Deep Learning for Online Fashion: A Novel Solution for the Retail E-Commerce Industry

Zachary O. Harris
Southern Methodist University, zharris@smu.edu

Gowtham G. Katta
Southern Methodist University, gkatta@mail.smu.edu

Robert Slater
Southern Methodist University, rslater@mail.smu.edu

Joseph L. Woodall IV
Southern Methodist University, woodallj@mail.smu.edu

Recommended Citation

Deep Learning for Online Fashion: A Novel Solution for the Retail E-Commerce Industry

Zachary Harris[1], Gowtham Katta[1], Joseph Woodall[1], Advisor: Dr. Robert Slater[1]

[1] Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
{zharris, gkatta, woodallj, rslater}@mail.smu.edu

Abstract. The online shopping experience for clothing can be further enhanced by implementing Deep Learning techniques, such as Computer Vision and personalized recommendation systems. Automation, as a principle, can be applied to solving problems surrounding efficacy, efficiency, and security. It also provides a layer of abstraction for the user during the online shopping experience. This research aims to apply Deep Learning methods and principles of automation to augment the e-commerce fashion market in a novel way. After using these methods, it was found that Convolutional Autoencoders and Item-to-Item Based Recommenders may be used to accurately and precisely recommend articles of clothing based on a users' styling preferences.

1 Introduction

As the number of online clothing shoppers has significantly increased in the past decade, there has been an increase in the implementation of automated recommender systems on e-commerce platforms (Lee & Park, 2009). Currently, the market has responded to this increase in online shopping with the emergence of companies like Stitch Fix, Amazon Wardrobe, and Trunk Club, which rely on a human stylist to make recommendations for each user. However, this process can be slow, and human stylists often fail to provide suggestions tailored to the user's tastes and needs, leading to dissatisfaction with the service. To enhance the online shopping experience, an algorithm that uses images of a user's clothing should be created and implemented on e-commerce platforms. This algorithm, using a combination of Computer Vision methods and recommender frameworks, would help enhance the online shopping experience for the users. Furthermore, the algorithm could lead to more significant profits and better brand exposure for the participating clothing brands.

The physical and digital frontiers have become more intertwined in the recent decade, especially in the fashion, retail, and clothing industries (Sarker, 2021). With the increased availability of information, Machine Learning methods are becoming more widely used to analyze data effectively. Many companies in the retail industry already use Artificial Intelligence techniques to predict a consumer's behavior to

attract more consumers. For example, one company uses closed circuit television (CCTV) video footage in its shopping locations to examine products most often sought out based on consumer movement (Nguyen et al., 2022). The practice involves a continuous cycle of sense, analysis, and implementation to improve the store layout. This research could use this cycle to improve clothing recommender systems.

In addition to Machine Learning, Deep Learning methods have also become popular. More specifically, Computer Vision is one area of Deep Learning that is increasingly being used and has a variety of applications. Convolutional Neural Networks (CNN) and Computer Vision algorithms perform very well in facial recognition, object detection, and vision construction (Ferreira et al., 2018). When appropriately trained, Computer Vision can be used to predict images based on information included in a dataset. Overall, the most significant advantage of Deep Learning methods is item retrieval. Item retrieval allows one to identify identical or related items (Cheng et al., 2021).

As the online clothing shopping industry continues to grow, the amount of data about e-commerce concerning clothing is increasing. Therefore, algorithms and models must sort through and process copious amounts of information. Implementing recommendation or recommender systems is valuable for handling copious amounts of data. Recommendation algorithms became popular to decrease information overload by recovering the most relevant information from copious amounts of data (Patel et al., 2017). Ultimately, a recommendation system provides a method of information filtering and aims to give suggestions for items most pertinent to a particular user. Typically, the recommendations result from various decision-making processes based on variables and input from the user. Recommender systems are advantageous when an individual needs to choose an item from a potentially overwhelming number of products the platform may have to offer.

This research aims to provide meaningful and relevant recommendations vis-a-vis various Deep Learning techniques based on images of the users' personal clothing to a group of users for products that may interest them. To provide a tangible and novel solution to this problem and supply meaningful value-add to small and large retailers. This research will outline various Deep Learning methods in Computer Vision and recommendation algorithms, giving comparable measures between methods. This research will utilize Convolutional Neural Networks and Apriori Recommendation algorithms to achieve its goal, comparing its results on well-known data sets on the various components of the service offering to position its effectiveness using multiple metrics.

2 Literature Review

The literature review focuses on four crucial aspects of the clothing recommendation system: Online Shopping, Application of Machine Learning/Deep

Learning in E-Commerce, Computer Vision, and Recommendation Algorithms.

