INTRAINZ INNOVATION PRIVATE LIMITED

INTERNSHIP PROGRAM REPORT

Submitted by

AKSHARA SRI (201501005)

in partial fulfillment for the award of degree of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND

MACHINE LEARNING

RAJALAKSHMI ENGINEERING COLLEGE,

ANNA UNIVERSITY,CHENNAI –600025

2023-2024

INDUSTRY DETAILS

NAME OF THE INDUSTRY WITH ADDRESS	Intrainz Intrainz Innovation Private Limited - Evoma Business Centre, Old Madras Rd, Battarahalli, Bengaluru, Karnataka 560049
INTERN COURSE NAME	Data Science using Python
INTERNSHIP DURATION	1st February 2023 to 31st March 2023 (8 weeks)

CHAPTER 1
COMPANY PROFILE

Intrainz is a leading provider of comprehensive training and internship programs designed to prepare university students for successful entry into the corporate world. Their primary objective is to equip students with specialized skill sets in high-demand domains, enabling them to become industry-ready professionals. Intrainz Innovation Private Limited is an unlisted private company incorporated on 22 September 2022. It is classified as a private limited company and is located in , Karnataka. Its authorized share capital is INR 1.00 lac and the total paid-up capital is INR 10,000.00 . The current status of Intrainz Innovation Private Limited is - Active. Intrainz Innovation Private Limited has three directors - Batchu Rohith, Vishnu P Nair, and others. The Corporate Identification Number (CIN) of Intrainz Innovation Private Limited is U80903KA2022PTC166349. The registered office of Intrainz Innovation Private Limited is at No 47, K No 66 190 47, Teachers Colony, 6th Main Road Bangalore Bangalore, Karnataka.

Mission:

At Intrainz, The mission is to bridge the gap between academia and industry by offering a unique blend of in-depth industrial training and certified live internship projects. The company strives to empower students with practical experience and knowledge, equipping them with the necessary tools to secure promising careers in their respective fields.

Vision:

Intrainz envisions a future where every student is empowered with the necessary skills and experience to excel in their chosen careers and aims to be the preferred choice for students seeking industry readiness, recognized for their commitment to excellence, and admired for a stronger ties with the corporate world.

Services:

1) In-depth Industrial Training Program:

They offer a meticulously designed training curriculum that provides us with a deep understanding of industry practices, trends, and technologies. The expert trainers, who possess extensive industry experience, guide the students through hands-on sessions and interactive workshops.

2) Certified Live Internship Projects:

Intrainz provides valuable opportunities to work on real-time projects in collaboration with established industry partners. These internships offer the chance to apply their newly acquired skills in a professional setting, gaining practical experience and exposure.

3) Skill Development in High-Demand Domains:

They focus on domains that have a high market demand, ensuring that students develop skill sets that align with the needs of the industry. Comprehensive training and internship programs, enables us to enhance their expertise in areas such as technology, marketing, finance, human resources, and more.

4)Industry Connections:

Intrainz has established strong partnerships with leading companies across various sectors. These connections allow us to provide students with networking opportunities, industry insights, and potential employment prospects.

Benefits:

- Practical Experience: The programs offer students hands-on experience, simulating real-life work scenarios and challenges, which prepare them for the demands of the corporate world.
- Industry-Relevant Skills: Intrainz equips students with up-to-date skills and knowledge required by the industry, enhancing their employability.
- Networking Opportunities: Through an extensive network of industry partners, students can build connections with professionals and potential employers.
- Certification: Upon successful completion of programs, students receive certifications that validate their industry-relevant skills and enhance their resumes.

CHAPTER 2

INTERN PROJECT WORK

2.1 PROJECT DEFINITION

The two months internship from 01/02/23 to 31/03/23 in this project will work on different types of recommendation systems. For this, they have attached a dataset containing information about recommendation systems for online retail data, so that we can understand what type of product can be recommended. Also, they are providing a dataset from Kaggle, which contains historical information about online retail data which can be used to detect which product is highly recommended.

2.2 MAIN OBJECTIVE

The objective of the code provided is to develop a recommendation system for a retail dataset. The system aims to provide personalized recommendations to users based on their estimated rating and user ratings. Additionally, it generates global recommendations, country-wise recommendations, and month-wise recommendations to enhance the user experience and assist users in finding relevant and appealing items. The code performs the following tasks:

3.2.1 Data Preprocessing:

The code reads the retail dataset from an Excel file and preprocesses it by dropping missing values and filtering out non-positive quantities.

3.2.2 Exploratory Data Analysis:

The code analyzes the dataset to provide insights into the available countries and items. It displays a table showing the count of transactions by country and identifies and displays the most popular items globally and country-wise.

3.2.3 Association Rule Mining:

The code utilizes the Apriori algorithm to perform association rule mining on monthly transaction data. It identifies frequent itemsets and association rules based on minimum support and lift thresholds.

3.2.4 Visualization:

The code visualizes the most popular items globally using a histogram plot. It showcases the count of repeated products each month.

3.2.5 Collaborative Filtering:

The code employs the Surprise library's SVD algorithm to build a collaborative filtering recommendation model. It trains the model using the retail dataset and generates predictions for user-item interactions.

