A Mini Project Synopsis on

Ecommerce Platform with Recommendation engine and price optimization engine

T.E. – Computer Science and Engineering-Data Science

Submitted By

Rohan Waghode	**21107008**
Varad Joshi	**21107002**
Meet Jamsutkar	**22207004**
Aman Dhumal	**21107036**

Under The Guidance Of

Prof. Poonam Pangarkar

DEPARTMENT OF CSE-DATA SCIENCE

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year: 2023-24

CERTIFICATE

This is to certify that the Mini Project report on the **E-commerce store using Prize optimization and tag-based recommendation engine** has been submitted by **Rohan Waghode(21107008), Varad Joshi(21107002), Meet Jamsutkar(22207004) and Aman Dhumal(21107036)** who are bonafide students of A. P. Shah Institute of Technology, Thane as a partial fulfillment of the requirement for the degree in **Computer Science Engineering (Data Science)**, during the academic year **2023-2024** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Poonam Pangarkar
Guide

Ms. Anagha Aher
HOD, CSE(Data Science)

Dr. Uttam D. Kolekar
Principal

External Examiner:
1.

Internal Examiner:
1.

Place: A. P. Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Ms. Poonam Pangarkar**. Expressing gratitude towards our HoD, **Ms. Anagha Aher**, and the Department of Computer Science Engineering (Data Science) for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our project coordinator **Ms. Poonam Pangarkar** who gave us his/her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

Abstract

The digital marketplace has undergone a profound transformation propelled by technological advancements and shifting consumer behaviors. In this landscape, pricing strategies emerge as pivotal determinants of success, influencing both seller competitiveness and buyer decisions. This report introduces an innovative application poised to revolutionize traditional pricing methodologies within the ecommerce sector. At its core, the application seeks to redefine how pricing strategies are conceived, implemented, and optimized in response to the dynamic nature of the digital marketplace. With a focus on agility, intuitiveness, and data-driven decision-making, the application empowers sellers to adapt pricing strategies swiftly in real time, thereby enhancing their competitiveness and driving conversions on buyer applications. Leveraging advanced algorithms and real-time data analytics, it transcends conventional approaches by considering a multitude of factors, including consumer demands, product shelf life, and inventory dynamics. The application's efficacy lies in its targeted approach to categories where pricing adjustments wield significant influence over customer behavior, notably perishable goods and services. By honing in on these areas, it maximizes its impact on seller performance and buyer satisfaction. Moreover, it equips buyers with comprehensive pricing information and comparative analysis, enabling informed choices aligned with their preferences and budgetary constraints. The project's primary aim is to optimize the overall value proposition within the ecommerce ecosystem by enhancing seller performance, improving buyer satisfaction, and fostering a mutually beneficial relationship between sellers and buyers. To achieve this, the report addresses the multifaceted challenges of revolutionizing pricing methodologies and enhancing recommendation engines. By integrating advanced algorithms and machine learning techniques, the application aims to provide personalized recommendations that resonate with users, thereby enhancing engagement and satisfaction. In essence, the report presents a forward-thinking solution to the complexities and challenges inherent in pricing within the ecommerce sector. By leveraging technological innovation and real-time data insights, the application offers a glimpse into the future of ecommerce strategy, promising a more dynamic, personalized, and efficient marketplace for all stakeholders.

Chapter 1

Introduction

The ecommerce landscape has undergone a seismic shift in recent years, driven by technological advancements and evolving consumer behaviors. Amidst this transformation, pricing strategies stand as a linchpin for success, dictating the competitive edge of sellers and influencing purchasing decisions on buyer applications. In response to this dynamic environment, an innovative application emerges, poised to revolutionize traditional pricing methodologies within the ecommerce sector. `At the heart of this application lies a fundamental objective: to redefine how pricing strategies are conceived and implemented in the digital marketplace. Sellers are confronted with the perpetual challenge of determining the most effective pricing strategies to stimulate conversions on buyer applications. Crucially, these prices must be time-bound, ensuring their relevance and competitiveness for a limited duration. The application represents a paradigm shift in pricing strategy, embodying traits of agility, intuitiveness, and data-driven decision-making. Its core functionality lies in the ability to swiftly adapt prices in real time, responding dynamically to shifts in market dynamics and consumer preferences. This agility is underpinned by a multifaceted approach that considers an array of factors, including the ever-changing demands of consumers, the intricate dynamics of product shelf life, and the real-time status of inventory levels.

Central to the application's efficacy is its targeted focus on products and categories where pricing adjustments wield significant influence over customer behavior. Perishable goods and services represent prime examples of such categories, where time-sensitive pricing strategies can directly impact purchasing decisions. By honing in on these areas, the application seeks to maximize its impact on seller performance and buyer satisfaction. Within the application, customers are empowered with a wealth of information to aid in their purchasing decisions. They can compare pricing from various seller platforms, enabling them to make informed choices based on a multitude of factors. Inventory availability, trade terms, profit margins, and strategic considerations all play a pivotal role in shaping the perceived value of a product or service. Moreover, the time of day and prevailing market prices further contribute to the dynamic nature of pricing within the application.

In essence, the application serves as a conduit for seamless interaction between sellers and buyers, facilitating transactions based on transparency, efficiency, and value optimization. By leveraging advanced algorithms and real-time data analytics, it transcends traditional pricing methodologies, offering a glimpse into the future of ecommerce strategy. As the digital marketplace continues to evolve, the significance of innovative pricing strategies cannot be overstated. The application represents a forward-thinking solution to the complexities and challenges inherent in pricing within the ecommerce sector. Its ability to adapt to changing market dynamics, anticipate consumer preferences, and optimize pricing strategies in real time positions it as a game-changer in the realm of ecommerce.

1.1 Purpose:

The primary aim of this project is to fundamentally transform the way pricing strategies are conceived, implemented, and optimized within the ecommerce sector. In response to the dynamic and ever-evolving nature of the digital marketplace, the project endeavors to introduce an innovative application that goes beyond traditional pricing methodologies. By harnessing the power of advanced algorithms and real-time data analytics, the application aims to empower sellers with unprecedented agility and flexibility in adjusting their pricing strategies.

One of the key objectives of the project is to enhance seller performance by providing them with the tools and insights necessary to adapt to changing market dynamics swiftly. Sellers face the constant challenge of determining the most effective pricing strategies to maintain a competitive edge and drive conversions on buyer applications. This project seeks to address this challenge by equipping sellers with the ability to make informed decisions based on real-time data and insights. The project aims to improve buyer satisfaction by offering transparent pricing information, comparative analysis, and real-time updates. In today's digital age, consumers demand transparency and convenience when making purchasing decisions. By providing buyers with access to comprehensive pricing information from various seller platforms, the application seeks to empower them to make informed choices that align with their preferences and budgetary constraints.

The project aims to optimize the overall value proposition within the ecommerce ecosystem. By considering a diverse range of factors, including inventory dynamics, product shelf life, and market trends, the application seeks to maximize customer value while ensuring profitability for sellers. This symbiotic relationship between sellers and buyers fosters a mutually beneficial ecosystem where all stakeholders stand to gain.

1.2 Problem Statement:

In the rapidly evolving landscape of ecommerce and digital content consumption, the challenges of optimizing pricing strategies and enhancing user experiences through personalized recommendations have become increasingly pressing. Traditional pricing methodologies in ecommerce often fall short in adapting to the dynamic demands of consumers and the intricate dynamics of product lifecycle management and inventory management. Sellers struggle to set optimal prices that maximize conversions while remaining competitive in the market. Meanwhile, recommendation engines, although essential for guiding users to relevant content, face obstacles in delivering precise recommendations due to the complexity of user preferences and the vast amount of available data. The problem statement for this project is multifaceted. On one hand, there is a need to revolutionize traditional pricing methodologies within the ecommerce sector to address the shortcomings of existing approaches. This involves developing an innovative application that embodies agility, intuitiveness, and data-driven decision-making. Such an application should enable sellers to make swift, real-time adjustments to prices based on a comprehensive understanding of consumer behavior, product shelf life, and inventory levels. By leveraging advanced algorithms and machine learning techniques, this application aims to optimize pricing strategies to drive conversions while ensuring competitiveness and profitability for sellers.

On the other hand, there is a parallel need to enhance the effectiveness of recommendation engines to provide personalized suggestions that resonate with users. This requires refining recommendation processes by analyzing vast amounts of user data, including behavior, preferences, and interactions. The application must leverage advanced algorithms and machine learning techniques to generate tailored recommendations that align with individual user preferences and interests. By delivering more accurate and relevant suggestions, the aim is to enhance user satisfaction and engagement in the digital marketplace. Overall, by addressing these challenges, the project aims to optimize seller performance, drive conversions, and ultimately enhance user satisfaction in the digital marketplace. Through the development of an innovative application that integrates advanced pricing strategies and recommendation engines, the goal is to create a more dynamic and personalized ecommerce experience for both sellers and buyers.