2.1 Online Shopping

The advancement of technology, along with increased internet usage, has brought local, regional, and niche markets to the global stage, allowing geographically distant private and public actors to engage in commerce. The internet has become a platform where businesses can render and interact with goods and services in this new globalized market. Multinational e-tailers have changed their regional business models to account for this increase in consumer demand. The drive for this increase in consumer preference for a digital storefront is due to the ease of use of online shopping, the elimination of the perceived risk of uncertainty and adverse consequences of purchasing a good or service, and the consumer's perceived enjoyment of online shopping. These principles are primarily true regardless of demographic variables like gender, age, and qualification to purchase online goods or services. New internet-focused businesses and vendors have adopted these principles to prevent a lack of trust in the technology and declining sales because of the increasing number of distrusting consumers. A positive association was found with consumer intention for online shopping (Akhlaq & Ahmed, 2014). The consumer has even noticed the perceived effectiveness of the legal framework of online shopping, opting to trust online storefronts more than in-person transactions. Akhlaq & Ahmed found that distrust is negatively associated with consumer intention in online shopping (Akhlaq & Ahmed, 2014). Furthermore, and most importantly, the perceived ease of use was determined to be the most significant factor in capturing a sale of a good or service, which leads to a higher likelihood of returning customers with positive intentions to online shops (Akhlaq & Ahmed 2014). Similarly, these businesses have adapted their websites to match user preferences to maintain positive correlations between forecasted consumer behavior and psychology (Akhlaq & Ahmed, 2014).

The emerging digital market space is a fascinating phenomenon, with multidimensional interactions between producers and consumers of large-scale goods and services. The inherent risks of engaging with a market on such a scale have increased. Based on a web-based survey, similar to Akhlaq & Ahmed's study, Almousa found that four risk factors influence consumers before engaging with a web-based transaction: time risks, performance risks, privacy risks, and social risks. Apparel internet shopping is no different, as it is a category in the digital market space (Almousa, 2011). Almousa found that while physical infrastructure chiefly impacts the delivery effectiveness of goods, it has no role in the consumer's likelihood of completing an online transaction (Almousa, 2011). Instead, the business can reliably communicate and execute the delivery of the goods and services to the customer in each step of the transaction, which chiefly impacts transaction completion. Internet-focused businesses may maintain this reliability and ability to execute by adopting an e-commerce business model to sell their products online (Almousa, 2011). Though the

absence of a proper metric to evaluate these products in the digital marketplace creates a significant risk for apparel customers (Almousa, 2011). This contradicts Akhlaq & Ahmed in the perceived ease of use of internet-focused businesses. However, businesses may circumnavigate this dilemma vis-a-vis the ranking mechanism through the introduction of evaluation metrics, such as satisfaction ratings for the customers, descriptions of products, and reviews from the consumers (Almousa, 2011). The arguments in consumer decision-making in shopping can ultimately be diluted down to online personalization (Almousa, 2011).

By providing online personalization in the digital market space, specifically in the apparel retailing context, customer discernment for ease of use can be dramatically affected. As such, depending on consumer satisfaction, there could be a high increase or decrease in sales. Three types of widely used online personalization mechanisms were discovered as crucial in retaining the possibility of an online sale: offer, recognition, and personal advice (Lee & Park, 2009). The offer pertains to the business' ability to present the goods or services to the consumer in its online storefront in a clear, concise, and organized manner. Recognition refers to the consumers' ability to understand and recognize the good or service being presented and its purpose. Lastly, personal advice refers to the dynamic relationship between producers and consumers, and the producers' ability to enhance their consumers' online shopping intentions, leading to returning customers (Lee & Park, 2009).

2.2 Application of Machine/Deep Learning in E-Commerce

The frontiers between the physical and digital realms continue to mesh. This is often referred to as the fourth industrial revolution. The amount of data gathered and accumulated has caused a shift in new technological applications. One such application, Machine Learning, has become widely popular since these techniques can analyze and process substantial amounts of data in an efficient manner. Specifically, there has been a significant increase in the usage of Reinforcement, Supervised, and Unsupervised Learning algorithms to proficiently decompose the information in the data (Sarker, 2021). These algorithms find patterns within data that would typically go unnoticed.

Regarding the e-commerce industry, Machine Learning techniques can be used to analyze any trends by the consumer. For example, predictive analysis can suggest products from previous purchase history. This method of analysis and decision-making would examine the relationship between the explanatory and response variables and produce a related recommendation of an unknown outcome. A more advanced recommendation system could suggest products to the consumer based on product purchase frequency and associated trends in the data. The application of Machine Learning not only recommends suggestions for the consumer but also aids the e-commerce business in keeping inventory and stock up to date (Sarker, 2021).