3.2.6 Top-N Recommendations: The code generates the top-N recommendations for each user based on the collaborative filtering model. It creates a dictionary containing the recommended items for each user, sorts them based on estimated ratings, and displays the most recommended items overall, considering the number of recommendations.

The objective of this code is to lay the foundation for a retail recommendation system by performing data preprocessing, exploratory data analysis, association rule mining, collaborative filtering, and generating top-N recommendations. These steps contribute to enhancing the user experience, increasing user engagement, and facilitating item discovery for the users of the retail platform.

2.3 NEED FOR PROPOSED SYSTEM

The need for recommendation systems in data science varies globally, monthly, and country-wise based on several factors.

GLOBAL NEED:

a. Personalization: With an ever-increasing amount of data being generated, users expect personalized experiences. Recommendation systems help analyze user preferences, behaviors, and historical data to provide tailored recommendations, enhancing user satisfaction.

b. Information Overload: The digital era has led to an abundance of choices and information. Recommendation systems help alleviate information overload by suggesting relevant items, products, or content to users, saving time and effort.

c. Business Competitiveness: Many businesses rely on recommendation systems to enhance customer engagement and boost sales. By leveraging customer data and behavior, companies can deliver targeted recommendations, resulting in improved customer satisfaction and increased revenue.

MONTHLY NEED:

a. Seasonal Trends: User preferences and behaviors often change based on seasonal factors, such as holidays, festivals, or events. Recommendation systems need to adapt to these changes to provide timely and relevant recommendations.

b. Content Updates: In various domains like media streaming, e-commerce, or news platforms, new content is frequently released or updated. Recommendation systems must continuously analyze and incorporate the latest content to provide up-to-date recommendations.

COUNTRY-WISE NEED

a. Cultural Differences: Different countries have unique cultural preferences and consumption patterns. Recommendation systems must consider these differences to provide culturally relevant recommendations and ensure user satisfaction.

b. Localized Content: In certain countries, there is a significant demand for localized content. Recommendation systems need to be equipped with the ability to understand and recommend country-specific items or content, tailored to the local audience.

c. Market Dynamics: Market dynamics can vary from country to country, influenced by factors such as economic conditions, industry trends, and user behavior. Recommendation systems must adapt to these dynamics to provide accurate and effective recommendations in each specific market.

2.4 METHODOLOGY

The proposed methodology for developing a recommendation system for a retail dataset involves preprocessing the data, performing exploratory data analysis, association rule mining, visualization, collaborative filtering, and generating top-N recommendations. This methodology is scalable, flexible, and reliable, and can be used to generate personalized recommendations for users, enhance the user experience, increase user engagement, and facilitate item discovery.

1. Data Preprocessing

The first step is to preprocess the data by dropping missing values and filtering out non-positive quantities. This ensures that the data is clean and consistent, which is necessary for the analysis to produce accurate results.

2. Exploratory Data Analysis

The next step is to perform exploratory data analysis (EDA) on the dataset. This involves analyzing the data to identify trends and patterns. For example, the code can be used to identify the most popular items globally and by country. This information can be used to create targeted recommendations.

3. Association Rule Mining

Association rule mining is a technique that can be used to identify items that are frequently purchased together. This information can be used to generate recommendations for users who have purchased similar items in the past. For example, if a user has purchased a book about cooking, the code can recommend other cooking-related books or kitchen utensils.

4. Visualization

The code can be used to visualize the data to make it easier to understand. For example, the most popular items globally can be visualized using a histogram plot. This makes it easy to see which items are most popular and how they compare to each other.

5. Collaborative Filtering

Collaborative filtering is a technique that can be used to recommend items to users based on their past ratings. This is a personalized approach that can help users discover new items that they are likely to enjoy. The code can be used to train a collaborative filtering model using the retail dataset. This model can then be used to generate recommendations for users.

6. Top-N Recommendations

The final step is to generate top-N recommendations for each user. This involves sorting the recommended items for each user based on estimated ratings. The most recommended items overall, considering the number of recommendations, can also be displayed.

This proposed methodology is a scalable, flexible, and reliable approach that can be used to develop a recommendation system for a retail dataset. It can be used to generate personalized recommendations for users, enhance the user experience, increase user engagement, and facilitate item discove

2.5 PLATFORM

The platform used in this code is Streamlit. Streamlit is a Python library that is used to create web applications. It is a popular choice for data science and machine learning projects because it is easy to use and can be deployed quickly.

The code for this project is written in Python and uses the following libraries:

- Pandas for data manipulation
- NumPy for numerical computing
- Matplotlib for visualization
- Surprise for collaborative filtering
- Streamlit for web application development

The code can be run on any platform that has Python installed. However, it is recommended to use a cloud platform such as Google Cloud Platform or Amazon Web Services to deploy the application.

Here are some of the benefits of using Streamlit for this project:

- Ease of use: Streamlit is a very easy-to-use library. The code for this project is relatively short and straightforward.
- Quick deployment: Streamlit applications can be deployed quickly and easily. This makes it a good choice for projects that need to be deployed quickly.
- Visualization: Streamlit makes it easy to visualize data. This is important for this project because it allows users to see the results of the analysis.

In addition to the benefits mentioned above, Streamlit also has a number of other features that make it a good choice for this project. For example, Streamlit is,

- Interactive: Streamlit applications are interactive, which means that users can interact with the data and the results of the analysis. This makes it a more engaging experience for users.
- Extensible: Streamlit is extensible, which means that it can be customized to meet the specific needs of a project. This makes it a flexible platform that can be used for a variety of different projects.

