Creating an ML-based Product Recommendation System using Content-Based Filtering Algorithms

MSCI 446 - Project Proposal

February 14, 2023

Group 9

Benjamin Luo (b33luo)	20890448
Celine Tran (cyptran)	20759583
Michael Lam (mswlam)	20762462
Runyi (Harry) Yao (r28yao)	20822000

1.0 Introduction

The online retail market, or e-commerce, is a $6.3 billion dollar industry that is estimated to grow by 10.4% [1]. Dr. Katherine Taken Smith of Murray State University states that "the value of e-commerce includes its instrumental role in the global marketplace, the evolution of virtual businesses, and the unique opportunities it provides for linking marketers with consumers" [2]. The emergence of e-commerce has led to the globalization of direct consumer retail and decreased the trade barriers between distant retail merchants [2].

Many e-commerce websites have recommendation systems personalized for users when they are browsing or purchasing a product. The implementation of recommendations systems helps customers by "[reducing] consumer search costs and uncertainty associated with the purchase of unfamiliar products" [3]. Furthermore, it was found that "recommendations not only improve sales, but they also provide added flexibility to retailers to adjust their prices" [3]. This increase in sales implies that a robust recommendation system is a strong asset to have in the online retail market.

When discussing recommendation systems, there are two main categories: collaborative filtering and content filtering [4]. Collaborative filtering creates recommendations based on user data and can be thought of as a word-of-mouth recommendation between two people. In contrast, content filtering recommendation systems make recommendations based on characteristics of the product. For this project, a content-based filtering recommendation system will be implemented to analyze the effectiveness of purely content-based algorithms on a large dataset.

The objective of this project is to gain business insights in game products sold on Amazon; more specifically, to evaluate the performance of content-based filtering recommendation models on a market environment with new users who do not have established behaviours, past reviews, or purchase data. As such, unsupervised machine learning will be used to find products that are similar or related to a product that a customer has recently looked at.

As the dataset contains mostly text-based information, the data will be processed in the form of document clustering with the varying categories: product name, categories, and description. By clustering documents, or in this case, products, the model will be able to find relevant products with similar characteristics or features to the products that the user is interested in. However, the data first needs to be cleaned with text processing (e.g. stemming, removing stop words) for proper retrieval.

2.0 Problem Definition & Objective

Since the onset of the COVID-19 pandemic, lockdowns and social distancing measures have turned many consumers towards online shopping to meet their needs. According to the U.S. Department of Commerce, online sales in the U.S. grew by 44.5% in the second quarter of 2020 [5]. This statistic was matched worldwide, with e-commerce sales increasing by 27.6% overall [6]. Following this increase in popularity, e-commerce sites received a surge in new customers who had not shopped online previously. A end-of-year 2020 survey performed by Salesforce reported that 36% of U.S. consumers tried a new online shopping app or website for the first time during the pandemic [7]. Moreover, the pandemic also led to a shift in consumer behaviour, with people choosing to shop online for a wider range of products, such as groceries, personal care items, and household goods.

As of 2021, Amazon is the world's largest online marketplace, with a capitalization of over $1.6 trillion [8]. The company has a dominant share in the U.S. market, accounting for over 40% of all online sales in 2020 [9]. Its comparative standing is achieved by its immense buying power: Amazon purchases products from manufacturers in immense quantities, allowing more flexibility in price. The undercutting of the traditional retail prices has caused many brick-and-mortar stores to close, dislocating an increasing number of customers from physical retail to e-commerce.

Much like other e-commerce giants such as eBay and Walmart, Amazon relies on machine learning algorithms to provide personalized product recommendations to their customers. The impact of machine learning models for product recommendations in e-commerce is significant. In a 2018 study by Barilliance, a developer of e-commerce recommendation machine learning models, personalized product recommendations accounts for up to 31% of e-commerce site revenue [10]. Another 2018 study by Epsilon, a global marketing company, indicated that approximately 80% of customers were more likely to make a purchase if the site offered personalized recommendations [11]. Through ongoing improvements in the accuracy and relevance of product recommendation algorithms, recent models are likely to further bolster customer engagement, reduce bounce rates, and drive more sales for e-commerce sites. Lastly, product personalization was found to deliver five to eight times the return on investment on marketing spend, and lift sales by more than 10%, on average [12]. Market Research Future estimates that the global market for machine learning in e-commerce will grow at a compound annual growth rate (CAGR) of 38.8% between 2020 and 2030 [13]. Ultimately, recommendation machine learning models typically lead to higher revenue, increased customer retention, and a better overall shopping experience for customers.