Before the technological revolution, the placement of products in retail stores

to attract customer attention was based on business intuition and discretion. The business would adjust the products through trial and error depending on consumer interaction. Businesses would often examine the data from product purchases, inventory changes, and overall sales to find the most attractive products. Artificial Intelligence (AI) has now been implemented to quickly decide price adjustments, product predictions, and supply chain management (Nguyen et al., 2022). Deep Learning methods such as Computer Vision are used to study consumer behavior in detail. For example, Computer Vision has been used to process and interpret CCTV videos and footage to gain insight into consumer shopping activity and trends. A two-stage Computer Vision detector can be used to extract regions (categories) that are most interacted with and apply Deep Learning classification for the features of each region or category. As such, there are a lot of applications for artificial intelligence in aiding product selection. Similar to the findings from Sarker, the continuous cycle of sense (detect), analyze, and act (implement) can be used to improve product selection until no further improvement can be made (Nguyen et al., 2022). The findings from Sarker, the continuous cycle of sense (detect), analyze, and act (implement) can be used to improve product selection further until no further improvement can be made (Nguyen et al., 2022). improve product selection until no further improvement can be made (Nguyen et al., 2022). The findings from Sarker, the continuous cycle of sense (detect), analyze, and act (implement) can be used to improve product selection further until no further improvement can be made (Nguyen et al., 2022).

2.3 Computer Vision

Computer Vision is often described as the capability of computers and systems to extract meaningful and valuable information from digital and visual inputs. A variety of techniques and methods exist to allow computers to do this.

Deep Learning methods are becoming increasingly popular and, in some cases, perform better than other Machine Learning techniques and methods. One area in which deep-learning methods perform well is Computer Vision. Computer Vision is known to be helpful when solving visual recognition problems. Some techniques include the more familiar Convolution Neural Network (CNN) and the lesser-known Stacked Denoising Autoencoders (SdA), Deep Belief Networks (DBN), and Deep Boltzmann Machines (DBM). Each technique has advantages and disadvantages (Ferreira et al., 2018).

Convolutional Neural Networks are incredibly beneficial in solving Computer Vision problems. Some applications of this method include face recognition, object sensing and detection, and implementation vision in robotics. The applications are almost endless. One reason for the popularity of this method is the ability to be operated in a diverse collection of applications. However, it is computationally demanding to train CNN models. Like CNN models, DBN models and DBM models also have the same flaw. They are both versatile in the number of

problems they can solve but involve high computational demand (Ferreira et al., 2018). If high computational demand is an issue, using SdAs (Stacked Denoising Autoencoders) is an acceptable alternative. SdAs can significantly reduce the amount of learning and training time needed for a data set. The decreased learning and training time is partially due to SdAs lowering the required dimensionality of the data. So, if one has a large dataset, using SdAs will allow the data to be understood from a smaller portion of the data, provided that small portion is representative of the larger dataset (Ferreira et al., 2018).

As the world shifts and becomes more digital, the fashion sector in e-commerce continues to grow. To keep up with this growth, many algorithms and techniques have been created to improve the user experience of buying clothes online. These techniques include clothing identification, popularity prediction, and fashion synthesis, which can be described as creating an outfit based on readily available data. Popularity prediction is an essential element for both fashion designers and consumers. Fashion designers want to know the latest trends to be competitive in the fashion market. Consumers want the same knowledge but want to ensure they are buying and wearing the latest designs and styles. This feature can also play a role in fashion synthesis, which involves the visual input of a person and then synthesizing a realistic possibility for an outfit based on that visual input. A significant component required to implement all the techniques listed above is image classification.

Computer Vision can be used to enhance the prediction of image classification problems. Prior structural knowledge, which is knowledge observed from prior decision-making, can be combined with relational information to produce appropriate product recognition and improve the classification of an image. Specifically, Convolutional Neural Networks can be used to promote and produce visual recognition. A category tree can be implemented for the fashion classification model with five distinct areas: attribute, category, sub-category, gender, and family. The mutually exclusive factors would be the families, categories, and the sub-categories, which are independent of one another. For example, to identify if a piece of fashion is a dress, the category can be defined as dress, top, or skirt, the sub-category can be defined as day-dress or evening-dress, and the attribute can be defined as length (short or long) or neckline (round or squared) (Ferreira et al., 2018). The Res-Net, a residual neural network, can be used to promote the classification and typically performs better than others. For the fashion classifier, a custom Res-Net-50, a convolution neural network with 50 deep layers, is used to process and recognize the image. Only the final 11 layers are used to train the models for the training, with the weights being re-trained each time. As such, the architecture of the network is adapted in a way in which the attribute is affected by the category (and vice-versa). Likewise, the sub-category is also streamed from the category. Further promoting each level's features enhances the current and future output (Ferreira et al., 2018).