Overall, Streamlit is a powerful platform that can be used to create interactive web applications. It is a good choice for this project because it is easy to use, quick to deploy, and interactive.

2.6 ARCHITECTURE DIAGRAM FOR THE PROPOSED MODEL

The flowchart of the proposed model is shown in Figure 1

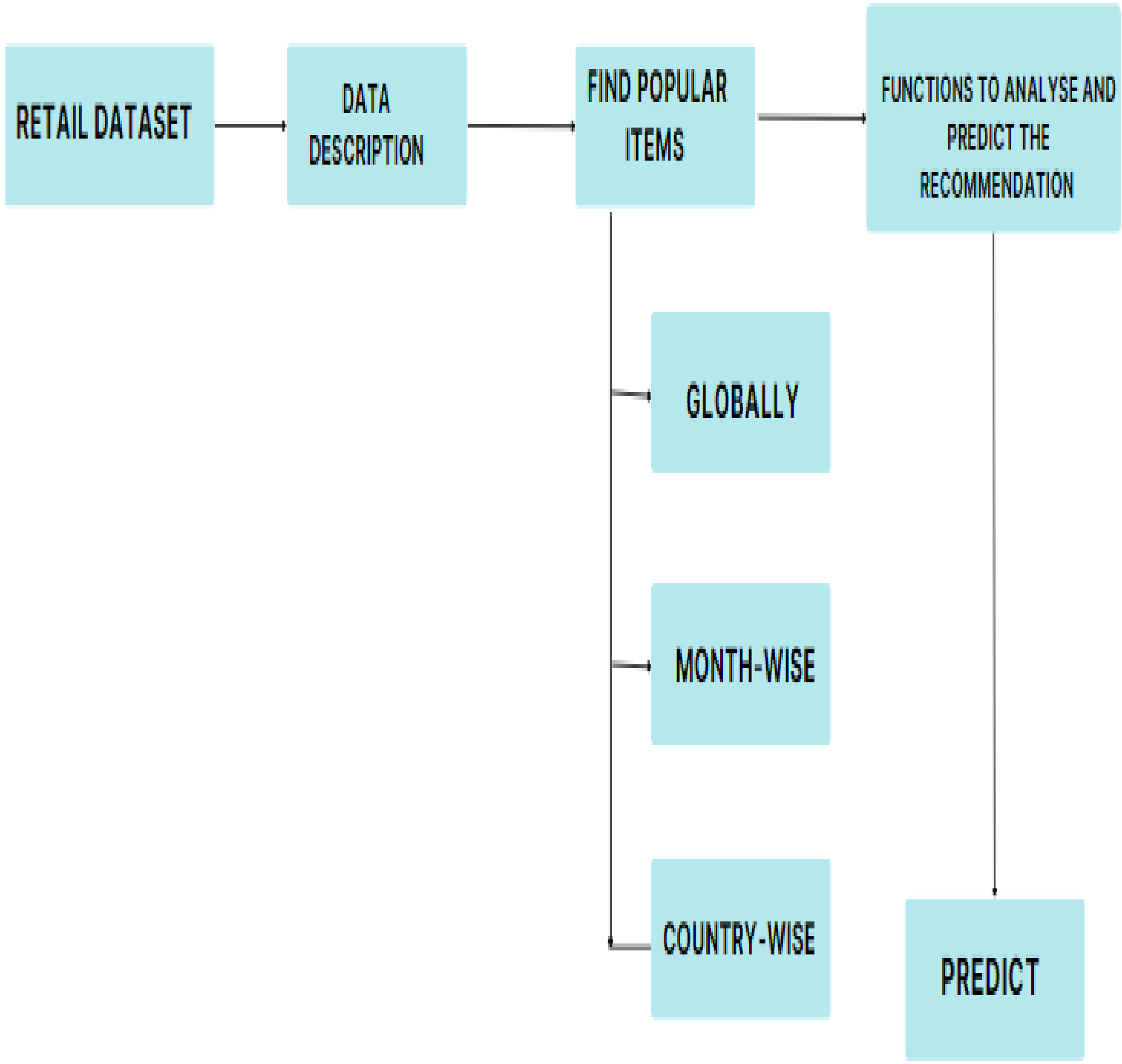


Fig 1: Architecture diagram for the proposed model

2.7 CODE

 # IMPORT THE NECESSARY LIBRARY

 # LOAD THE DATASET

```
import pandas as pd
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from surprise import SVD
from surprise import Dataset
from surprise import Reader
from mlxtend.frequent_patterns import apriori, association_rules
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
df = pd.read_excel('/kaggle/input/onlineretail/OnlineReta
df = df.loc[df['Quantity'] > 0]
```

#LIST OF COUNTRIES AVAILABLE IN THE DATASET

```
df.value_counts(['InvoiceDate'])
df.dropna(inplace=True)
```

#MOST POPULAR ITEMS GLOBALLY

```
df1 = df['Description']
duplicates = df1.duplicated()
```

```
print("Number of duplicates:", duplicates.sum())
duplicated_rows = df1[duplicates]
most_common_duplicates = duplicated_rows.value_counts().head(10)
print(f"Most popular items Globally:\n{most_common_duplicates}")
```