E-commerce models are trained and regularly updated on large amounts of customer data, including customer behavior and past purchase history, to generate personalized product recommendations. The more data the model has access to, the more accurate and personalized the recommendations can be. Fortunately, the recent influx of online shoppers due to the pandemic has introduced numerous data points that not only describe the e-commerce regulars, but also the behaviour of new consumers as well as a general shift in product range compared to the pre-pandemic environment.

Recommendation models can use various algorithms such as collaborative filtering, content-based filtering, and hybrid filtering to suggest products that a customer is more likely to purchase based on their preferences and behavior [14]. Of these algorithms, content-based filtering has key advantages in the current data climate. The cold start problem occurs when new consumers with no historical data interact with the system. Content-based filtering can overcome this by using product features to recommend items that are similar to ones that a new user has interacted with. Moreover, content-based filtering has reduced dependence on user ratings, unlike collaborative filtering, which relies on user ratings and reviews. Content-based filtering generates recommendations based solely on the characteristics of a product, such as its attributes and features. Consequently, content-based filtering can recommend products that are new to the market, or have not yet been rated by customers, making it a useful approach for Amazon, an e-commerce site that offers a wide variety of products.

One limitation of current machine learning models for product recommendations is their reliance on past behavior and preferences while using the site [14]. This often leads to a lack of serendipity and discovery for customers, resulting in lower satisfaction and engagement. As such, the goal of this project is to develop a machine learning model that offers customers personalized game recommendations on Amazon based on the product's name, categories, and description. By utilizing content-based filtering, the model can provide recommendations that circumvent the cold start problem and are new or niche products that have little to no user ratings and reviews. Thus, the model aims to provide business insights into the accuracy, sensitivity, and adaptability of content-based filtering algorithms in an e-commerce environment largely represented by new users, described by a shift in consumer behaviour with regards to a specific product range – a representation of the current post-pandemic environment.

3.0 Dataset Description

The dataset is collected from *data.world*, which is a platform for hosting, managing, sharing and collaborating on datasets. The dataset is publicly accessible using the link below:

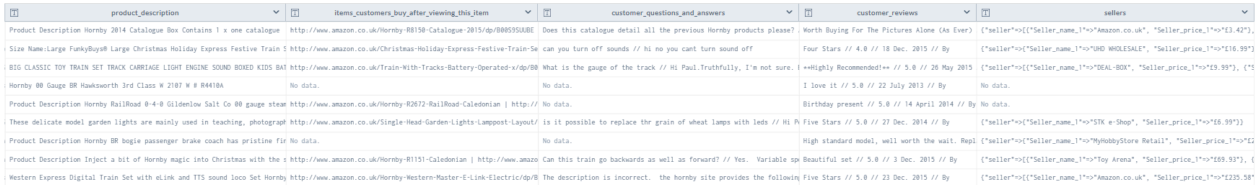
https://data.world/promptcloud/fashion-products-on-amazon-com

PromptCloud's web-crawling service created a pre-crawled dataset of 22,000 game products from Amazon, which is a subset of a larger dataset of over 7 million game products. By analyzing ratings, prices, and reviews, the dataset will be utilized to determine products that are similar and likely to be of interest to potential buyers. This dataset contains detailed information about products on Amazon.com, including product details, prices, description, sellers, etc. It has 17 columns and 10,000 rows in total.

