WGU C951

Task 3

MACHINE LEARNING PROJECT PROPOSAL

Name: Joshua James

Student ID #: 011101737

Date: 11/29/2023

Table of Contents

A. Project Overview............................................................................................................... Page 3 A.1. Organizational Need ...................................................................................................... Page 3 A.2. Context and Background................................................................................................ Page 3 A.3. Outside Works Review .................................................................................................. Page 3 A.4. Solution Summary.......................................................................................................... Page 5 A.5. Machine Learning Benefits............................................................................................ Page 5 B. Machine Learning Project Design..................................................................................... Page 5 B.1. Scope ............................................................................................................................. Page 5 B.2. Goals, Objectives, and Deliverables .............................................................................. Page 6 B.3. Standard Methodology .................................................................................................. Page 7 B.4. Projected Timeline ......................................................................................................... Page 7 B.5. Resources and Costs....................................................................................................... Page 9 B.6. Evaluation Criteria ......................................................................................................... Page 9 C. Machine Learning Solution Design................................................................................. Page 10 C.1. Hypothesis.................................................................................................................... Page 10 C.2. Selected Algorithm....................................................................................................... Page 10 C.2.a Algorithm Justification............................................................................................... Page 10 C.2.a.i. Algorithm Advantage ............................................................................................. Page 11 C.2.a.ii. Algorithm Limitation ............................................................................................ Page 11 C.3. Tools and Environnent ................................................................................................ Page 11 C.4. Performance Measurement........................................................................................... Page 12 D. Description of Data Sets ................................................................................................. Page 12 D.1. Data Source ................................................................................................................. Page 12 D.2. Data Collection Method............................................................................................... Page 12 D.2.a.i. Data Collection Method Advantage........................................................................ Page 12 D.2.a.ii. Data Collection Method Limitation ...................................................................... Page 13 D.3. Data Normalization...................................................................................................... Page 13 D.4. Data Security................................................................................................................ Page 13 References............................................................................................................................ Page 14

**A. Project Overview**

**A.1. Organizational Need**

Currently, Fred’s Pet Supplies website recommends the same top 5 most popular products to every customer. Fred’s Pet Supplies believes they will be able to increase sales by using machine learning to make a product recommendation system that provides personalized recommendations for each customer.

**A.2. Context and Background**

Fred’s Pet Supplies has seen a dramatic increase in online orders due to recently being positively featured on a national morning show. While in the past, Fred’s Pet Supplies would feature the top 5 selling products as recommendations to each customer, the company believes they could increase their sales, especially with the dramatic recent increase in customers, by offering personalized product recommendations for each customer.

**A.3. Outside Works Review**

Multiple outside works were reviewed in preparation for this project to better understand current product recommendation systems and methodologies. The following three outside works related to using machine learning for product recommendations were reviewed:

1. “Product Recommendation System a Comprehensive Review” by Jatin Sharma, Kartikay Sharma, Kaustubh Garg, and Avinash Kumar Sharma (1)

This article provided a comprehensive review of product recommendation systems. The article first focuses on a product’s recommendation system’s filtering system, including content-based filtering, collaborative-based filtering, or hybrid filtering. The article then outlines popular algorithms commonly used by product recommendation systems. As product recommendation systems rely on machine learning, supervised, unsupervised, and reinforcement learning is discussed. The article provides several examples of companies currently using recommendation systems, including Amazon, Netflix, and Best Buy. Finally, the article concluded that the best product recommendation systems don’t rely solely on the most popular items but offer diverse and personalized recommendations.

1. “Toward Improving the Prediction Accuracy of Product Recommendation System Using Extreme Gradient Boosting and Encoding Approaches” by Zeinab Shahbazi, Debapriya Hazra, Sejoon Park and Yung Cheol Byun (2)

The article proposes how the XGBoost algorithm can be used to increase product recommendations accuracy when compared to other commonly used algorithms such as “Random Forest”, “Support Vector Regressor”, and “Linear Regression”. The XGBoost algorithm can “predict, classify, and optimize” product recommendations with the highest accuracy due to applying the collaborative filtering technique to both “classification and prediction results”. In other words, product recommendations are the nearest products based on their product prediction weight. The article concludes that out of multiple commonly used algorithms, based on the increased recommendation accuracy, the XGBoost algorithm should be the preferred choice for product recommendation systems.