 # THE NEXT IS TO FIND THE MMOST POPULAR ITEM AMONG COUNTRIES

```
most_popular_items=df.groupby(['Country'])['Description'].sum().sort_values(ac
ending=False).reset_index()
print(most_popular_items.head(25))
```

#THE BELOW STEP IS TO FIND THE MOST POPULAR ITEMS MONTHLY WISE

```
df['InvoiceDate'] = pd.to_datetime(df['InvoiceDate'])
df['date_new'] = df.InvoiceDate.dt.strftime('%Y-%m')
top_items_monthly = []
for month in df.date_new.unique():
    trans_month = df.loc[df.date_new == month]
    trans_month = (trans_month.groupby(['InvoiceNo', 'Description'])['Quantity']
                   .sum().unstack().reset_index().fillna(0)
                   .set_index('InvoiceNo'))
trans_month[trans_month >= 1] = True
trans_month[trans_month.isna()] = False
import warnings
warnings.filterwarnings('ignore')
frequent_itemsets=apriori(trans_month, min_support=0.03,use_colnames=True)
associationRules=association_rules(frequent_itemsets,metric="lift",min_thresh
```

```
 ld=1)
top_k=associationRules.sort_values(by=['support'],ascending=False).iloc[:10][['a
ntecedents','support']].reset_index(drop=True)
top_items_monthly.append((month, top_k))
pivot_dfs = []
 for i, montly_pairs in enumerate(top_items_monthly):
    month, data = montly_pairs
    inv_map = {k: v for k, v in enumerate(data.antecedents)}
    rows = []
    for index, row in df.loc[(df.date_new == month)].iterrows():
        keys = [inv_map[k] for tup in str(row['Description']).split(',') for k,v in
 inv_map.items() if str(row['Description']) in list(v)]
        for key in keys:
            rows.append([month, key])
    pivot_df = pd.DataFrame(rows, columns=['month','Item'])
    pivot_df.head()
 pivot_dfs.append(pivot_df.pivot_table(values=["Item"],index=["month"],aggfun
 c="count",fill_value=0))
 df['InvoiceDate'] = pd.to_datetime(df['InvoiceDate'])
 df.set_index('InvoiceDate', inplace=True)
monthly_df = df.groupby(pd.Grouper(freq='M'))['Description'].value_counts()
 repeated_products = monthly_df[monthly_df > 1]
for month, counts in repeated_products.groupby(level=0):
    print(f"Month: {month.strftime('%B %Y')}")
    print(counts)
```

#THE FOLLOWING STEPS WILL BE IMPLEMENTING THE RECOMMENDATION PREDICTOR USING USER RATINGS AND ESTIMATED RATINGS

```
df1 = df[['CustomerID', 'Description','StockCode', 'Quantity']]
ratings_matrix=df1.pivot_table(index=['CustomerID'],columns=['StockCode'],
values='Quantity', fill_value=0)
algo = SVD()
reader = Reader(rating_scale=(1, 5))
surprise_data = Dataset.load_from_df(df1[['CustomerID', 'StockCode',
'Quantity']], reader)
trainset = surprise_data.build_full_trainset()
testset = trainset.build_anti_testset()
algo.fit(trainset)
predictions = algo.test(testset)
top_n = {}
for uid, iid, true_r, est, _ in predictions:
    if uid not in top_n.keys():
        top_n[uid] = [(iid, est)]
    else:
        top_n[uid].append((iid, est))
df1.dropna(subset=["StockCode", "Description"], inplace=True)

descriptions = df1.groupby("StockCode").first()["Description"]
desc_dict = descriptions.to_dict()
```

#THE BELOW CODE DISPLAYS THE PREDICTIONS BASED ON THE RECOMMENDED ITEMS

```
global_top_n = {}
for uid, user_ratings in top_n.items():
    user_ratings.sort(key=lambda x: x[1], reverse=True)
```