1.3 Objectives

In the rapidly evolving landscape of ecommerce and digital content consumption, the challenges of optimizing pricing strategies and enhancing user experiences through personalized recommendations have become increasingly pressing. To address these challenges, the following objectives are outlined:

1. Revolutionize Pricing Methodologies:
    - Develop pricing algorithms that consider real-time consumer behavior, product lifecycle stages, and inventory levels.
    - Implement dynamic pricing strategies that allow for rapid adjustments based on market conditions and demand fluctuations.
2. Enhance Competitiveness:
    - Enable sellers to analyze competitor pricing and adjust their own prices accordingly to remain competitive.
    - Provide insights and recommendations to sellers to help them understand market trends and consumer preferences better.
3. Drive Conversions:
    - Utilize advanced algorithms and machine learning techniques to identify pricing strategies that maximize conversion rates.
    - Implement A/B testing methodologies to evaluate the effectiveness of different pricing strategies in driving conversions.
4. Personalize Recommendations:
    - Develop recommendation algorithms that leverage user data to provide personalized product suggestions.
    - Utilize collaborative filtering and content-based filtering techniques to improve the accuracy of recommendations based on user preferences.
5. Improve User Satisfaction:
    - Monitor user feedback and engagement metrics to continuously refine and improve recommendation algorithms.
    - Implement feedback loops to allow users to provide input on the relevance and usefulness of product recommendations.
6. Increase Seller Performance:

- Provide sellers with actionable insights and analytics to optimize their pricing and recommendation strategies.
- Offer training and resources to help sellers understand and utilize the application effectively to improve their performance.

7. Create Dynamic Ecommerce Experience:
   - Design an intuitive and user-friendly interface that allows sellers to easily navigate and interact with pricing and recommendation features.
   - Incorporate interactive dashboards and visualizations to present insights and recommendations in a clear and understandable manner.

1.4 Scope

In order to achieve the objectives outlined, the project encompasses the following key areas of focus:

1. Algorithm Development:
   - Developing and refining algorithms for dynamic pricing and personalized recommendations.
   - Implementing scalable algorithms capable of handling large volumes of data as the platform grows.
2. Machine Learning Integration:
   - Integrating machine learning models for analyzing user data and generating personalized recommendations.
   - Ensuring compatibility and interoperability with various machine learning frameworks and libraries.
3. Real-Time Adjustment:
   - Implementing mechanisms for real-time price adjustments based on market dynamics and user behavior.
   - Developing algorithms for predicting demand and adjusting inventory levels in real-time.
4. User Preference Analysis:
   - Utilizing data analytics techniques to analyze user behavior, preferences, and interactions.
   - Identifying patterns and trends in user data to enhance the accuracy of personalized recommendations.
5. Integration with Ecommerce Platforms:

- Ensuring seamless integration with popular ecommerce platforms like Shopify, Magento, and WooCommerce.
- Providing APIs and SDKs for easy integration with third-party applications and services.

6. User Interface Design:
    - Designing a responsive and user-friendly interface for sellers and buyers.
    - Conducting usability testing to gather feedback and iterate on the design for an improved user experience.
7. Testing and Optimization:
    - Conducting comprehensive testing to ensure the reliability and accuracy of pricing and recommendation algorithms.
    - Continuously optimizing algorithms based on performance metrics and user feedback.
8. Security:
    - Implementing robust security measures to protect user data and ensure compliance with privacy regulations.
    - Encrypting sensitive data in transit and at rest, and conducting regular security audits to identify and address vulnerabilities.
9. Documentation and Support:
    - Offering responsive customer support to address any issues or concerns raised by users, ensuring a smooth user experience.

Chapter 2

Literature Review

In the ever-evolving landscape of recommendation systems, a multitude of research papers have emerged to advance our understanding and improve user experiences.

"Recommendation Systems: Principles, Methods, and Evaluation" [1] offers an in-depth exploration of recommendation systems, emphasizing their indispensable role in enhancing user experiences, particularly in the realms of e-commerce and content streaming. This paper delves into various aspects, including the fundamental principles, methodological approaches, and the critical evaluation criteria that underpin these systems.

"Content-Based Filtering for Recommendation Systems Using Multiattribute Networks" [2] introduces an innovative methodology leveraging multiattribute networks to alleviate the challenges posed by information overload in recommendation systems. Empirical evidence provided in the study showcases the notable improvement in recommendation accuracy achieved through this approach.

In "Learning similarity with cosine similarity ensemble" [3], the authors introduce an ensemble technique tailored to enhancing similarity assessments, with a specific focus on the widely used cosine similarity measure.

Dehak et al.'s 2010 research [4] delves into cosine similarity scoring without the need for score normalization, particularly in the context of speech and language processing.

Li, Guo, and Zhao's 2008 paper [5] explores the concept of "Tag-based Social Interest Discovery," shedding light on how this approach aids in comprehending user interests within online social environments.

"Neural Collaborative Filtering" [6] represents a significant advancement in recommendation systems, integrating neural networks into collaborative filtering to optimize personalized recommendations. This approach leverages the power of deep learning to enhance the quality of recommendations, aligning them more closely with individual user preferences and behaviours.

Customers' demand for products and the optimization of prices to maximize profit have always been important aspects in business (Narangajavana et al., 2014)[8]. The study of demand models and price optimization performance has gained increasing attention in recent years. This interest has been fueled by advancements in available data, econometric methods, and computing power, which have revolutionized demand modeling. Various sources have contributed to the understanding of demand models and price optimization performance. One such source is a paper titled "Models of Consumer Demand for Differentiated Products" by Céline Bonnet and Timothy J. Richards[8]. Advances in available data, econometric methods, and computing power have created a revolution in demand modelling over the past two decade

Chapter 3

Proposed System

The proposed system aims to revolutionize the e-commerce landscape by introducing advanced pricing strategies and personalized recommendation systems. It is designed to address the challenges faced by sellers in setting optimal prices and delivering relevant product suggestions to users. The system leverages cutting-edge technologies such as machine learning and real-time analytics to provide dynamic and data-driven solutions.

3.1 Features and Functionality:

1. Dynamic Pricing Algorithms:
   - The system will develop and refine algorithms for dynamic pricing based on real-time data streams. These algorithms will take into account various factors such as consumer behavior, product lifecycle stages, inventory levels, and competitor pricing to determine optimal prices.
   - Sellers will have the ability to set pricing rules and thresholds, allowing for automated price adjustments based on predefined criteria.
2. Personalized Recommendation Engine:
   - The recommendation engine will utilize machine learning models to analyze user data and generate personalized product suggestions.
   - By analyzing user behavior, preferences, and interactions, the system will identify patterns and trends to deliver accurate and relevant recommendations.
   - Recommendations will be tailored to each user's individual preferences, increasing the likelihood of conversion and enhancing user satisfaction.
3. Real-Time Adjustment Mechanisms:
   - The system will implement mechanisms for real-time price adjustments based on market dynamics and user behavior.
   - By continuously monitoring changes in demand, competitor prices, and user interactions, the system will dynamically adjust prices to remain competitive and maximize conversions.

4. User Preference Analysis:
    - Data analytics techniques will be employed to analyze user behavior, preferences, and interactions.
    - Insights gained from user data analysis will be used to improve the accuracy of personalized recommendations and enhance user engagement.
5. Integration with Ecommerce Platforms:
    - The system will ensure seamless integration with popular ecommerce platforms such as Shopify, Magento, and WooCommerce.
    - APIs and SDKs will be provided to facilitate easy integration with third-party applications and services, allowing sellers to leverage the system's capabilities within their existing workflows.
6. User Interface Design:
    - A responsive and user-friendly interface will be designed for both sellers and buyers.
    - The interface will provide intuitive access to pricing and recommendation features, allowing sellers to easily navigate and interact with the system.
7. Testing and Optimization:
    - Comprehensive testing will be conducted to ensure the reliability and accuracy of pricing and recommendation algorithms.
    - Continuous optimization based on performance metrics and user feedback will ensure the system remains effective and efficient.
8. Scalability and Security:
    - The system architecture will be designed to be scalable and resilient to handle increasing data volumes and user traffic.
    - Robust security measures will be implemented to protect user data and ensure compliance with privacy regulations.
9. Documentation and Support:
    - Responsive customer support will be available to address any issues or concerns raised by users, ensuring a seamless user experience.

Chapter 4

Requirement Analysis

1. **User Requirements:**

- A simple intuitive and Responsive interface
- Personalized pricing for every user as per their shopping habits
- Effective personalized recommendations for users as per their interests and past interactions
- Transparency and trust
- Search and Filters

2. **Preprocessing and Analysis Requirements:**

- Clean Data: Ensure data quality and cleanliness by handling missing values, removing outliers, normalizing data, and encoding categorical variables1.
- Transform Data: Prepare data in a suitable format for analysis, considering features like product attributes, user behavior, and pricing.

3. **Model Evaluation Requirements:**

- Accuracy Metrics: Evaluate model performance using accuracy metrics to ensure reliable and effective phishing detection.
- Precision and Recall: Measure precision and recall to assess the model's ability to correctly identify phishing attempts while minimizing false positives.
- Confusion Matrix: Generate confusion matrices to gain insights into the model's performance across different classes (phishing vs. legitimate).

4. **Deployment Requirements:**

- EC2 instance t2 micro or higher
- RDS small instance
- S3 bucket storage

Chapter 5

Project Design

The design of the proposed system encompasses a comprehensive approach to revolutionize the ecommerce landscape by addressing key challenges faced by sellers and enhancing user experiences. Through advanced pricing strategies, personalized recommendation systems, and cutting-edge technologies such as machine learning and real-time analytics, the project aims to create a dynamic and data-driven ecommerce platform. The project design focuses on developing robust algorithms, seamless integration with existing ecommerce platforms, user-friendly interfaces, and rigorous testing and optimization processes. By leveraging these design elements, the system will empower sellers to set optimal prices, deliver personalized product recommendations, and ultimately drive conversions and enhance user satisfaction in the digital marketplace.