Here are a few sample rows from the dataset:

	uniq_id	product_name	manufacturer	price	number_available_in_stock	number_of_reviews	number_of_answered_questions	average_review_rating	amazon_category_and_sub_category
1	eac7efa5dbd3d667f26eb3d3ab504464	Hornby 2014 Catalogue	Hornby	£3.42	5 new	15	1	4.9 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
2	b17540ef7e86e461d37f3ae58b7b72ac	FunkyBuys® Large Christm	FunkyBuys	£16.99	No data.	2	1	4.5 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
3	348f344247b0c1a935b1223072ef9d8a	CLASSIC TOY TRAIN SET TR	ccf	£9.99	2 new	17	2	3.9 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
4	e12b92dbb8eaee78b22965d2a9bbbd9f	HORNBY Coach R4410A BR H	Hornby	£39.99	No data.	1	2	5.0 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
5	e33a9adeed5f36840ccc227db4682a36	Hornby 00 Gauge 0-4-0 Gi	Hornby	£32.19	No data.	3	2	4.7 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
6	cb34f0a84102c1ebc3ef6892d7444d36	20pcs Model Garden Light	Generic	£6.99	No data.	2	1	5.0 out of 5 stars	Hobbies > Model Trains & Railway Sets > Lighting & Signal Enginee
7	f74b562470571dfb689324adf236f82c	Hornby 00 Gauge 230mm BR	Hornby	£24.99	No data.	2	1	4.5 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
8	87bbb472ef9d90dcef140a551665c929	Hornby Santa's Express T	Hornby	£69.93	3 new	36	7	4.3 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains
9	7e2aa2b4596a39ba852449718413d7cc	Hornby Gauge Western Exp	Hornby	£235.58	4 new	1	1	5.0 out of 5 stars	Hobbies > Model Trains & Railway Sets > Rail Vehicles > Trains

	customers_who_bought_this_item_also_bought	description	product_information
1	http://www.amazon.co.uk/Hornby-R8150-Catalogue-2015/dp/B00S9SUUBE	Product Description Hornby 2014 Catalogue Box Contains 1 x one catalogue	Technical Details Item Weight640 g Product Dimensions29.6 x 20.8 x 1 cm Manufacturer recommended age:6 years
2	http://www.amazon.co.uk/Christmas-Holiday-Express-Festive-Train-Se	Size Name:Large FunkyBuys® Large Christmas Holiday Express Festive Train Se	Technical Details Manufacturer recommended age:3 years and up Item model numberSI-TY1017-B Additional Info
3	http://www.amazon.co.uk/Classic-Train-Lights-Battery-Operated/dp/B	BIG CLASSIC TOY TRAIN SET TRACK CARRIAGE LIGHT ENGINE SOUND BOXED KIDS BAT	Technical Details Manufacturer recommended age:3 years and up Additional Information ASINB00E5MNXJ4 Best S
4	No data.	Hornby 00 Gauge BR Hawksworth 3rd Class W 2107 W # R4410A	Technical Details Item Weight259 g Product Dimensions31.6 x 9.2 x 4.6 cm Manufacturer recommended age:3 years
5	http://www.amazon.co.uk/Hornby-R6367-RailRoad-Gauge-Rolling/dp/B00	Product Description Hornby RailRoad 0-4-0 Gildenlow Salt Co 00 gauge steam	Technical Details Item Weight159 g Product Dimensions18.4 x 10.2 x 6 cm Manufacturer recommended age:4 years
6	http://www.amazon.co.uk/Single-Head-Garden-Lights-Lamppost-Layout/	These delicate model garden lights are mainly used in teaching, photography	Technical Details Manufacturer recommended age:3 years and up Additional Information ASINB008MRH3Y4 Best S
7	http://www.amazon.co.uk/Hornby-R4388-RailRoad-Composite-Gauge/dp/B	Product Description Hornby BR bogie passenger brake coach has pristine fin	Technical Details Item Weight222 g Product Dimensions31 x 9.2 x 4.6 cm Manufacturer recommended age:3 years a
8	http://www.amazon.co.uk/Hornby-R8221-Gauge-Track-Extension/dp/B000	Product Description Inject a bit of Hornby magic into Christmas with the sp	Technical Details Item Weight1.2 Kg Product Dimensions40 x 29.8 x 8 cm Manufacturer recommended age:8 years a
9	http://www.amazon.co.uk/Hornby-Western-Master-E-Link-Electric/dp/B	Western Express Digital Train Set with eLink and TTS sound loco Set Hornby	Technical Details Item Weight2.3 Kg Product Dimensions80.4 x 24 x 8.2 cm Manufacturer recommended age:8 years