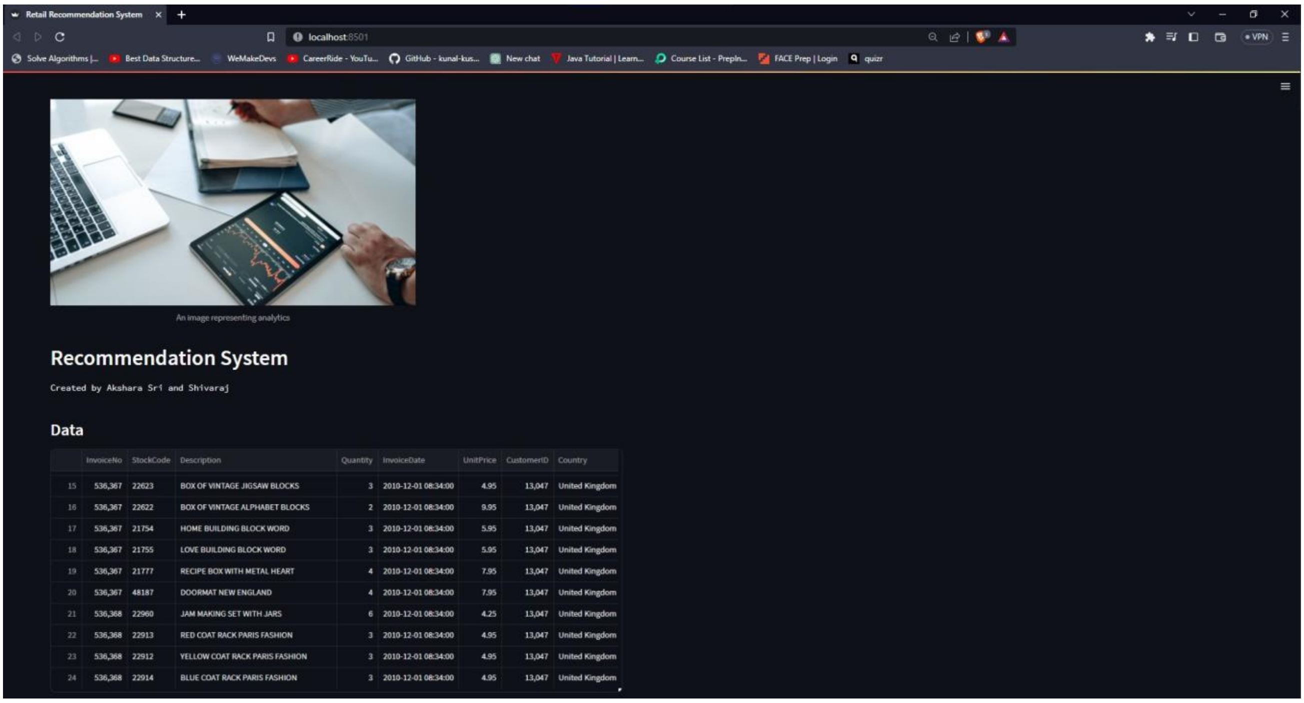
```
    global_top_n[uid] = []
    for iid, est_rating in user_ratings[:5]:
        if iid in desc_dict:
            global_top_n[uid].append((iid, desc_dict[iid]))
all_items = [iid for uid in global_top_n for iid, desc in global_top_n[uid]]
item_counts = {iid: all_items.count(iid) for iid in set(all_items)}

print("Most Recommended Items (in number of recommendations):")
for item, count in sorted(item_counts.items(), key=lambda x: x[1],
reverse=True):
    if item in desc_dict:
        desc = desc_dict[item]
        print("\t", "Item ID:", item, "(\"" + str(desc) + "\")", f"recommended {count}
times")
```