1. “Product Recommendation Using Machine Learning a Review of the Existing Techniques” by Anam Naz Sodhar, Umair Ali Khan, Irum Naz Sodhar, Abdul Hafeez Buller, and Jahanzeb Sodhar1 (3)

The article summarized several existing machine learning product recommendation system techniques, including:

**Content-Based Filtering:** Shows recommendations based on the user’s profile information.

**Collaborative Filtering:** Shows recommendations by analyzing similarities between users’ behaviors and preferences. “User-User” collaborative filtering shows recommendations based on similar users, while “Item-Item” collaborative filtering shows recommendations based on similar items.

**Knowledge-based Recommendation Systems:** Shows recommendations based on specific queries made by a user. As an example, think of how when you visit a website selling houses, you typically must input certain information, such as city/location, specific number of rooms, bathrooms, and other house options, before being shown the matching houses.

**Hybrid Recommendation Systems:** Shows recommendations by combining content-based and collaborative-based predictions.

**A.4. Solution Summary**

A hybrid filtering product recommendation would best fit Fred’s Pet Supplies' desire to implement a product recommendation feature on their website. Hybrid filtering is a combination of both content-based and collaborative-based filtering. As an example, if a user purchased dog treats, the next time they visited Fred’s Pet Supplies website, they would be recommended a combination of similar items to the dog treats they purchased, as well as the best-selling or highest-rated dog treats. By using hybrid filtering, product recommendation accuracy is increased compared to using only content-based or collaborative-based filtering.

**A.5. Machine Learning Benefits**

Using machine learning will allow Fred’s Pet Supplies to provide personalized product recommendations for each user. These recommendations will also continue to update over time and become more and more accurate, as more customer information and products are added, all without the need for human intervention. Attempting to do this without machine learning would be virtually impossible without an unrealistically large investment of resources, and even then, would likely be a significantly slower and less accurate process.

**B. Machine Learning Project Design**

**B.1. Scope**

The scope of our proposed solution for the new recommender system for Fred’s Pet Supplies will cover the following:

1. Collect customer information, including customer profile information, transactions, search and view histories.
2. Design a recommender service that will provide personalized product recommendations to customers who visit Fred’s Pet Supplies' website.
3. Implement the recommender service to Fred’s Pet Supplies website with minimal system downtime.

The following is considered “out of scope” for our proposed solution for the new recommender system for Fred’s Pet Supplies:

1. No physical equipment, such as computers, servers, network cables, etc., will be replaced, updated, or upgraded during this project.

**B.2. Goals, Objectives, and Deliverables**

**Goals**

* Offer personalized product recommendations to customers who visit Fred’s Pet Supplies’ website.

**Objectives**

* Design and implement a product recommender system for Fred’s Pet Supplies’ website that offers a minimum of 5 and a maximum of 20 product recommendations, which are personalized for each customer who visits the website.
* Due to the introduction of personalized product recommendations, Fred’s Pet Supplies expects to see sales increase by at least 20 percent in the first six months.
* Increased customer satisfaction due to enhancing customer experience with personalized product recommendations.

**Deliverables**

* New product recommender system designed and deployed on Fred’s Pet Supplies website.

**B.3. Standard Methodology**

The development of our proposed recommender system will follow the commonly used SEMMA methodology (4).

* **Sample:** A representative sample of customer information, including customer profiles, transactions, and search and view histories, is extracted.
* **Explore:** The customer information data is analyzed to help identify trends and groupings that will assist in making accurate product recommendations.
* **Modify:** The customer information data is increased, and organized into groups and sub-groups, to provide more recommendations and recommendation variations.
* **Model:** The customer information data is data mined and is used to train the product recommendation system.
* **Assess:** The product recommendation system will be tested to ensure it is meeting the before-stated objectives, including recommending a minimum of 5 and a maximum of 20 product recommendations, that are personalized for each customer.

