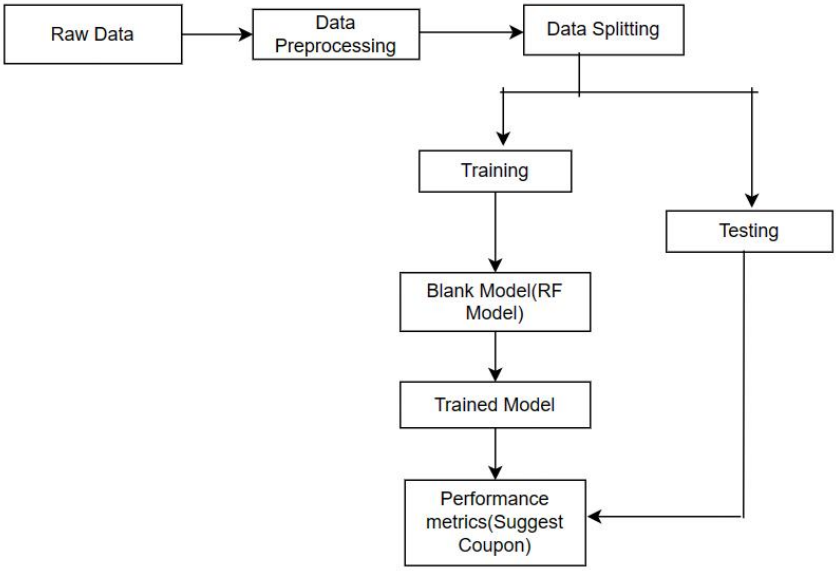

Fig:3 Coupon Recommendation Model

SYSTEM MODULES

Implementing data curation and quality evaluation for a Machine Learning-based cyber intrusion detection project involves a series of practical steps.

Data Collection:

Data collection is a fundamental process in research and analysis, involving the systematic gathering of raw information from various sources. This process serves diverse purposes, from understanding trends and patterns to informing decision-making. Researchers employ a range of methods, including surveys, interviews, observations, and data mining, tailored to their specific objectives and the nature of the data sought. Ethical considerations plays a critical role, necessitating adherence to guidelines to protect participant privacy and rights. Quality assurance measures are implemented to ensure the reliability and validity of the collected data, encompassing meticulous planning, training of personnel, and ongoing monitoring. Once collected, data undergoes organization and management, readying it for analysis where it is transformed into valuable insights to inform further actions or research directions. Thus, data collection forms the bedrock of evidence-based inquiry and decision-making processes across diverse fields and disciplines.

Data Pre-processing:

Data preprocessing is crucial step in the data analysis pipeline, involving the transformation and manipulation of raw data to make it suitable for analysis. This process encompasses several tasks aimed at enhancing the quality and usability of the data.Common preprocessing steps include handling missing values, removing duplicates, and dealing with outliers to ensure the integrity of the dataset. Additionally,

data may be normalized or standardized to bring features to a similar scale, facilitating comparison and analysis. Categorical variables are often encoded into numerical representations, while text data may undergo tokenization and vectorization for machine learning applications. Feature selection or dimensionality reduction techniques may be applied to reduce the complexity of the dataset and improve model performance. Throughout this process, careful consideration is given to maintain the integrity and representativeness of the original data while preparing it for analysis. Ultimately, effective data preprocessing lays the foundation for accurate and meaningful insights to be derived from the data analysis process.

Labeling:

Labeling, in the context of data analysis and machine learning, refers to the process of assigning categories or classes to data points. This is typically done to train supervised learning algorithms, where the algorithm learns to associate input data with corresponding output labels. Labeling can be manual, where human annotators assign labels to data points based on their knowledge or observation. Alternatively, labeling can be automated or semi-automated using techniques such as clustering, rule-based methods, or natural language processing for text data. High-quality labeling is essential for the success of supervised learning algorithms, as it directly impacts the accuracy and performance of the resulting models. It's crucial to ensure consistency, correctness, and representativeness in the labeling process to avoid introducing biases or errors that may adversely affect the model's performance. Additionally, labeled data is often split into training and testing sets to evaluate the model's performance on unseen data and prevent overfitting.

Train / Validation / Test Split:

After feature encoding is done, our dataset is ready for the exciting machine learning algorithms. But before we start deciding the algorithm which should be used, it is always advised to split the dataset into 2 or sometimes 3 parts. Machine Learning algorithms, or any algorithm for that matter,has to be first trained on the data distribution available and then validated and tested, before it can be deployed to deal with real-world data.