Computer Vision applications in the e-commerce fashion market vary across use cases. Based on a survey of more than 200 fashion related works, there appears to

be four main use cases: fashion detection, analysis, synthesis, and recommendation. (Cheng et al., 2021). For the first two, landmark detection, fashion parsing, and item retrieval correspond to fashion detection while popularity prediction and attribute recognition fall under fashion analysis. For the last two, physical simulation and style transfer comprise fashion synthesis while outfit suggestion and fashion compatibility represent fashion recommendation. This research will pertain to the first and second use cases.

Landmark detection is the discipline aimed at predicting positions of functional points in image data, in this case, images of articles of clothing. As such, the boundary and design pattern of these surrounding boxes are captured and evaluated. After, a categorical value of each image is generated, and a discriminative feature representation is produced. The architecture for a model similar to this includes a CNN with three primary levels, with each stage subsequently adjusted and polished from the subsequent predictions. A model of this caliber performed well on benchmark datasets.

Fashion parsing is when the labels from each particular class are formulated from clothing items. The difference between scene segmentation problems lies in the granularity of categorization. There are more landmarks, a higher level of granularity with the semantics of clothing images, a deforming structure, and a larger number of possible classes (Cheng et al., 2021).

When aiming to find similar or even exact replicates of certain items from a group, item retrieval can be used (Cheng et al., 2021). Some obvious advantages of deep learning in regard to item retrieval are the representation of in-depth features guided by attributes and continuous embedding. As such, the visual similarity constraints and the semantic attributes are constantly embedded during the learning stage while also maintaining appropriate discrepancies between the domains (Cheng et al., 2021). Along with attribute detection, multi-class classification problems directed at properly identifying which clothing elements are associated with one another becomes a trivial problem (Cheng et al., 2021). This research aims to build upon these principles to obtain optimal results found by (Cheng et al., 2021). A comprehensive list of these principles and their corresponding benchmark statistics may be found in the article (Cheng et al. 2021).

2.4 Recommendation Algorithms

A recommender system is an algorithm in which at least one useful suggestion or recommendation of an item is offered to an output based on the output's previous past behavior. Recommendation algorithms emerged as an individualistic research area in order to minimize the overall computation load. Only the most pertinent data is evaluated by the algorithm from a large selection of information (Patel et al., 2017). This research will define a recommendation algorithm in which at least one useful item is suggested based on the consumer's previously defined

preferences or interests in the e-commerce fashion market. These preferences or interests must be captured and stored in a data set for other use. Researchers have developed recommendation algorithms in the past, each with benefits and drawbacks (Patel et al., 2017). Many of these recommendation systems face user preference challenges. For instance, since the algorithm does not take into account the preferences or interests of the consumer, the expected output does not become as relevant. Because the quantity of the customers and the number of rated items may not always be equivalent and disproportionate, the algorithm might operate and pull from sparse data. Moreover, since the algorithm might not scale well, there would be very limited infrastructure or capacity for potential growth. This would be a problem when new information, such as more users and items, is introduced. Furthermore, the algorithm might not have strong synonymy, in which the likelihood of near items has dissimilar names or entities. Lastly, the algorithm might not protect the user demographic data and would pose a risk in exposing private user data (Patel et al., 2017). Similar to the number of challenges any given recommendation algorithm might face, there is a relative amount of methods of recommender systems, each with its advantages and disadvantages. Patel et al. outline three main methods in which various sub-level methods are explored: a hybrid, a collaborative, and a content-based filter (Patel et al., 2017).

The content-based filter makes use of the text-based descriptions from items and then provides a single user recommendation on a case-by-case basis for each user. The recommendations are separated through the similarities found in each user. Furthermore, the content-based filter also operates on a single-user basis per instance where each user is independent. As a result, the filter provides transparency in the algorithm's function and reduces the content analysis and user profiles for the data set. Although the filter is the most mathematically simple yet basic recommendation method, it may not be the best performing method. Other methods, such as the collaborative filter, may be better suited for larger platforms.

Collaborative filtering is the most widely used recommendation algorithm in the industry, and it usually offers the best performance. Collaborative filtering is based on the comparison of active users in a dataset, rather than only a single user. Using a comparison of active users boosts the algorithm's performance. However, it greatly suffers from data sparsity, scalability, and synonymy. It reviews more than one common item to generate a set of users. Furthermore, subfields of collaborative filtering may be expanded upon vis-a-vis memory-based, model-based, and hybrid filtering techniques.

The hybrid filter method, in which other recommendation algorithms are merged, not avoid the problems and limitations from the other methods (mentioned above), but also improves the overall optimization for the process. The basic idea of the hybrid filter method is to maximize the advantages and minimize the computations as much as possible. The most common approach to the hybrid filter method is to merge the content-based with collaborative filter methods to avoid data sparsity,

scalability, and synonymy. From this method, more of the user preferences are taken into account. As such, the cold-start and data sparsity problem would be overcome, though at the expense of an increase in complexity since the method may be computationally more expensive to implement (Patel et al., 2017).