product_description	items_customers_buy_after_viewing_this_item	customer_questions_and_answers	customer_reviews	sellers
Product Description Hornby 2014 Catalogue Box Contains 1 x one catalogue	http://www.amazon.co.uk/Hornby-R8150-Catalogue-2015/dp/B00S9SUUBE	Does this catalogue detail all the previous Hornby products please?	Worth Buying For The Pictures Alone (As Ever)	{"seller"=>[{"Seller_name_1"=>"Amazon.co.uk", "Seller_price_1"=>"£3.42"},
Size Name:Large FunkyBuys® Large Christmas Holiday Express Festive Train S	http://www.amazon.co.uk/Christmas-Holiday-Express-Festive-Train-Se	can you turn off sounds // hi no you cant turn sound off	Four Stars // 4.0 // 18 Dec. 2015 // By	{"seller"=>{"Seller_name_1"=>"UHD WHOLESALE", "Seller_price_1"=>"£16.99"}
BIG CLASSIC TOY TRAIN SET TRACK CARRIAGE LIGHT ENGINE SOUND BOXED KIDS BAT	http://www.amazon.co.uk/Train-With-Tracks-Battery-Operated-x/dp/B0	What is the gauge of the track // Hi Paul.Truthfully, I'm not sure.	**Highly Recommended!** // 5.0 // 26 May 2015	{"seller"=>[{"Seller_name_1"=>"DEAL-BOX", "Seller_price_1"=>"£9.99"}, {"S
Hornby 00 Gauge BR Hawksworth 3rd Class W 2107 W # R4410A	No data.	No data.	I love it // 5.0 // 22 July 2013 // By	No data.
Product Description Hornby RailRoad 0-4-0 Gildenlow Salt Co 00 gauge steam	http://www.amazon.co.uk/Hornby-R2672-RailRoad-Caledonian | http://	No data.	Birthday present // 5.0 // 14 April 2014 // By	No data.
These delicate model garden lights are mainly used in teaching, photograph	http://www.amazon.co.uk/Single-Head-Garden-Lights-Lamppost-Layout/	is it possible to replace thr grain of wheat lamps with leds // Hi P	Five Stars // 5.0 // 27 Dec. 2014 // By	{"seller"=>{"Seller_name_1"=>"STK e-Shop", "Seller_price_1"=>"£6.99"}}
Product Description Hornby BR bogie passenger brake coach has pristine fir	No data.	No data.	High standard model, well worth the wait. Repl	{"seller"=>{"Seller_name_1"=>"MyHobbyStore Retail", "Seller_price_1"=>"£2
Product Description Inject a bit of Hornby magic into Christmas with the s	http://www.amazon.co.uk/Hornby-R1151-Caledonian | http://www.amazo	Can this train go backwards as well as forward? // Yes. Variable sp	Beautiful set // 5.0 // 3 Dec. 2015 // By	{"seller"=>[{"Seller_name_1"=>"Toy Arena", "Seller_price_1"=>"£69.93"}, {
Western Express Digital Train Set with eLink and TTS sound loco Set Hornby	http://www.amazon.co.uk/Hornby-Western-Master-E-Link-Electric/dp/B	The description is incorrect. the hornby site provides the followin	Five Stars // 5.0 // 23 Dec. 2015 // By	{"seller"=>[{"Seller_name_1"=>"Amazon.co.uk", "Seller_price_1"=>"£235.58"

Figure 1: Sample rows from dataset.