Training data:

This is the part on which your machine learning algorithms are actuallytrained to build amodel. Model tries to learn the dataset and its various characteristics.

Validation data:

This is the part of the dataset which is used to validate our various model fits. In simpler words, we use validation data to choose and improve our model hyperparameters. The model does not learn the validation set but uses it to get to a better state of hyperparameters.

Test data:

This part of the dataset is used to test our model hypothesis. It is left untouched and unseen until the model and hyperparameters are decided, and only after that the model is applied on the test data to get an accurate measure of how it would perform when deployed on real-world data.

Fig 4.2 discusses about Data splitting.

Data Split into parts:

Split Ratio:

Data is split as per a split ratio which is highly dependent on the type of model we are building and the dataset itself. If our dataset and model are such that a lot of training is required, then we use a larger chunk of the data just for training purposes. If the model has a lot of hyper parameters that can be tuned, then keeping a higher percentage of data for the validation set is advisable. Models with a smaller number of hyperparameters are easy to tune and update, and so we can keep a smaller validation set. In this project the split ratio is 75:25.

Algorithms:

Machine learning is the latest buzzword in our world. We are using machine learning in many sectors like face recognition in our social media, applications like Siri, disease recognition in the healthcare industry, prediction on weather, analyzing customer behaviour, etc. It actually refers to the feeding process to the machine to act intelligently without explicit programming, analysis the previous data to do the prediction of new data. We are also using machine learning to detect malware. Here, we are using supervised machine learning which refers to the events when we have all information labelled or in another way, we can say we know how all information is classified.

Classification and regression are types of supervised learning where we are aware of the classes of classification and values in regression. However, in the case of unsupervised learning, we are not provided historical data so we have to do association or clustering to make classes.

In this project we have considered 4 supervised machine learning algorithms with precision, recall and Area under roc curve (AUC) as metrics and we will compare all the considered algorithm with the help of metrics and finally we predict the efficient model to suggest coupons.

Now, the algorithms we used are described below

1. Logistic Regression:
2. Random Forest
3. XG-Boost
4. K-Means

Results & Analysis

Evaluation Criteria

In this study we construct a system for segmenting and coupon issuance prediction ,we use it to improve the shopping mall revenue and increase the product the brand name.In this method we a)Use K-means clustering techniques in order to group the customers data.b) We predict the coupons based on the labaled data.Finally the system suggest the coupons to customer.

The dataset which we have taken into the consideration ,forms clusters by using K-Means technique which is shown below.

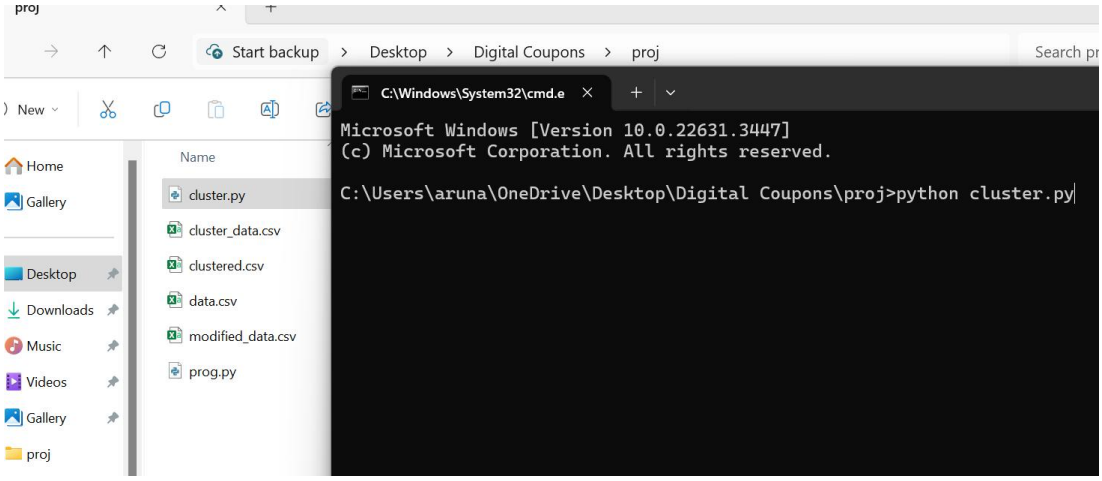

Fig-1: Open digital coupons folder and navigate to command prompt and enter the cluster file to run.