Despite introducing these new methods, the research faced some challenges. First, the system needed to offer suggestions and recommendations at the same level of expertise as that of an ambassador that routinely aims to match the consumer's tastes. Second, assigning a priority to the suggestions in the amidst of the vast number of diverse items would be a disadvantage. However, to overcome these challenges, Sekozawa et al. implemented a hierarchical analytical process to numerically assign values to attributes of clothing in order to digitize human perceptions and taste (Sekozawa et al., 2011). Likewise, a cluster of apparel was implemented based on the hierarchical processing vis-a-vis k-means clustering to provide item recommendations based on consumers with similar tastes (Sekozawa et al., 2011). The approach to recommendations as one-to-one functioned in unison. While this approach proved fruitful, it illustrates a problem in applying recommendation algorithms in the e-commerce fashion market. The availability of data is always a concern in algorithms requiring vast amounts of variance. Anticipating and solving tomorrow's problems while solving today's problems is likely to sell in any market, whether digital or physical. The analytical hierarchical process was not the best suitable recommendation algorithm to meet this business requirement, as it can only produce results with the available data. A network-based recommendation algorithm, however, can meet this business requirement.

Network-based recommendation algorithms serve the same functionality as the three traditional approaches but differ in their overall architecture at the most basic level. The traditional recommendation algorithms represent input data with bipartite user-item networks where users are connected with their collected items (Yu et al., 2016). A brand-new recommendation model was proposed, in which a probabilistic spread and random walk process were integrated in the system. The probabilistic spread assigns the items at each node and evenly separates the partitions transmitted from the user while the random walk method builds upon the user item network (Yu et al., 2016). However, the probabilistic spread becomes problematic for the recommendation methods with parameters but can effectively identify items that the user prefers vis-a-vis ranking scores. This method is particularly effective in system environments where relatively few items receive high user rankings, otherwise known as the "long-tail problem", where long-term diversity is relatively low (Yu et al., 2016). Depending on the business context, any of the methods mentioned above might be a suitable and effective technique in recommending items to a user base. The deciding factors for any business will be the requirements that govern the system, the characteristics of the ecological user system, and its interaction with the offered items.

Automation can potentially increase the efficiency of any given task. Particularly in the current generation, the people around us want a simple yet fast

solution to their everyday problems. As such, this study aims to increase the efficiency of everyday online retail for the users (the customers). Using Machine Learning classification methods, this study aims to alleviate the time necessary for selecting and purchasing clothing. The algorithm produces recommendations and results to enhance the online shopping experience for the user to make life more efficient for the everyday person.

3 Data

3.1 Data

Obtaining data for this study will be done by using public datasets published by clothing companies and other fashion industries. One dataset used for this study contains information on various articles of clothing that are produced and manufactured by a clothing company originally established in Sweden. This company is known as Hennes & Mauritz AB, or H&M Group. The dataset contains twenty-five variables and 45,875 observations which include images and samples of each article of clothing. The description of the variables can be seen in the Appendix of this research paper in Table 3. For model testing and training purposes, an MNIST Fashion dataset will also be used. This dataset contains over 60,000 examples of clothing articles that are formatted to be used with CNNs and other Computer Vision techniques. While the datasets for this study provide the variables and observations for the articles of clothing being sampled, additional data required by the algorithm will include sample inputs of user preferences and images of users' personal clothing. To create this data, multiple user personas will be created along with sample user preferences and image data to test the models. User preference inputs will include a variety of variables of different data types, while the photograph for personal clothing will be of common universal image types. Before introducing the data to the model, it will have to be cleaned to account for any errors or missing data within each observation. User images within the personas will also have to be reformatted to be accepted by the Deep Learning frameworks chosen for the study.

4 Methods

4.1 Aggregated Dataset

Initially, the goal was to combine the Fashion MNIST and H&M datasets. However, after the preliminary evaluation of the datasets, it was found that the H&M dataset would be more suitable for this model due to the size, data integrity, and data quality. For all model creation and execution pertaining to this study, the H&M

dataset will be used. The Fashion MNIST dataset will be reserved for additional analysis and further studies.

4.2 Execution

To execute the initial parts of the study, a Convolutional Autoencoder will be created to analyze visual inputs from the user. During this stage, input and output types will be analyzed to understand the requirements for feeding data through the network. The first objective is to interpret the submitted image of the article of clothing and analyze the submission, eventually comparing it to other pieces of clothing within the given dataset. The second objective is to recommend other articles of clothing that would complement the submitted image. This recommendation should consider the users' preferences and provide relevant outputs of items that interest the user. In the ideal execution of these procedures, the user would receive a tailored recommendation from the sequence of events that results in an output that complements the user's input, potentially completing an outfit.