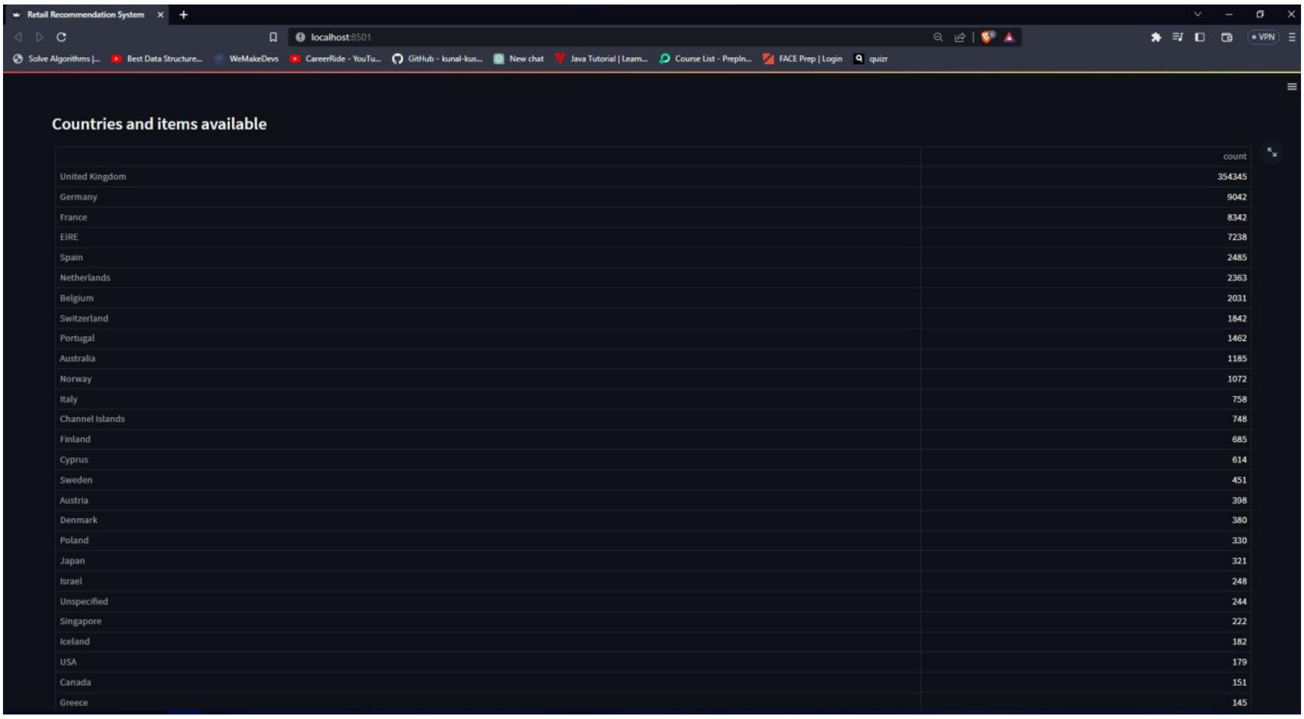
CHAPTER 3

RESULTS AND DISCUSSIONS

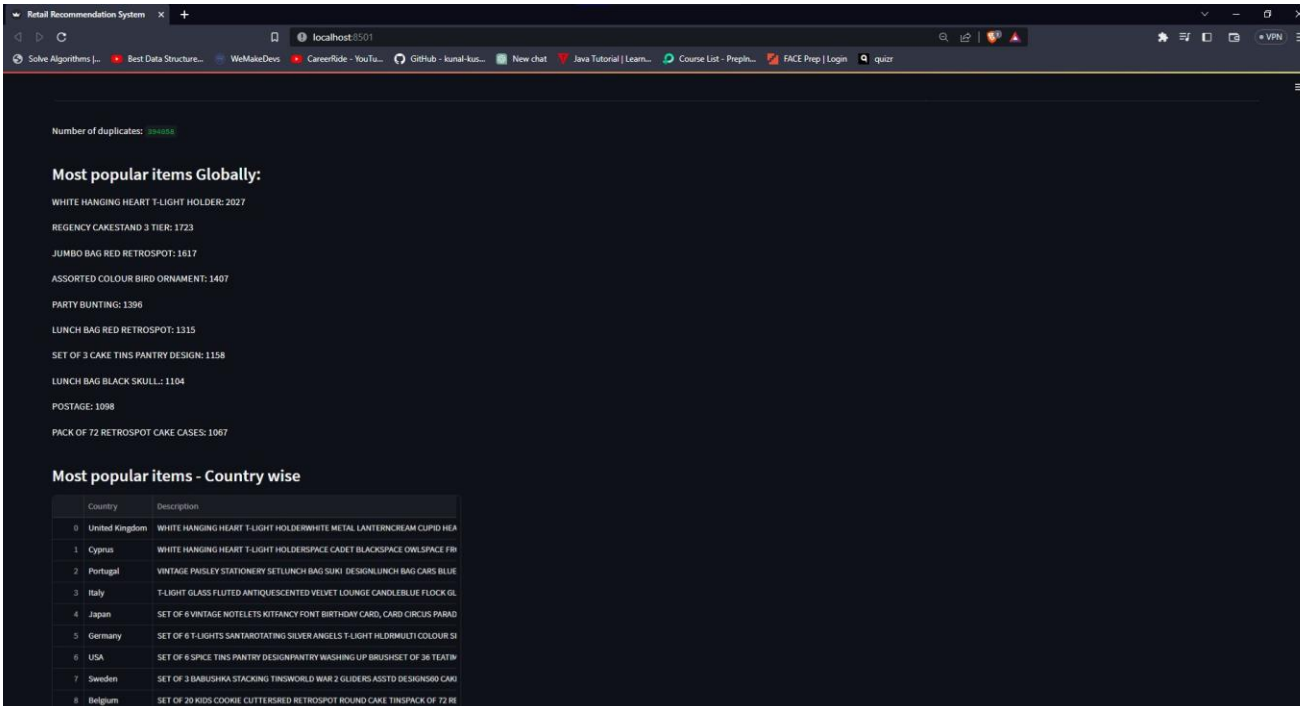
3.1 HOMEPAGE



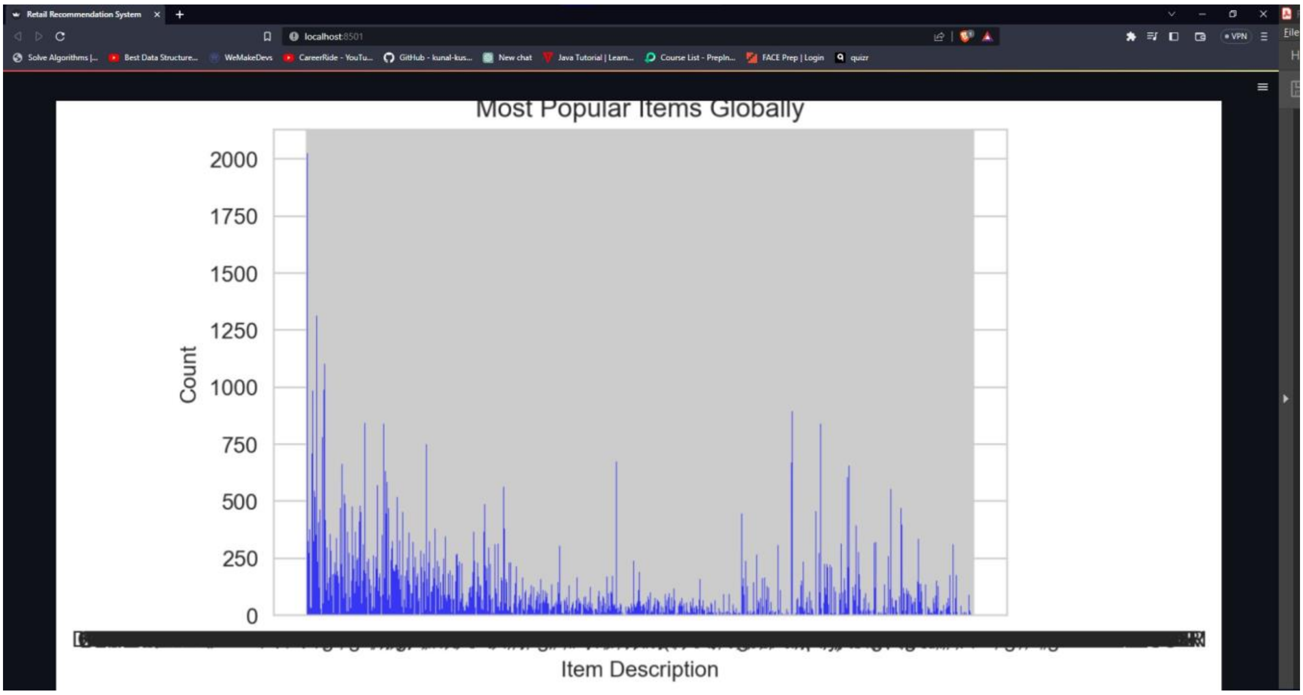
3.2 COUNTRIES AND ITEMS AVAILABLE



3.3 MOST POPULAR ITEMS GLOBALLY AND COUNTRY WISE



3.4 MOST POPULAR ITEMS GLOBALLY USING SEABORN LIBRARY



CHAPTER 4

CONCLUSION AND FUTURE ENHANCEMENTS

In conclusion, the recommendation system developed using Seaborn and pivot tables provides valuable insights for making item recommendations on a global, country-wise, and month-wise basis. These items have proven to be popular and well-received by users worldwide, making them strong candidates for global recommendations. Comparing estimated ratings with user ratings allows us to identify items that have the potential for greater popularity and user satisfaction. By recommending such items, we can introduce users to new and exciting options that align closely with their preferences, based on both their own ratings and the projected ratings of others.

However, there is still room for enhancement to incorporate social network analysis, which can take into account the relationships between users and their interactions with each other. This can provide insights into the social dynamics of the platform and help identify influencers and trends.

In summary, the recommendation system powered by Seaborn's visualization capabilities and pivot tables offers a comprehensive approach to item recommendations. By considering global, country-wise, and month-wise factors, as well as combining estimated and user ratings, the system ensures that recommendations are tailored to individual user preferences, geographic context, and the time of year. This results in a more personalized and engaging user experience, increasing the likelihood of user satisfaction and engagement.