5.1 Use Case Diagram :

The Engine leverages user interactions in tandem with existing user interests to keep creating and updating interests. The user interests would consist of the tags user is interested in, and the score of each tag, i.e. the score of how much a user is interested in a particular tag. The user interactions and user interests are compared using cosine similarity to generate new User interests. These User interests come into effect later into the recommendation part.

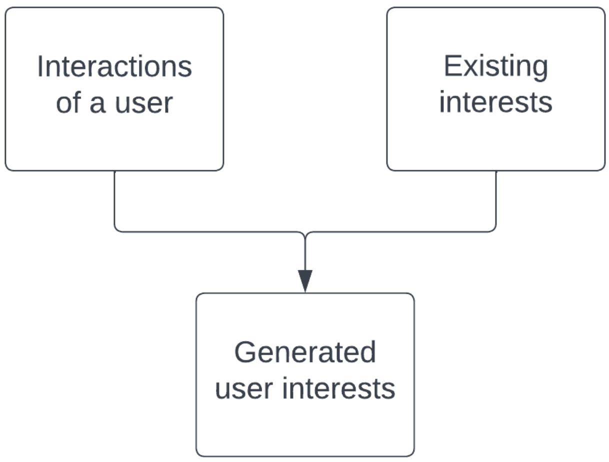

Figure 1: User Interest Generation

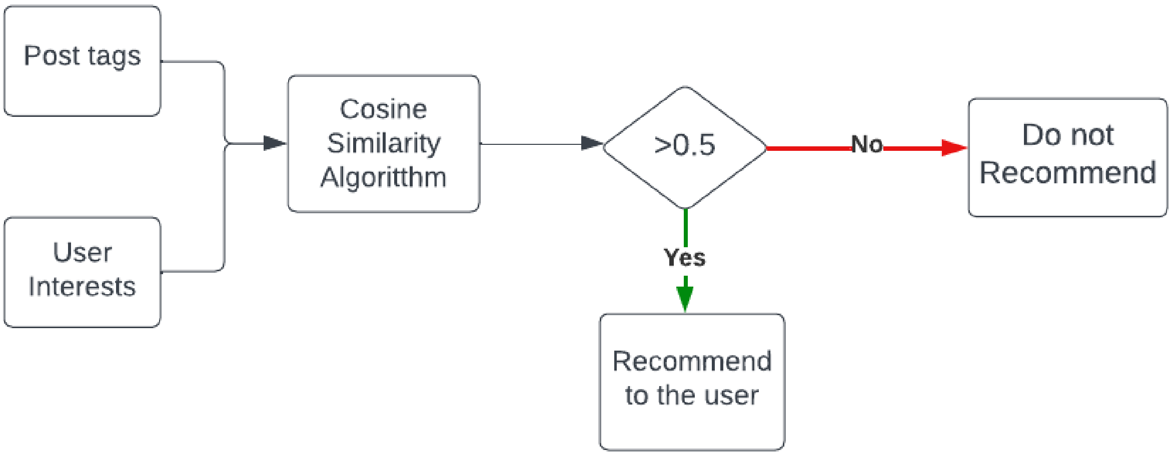

Figure 2: Recommendations Generation

2. Recommendation Generation:

The aforementioned User interests are then compared with Post tags(tags that a certain post has) using cosine similarity to obtain a score that is between 0 and 1. In this process every single user interest is compared with every single post tag and posts with post tags with a score greater than 0.5 are recommended to the User.

5.2 DFD (Data Flow Diagram):

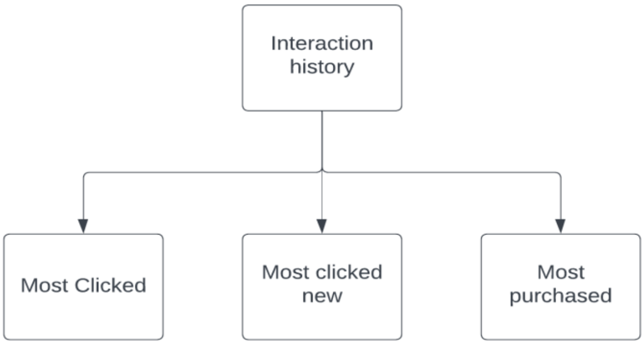

Figure 3: User Interaction

` We delve into the core components and mechanisms of our recommendation system. We'll explore how user interests are dynamically generated and updated through the Score Generator and how user feedback is incorporated to enhance accuracy. Additionally, we'll uncover the intricate process of recommendation generation, which leverages both collaborative filtering and content-based filtering for optimal results. This chapter provides a comprehensive understanding of the system's architecture, setting the stage for a detailed analysis of its performance and effectiveness in subsequent sections.

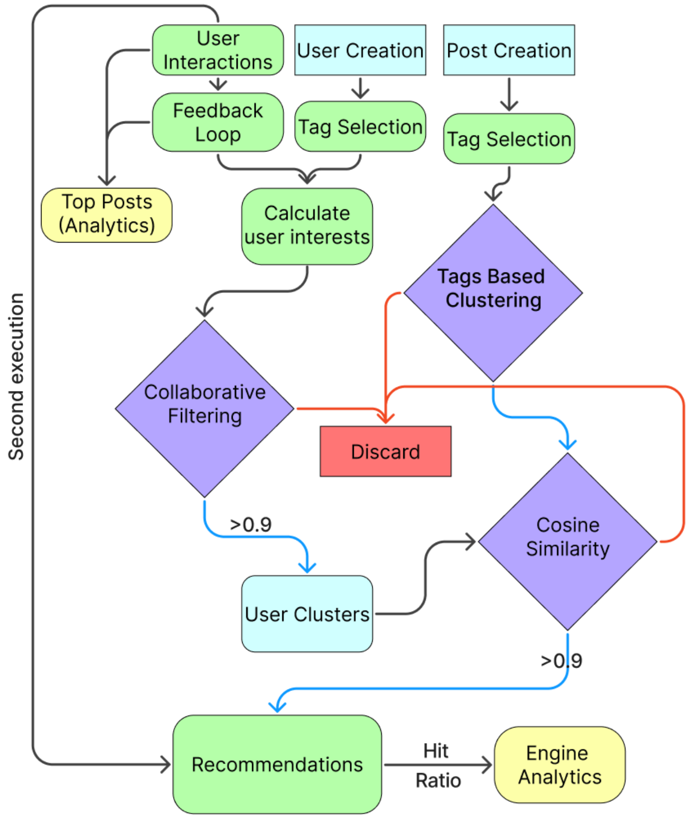

Figure 4: Recommendation Flow

To revolutionize pricing strategies in the ecommerce industry, the application will employ Deep Neural Networks (DNNs) for price optimization, utilizing a streamlined architecture with only two hidden layers. DNNs offer the capability to analyze complex, multidimensional data and derive insights that traditional methods may overlook. By leveraging DNNs, the application can swiftly adapt prices in real time, considering factors such as consumer demand fluctuations, product shelf life, and inventory levels. This agility ensures that sellers can respond promptly to market dynamics and optimize prices for maximum conversions. Moreover, DNNs can handle a multifaceted array of factors, including the perishability of products and service offerings, ensuring that pricing decisions align with the strategic goals of sellers. The application will utilize data from various sources, including inventory status, terms of trade, margins, and real-time market prices, to provide customers with the best value for their purchases.

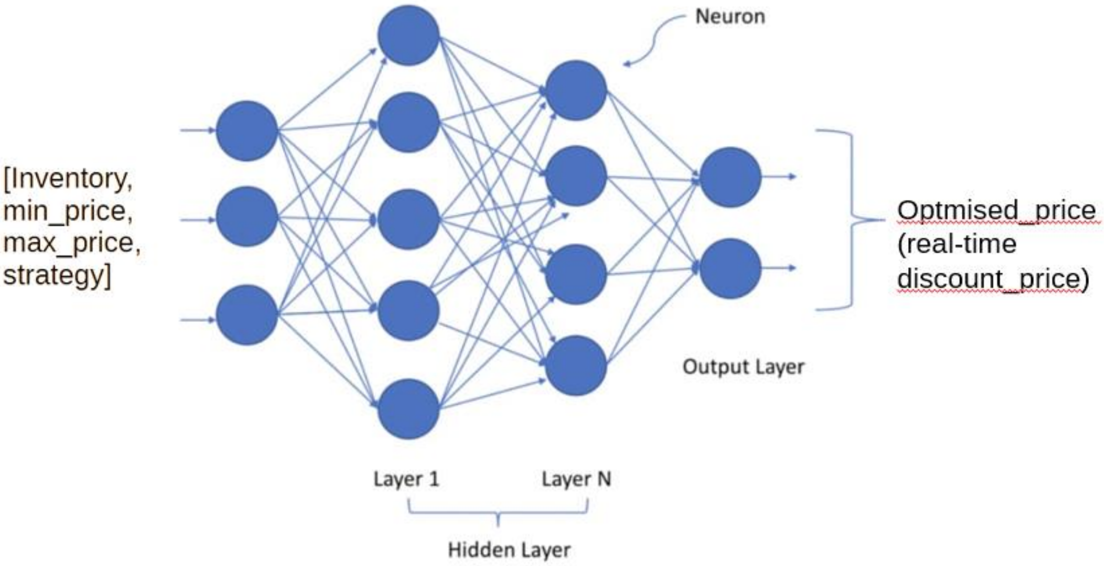

Figure 5 : Prize optimization using DNN

Through the integration of DNNs, the application will enable sellers to make data-driven pricing decisions that not only enhance competitiveness but also cater to the evolving needs and preferences of consumers. This approach reflects the core principles of agility, intuitiveness, and data-driven decision-making, essential for transforming pricing strategies within the ecommerce industry.