5 Results

5.1 Stylist Module

The researchers discovered interesting findings from the data. The researchers objectively met the goal stated in the introduction of this research article. The two modules created by the researchers were successfully run and returned robust results. For the stylist component, a mean squared error loss (MSE) was used to evaluate the performance of the model across multiple epochs. After each epoch, the learning parameter passed a portion of the data under the model algorithm. The stylist module passed through ten epochs until the entire dataset was processed. The performance of the model was evaluated by examining the training loss and the validation loss after each epoch. Although the training loss was more sporadic than the validation loss, mean squared error loss was under 0.04 for most of the epochs. When looking at the validation loss, the mean squared error loss continually decreased after each epoch. As such, the stylist model performed the best during the last epoch, Epoch 9. During Epoch 9, the training loss was found to be 0.0358 and the validation loss was found to be 0.0344. The results for the stylist component can be found in Figure 1 and Table 1, which help support the researchers' conclusion.

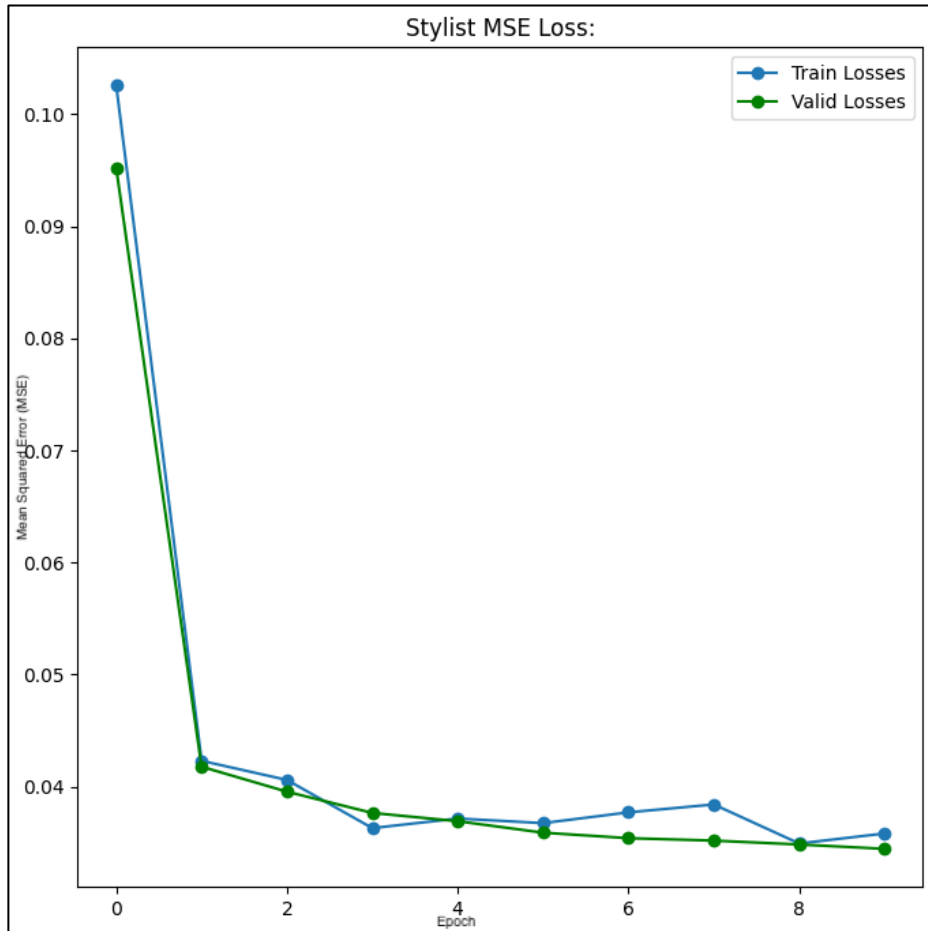**Figure 1**: Stylist Mean Squared Error Loss

Table 1: Stylist Training and Validation Loss

Epoch	Loss	Mean Squared Error
Epochs = 0	Training Loss	0.10256695747375488
Epochs = 0	Validation Loss	0.09520003944635391
Epochs = 1	Training Loss	0.042314495891332626
Epochs = 1	Validation Loss	0.04178624227643013
Epochs = 2	Training Loss	0.040604978799819946
Epochs = 2	Validation Loss	0.03955893963575363
Epochs = 3	Training Loss	0.03631202504038811
Epochs = 3	Validation Loss	0.037663765251636505
Epochs = 4	Training Loss	0.037166204303503036
Epochs = 4	Validation Loss	0.03692710027098656
Epochs = 5	Training Loss	0.03675476461648941
Epochs = 5	Validation Loss	0.035901427268981934
Epochs = 6	Training Loss	0.0377175472676754
Epochs = 6	Validation Loss	0.03541313484311104
Epochs = 7	Training Loss	0.03842123597860336
Epochs = 7	Validation Loss	0.035194166004657745
Epochs = 8	Training Loss	0.03493717312812805
Epochs = 8	Validation Loss	0.034848593175411224
Epochs = 9	Training Loss	0.03583287075161934
Epochs = 9	Validation Loss	0.03445950895547867

