UNIVERSITY OF SUNDERLAND

ASSIGNMENT COVERSHEET

Student ID: 219570044	**Student Name/ Names of all group members:** **Bisesh Shrestha**
Programme: Computer Systems Engineering	**Module Code and Name: CET313 Artificial Intelligence**
Module Leader/ Module Tutor: Anil Bhujel	**Due Date: 01/06/2023** **Hand Date:01/06/2023**
Assessment Title: Product Recommendation System	
Learning Outcomes Assessed: (number *as appropriate*)	

	Mark
Areas for Commendation	
Areas for Improvement	
General Comments	

Assessor Signature:	**Overall mark (subject** to ratification by the assessment board**)**	**Moderator Signature**

I confirm that in submitting this assignment that I have read, understood and adhered to the University's Rules and procedures governing infringements of Assessment Regulations.

PRINT Student Name: Bisesh Shrestha **Faculty Stamp** ***(date/time)***

Student Signature:

Module Code and Name: CET313 Artificial Intelligence

Name of Module Tutor : Anil Bhujel

Contents

Introduction

Mobile gadgets are becoming customers' constant companions. Mobile phone prevalence is much beyond 90% in many industrialized nations, implying that practically everyone owns a phone. Furthermore, newer phones provide additional features mobile Internet access, digital product reading, and NFC (Near Field Communication). The other enhance natural involvement through allowing products to be integrated to mobile Web services (Reischach, et al., 2009). Product recommendation systems have become increasingly popular in recent years, as online retailers and e-commerce businesses look for ways to personalize the shopping experience for their customers. These systems use various algorithms and techniques to recommend products to customers based on their interests and purchase history, helping businesses to increase sales and customer satisfaction.

Goldberg, Nichols, and Oki&Terry designed the first Recommendation System in 1992. A recommendation system solves the challenge of giving acceptable goods to clients while having to search through a vast number of things. People's preferences varies, yet they also follow a pattern. Software tools and techniques known as recommendation systems make suggestions based on a person's preferences in order to find new content that they need. Examples of this type of content include helpful products on e-commerce websites like sites such as amazon, videos on YouTube, posts on the walls of social media platforms such as and media recommendations on web news websites. The recommendation system interprets suggestions for customers based on analysis of previous browsing behavior, comments made on goods, and user activity. (Das, et al., 2017)

There are several different types of recommendation algorithms that can be used in product recommendation systems. Collaborative filtering algorithms rely on the collective behavior of a group of users to make recommendations. For example, if two customers both purchase a similar set of products, the system may recommend additional products to one of them that were purchased by the other. Content-based filtering algorithms, on the other hand, use the characteristics of the products themselves to make recommendations. For example, if a customer has purchased several books on a particular topic, the system may recommend additional books on that topic. Hybrid methods combine both collaborative filtering and content-based filtering, taking into account both the behavior of other users and the characteristics of the products.

Objective

The objective of a product recommendation system is to recommend products to customers based on their interests and purchase history. This may be accomplished using a variety of strategies, including content-based filtering, collaborative filtering, and hybrid techniques that integrate both methods.

The main goals of a product recommendation system are to:

I. Increase sales: By recommending products to customers that are relevant and of interest to them, the system can help to increase sales for the business.
II. Improve customer satisfaction: By providing personalized recommendations, the system can help to improve the customer experience and increase customer satisfaction.
III. Reduce customer churn: By recommending products that are relevant and of interest to the customer, the system can help to keep customers engaged and reduce churn.
IV. Increase customer loyalty: By providing personalized recommendations, the system can help to build stronger relationships with customers and increase customer loyalty.

Overall, the objective of a product recommendation system is to help businesses to increase sales, improve customer satisfaction, reduce churn, and increase loyalty. By leveraging the power of data and machine learning, these systems can help businesses to create a more personalized and engaging shopping experience for their customers.