5.3 System Architecture

A. Recommendation

We will now take an in-depth overview of how each component works together.

1. Post:

   A Post is a piece of content that will be recommended to the User based on the decision made by the Recommendation Engine.

post_tags={postid=post1: tag1=yes, tag2=no, tag3=no ... tagn=yes}

While processing, all "yes" values become 10 and no become 0 for cosine similarity calculation.

2. Tags:

Tags are the attributes given to our Posts which help us establish an analogy between a post and the User. Each post can have multiple tags describing itself. Tags can either be set by reviewers, post author or automatically by using NLP based on different use cases. In our testing, the Post Author gets to set its Tags.

3. User Interests:

User interests track the interest of a user in different categories of Posts by assigning a value, hereon referred to as Score, to each Tag. This is generated by the "Score Generator" part of the recommendation engine and serves as a basis for generating Recommendations.

user_interests={user_id=u1: tag1=score, tag2=score, ... tagn=score}

where the value of score ranges between 0 to 10.

4. User Interactions:

User Interactions refer to the various things a user might do on the platform. In our case, we are using a simple implementation which records the timestamp, post id, user id and action. Based on what action is being done, there is provision in the recommendation engine to provide more or less weightage to the interest of the user in that post. Interactions are also our main basis for calculating Analytics, both Post analytics as well as Engine analytics. User interactions are also used by the Score generator to generate User Interests. For more advanced analytics, we can add more user data such as region, language, gender, age group and get analytics based on those factors.

user_interactions={userid=u1, postid=post1, event:click}

5. Score Generator:

The score generator is the first part of our recommendation engine which generates the User Interests which can then be compared with Post Tags in order to decide whether to recommend a post or not. Following is the example of how the score generator generates user interests.

Iterate over users:

Retrieve existing User Interests from the database and store them in array A.

Let us assume some values for array A.

A = {0, 1, 5, 3, 10} ------- *(3.5.1)*

Next, we check the database for user interactions and retrieve the tags of every post the user has interacted with since the last score generation. Let us assume this data is saved in a two-dimensional array B where the rows represent Posts and columns represent tags. Let us assume data for array B

B = {0, 10, 0, 10, 0

10, 0, 10, 0, 10

0, 0, 10, 0, 0

10, 10, 0, 0, 0

10, 0, 10, 0, 0} ------- *(3.5.2)*

Now, the score generator takes summation of all rows and stores it in an array C.

C = {30, 20, 30, 10, 10} ------- *(3.5.3)*

Array C represents the latest interests of the User. The array C is now normalized to 10 by dividing throughout by 3.

C = {10, 6.67, 10, 3.34, 3.34} ------- *(3.5.4)*

Now, we have array A representing the user's historical interests and array C representing the user's latest interests. We average both the arrays, and the resulting data is updated into the database. In this case, let us assume the result is array D.

D = {5, 8.34, 7.5, 3.17, 1.67} ------- *(3.5.6)*

Rounding to nearest integer,

D = {5, 8, 7, 3, 2} ------- *(3.5.7)*

This way, the scores never become stagnated, and the user never gets recommended something that he is no longer interested in.

6. Feedback Loop:

Our primary way of generating latest user interests is by aggregating past user interests and their latest interactions. However, sometimes a stray tag might get too weighted while the user isn't really interested in it. This is why we introduced a Feedback Loop which the user can use for letting the system know that they are not interested in a particular post/tag. These posts/tags are negated while calculating user interests by the score generator. Doing this improves the accuracy of User Interests, thereby reducing the error factor while generating recommendations and enhancing prediction quality.

7. Recommendation Generation:

For generating recommendations, we use the cosine similarity algorithm along with clustering of Users as well as Posts, thus making it a hybrid system. A hybrid system is the one that implements both, collaborative filtering as well as content-based filtering.

Cosine Similarity: Cosine similarity is a metric used to measure the similarity between two vectors. We can use Post Tags and User Interests as the dimensions for 2 distinct vectors and compare them using cosine similarity, as follows.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Figure 6 : Cosine Similarity Formula

Let array A represent User Interests and array B represent Post Tags. Assuming data,

$A = \{0, 10, 10, 0, 10, 0, 0, 10\}$ ------- (5)

$B = \{0, 4, 6, 1, 4, 0, 3, 10\}$ ------- (6)

By formula, we first need to find the dot product between A and B vectors as well as the magnitude of A and B.

$A.B = 240$

$|A| = 20$

$|B| = 13.341$

Now, $cosine\ similarity = 240/(20*13.341) = 0.8988$

Based on the value of cosine similarity (0.8988), we can assume that vectors A and B are similar. The closer the value of cosine similarity is to 1, the better fit it is.

Collaborative Filtering: Collaborative filtering is a method which analyses the history of Users with similar interests and recommends them the same Posts. In our implementation, we use collaborative filtering to divide Users into categories by using cosine similarity to find users with similar interests.

Tag Based Clustering: Tag-based clustering is a data analysis technique used to group or categorize Posts on the basis of Tags. We already have categories in place and hence do not need to do any additional processing for this.

Using the methods given above, we obtain User clusters and Categories. In order to generate recommendations, the first step is to generate scores for Categories using the same score generator used for users. Once obtained, we can compare User clusters with Categories using cosine similarity and rank the outcomes on the basis of the outcome of cosine similarity. These recommendations can be served to all the users from a cluster.

B. Prize Optimization

1. Data Generation:
    - The system begins by generating random data for product details such as product ID, category, minimum and maximum prices, rating, price difference, and discount.

- The `generate_data()` function is used to generate random data for each product, considering factors such as category, price range, and rating.

2. Data Preprocessing:
   - Once the data is generated, it is preprocessed for training the neural network model.
   - The generated data is split into features (X) and target (y) arrays. Features include product ID, minimum and maximum prices, rating, and price difference, while the target is the discount.
3. Neural Network Model Definition:
   - A neural network model is defined using TensorFlow's Keras API. The model architecture consists of several dense layers.
   - In this implementation, the model comprises four hidden layers with 64, 32, 16, and 8 neurons, respectively, followed by an output layer.
4. Model Compilation:
   - The model is compiled using the Adam optimizer and mean absolute error (MAE) as the loss function.
   - Additionally, mean absolute percentage error (MAPE) is included as a metric to monitor model performance during training.
5. Model Training:
   - The compiled model is then trained using the generated data.
   - Training is conducted over 50 epochs with a batch size of 32, and 20% of the data is used for validation.
6. Prediction for New Data:
   - After training, new random data for product details is generated using the `generate_and_print_data()` function.
   - This new data is preprocessed, and the trained model is used to predict discounts for these products.
   - Predicted discounts are obtained by passing the new data through the trained model.
7. Displaying Results:
   - The predicted discounts for the new data are printed, along with the corresponding product IDs.
   - Additionally, the training history of the model is visualized using Matplotlib, showing the training loss and validation loss over epochs.

- This visualization provides insights into the convergence and performance of the model during training.
- The system demonstrates the use of deep neural networks for price optimization in the ecommerce industry, showcasing the capability to predict discounts based on various product attributes.
- By training on generated data and predicting discounts for new data, the system illustrates how machine learning techniques can enhance pricing strategies and decision-making processes in ecommerce.