The column names are (the bolded text is the datatype of each column):

1. uniq id (**string**): a unique identifier for each product
2. product name (**string**): the name of the product
3. manufacturer (**string**): the company that makes the product
4. price (**string**): the price of the product (in Euro)
5. number available in stock (**string**): the number of units available in stock
6. number of reviews (**integer**): the total number of reviews for the product
7. number of answered questions (**integer**): the number of questions about the product that have been answered
8. average review rating (**string**): the average rating (in stars) for the product
9. amazon category and sub category (**string**): the Amazon category and subcategory for the product
10. customers who bought this item also bought (**string**): a link to other products that customers who bought this item also bought
11. description (**string**): a brief description of the product
12. product information (**string**): additional information about the product which includes the item weight, product dimensions, recommended age, etc.
13. product description (**string**): a longer description of the product
14. items customers buy after viewing this item (**string**): a link to other products that customers bought after viewing this product
15. customer questions and answers (**string**): questions from customers about the product and answers to the question
16. customer reviews (**string**): reviews of the product
17. sellers (**string**): information about the sellers of the product, including their names and price of related product

4.0 Machine Learning

The objective of the machine learning (ML) model is to identify similar products [class label] to a given product. If users are interested in a given item, then they will also be interested in a product with similar features or price point. Consumer preferences are subjective in that there are innumerable variables when discerning pathways in which humans make decisions, and these variables vary considerably between people. For example, a person may only be interested in the parody of which a product is based off (e.g. Star Wars) instead of the product itself (e.g. Lego). The user's age group, gender, location, and external influences may also play a role, but this data is difficult to obtain without administrative access to an e-commerce website. Therefore, the feature selection is targeted toward identifying similarities between products rather than patterns in consumer demographics.

The most important features for the initial implementation are:

- *product_name*: The name of the product
- *amazon_category_and_sub_category*: The product's category
- *description*: A explanation of what the product is and its specifications

The returned data is a list of *product_name* titles from the dataset that best match the user inputted data.

4.1 Data Collection

The dataset(s) will contain information of the product name, description, and ideally the product's category. Due to the diverse range of products, any selected datasets must be large enough to ensure reasonable matches can be identified.

The selected Amazon dataset on products in the 'Games' section satisfies both requirements and contains additional features on review data, which can be analyzed to ensure the items have a positive user experience prior to recommending them to prospective buyers. Further, because it is constrained to the 'Games' section, the 10,000 data points will likely form more recognizable clusters of products such as 'Lego,' 'Puzzles,' and 'Stuffed Animals,'' consequently producing more specific and accurate recommendations.

4.2 Data Cleaning

Provided that the initial analysis will focus on only three features (title, description, category), the remaining columns can be removed. All three of these features are text-based, and as such, they must be decomposed to reduce noise. Pre-processing can be done through stemming, lemmatization, and removing redundant words/symbols/whitespace, and then tokenizing the remaining 'key words.' To avoid overfitting and an overly intricate model, only the top *n* 'key words' will be kept. For example, keeping 'Lego, Star Wars, Children' while excluding 'Easy, Toy, Fun.'

Afterward, three text-processing methods can be attempted to assign numeric values to the text: bag of words, term frequency, and inverse document frequency. All three will be considered and compared for their accuracy.

In later stages of the project, additional features may be considered to assess and increase accuracy. In particular, the manufacturer, price, number of reviews, ratings, and user reviews. Similar text processing would be done to the 'reviews' and numeric features would be scaled accordingly.

4.3 Algorithm Selection

The machine learning component will focus on clustering algorithms to categorize similar products based on their text features. Clustering will be used, as the algorithm is trying to predict a categorical value (a list of recommended products) and the dataset is labelled. Specifically, K-Means clustering will be used, due to the large size of the dataset (~10K rows) and because the number of categories can be fixed.