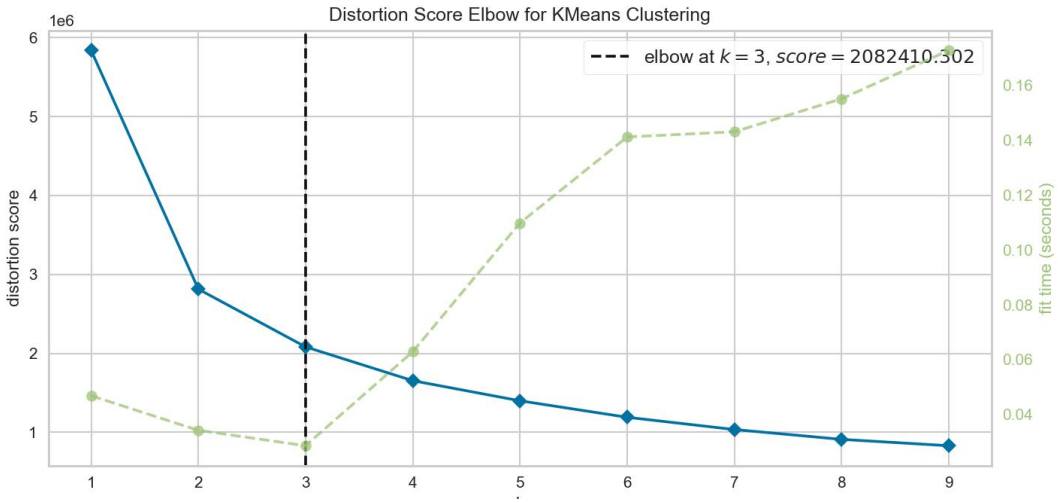

Fig-2 : Data Clusters will be displayed.

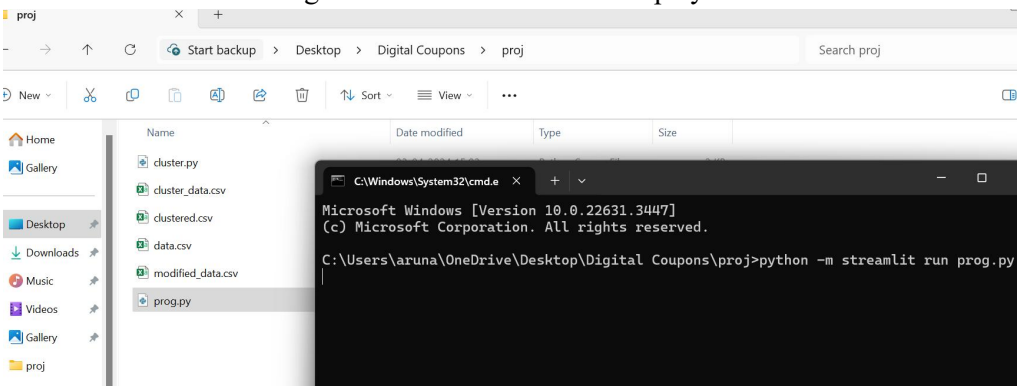

Fig-3: Again open the folder and navigate to command prompt to run prog.py file

Fig-4 : The Home page is displayed first.

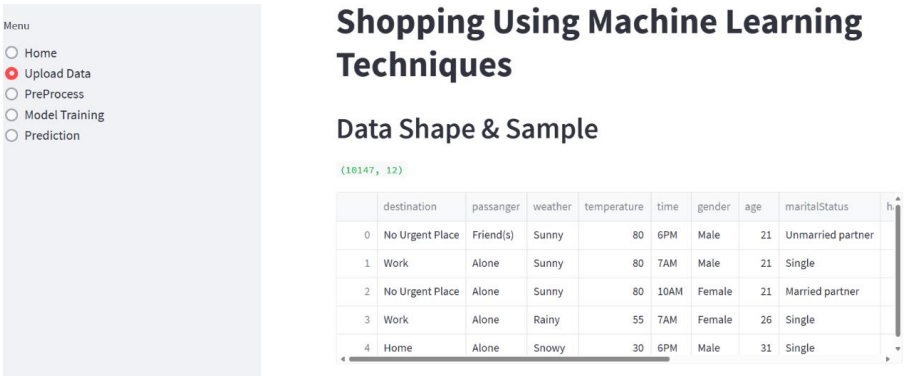

Fig-5: Click on Upload Data.The data set is uploaded to model.