5.4 Implementation

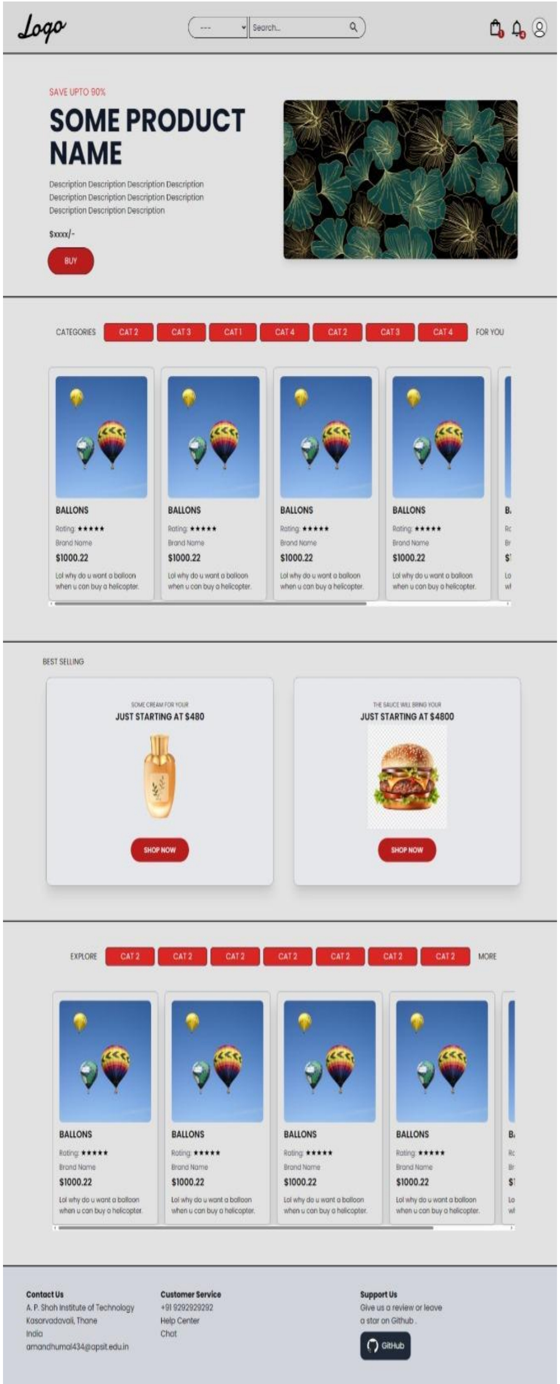

Figure 7: Landing Page

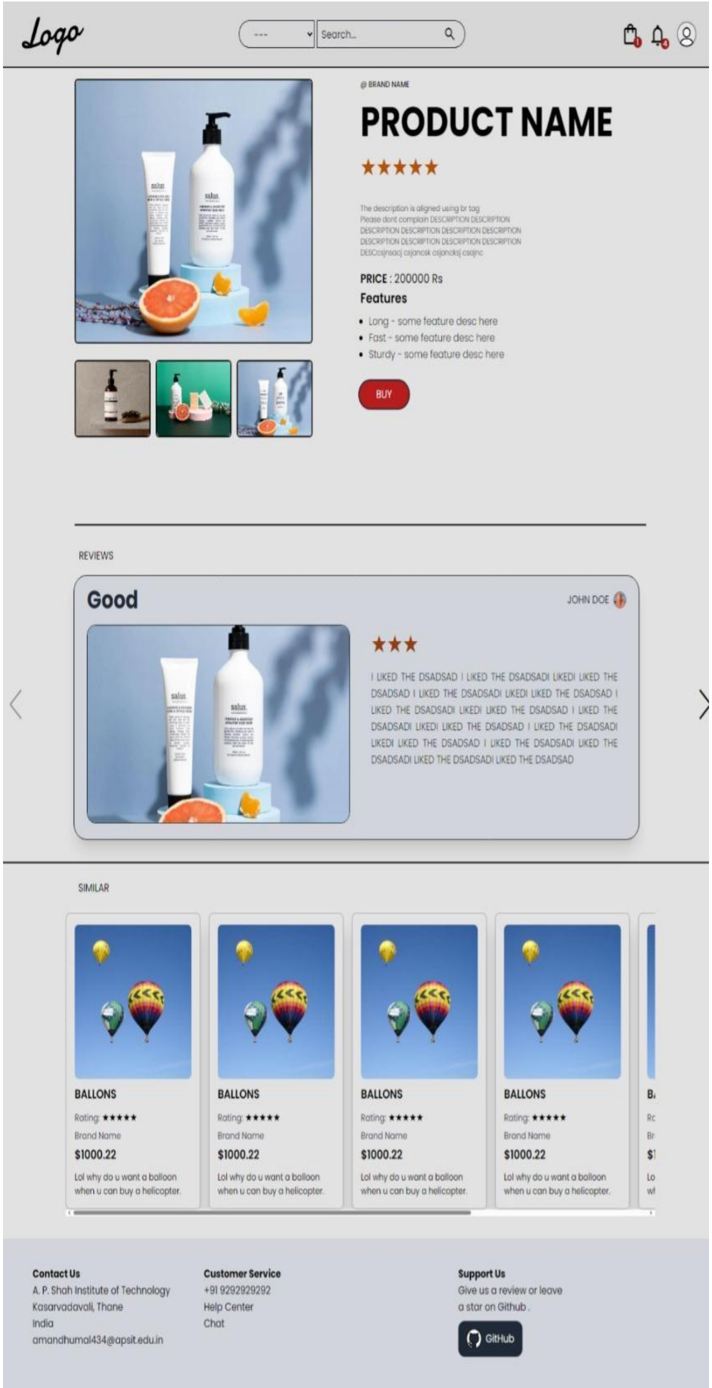

Figure 8: Product page(reviews and similar products)