5.2 Recommender Module

For the recommender component, a mean cross-entropy loss was used to evaluate the performance of the model. The recommender module experienced ten epochs in order to fully process the data. When examining the mean cross-entropy loss of the module, the model performed the best during the last three epochs: Epoch 8, Epoch 9, and Epoch 10. During Epochs 8, 9, and 10, the module returned diminishing results and had a mean cross-entropy loss of 0.0833. More information about the mean cross-entropy loss can be viewed in Figure 2 and Table 2.

Figure 2: Recommender Mean Cross Entropy Loss
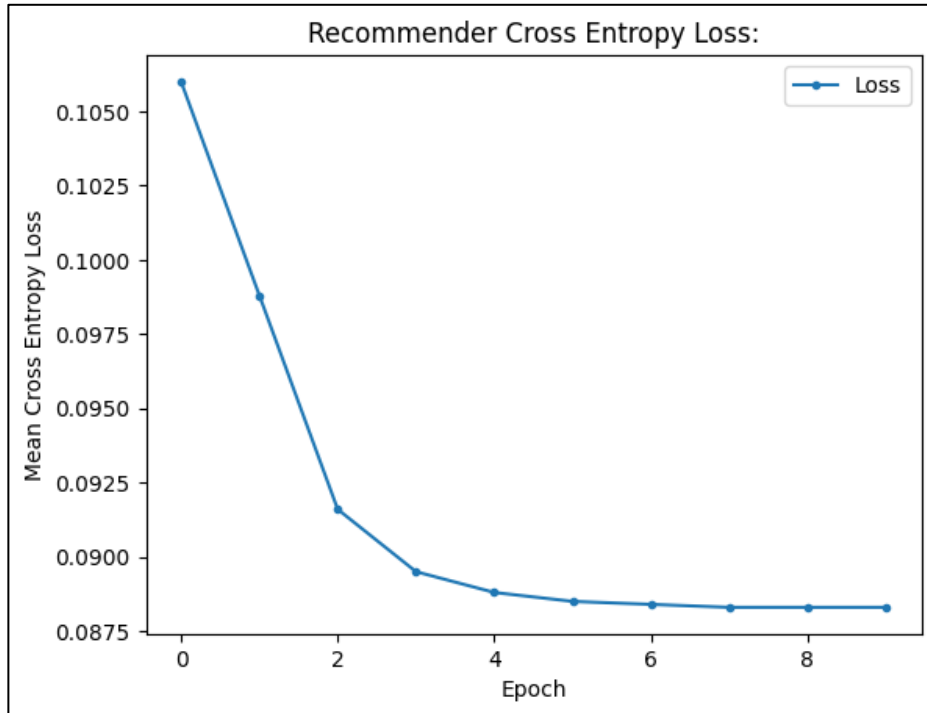
Table 2: Recommender Mean Cross Entropy Loss

Epoch	Mean Cross Entropy Loss
Epochs = 1	0.1060
Epochs = 2	0.0988
Epochs = 3	0.0916
Epochs = 4	0.0895
Epochs = 5	0.0888
Epochs = 6	0.0885
Epochs = 7	0.0884
Epochs = 8	0.0883
Epochs = 9	0.0883
Epochs = 10	0.0883

After exploring the mean cross-entropy loss, the binary cross entropy loss (BCE) and the mean average precision score (MAP) were used to determine the precision of the model. These metrics were used to determine how well the model performed when pairing the articles of clothing within the dataset. Again, the recommender module had iterated through ten epochs in order to process the entire

dataset. The model had the smallest binary cross entropy loss at the last epoch, Epoch 10. The BCE was the choice loss function for this problem, as it combined a sigmoid layer with BCE in one single class. The model experienced the lowest mean average precision score during the second and the last epochs, which were Epoch 2 and Epoch 10. The mean average precision score for Epoch 2 was found to be 0.00455, while the mean average precision score for Epoch 10 was found to be 0.005216. The mean average precision score is used to determine if the model is well balanced. A detailed insight into the mean average precision score for the recommender component can be found in Figure 3, Figure 4, and Table 3.