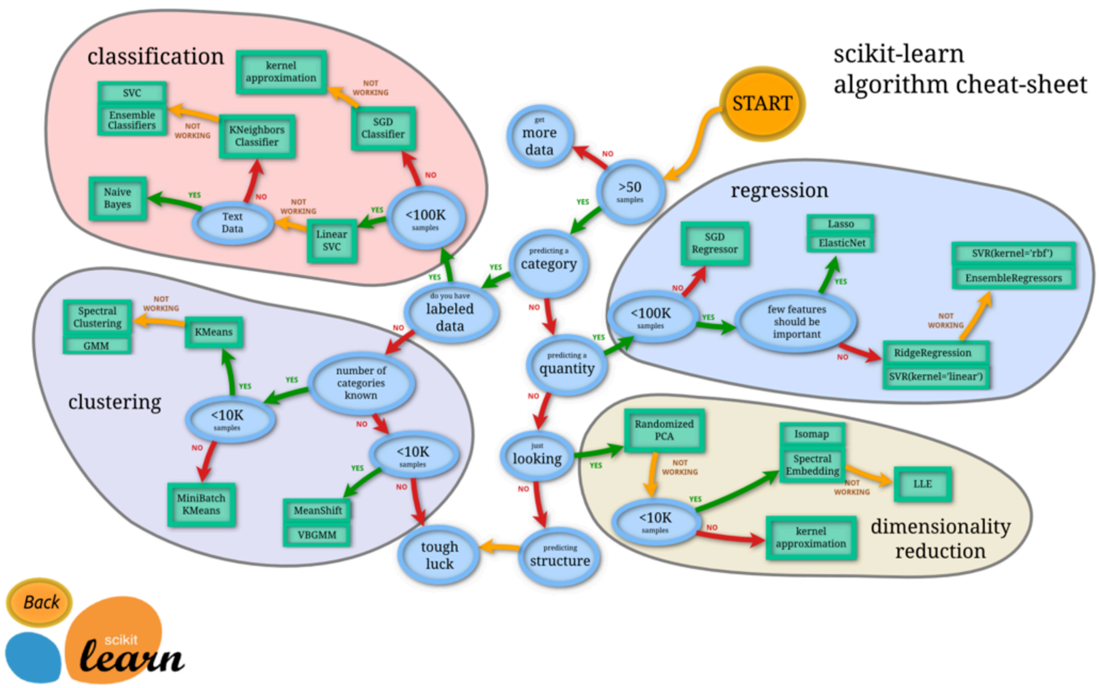


Figure 2: Scikit-learn algorithm selection guide.

4.4 K-Means Clustering

K-Means clustering is an unsupervised machine learning method that clusters data points into 'K' number of groups based on similarities between their features. In this context, the clusters would represent subcategories of products such as 'Lego products'. If higher 'K' values are selected, then the subcategories would become more specific (while risking overfitting). An instance of this would be 'Lego' branching into 'Star Wars Lego,' or in a case of overfitting, 'DIY Lightsaber Lego.'.

The mathematical objective of the K-means algorithm is to minimize the sum of squares distances between each point and its cluster mean. For the analysis of textual data, pre-processing is required to extract the key words from each product and convert them into a vector. When mapped onto a graph, each key word would represent a dimension with the magnitude being a numeric value, such as its frequency. For instance, a graph may have categorical values as the axes, and the vectors *a*: [Child, Lego, Star Wars], *b*: [Child, Stuffed Animals, Dog], and *c*: [Teenager, Lego, Avengers].

Clusters in this graph can be roughly identified as 'Child' and 'Lego.' The algorithm would determine these groupings by arbitrarily assigning mean values and recursively optimizing their placement until the squared distance metric is minimized.

The distance can be calculated exactly through a Euclidean distance or as a Manhattan distance. The latter is most likely more reasonable for this form of categorical data, as it has no ordinal value. For the above example, vectors *a* and *b* have a distance of two. Data points that are closer together have higher correlation and in the real-world context, would indicate that they are likely related products that can be recommended by the algorithm.

4.5 Model Evaluation

The dataset contains two crucial features generated by Amazon's suggestion engine that can be used to evaluate the model:

- *customers_who_bought_this_item_also_bought*
- *items_customers_buy_after_viewing_this_item*

Ideally, the recommendations outputted by the algorithm should be located in the same cluster(s) as the Amazon-recommended items held within these two features. Therefore, accuracy can be approximated as the distance between the predicted label(s) and the actual recommendations in the cluster.

5.0 Bibliography

[1] Oberlo. "Global Ecommerce Sales Growth (2021–2026) [Oct 2022 Update]," 2022. https://www.oberlo.ca/statistics/global-ecommerce-sales-growth.

[2] Smith, K. T. "Consumer Perceptions Regarding E-Commerce and Related Risk," *EBSCO Industries*, 2011.

[3] Pathak, B., et al. "Empirical Analysis of the Impact of Recommender Systems on Sales." *Journal of Management Information Systems*, vol. 27, no. 2, pp. 159–88, 2010, doi: 10.2753/MIS0742-1222270205.