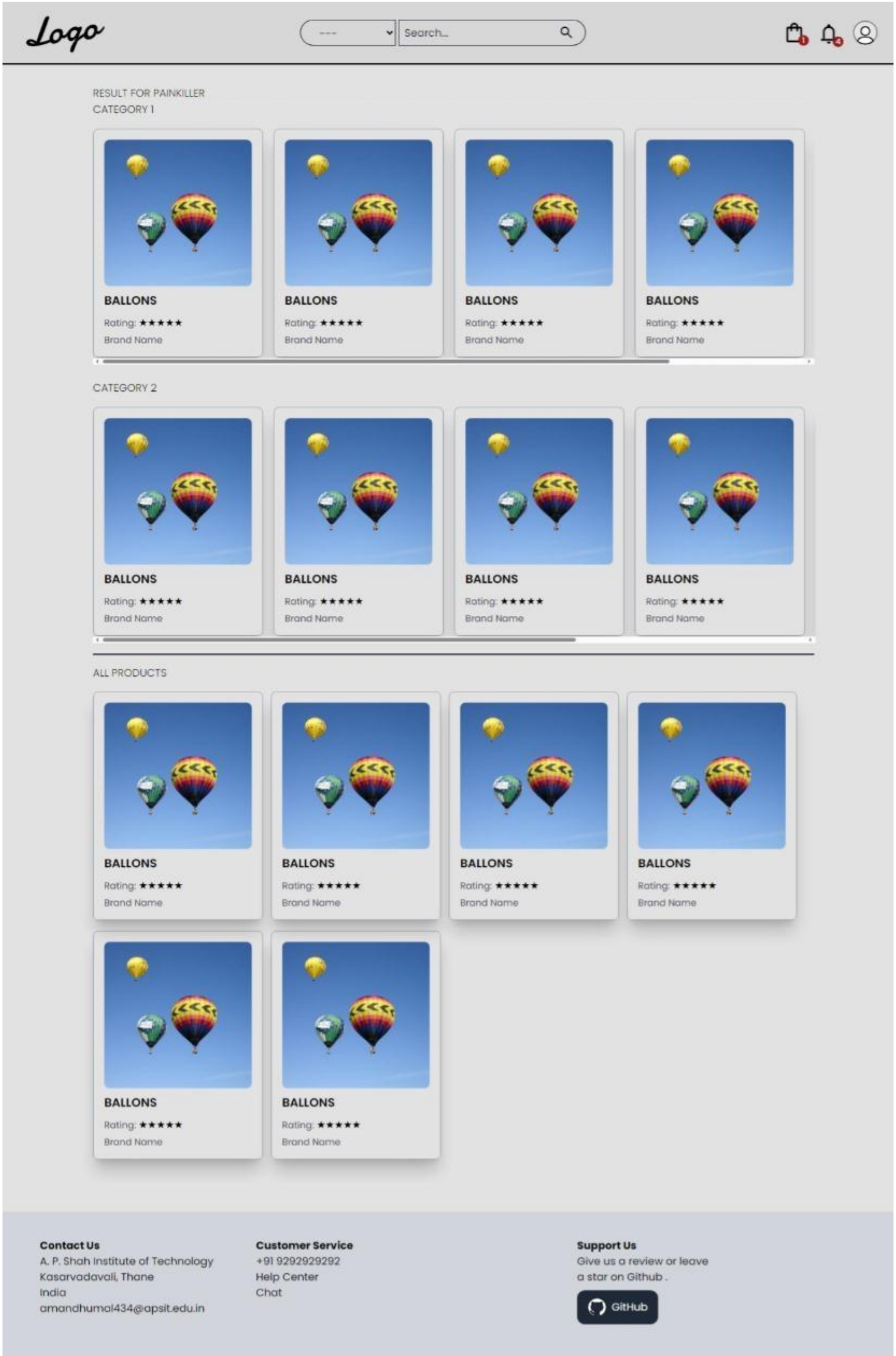

Figure 9: Search page with top charts and all products

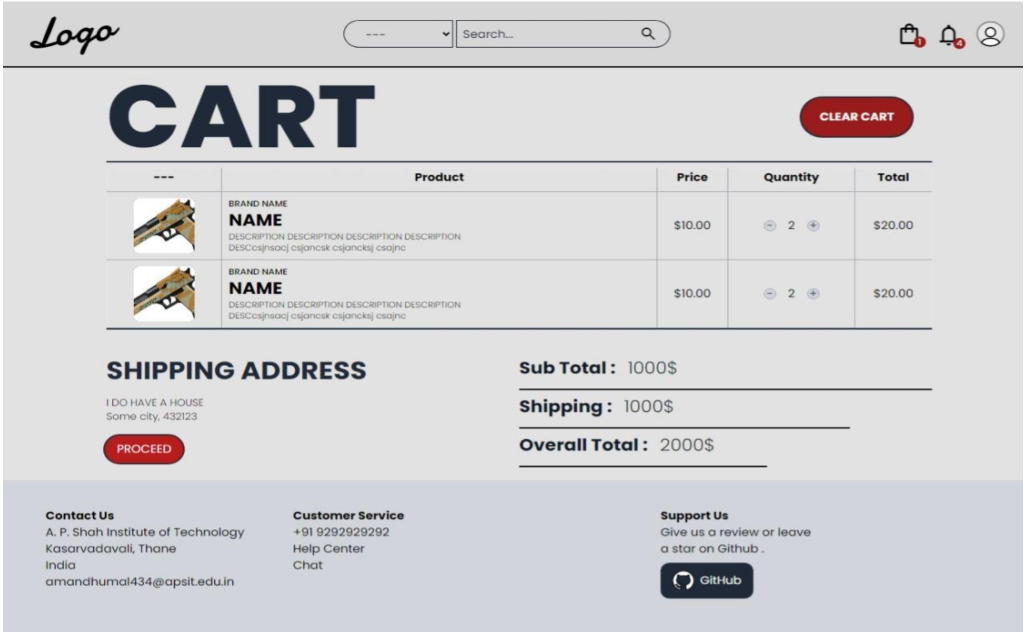

Figure 10: Cart page of selected products by user

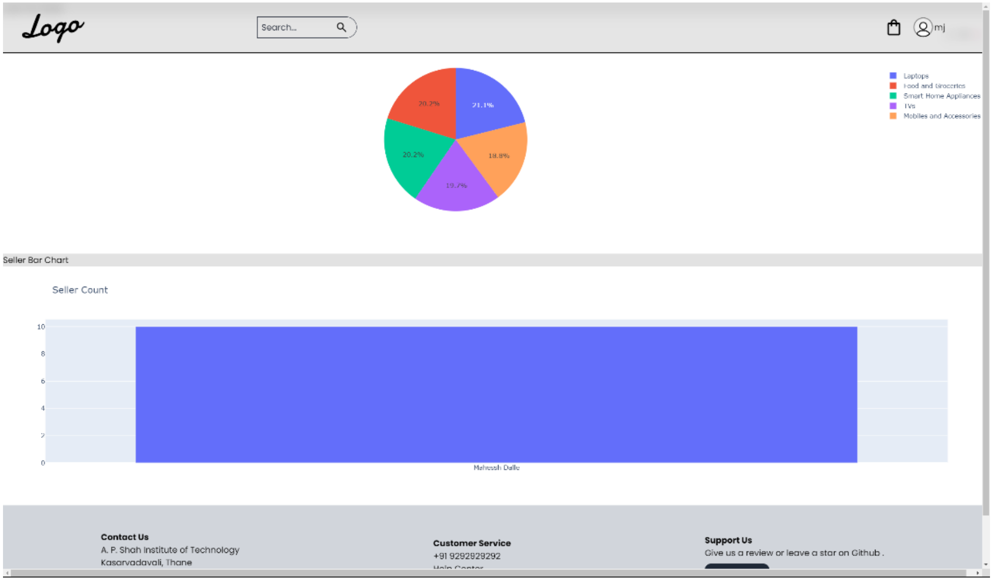

Figure 11: Analytics Page for Seller

Chapter 6

Technical Specification

The technical specifications outline the hardware and software requirements for executing a machine learning project, encompassing components such as CPU, GPU, RAM, operating system, development tools, programming languages, frameworks, databases, and additional libraries. These specifications ensure the system's capability to support data processing, model training, and deployment tasks effectively.

6.1 Hardware Specification

- CPU: Equivalent to AMD Ryzen 5 5500U or higher
- GPU: Equivalent to NVIDIA Geforce 1450 or higher
- RAM: 8GB LPDDR 5 or higher

6.2 Software Specification

- Ubuntu Linux 22.0.4 or Higher: Provides a stable and customizable environment for machine learning development, with extensive support for libraries and tools required for data processing and model training.
- VS Code, Jupyter Notebook: Versatile code editors offering features like interactive coding and visualization, crucial for prototyping, experimenting, and documenting machine learning workflows.
- Python 3.10, Javascript: Supported programming languages enabling seamless integration with machine learning libraries and frameworks, facilitating development of both backend algorithms and frontend applications.
- SQLite 3.45.3: Lightweight and efficient database solution for managing datasets and model outputs, suitable for small to medium-scale machine learning projects with moderate data storage requirements.
- Django 5.0.2: Robust web framework providing tools for building scalable and secure machine learning applications, offering features like ORM, authentication, and REST API support.
- TensorFlow 2.16.1: Powerful machine learning framework enabling efficient implementation of neural networks and deep learning algorithms, essential for training and deploying complex models in various ML tasks.

- Git 2.42.0, GitHub: Version control and collaborative development platforms facilitating team coordination, code sharing, and project management, crucial for maintaining and iterating machine learning codebases.
- Plotly 5.20.0, NumPy 1.26.4, Django-Allauth 0.61.1, Pillow 10.2.0: Additional libraries enhancing data visualization, numerical computation, user authentication, and image processing capabilities, augmenting the functionality and performance of machine learning projects.

Chapter 7

Project Scheduling

In the context of the E-commerce Store, scheduling plays a vital role in organizing and managing the development process. The project schedule comprises a comprehensive list of milestones, tasks, and deliverables, serving as a roadmap for the project's execution. It outlines the timeline for task completion, allocation of resources, and dependencies between activities.

<table>
<tr><th>Sr. No</th><th>Group Member</th><th>Time Duration</th><th>Work to be done</th></tr>
<tr><td rowspan="4">1</td><td rowspan="4">Rohan Waghode
Varad Joshi
Meet Jamsutkar
Aman Dhumal</td><td rowspan="2">1st week of January</td><td>Group formation and Topic finalization. Identifying the scope and objectives of the Mini Project.</td></tr>
<tr><td>Discussing the project topic with the help of a paper prototype.</td></tr>
<tr><td rowspan="2">3rd week of January</td><td>Identifying the functionalities of the Mini Project.</td></tr>
<tr><td>Designing the Graphical User Interface (GUI).</td></tr>
</table>

			Scraping data and building dataset for price optimization engine
2	Meet Jamsutkar Varad Joshi Rohan Waghode	**2nd week of February**	Database Design and Recommendation engine processing and working.
3	Rohan Waghode Varad Joshi Meet Jamsutkar	**1st week of March**	Training and optimization of price optimization model
4	Rohan Waghode Varad Joshi Meet Jamsutkar Aman Dhumal	**Last week of March**	Integration of all modules, User Interfaces and Report Writing.

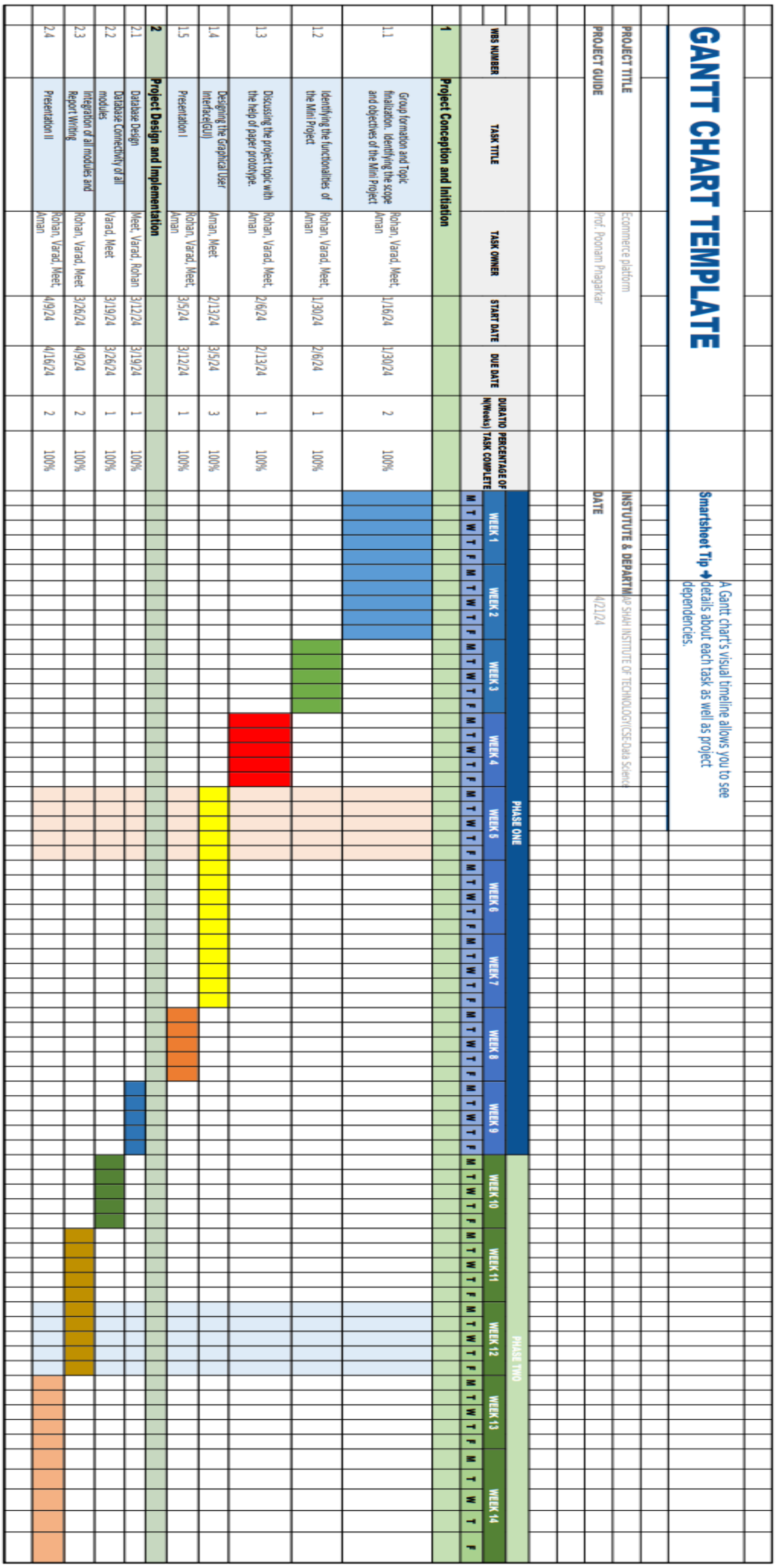

Figure 12: Gantt Chart

Chapter 8

Result

In this section, we delve into the results and analysis of our recommendation engine based on the system architecture described in the previous section. We begin by presenting the outcomes of the score generation process, followed by an examination of the feedback loop's impact on user interests. Finally, we discuss the recommendation generation process, including the application of cosine similarity, collaborative filtering, and tag-based clustering to provide users with personalized recommendations. We will also discuss how the recommendation engine provides insights regarding its own accuracy for further enhancing its effectiveness.