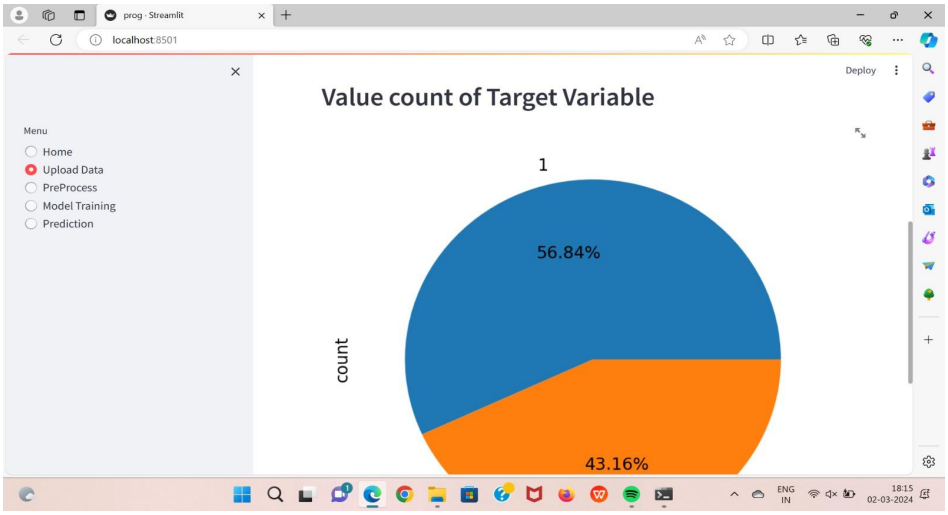

Fig-6: On clicking the upload data ,it displays the value count of target data.

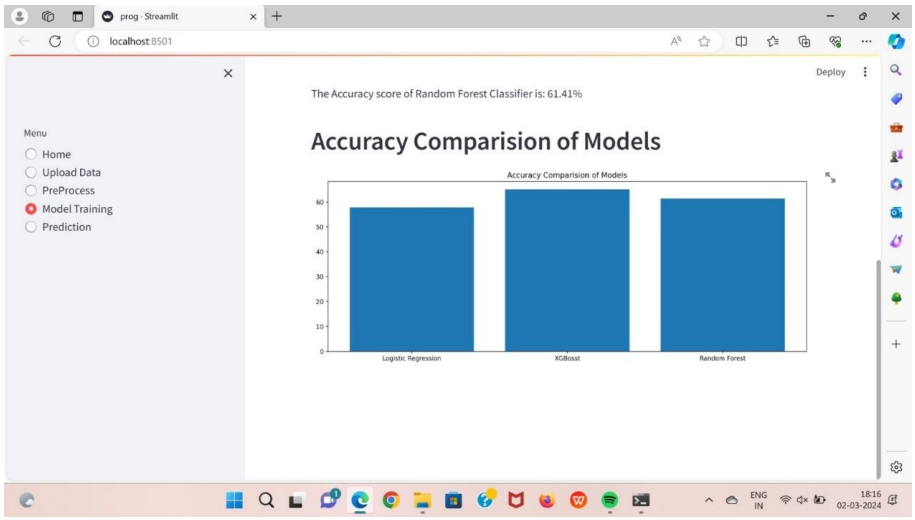

Fig-7: Click on Model training,it displays the accuracy of the models

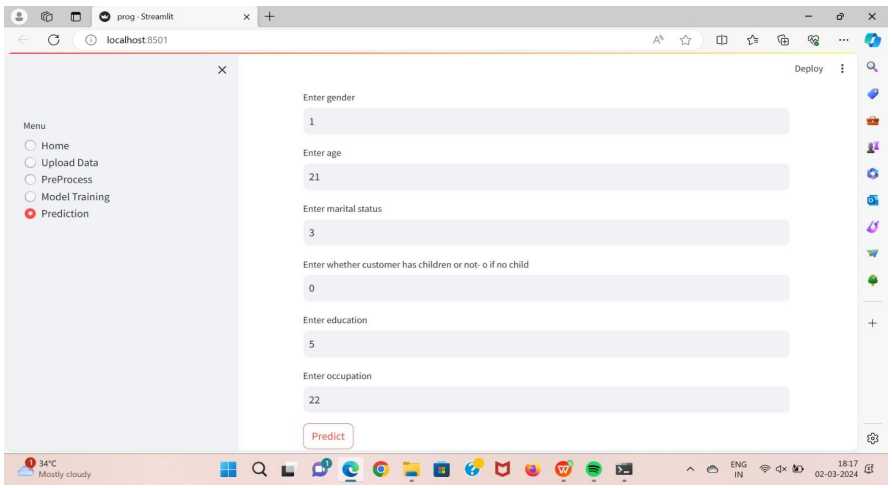

Fig-8: Click on prediction button ,it displays to enter details .