CHAPTER 5
REFERENCES

1. S. Bahulikar, "Analyzing recommender systems and applying a location-based approach using tagging", *2017 2nd International Conference for Convergence in Technology (I2CT)*, 2017. https://ieeexplore.ieee.org/abstract/document/8226120

2. Y. C. Lee et al., "Recommendation of research papers in DBpia: A Hybrid approach exploiting content and collaborative data", *2016 IEEE International Conference on Systems Man and Cybernetics SMC 2016 - Conference Proceedings*, pp. 2966-2971, 2017. https://ieeexplore.ieee.org/abstract/document/7844691

3. M. Ohta, T. Hachiki and A. Takasu, "Related paper recommendation to support online-browsing of research papers", *4th International Conference on the Applications of Digital Information and Web Technologies ICADIWT 2011*, pp. 130-136, 2011. https://ieeexplore.ieee.org/abstract/document/6041413

4. H. Jeon and C. Jeon, "UserProfile-Based Personalized Research Paper Recommendation System", *Comput. Netw. Technol. (ICCNT) 2012 8th Int. Conf.*, pp. 134-138, 2012. https://ieeexplore.ieee.org/abstract/document/6418639

5. J. Chen and Z. Ban, "Literature recommendation by researchers' publication analysis", *Information and Automation (ICIA) 2016 IEEE International Conference on*, 2016.

https://ieeexplore.ieee.org/abstract/document/783214

6. C H Chen, S D Mayanglambam, F Y Hsu et al., "Novelty Paper Recommendation Using Citation Authority Diffusion[C]", *Technologies and Applications of Artificial Intelligence (TAAI) 2011 International Conference on. IEEE*, pp. 126-131, 2011. https://ieeexplore.ieee.org/abstract/document/6120731

7. LIANG, H., XU, Y., LI, Y., AND NAYAK, R. Collaborative filtering recommender systems using tag information. In IEEE/WIC/ACM International Conference on Web Intelligence (IAT)(2009),vol.3,IEEE,pp.59-62.