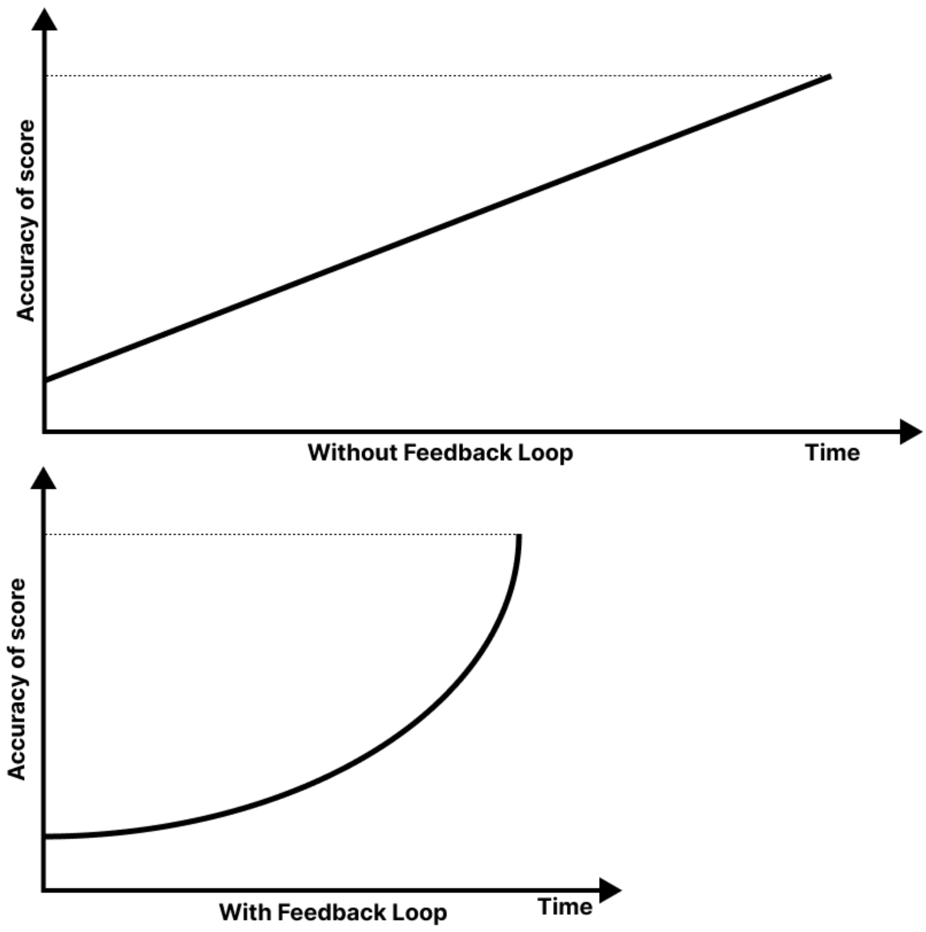

Figure 13: Accuracy over time graph

The process of score generation incorporates User inputs to generate User Interests and feedback loop assists in reducing the factor of error that might be prevalent in such a system. Furthermore, since every interaction gets weighted differently, the chances of having error diminish even further. We can see the difference between a system without feedback loop and a system with feedback loop in figure (Feedback Loop). It can be observed that the system with a constant feedback loop will achieve the same accuracy in less time as the user has more control over their interests.

The recommendation generation process incorporates three different levels of filtering and analysis in order to generate most accurate recommendations with least number of comparisons. If we consider it to just use cosine similarity, the number of comparisons will equal (m*n) where m represents the number of active users and n represents the number of posts. If we consider a small platform with 100 active users and 1000 posts, the number of comparisons becomes 1,00,000. As we can observe, this will grow even more for mid-sized platforms and is literally unusable for platforms with a larger audience. Furthermore, in order to get a reasonable number of recommendations, we need to use a threshold for cosine similarity around 0.5. This makes the mean accuracy, ignoring errors in tag assignment and user interests to be around 50%. The distribution of cosine accuracy can be seen in Figure 4. 2.

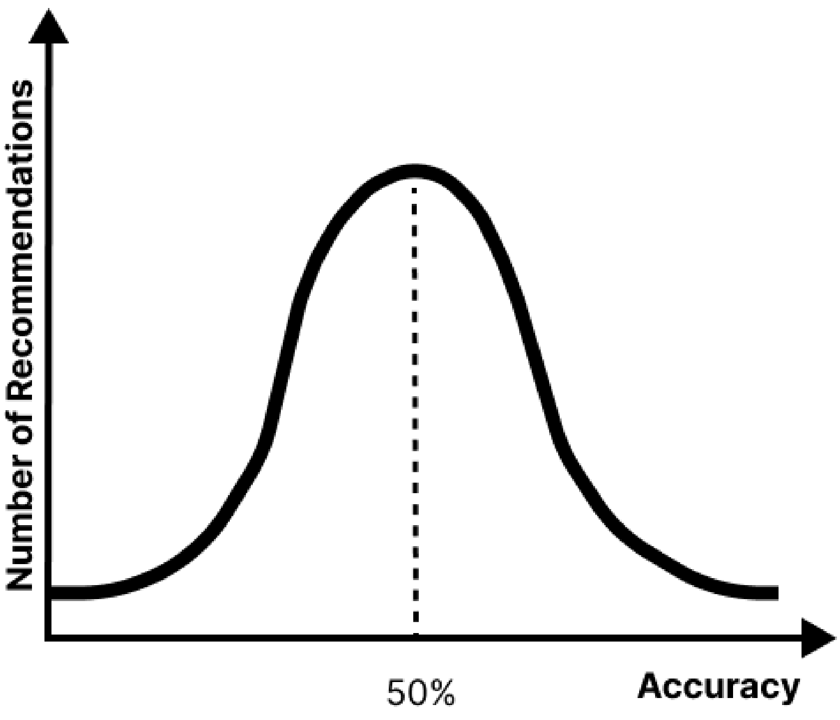

Figure 14: Accuracy distribution of cosine similarity with threshold 0.5

In order to optimize this, we first categorize our Posts based on their Tags. This reduces the number of comparisons to (m*n) where m represents the number of active users and n represents the number of categories. The number of categories will always will less than the total number of Posts. Along with these benefits, we also have higher accuracy in cosine similarity because our Category scores are also in a range of 0-10 as opposed to each Post tags, which are either 0 or 10. In order to get recommendations nearly the same as with just using cosine similarity, we can now use tighter thresholds for cosine similarity and hence our simulated mean accuracy, ignoring errors in tag assignment and user interests to be higher than before.

In our system, we take this a step further and consolidate Users into clusters as well. This further reduces the number of comparisons. Also, we can use much tighter thresholds now and still get a large amount of recommendations. For example, we can set a threshold of 0.9 cosine similarity to categorize Users into similar clusters and then categorize the results of comparing User cluster with Category in ranges like 0.7-0.79 cosine similarity, 0.8 to 0.89 cosine similarity and 0.9 to 1.00 cosine similarity, we get a large amount of recommendations and the simulated mean accuracy for each range will be as follows:

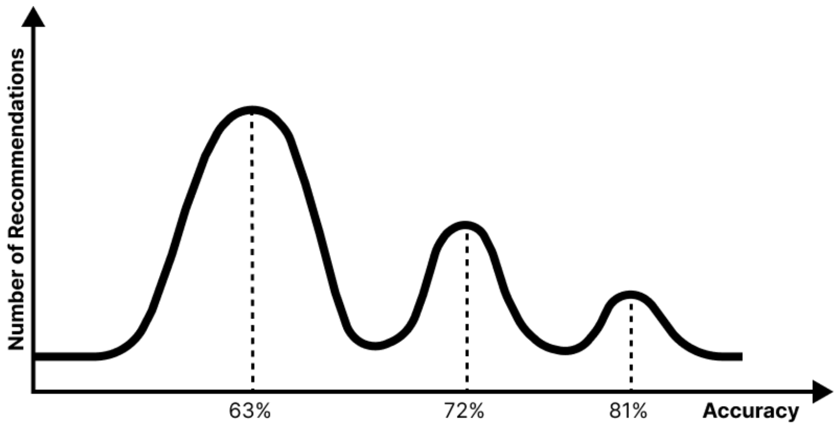

Figure 15 : Accuracy distribution of our system

For 0.7 to 0.79, it will be (0.7) *(0.9) =63%

For 0.8 to 0.89, it will be (0.8) *(0.9) =72%

For 0.9 to 1.00, it will be (0.9) * (0.9) =81%

Apart from the advantages discussed above, another huge advantage of our system is that there are 2 levels of processing happening with different levels of complexity. At one level, User clusters and Post categories are processed to generate recommendations in general for all clusters. This takes much less processing power but only makes sense to run after there has been significant interactions. Meanwhile, we can generate thresholds to fit even new users into these clusters and provide them with recommendations in real time as score generation and comparisons can be offloaded to client side.