**B.4. Projected Timeline**

While the project timeline may change based on the proposal acceptance date, team size, budget, etc., please see below for our current projected timeline for our proposed recommender system:

|  |  |  |
| --- | --- | --- |
| **Date** | **Description** | **Length** |
| January 1, 2024 | Fred Pet Supplies’ product recommender proposal is accepted. Project official start date. | 1 Day |
| January 2, 2024 | Collect and prepare customer information data, including customer profile information, transaction history, and search and view histories. | 2 Weeks |
| January 17, 2024 | Design and develop product recommender system. | 4 Weeks |
| February 18, 2024 | Use customer information data for training and testing. | 6 Weeks |
| April 1, 2024 | User testing is performed and feedback is obtained. | 2 Weeks |
| April 16, 2024 | Based on the feedback received, the product recommender system is updated and optimized. | 2 Weeks |
| May 1, 2024 | The product recommender system will be added to Fred’s Pet Supplies’ website. Its performance will be reviewed and evaluated to ensure it is functioning fully in its live environment. | 2 Weeks |
| May 16, 2024 | End-of-project reporting is performed. | 1 Week |
| May 24, 2024 | The project is officially completed. | 1 Day |

**B.5. Resources and Costs**

|  |  |  |
| --- | --- | --- |
| **Resource** | **Description** | **Cost** |
| Design and Development | Labor costs include the hiring of a project manager, UI/UX designers, and software developers. | $25,000 |
| Development Tools | The cost of the IDE and development tools used to develop the proposed product recommender system. | $750 |
| Third-Part Cloud Service | Monthly cost to store customer information through third-party cloud service. | $1,000 per month |
| Machine Learning Services and Server | Machine learning service costs that include the training and tuning of real-time, personalized product recommendations. | $4,250 per month |
| Software Testing | The overall cost for software testing. | $4,000 |
|  | **Total** | $35,000 |

**B.6. Evaluation Criteria**

|  |  |
| --- | --- |
| **Objective** | **Success Criteria** |
| New product recommender system | Design and implement a product recommender system for Fred’s Pet Supplies’ website that offers a minimum of 5 and a maximum of 20 product recommendations, which are personalized for each customer who visits the website. |
| Increase Sales | Due to the introduction of personalized product recommendations, Fred’s Pet Supplies expects to see sales increase by at least 20 percent in the first six months. |
| Increase Customer Satisfaction | Increased customer satisfaction due to enhancing customer experience with personalized product recommendations. This will be monitored through tracking online and social media sentiment, as well as traditional customer surveys and feedback groups. |

**C. Machine Learning Solution Design**

**C.1. Hypothesis**

By introducing a new recommender system that uses machine learning to provide personalized product recommendations to customers, Fred’s Pet Supplies will see an increase in sales and customer satisfaction. This is due to customers seeing more relevant recommendations, which positively affects their customer experience, and leads to increasing both customer satisfaction and sales. Sales will be monitored and tracked to show a positive increase after the new product recommendation system is deployed to the website. Customer satisfaction will be tested through user experience tests, tracking online and social media sentiment, along with traditional customer surveys, and feedback sessions.

**C.2. Selected Algorithm**

LightFM is one of the most popular hybrid algorithms and is part of the Python library. Hybrid filtering uses a combination of both content-based and collaborative-based filtering. For this project, the product recommender system using hybrid filtering, and a hybrid algorithm like LightFM would show a combination of items that a specific user may be interested in, as well as the most popular items (the items most “liked” by users).

**C.2.a Algorithm Justification**

The LightFM hybrid algorithm used for this project’s hybrid filtering product recommender system is an unsupervised machine learning algorithm. Unsupervised machine learning can learn and make connections from data without any human intervention (5). In this project, the LightFM hybrid algorithm combines customer information, such as their customer profile information, search history, view history, and transaction data, as well as the most well-liked items from all users, to make personalized product recommendations for each customer.

**C.2.a.i. Algorithm Advantage**

Hybrid filtering, like the LightFM hybrid algorithm used for this project, is more accurate when compared to using either content-based filtering or collaborative-based filtering. This is due to Hybrid filtering combining content-based filtering and collaborative-based filtering, which leads to making better and more accurate recommendations. Another advantage that Hybrid filtering offers is a solution to the common “cold start” problem collaborative filtering recommender systems commonly encounter. An example of a “cold start” problem, would be a brand-new visitor to Fred’s Pet Supplies website. Since the brand-new visitor has no interactions or data, a collaborative filtering recommender system does not have any information to make recommendations with. Hybrid filtering fixes this problem by using both collaborative-based and content-based filtering, meaning a brand-new visitor to the website would see content-based recommendations (the most popular or most liked items).