[4] Ansari, A., et al. "Internet Recommendation Systems." *Journal of Marketing Research*, vol. 37, no. 3, pp. 363–75, 2000, doi: 10.1509/jmkr.37.3.363.18779.

[5] Brewster, M. "Annual Retail Trade Survey Shows Impact of Online Shopping on Retail Sales During COVID-19 Pandemic." United States Census Bureau. Apr. 2022. https://www.census.gov/library/stories/2022/04/ecommerce-sales-surged-during-pandemic.html.

[6] Cramer-Floor, E. "Worldwide ecommerce will approach $5 trillion this year." *Global Ecommerce Update 2021*. Insider Intelligence. Jan. 2021. https://www.insiderintelligence.com/content/worldwide-ecommerce-will-approach-5-trillion-this-year

[7] Salesforce. "Salesforce Data Shows U.S. Digital Sales Up 36% in December as Retailers Close Out Virtual Holiday Season." *News & Insights*. Dec. 2020. https://www.salesforce.com/news/stories/salesforce-data-shows-u-s-digital-sales-up-36-in-december-as-retailers-close-out-virtual-holiday-season/

[8] Companies Marketcap. "Market capitalization of Amazon (AMZN)." https://companiesmarketcap.com/amazon/marketcap/ (accessed Feb. 2023).

[9] Lebow, S. "Amazon clobbers competition, accounting for over 40% of US retail ecommerce sales in 2021." Insider Intelligence. Aug. 2021. https://www.insiderintelligence.com/content/amazon-clobbers-competition-us-retail-ecommerce-sales-2021.

[10] Serrano, S. "[Guide] Personalized Product Recommendations Tactics for Profits." Barilliance. Sept. 2021. https://www.barilliance.com/personalized-product-recommendations-stats/.

[11] Epsilon. "New Epsilon research indicates 80% of consumers are more likely to make a purchase when brands offer personalized experiences." Jan. 2018. https://www.epsilon.com/us/about-us/pressroom/new-epsilon-research-indicates-80-of-consumers-are-more-likely-to-make-a-purchase-when-brands-offer-personalized-experiences.

[12] Ariker, M., Heller, J., Diaz, A. & Perrey, J. "How Marketers Can Personalize at Scale." Harvard Business Review. Nov. 2015. https://hbr.org/2015/11/how-marketers-can-personalize-at-scale.

[13] Market Research Future. "Machine Learning Market Research Report: by Component (Hardware, Software), Organization Size (Large Enterprise, Small and Medium-Sized Enterprises) Vertical (BFSI, Media and Entertainment, Automotive, Telecommunication, Retail and E-commerce, Education, Healthcare, Government and Defense, others) and Region (North America, Europe, Asia-Pacific, Rest of the World) - Forecast to 2030." *AI and Analytics - Machine Learning Market*. Feb. 2020. https://www.marketresearchfuture.com/reports/machine-learning-market-2494.

[14] Glauber, R. & Loula, A. "Collaborative Filtering vs. Content-Based Filtering: differences and similarities." *Information Retrieval*, pp. 1-8, Dec. 2019, doi: 10.48550/arXiv.1912.08932.

[15] Checker S. "Amazon Product Recommendation System". Medium. Jul. 2021. https://medium.com/analytics-vidhya/amazon-product-recommendation-system-855c9d3df481

[16] Golab, L. (Jan. 2023). MSCI 446 – Unsupervised Learning [Online]. University of Waterloo. Available: https://learn.uwaterloo.ca/d2l/le/content/873305/viewContent/4720308/View

[17] promptcloud. "Fashion Products on Amazon.com." Data World. 2017. Available: https://data.world/promptcloud/fashion-products-on-amazon-com.

[18] Shivangi, C. & Golab, L. "Job Description Mining to Understand Work-Integrated Learning." University of Waterloo. Available: https://educationaldatamining.org/files/conferences/EDM2018/papers/EDM2018_paper_42.pdf

[19] Gajendragadkar, U. "Product Recommender using Amazon Review dataset." Towards Data Science. Jun. 2019. Available: https://towardsdatascience.com/product-recommender-using-amazon-review-dataset-e69d479d81dd.