The recommendation engine is equipped with the ability to analyse its own effectiveness and report it back to the system administrator every time it is run. This is reported in the form of hits/recommendations ratio which is calculated based on how many recommendations served to the user were actually clicked. Using this data, the thresholds can be reworked and experimented with till we have the most optimal results.

The price optimization system utilizes deep neural networks (DNNs) to predict discounts for products based on various attributes such as category, price range, and rating. It begins by generating random data for product details and preprocesses it for training. The neural network model, comprising four hidden layers, is defined using TensorFlow's Keras API and compiled with the Adam optimizer. Training is conducted over 50 epochs, with validation on 20% of the data. After training, new data is generated and used to predict discounts, demonstrating the system's ability to adapt pricing strategies in real-time. The predicted discounts are printed alongside product IDs, and the training history is visualized to assess model performance.

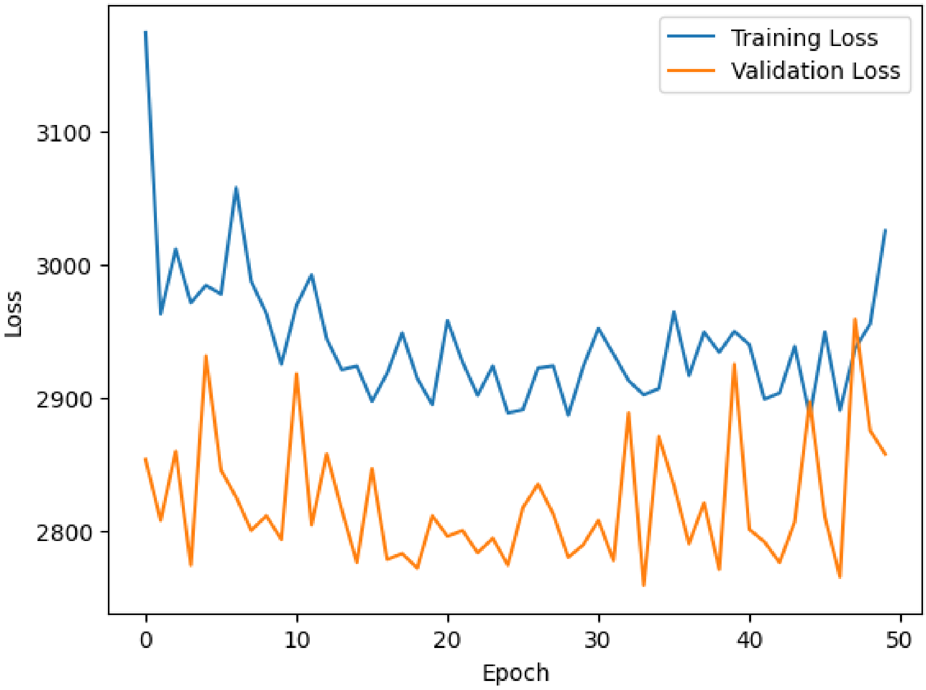

Figure 16: Training and Validation Loss Over Epochs for Price Optimization Model

Chapter 9

Conclusion

The emergence of the innovative pricing application marks a significant turning point in the realm of e-commerce strategy, signalling a departure from conventional approaches towards a more dynamic, data-centric, and customer-focused paradigm. At its core, the application embodies the principles of agility, adaptability, and responsiveness, enabling sellers to navigate the ever-evolving digital marketplace with precision and efficacy. By leveraging advanced algorithms and real-time data analytics, the application empowers sellers to make informed pricing decisions that resonate with consumer preferences and market trends.

One of the key features of the application is its ability to redefine traditional pricing methodologies. Instead of relying solely on static pricing models, the application introduces dynamic pricing strategies that adjust in real-time based on a multitude of factors, including consumer behavior, product shelf life, and inventory levels. This dynamic approach ensures that prices remain relevant and competitive, maximizing the potential for conversions while maintaining profitability for sellers. Moreover, the application places a strong emphasis on customer-centricity, recognizing that buyer satisfaction is paramount in driving long-term success in ecommerce. Through transparent pricing practices and enhanced information accessibility, buyers are empowered to make well-informed purchasing decisions that align with their preferences and priorities. Whether it's comparing prices across different seller platforms or evaluating trade terms and profit margins, buyers have access to the insights they need to make value-driven transactions. Furthermore, the application's targeted focus on time-sensitive pricing strategies for perishable goods and services highlights its strategic approach to maximizing impact on seller performance and buyer satisfaction. By recognizing the influence of time on purchasing decisions, the application enables sellers to capitalize on fleeting opportunities and optimize pricing strategies for maximum effectiveness. As e-commerce continues to evolve, the innovative pricing application stands as a beacon of progress and innovation, offering a glimpse into the future of digital commerce. Its ability to anticipate consumer preferences, adapt to changing market dynamics, and drive value for both sellers and buyers positions it as a trailblazer in shaping the future landscape of ecommerce. In essence, the application represents more than just a tool for pricing optimization—it is a catalyst for transformation, ushering in a new era of ecommerce innovation and excellence.

Chapter 10

Future Scope

Based on the innovative features and capabilities of the pricing application described above, the future scope is promising and multifaceted, encompassing various areas of advancement and expansion within the e-commerce landscape:

1. Enhanced Personalization: Future iterations of the pricing application can delve deeper into personalization by incorporating advanced machine learning algorithms to analyze individual customer preferences, behaviour patterns, and purchase history. By tailoring pricing strategies and recommendations at a granular level, sellers can further optimize conversions and enhance customer satisfaction.

2. Integration of Predictive Analytics: Integrating predictive analytics capabilities into the application can enable sellers to anticipate future market trends, demand fluctuations, and competitive dynamics. By leveraging predictive models, sellers can proactively adjust pricing strategies and inventory management to stay ahead of the curve and capitalize on emerging opportunities.

3. Expansion to New Verticals: While the initial focus may be on perishable goods and services, the application can expand its scope to cater to a broader range of verticals and industries. By adapting pricing strategies to suit the specific needs and characteristics of different product categories, the application can serve as a versatile tool for sellers across diverse sectors.

4. Global Market Penetration: With the increasing globalization of ecommerce, there is significant potential for the pricing application to penetrate global markets and cater to sellers operating on an international scale. By incorporating multi-currency support, localization features, and region-specific pricing optimization algorithms, the application can cater to the unique requirements of sellers in different geographical regions.

5. Integration with Emerging Technologies: As emerging technologies such as blockchain, augmented reality (AR), and virtual reality (VR) continue to reshape the ecommerce landscape, there is an opportunity to integrate these technologies into the pricing application to enhance the overall user experience. For example, blockchain can be leveraged to ensure transparency and security in pricing transactions, while AR and VR can provide immersive shopping experiences for buyers.

6. Partnerships and Collaborations: Collaborating with industry partners, ecommerce platforms, and technology providers can further enhance the capabilities and reach of the pricing application. By forging strategic partnerships, the application can access new markets, tap into additional data sources, and leverage complementary technologies to deliver greater value to sellers and buyers alike.

7. Continuous Innovation and Iteration: Finally, the future scope of the pricing application hinges on a commitment to continuous innovation and iteration. By soliciting feedback from users, monitoring industry trends, and staying abreast of technological advancements, the application can evolve and adapt to meet the evolving needs and expectations of the ecommerce ecosystem.

In summary, the future scope of the pricing application is characterized by expansion into new verticals, integration of advanced technologies, global market penetration, and a commitment to continuous innovation. By embracing these opportunities, the application can solidify its position as a leading solution for pricing optimization in the ever-evolving world of ecommerce.

References

1) Isinkaye, Folasade Olubusola, Yetunde O. Folajimi, and Bolande Adefowoke Ojokoh. "Recommendation systems: Principles, methods and evaluation." _Egyptian informatics journal_ 16.3 (2015): 261-273.
2) Son, Jieun, and Seoung Bum Kim. "Content-based filtering for recommendation systems using multiattribute networks." _Expert Systems with Applications_ 89 (2017): 404-412
3) Xia, Peipei, Li Zhang, and Fanzhang Li. "Learning similarity with cosine similarity ensemble." _Information sciences_ 307 (2015): 39-52.
4) Dehak, Najim, et al. "Cosine similarity scoring without score normalization techniques." _Odyssey_. 2010.
5) Li, Xin, Lei Guo, and Yihong Eric Zhao. "Tag-based social interest discovery." _Proceedings of the 17th international conference on World Wide Web_. 2008.
6) He, Xiangnan, et al. "Neural collaborative filtering." _Proceedings of the 26th international conference on world wide web_. 2017
7) Koren, Yehuda, Steffen Rendle, and Robert Bell. "Advances in collaborative filtering." _Recommender systems handbook_ (2021): 91-142

8) "Study of demand models and price optimization performance"- Georgia Institute of Technology, December 2011