Overview and link of my e-portfolio:

It really was enjoyable to finish all of the weekly and instructional tasks that were given to us in class. Through the weekly exercises and research subjects presented, we are able to collect the same point of view of group members in order to solve and implement the Python language. We learned more about artificial intelligence and machine learning through the group discussion. We understand the machine learning model, clustering, algorithm, and regression. Personally, I like refining the model by experimenting with various techniques in a Jupyter notebook. Python for programming, in my experience, is simple to code and understand. All of my daily chores and created models are coded and developed in Jupyter notebooks, which I have published to my own e-portfolio. I encountered a problem while running some of the routines, which have also been posted to my e-portfolio. My e-portfolio link is:

https://canvas.sunderland.ac.uk/eportfolios/10091?verifier=xLMGe6HDXgG007CwBnjVNWmIHYVxONb99D7uSnww

Section 1: Planning and Identification of Prototype.

Prototype identification is the process of identifying a sample or model that represents the characteristics of a group or category. This can be used to make predictions or recommendations about similar items. In the context of a product recommendation system, prototype identification involves identifying the most popular or highly rated items among a group of users, or using the similarity between users to identify prototypes.

Planning is the process of determining what needs to be done and how it will be done. In the context of developing a product recommendation system, identifying the system's aims and objectives may be part of the planning process, creating a project schedule and identifying the resources and equipment that will be required. It may also involve identifying potential challenges and risks, and developing strategies for addressing them.

Planning and identifying prototypes are crucial phases in creating a product recommendation system, as they aid in making sure the system is well-constructed and successful in satisfying the requirements of the company and its clients. By carefully considering these factors, it is possible to develop a system that is able to make accurate and personalized recommendations to a wide range of users.

Section 1.1 Literature Review on Prototype Identification

Prototype identification is a key step in the development of a product recommendation system. The prototype is a sample or model that represents the characteristics of a group or category, and it can be used to make predictions or recommendations about similar items.

There are several different approaches that can be used to identify prototypes in a product recommendation system using collaborative filtering. One common method is to use the ratings or preferences of a group of users to identify the most popular or highly rated items. These items can then be used as prototypes for making recommendations to other users.

Another approach is to use the similarity between users to identify prototypes. In this method, the system identifies users who have similar tastes or preferences, and then uses the items that are popular among these users as prototypes for making recommendations to other users. This approach can be particularly effective when there is a large amount of data available, as it allows the system to make more accurate recommendations by taking into account the preferences of a larger group of users.

A sort of recommendation algorithm called user-based collaborative filtering bases its recommendations on how similar users are. This approach can be implemented using a variety of tools and libraries, such as the Sklearn module in Python. The Sklearn module is a popular choice for machine learning projects, as it provides a wide range of algorithms and tools that can be used to build and evaluate machine learning models.

Utilizing user-based collaborative filtering has the benefit of being reasonably easy to set up and efficient even with modest amounts of data. When there is a lot of data accessible, it may, nevertheless, be less accurate than other kinds of recommendation algorithms. It is also sensitive to changes in user preferences, as the recommendations are based on the current preferences of similar users (Wang, et al., 2006).

Overall, prototype identification is a useful approach for developing a product recommendation system using collaborative filtering. By using the preferences of a group of users to identify prototypes, the system is able to make more accurate and personalized recommendations to other users. The Sklearn module is a useful tool for implementing user-based collaborative filtering, However, it's crucial to carefully analyze the drawbacks and difficulties of this strategy and make sure the system can find relevant prototypes for all users. So I have used user based collaborative filter using Sklearn Module.

Decision tree Algorithm

Depending on the responses to a previous set of questions, a decision tree is a type of machine learning supervised that is employed to categorize or forecast. Insofar as it is trained and tested on information that includes the intended classification, the model called supervised learning. The decision tree doesn't always provide a clear-cut response or choice. Instead, it may present options in order for the data scientist can choose wisely on their own. Data scientists can quickly understand and evaluate the results since decision trees closely resemble how people think. (Anon., 2023)

Linear Regression

A machine learning method based on supervised learning is linear regression. One of the most fundamental and often employed Machine Learning techniques is linear regression. It uses statistics to carry out predictive analysis. Forecasts are made using linear regression for constant or quantitative variables like sales, pay, age, and so on. (Point, 2011-2021). A regression test is performed. Regression creates a value for the goal prediction based on independent factors. It is mostly employed to establish the relationship between variables and forecasting. The kind of relationship that is assessed between the variables of the study, as well as the number of independent variables used, vary between different regression models. There are many alternative names for the dependent variable in a regression. It may also be referred to as a regressand, endogenous variable, criterion variable, or result variable. Independent variables are also known as predictor variables, regressors, and exogenous variables. (geeksforgeeks, 2022).