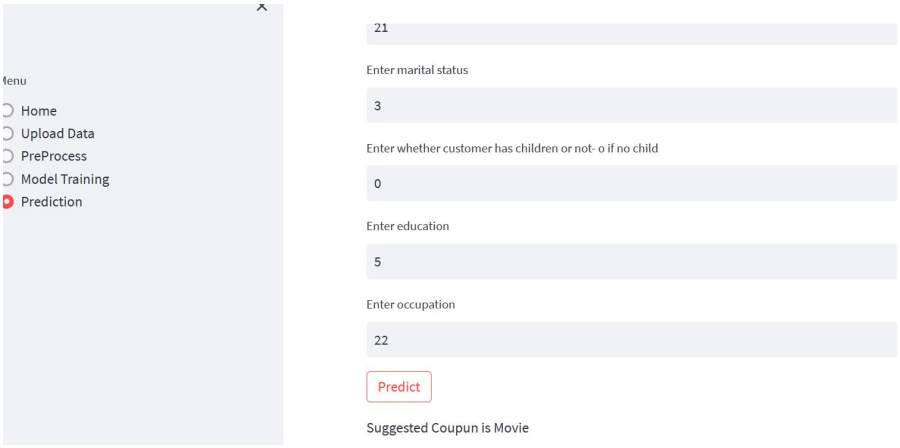

Fig-9: Click on predict button it predicts one of the coupon like paytm voucher,movie,restaurant

CONCLUSION

In this project we have detected, and segmented the customer data using Python, ML models. First ,we have to upload a dataset of customers data which is taken from shopping malls. The system is detected into two-dimensional analysis that are present in ML. After that the next phase which is recommending the coupons to customer which are paytm voucher, movie, restaurant to customer by using the scikit learn framework. We are suggesting the coupon to the customer based on the performance metrics like accuracy.

FUTURE ENHANCEMENTS

After many efforts we had successfully detected and recognized the customer behaviour in order to improve the shopping mall revenue. Recommendation model for customized E-Couponing System has parameters which depends on customer spending and nature metrics which will detect the speed very quickly and efficiently. But, for the real time purpose This system can be enhance with different categories of coupons like transport ,ecommerce ,etc. and also this system can be enhance by providing the large datasets to the model.

References

1. P. Naval and N. Pandey, ‘‘What makes a consumer redeem digital coupons?Behavioral insights from grounded theory approach,’’ J. Promotion Manage., vol. 28, no. 3, pp. 205–238, 2021.
2. M. Sarkar and A. D. Bruyn, ‘‘LSTM response models for direct marketing analytics: Replacing feature engineering with deep learning,’’ J. Interact.Marketing, vol. 53, pp. 80–95, Feb. 2021.
3. Y. Chen, M. Zhou, Z. Zheng, and D. Chen, ‘‘Time-aware smart object recommendation in social Internet of Things,’’ IEEE Internet Things J., vol. 7, no. 3, pp. 2014–2027, Mar. 2020.
4. G. Chandrasekaran and D. J. Hemanth, ‘‘An intelligent framework for online product recommendation using collaborative filtering,’’ in Proc. Int. Conf. Sustain. Expert Syst. (ICSES). Cham, Switzerland: Springer, 2021,p. 249.
5. G. Nie, ‘‘Finding the hidden pattern of credit card holder’s churn: A case of China,’’ in Proc. Int. Conf. Comput. Sci. Cham, Switzerland: Springer, 2009, pp. 561–569..
6. R. M. Gubela, S. Lessmann, and S. Jaroszewicz, ‘‘Response transformation and profit decomposition for revenue uplift modeling,’’ Eur. J. Oper. Res., vol. 283, no. 2, pp. 647–661, Jun. 2020.
7. S.-J. Lim, Z. Lee, L.-N. Kwon, and H.-W. Chun, ‘‘Medical health records based mild cognitive impairment (MCI) prediction for effective dementia care,’’ Int. J. Environ. Res. Public Health, vol. 18, no. 17, p. 9223, Sep. 2021.
8. Y. Yoo, T.-S. Heo, Y. Park, and K. Kim, ‘‘A novel hybrid methodology of measuring sentence similarity,’’ Symmetry, vol. 13, no. 8, p. 1442, Aug. 2021.
9. D. Koehn, S. Lessmann, and M. Schaal, ‘‘Predicting online shopping behaviour from clickstream data using deep learning,’’ Expert Syst. Appl., vol. 150, Jul. 2020, Art. no. 113342.