**C.2.a.ii. Algorithm Limitation**

Hybrid filtering is computationally more complex and taxing than using only collaborative-based filtering or content-based filtering. This leads to hybrid filtering being more expensive in both costs and resources.

**C.3. Tools and Environment**

The proposed hybrid filtering recommender system will be developed using the Windows operating system. Python will be the programming language used, as it is commonly used for machine learning. Finally, the LightFM algorithm used does not require any sort of license but will need to be obtained and incorporated into the hybrid filtering recommender system.

**C.4. Performance Measurement**

Quality and performance for the proposed recommender system project will be measured in several ways. First, the recommender system should always display a minimum of 5 recommendations and a maximum of 20 recommendations, per the project’s objectives. Second, due to increased personalization in recommendations, the company should see a 20 percent increase in sales during the first six months after the recommender system’s deployment. Third, customer satisfaction should increase based on enhancing the customer experience through personalized recommendations. Finally, as more products and customers are added to the recommender system’s dataset, the accuracy of recommendations should continue to increase and become more individualized and personalized.

**D. Description of Data Sets**

**D.1. Data Source**

The proposed recommender system’s data source is the customer information provided by the company. The customer information would include information such as customer profile information, search history, view history, and transaction data.

**D.2. Data Collection Method**

Customer information is collected by Fred’s Pet Supplies when customers visit and use their website. The company has provided access to this information. This customer information will be used to train and test our proposed recommender system before being deployed to the live website.

**D.2.a.i. Data Collection Method Advantage**

An advantage of the collection of customer information is that in this case, it is automatically done without any additional work needed to be performed by the customer. In addition, the more customer information obtained, the more accurate and personalized recommendations will become, thus further enhancing the customer experience and overall sales.

**D.2.a.ii. Data Collection Method Limitation**

A disadvantage of the collection of customer information is that in this case, a customer could potentially change their cookie settings and disable the ability for the company to collect certain information. This would limit the recommender system’s ability to provide accurate and personalized recommendations.

**D.3. Quality and Completeness of Data**

There are several steps that the customer information data will go through before it can be used to train and test our proposed recommender system. First, the customer information data must be “cleaned”. In other words, incomplete customer information data will either be completed or removed from the dataset, as well as any removing any data that is causing random errors. This is done to increase the accuracy of the recommendation results. The customer data is then integrated and reduced to only what’s needed by the recommender system. These steps are performed to decrease the computational time needed. Finally, the customer information data is transformed so that it can be used to train and test our recommender system.

**D.4. Precautions for Sensitive Data**

As our recommender system relies on sensitive customer information, the utmost precautions will be taken to safeguard this data. These steps include:

* Data encryption of sensitive data.
* Strong password policies.
* Limit access to customer information to only those who need it to complete the proposed recommender system project.
* Project communication must be performed through official channels.
* Work must be performed on company devices, and cannot be done on personal devices.

**References**

1. Sharma, Jatin, et al. "Product recommendation system a comprehensive review." IOP conference series: materials science and engineering. Vol. 1022. No. 1. IOP Publishing, 2021.
2. Shahbazi, Zeinab, et al. "Toward improving the prediction accuracy of product recommendation system using extreme gradient boosting and encoding approaches." Symmetry 12.9 (2020): 1566.
3. Sodhar, Anam Naz, et al. "Product Recommendation Using Machine Learning a Review of the Existing Techniques." IJCSNS 22.5 (2022): 523.
4. “Data Mining Process: Models, Process Steps & Challenges Involved.” Software Testing Help, <https://www.softwaretestinghelp.com/data-mining-process/#Data_Mining_Models>. Accessed 29 Nov 2023.
5. “What is unsupervised learning?” Google Cloud, <https://cloud.google.com/discover/what-is-unsupervised-learning#:~:text=Unsupervised%20learning%20in%20artificial%20intelligence,any%20explicit%20guidance%20or%20instruction>. Accessed 29 Nov 23.