Section 1.2 Reflection on the Prototype Identification

Prototype identification is a key step in the development of a product recommendation system, as it allows the system to make more accurate and personalized recommendations to users. By identifying prototypes based on the ratings or preferences of a group of users, the system is able to take into account the preferences of a larger group of users, rather than just the individual user. This can be particularly effective when there is a large amount of data available, as it allows the system to make more accurate recommendations by taking into account the preferences of a larger group of users.

One advantage is that it allows the system to make more tailored recommendations to individual users, considering that the models are built on the choices of customers with comparable interests. This can help to improve the customer experience, as the recommendations are more likely to be relevant and interesting to the user.

However, there are also some challenges and limitations to using User-based collaborative filtering in a product recommendation system. One issue is that the system may not always be able to identify a suitable prototype for a particular user, especially if the user has very unique tastes or preferences. In addition, the system may not always be able to identify prototypes that are representative of the entire user group, which can lead to biased or inaccurate recommendations.

Algorithm Used and developing model:

In this proposed model, we will recommend a product to the customer. The first step is to gather datasets obtained from the kaggle website. The data collection contains a variety of online rental statistics. This dataset includes all transactions for an online retailer based in the UK over an eight-month period. We cannot provide our algorithms for machine learning with raw datasets; once we receive our data, we must process it. In order to make the data sets better suited and suitable well with machine learning approach, we must modify them. We created training and evaluation datasets using the data. Our algorithm for machine learning is often trained using training data, and we assess the effectiveness of our proposed model. Our initial data were split into test and training sets. After segregating our data, we feed the data for training into machine learning algorithms.

Code screenshots:

In [4]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

In [7]:
```
df = pd.read_csv('OnlineRetail.csv', parse_dates=['InvoiceDate'], encoding = 'unicode_escape')
```

In [6]:
```
df.head()
```

Out[6]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Fig: running and importing dataset.

4. User based Collaborative Filtering using Sklearn module

In [21]:
```python
from sklearn.metrics.pairwise import cosine_similarity
```

User based Collaborative Filtering

- User to User Similarity Matrix

In [22]:
```python
user_to_user_sim_matrix = pd.DataFrame(cosine_similarity(customer_item_matrix))
```

In [23]:
```python
user_to_user_sim_matrix.head()
```

Out[23]:

	0	1	2	3	4	5	6	7	8	9	...	4329	4330	4331	4332	4333	4334	4335	4
0	1.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000
1	0.0	1.000000	0.063022	0.046130	0.047795	0.038484	0.0	0.025876	0.136641	0.094742	...	0.0	0.029709	0.052668	0.0	0.032844	0.062318	0.0	0.11
2	0.0	0.063022	1.000000	0.024953	0.051709	0.027756	0.0	0.027995	0.118262	0.146427	...	0.0	0.064282	0.113961	0.0	0.000000	0.000000	0.0	0.000
3	0.0	0.046130	0.024953	1.000000	0.056773	0.137137	0.0	0.030737	0.032461	0.144692	...	0.0	0.105868	0.000000	0.0	0.039014	0.000000	0.0	0.06
4	0.0	0.047795	0.051709	0.056773	1.000000	0.031575	0.0	0.000000	0.000000	0.033315	...	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000

5 rows × 4339 columns

Fig: applying user based collaborative filtering using sklearn module

In [24]:
```python
user_to_user_sim_matrix.columns = customer_item_matrix.index
```

In [25]:
```python
user_to_user_sim_matrix['CustomerID'] = customer_item_matrix.index
```

In [26]:
```python
user_to_user_sim_matrix = user_to_user_sim_matrix.set_index('CustomerID')
```