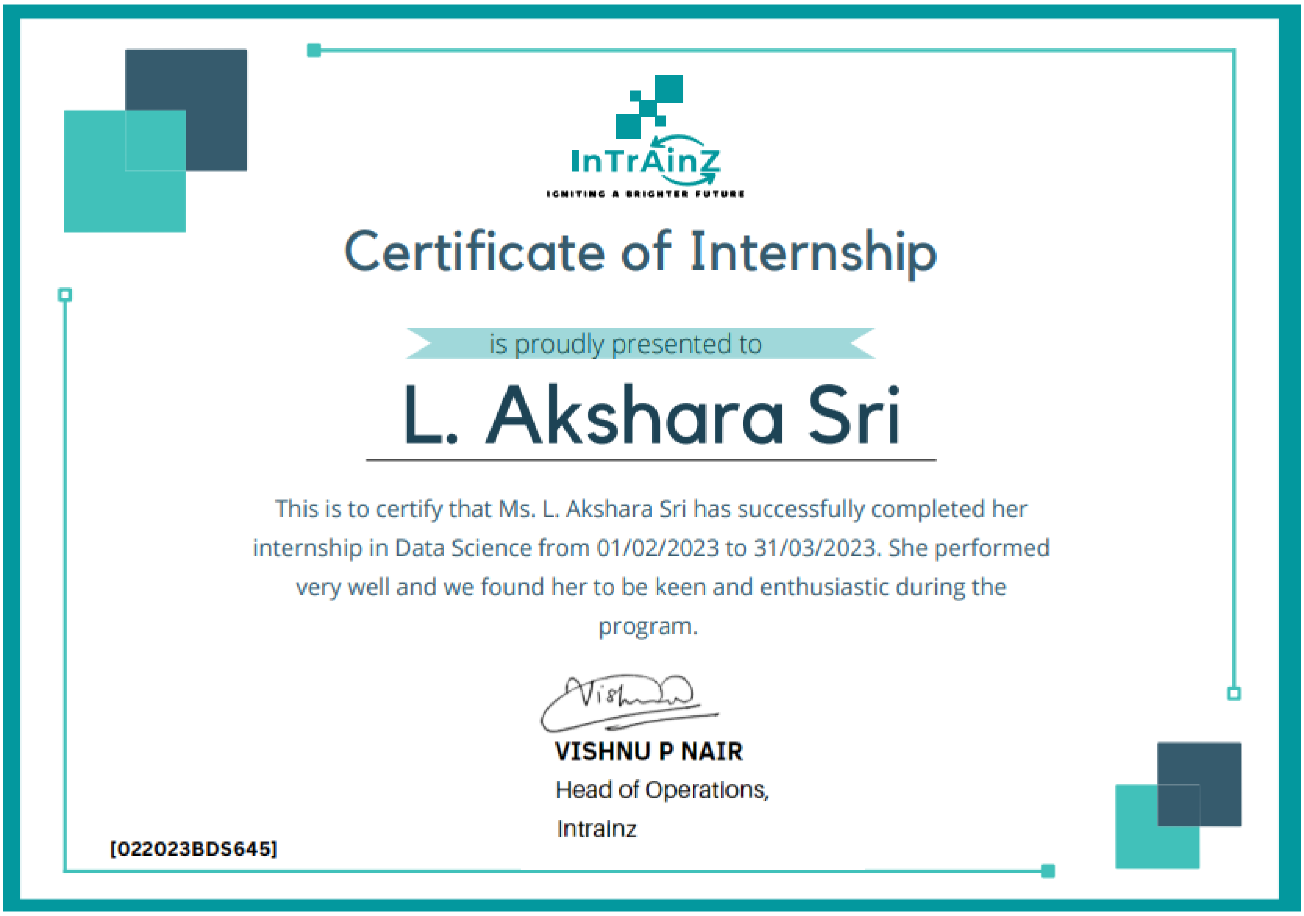
   https://ieeexplore.ieee.org/abstract/document/4740727

8. PAN, C., AND LI, W. Research paper recommendation with topic analysis. In Computer Design and Applications (ICCDA) (2010), vol. 4, IEEE, pp. V4-264 https://ieeexplore.ieee.org/abstract/document/5541170

9. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible Extensions", *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, pp. 634-749, 2005. https://ieeexplore.ieee.org/abstract/document/1423975

10. G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", *IEEE Internet Computing*, Jan./Feb. 2003. https://ieeexplore.ieee.org/abstract/document/1167344

INTERNSHIP CERTIFICATE



InTrAinZ

IGNITING A BRIGHTER FUTURE

Certificate of Internship

is proudly presented to

L. Akshara Sri

This is to certify that Ms. L. Akshara Sri has successfully completed her internship in Data Science from 01/02/2023 to 31/03/2023. She performed very well and we found her to be keen and enthusiastic during the program.

VISHNU P NAIR

Head of Operations,

Intrainz

[022023BDS645]

INDUSTRIAL TRAINING CERTIFICATE

CERTIFICATE

OF INDUSTRIAL TRAINING

This certificate is proudly presented to :

L. Akshara Sri

For completing her Industrial Training program with industry mentors from February 1st 2023 to March 31st 2023. We found her performance to be commendable

VISHNU P NAIR

Head of Operations,

Intrainz

InTrAinZ

IGNITING A BRIGHTER FUTURE

[022023BDS645]

LETTER OF RECOMMENDATION

InTrAinZ
IGNITING A BRIGHTER FUTURE

LETTER OF RECOMMENDATION

www.intrainz.com
hr.contact@intrainz.com

April 28th, 2023

TO WHOM IT MAY CONCERN

Ms. L. Akshara Sri completed an internship in Data Science with Intrainz from 1st February 2023 to 31st March 2023.

Her performance was excellent and we appreciate her sincere efforts in delivering quality work in the industrial projects assigned to her during the internship. We would like to restate our strong recommendation for her. If you have any further queries regarding her work, please dont hesitate to contact us.

Regards,

Vishnu

VISHNU P NAIR
Head of Operations,
Intrainz

Intrainz Innovation Private Limited - Evoma Business Centre, Old Madras Rd, Battarahalli, Bengaluru, Karnataka 560049
CIN : U80903KA2022PTC166349