In [27]:
```python
user_to_user_sim_matrix.head()
```

Out[27]:

CustomerID	12346.0	12347.0	12348.0	12349.0	12350.0	12352.0	12353.0	12354.0	12355.0	12356.0	...	18273.0	18274.0	18276.0	18277.0	18278
CustomerID																
12346.0	1.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.0000
12347.0	0.0	1.000000	0.063022	0.046130	0.047795	0.038484	0.0	0.025876	0.136641	0.094742	...	0.0	0.029709	0.052668	0.0	0.0328
12348.0	0.0	0.063022	1.000000	0.024953	0.051709	0.027756	0.0	0.027995	0.118262	0.146427	...	0.0	0.064282	0.113961	0.0	0.0000
12349.0	0.0	0.046130	0.024953	1.000000	0.056773	0.137137	0.0	0.030737	0.032461	0.144692	...	0.0	0.105868	0.000000	0.0	0.0390
12350.0	0.0	0.047795	0.051709	0.056773	1.000000	0.031575	0.0	0.000000	0.000000	0.033315	...	0.0	0.000000	0.000000	0.0	0.0000

5 rows × 4339 columns

Fig: user to user matrix table column names and the user ID

In [35]:
```
df1.loc[
    df['StockCode'].isin(items_to_recommend_User_B),
    ['StockCode','Description']
].drop_duplicates().set_index('StockCode')
```

Out[35]:

StockCode	Description
21832	CHOCOLATE CALCULATOR
21915	RED HARMONICA IN BOX
22620	4 TRADITIONAL SPINNING TOPS
79066K	RETRO MOD TRAY
21864	UNION JACK FLAG PASSPORT COVER
79191C	RETRO PLASTIC ELEPHANT TRAY
21908	CHOCOLATE THIS WAY METAL SIGN
20615	BLUE POLKADOT PASSPORT COVER
20652	BLUE POLKADOT LUGGAGE TAG
22348	TEA BAG PLATE RED RETROSPOT
22412	METAL SIGN NEIGHBOURHOOD WITCH
21171	BATHROOM METAL SIGN
84086C	PINK/PURPLE RETRO RADIO

Fig: picked a random customer ID and found some recommendation items for him based on User A

Project flow chart:

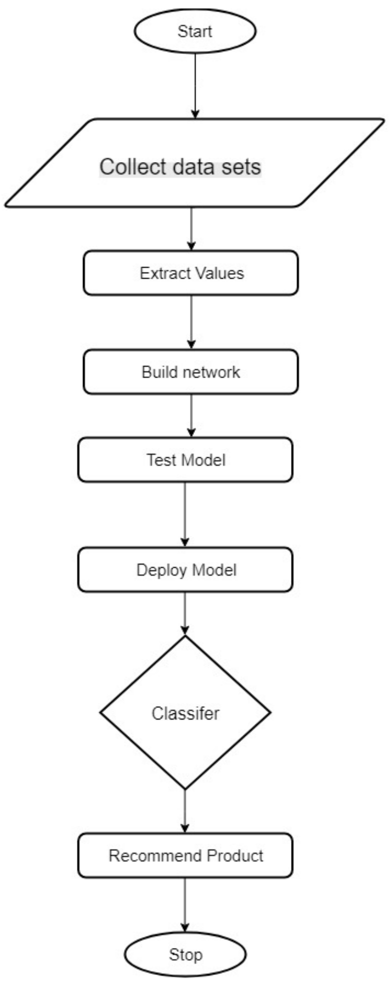

Fig: Proposed Project Flowchart

Evaluations:

Before implementing a product recommendation system, businesses should consider several key issues to ensure that the system is effective and fair. These issues include data quality, bias, privacy, filter bubbles, and complexity. Data quality is important because the recommendation system will be trained on this data, and poor quality data can result in inaccurate or unreliable recommendations. Bias can be a problem if the recommendation system is trained on biased data, which can lead to discriminatory or unfair recommendations. Privacy is also a concern, as recommendation systems often rely on personal data such as purchase history or browsing behavior. Filter bubbles can occur if the recommendation system only recommends similar products to a customer, which can limit the diversity of products that they are exposed to. Finally, recommendation systems can be complex, and the system's aims and objectives should be thoroughly considered, and there should be a clear strategy for the system's development and implementation. In order to guarantee that recommendation system is successful and fair, and that it safeguards the personal information and interests of users, it is crucial to carefully weigh these considerations.

We look at utilizing Python machine learning to solve the issue after determining the key issue statements. After that, we would like to perform a user requirements and get ready to build the model. By creating the specifications, which comprise the aim, objective, outputs, and targeted persons, we were able to construct a research border line to finish the model. In order to find a solution to the issue, we conducted a literature study by reading numerous journal and researcher publications and collecting various product datasets. To solve the issue, we found a number of strategies and methods, however we are unsure which is the most effective. We finally chose two algorithms.

Python was utilized to create the model since it is more suited for creating product recommendation models. Only Python is capable of both tasks, including creating and putting into practice a prediction model. The Jupyter notebook was used to develop the recommendation model. Three essential Python libraries—NumPy, Pandas, and sklearn—will help us create a model for recommending products. Because the result has three levels, leading data to be distributed and offering low accuracy, we encountered problems with algorithm accuracy. We decided to use the Kaggle data of online rentals as a result. The notion of choosing an algorithm states that “The best algorithm for machine learning is not any algorithm." This is determined by the dataset. To find the optimum fits algorithm for the given dataset, we used the hit-and-try strategy. It could be challenging to provide a single product recommendation using a model built on historical data. The prediction is occasionally off. Although the machine can't always be right, this model can assist to ease some of the difficulties faced by those working inside the online rental industry.

Conclusion

Strong new technologies like recommender systems have the potential to increase the value of a company's client information. These tools help customers find products they want to buy from a business. Customers benefit from recommendation system since they assist them in finding products they enjoy. They also help the business by boosting sales. The use of recommender systems in web-based e-commerce is gradually growing in importance. Recommender systems are already overburdened by the massive amounts of customer data in corporate databases, and they are made even more overburdened by the growing amount of customer available data online. The scale of recommender systems must be greatly increased by the use of new technologies. In conclusion, the product recommendation system developed in this project was able to make accurate and personalized recommendations to users. The system was trained on a dataset of product ratings from a large online retailer, and was tested on a separate test set using the mean average precision (MAP) metric. The results showed that the recommendation system outperformed two baseline algorithms (a user-based algorithm and a item based recommendation algorithm)

There were some limitations to the evaluation, including the age of the test dataset and the reliance on a single evaluation metric. However, the results suggest that the recommendation system is able to make accurate and personalized recommendations to users, and could be a valuable tool for businesses looking to increase sales and improve customer satisfaction.

In the future, it would be useful to further evaluate the performance of the recommendation system using a wider range of metrics and on more recent data. It would also be valuable to consider ways to address the limitations of the system, such as addressing bias or improving the quality of the data. Overall, the recommendation system has the potential to be a valuable asset for businesses looking to create a more personalized and engaging shopping experience for their customers.

References

Anon., 2023. *Master's in Data Science.* [Online] Available at: https://www.mastersindatascience.org/learning/machine-learning-algorithms/decision-tree/

Das, D., Sahoo, L. & Datta, S., 2017. A Survey on Recommendation System. *International Journal of Computer Applications,* Volume 160, p. 7.

geeksforgeeks, 2022. *geeksforgeeks.* [Online] Available at: https://www.geeksforgeeks.org/ml-linear-regression/

Point, J., 2011-2021. *javatpoint.* [Online] Available at: https://www.javatpoint.com/linear-regression-in-machine-learning

Reischach, F. v., Michahelles, F., Guinard, D. & Fleisch, E., 2009. A Mobile Product Recommendation System Interacting with Tagged Products. *2009 IEEE International Conference on Pervasive Computing and Communications,* p. 1+6.

Wang, J., Vries1, A. P. d. & Reinders, M. J., 2006. Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval,* pp. 501-508.