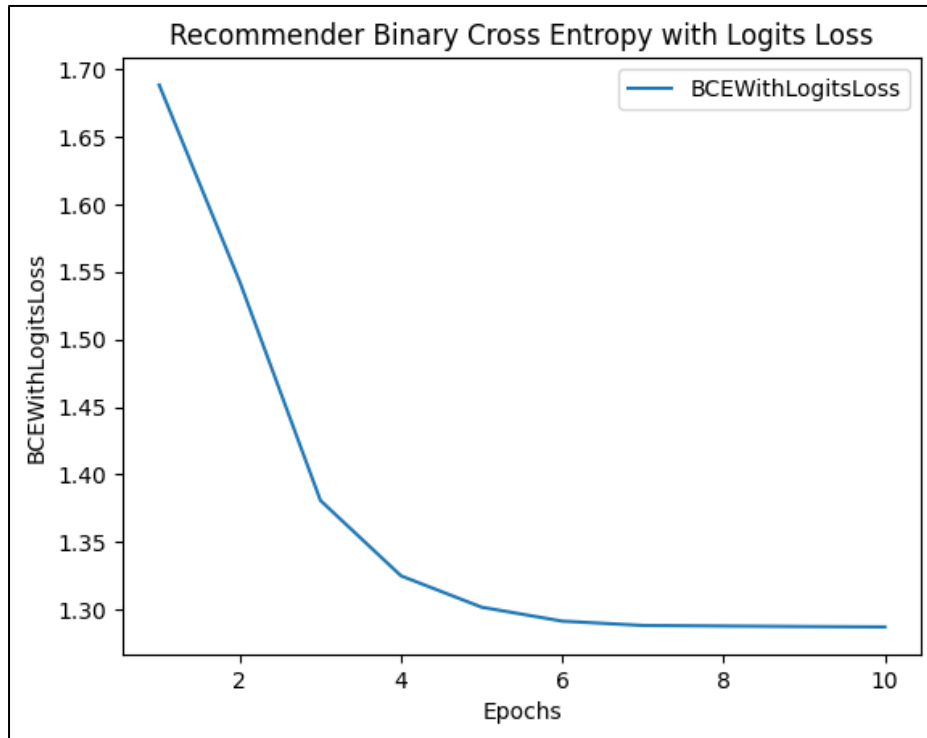**Figure 3:** Recommender Binary Cross Entropy (with Logits)

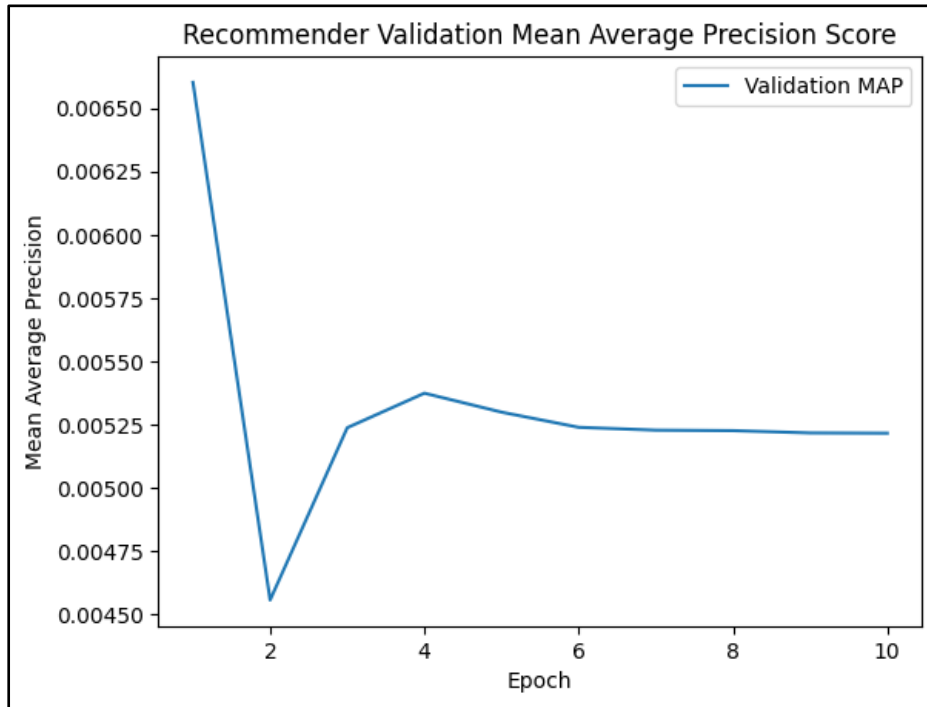**Figure 4:** Recommender Validation Mean Average Precision Score

Table 3: Recommender Binary Cross Entropy Loss & Mean Average Precision (MAP) Score

Epoch	BCE (with Logits)	MAP
Epochs = 1	1.6883	0.006602
Epochs = 2	1.5426	0.004557
Epochs = 3	1.3808	0.005238
Epochs = 4	1.3250	0.005374
Epochs = 5	1.3018	0.005299
Epochs = 6	1.2915	0.005239
Epochs = 7	1.2883	0.005228
Epochs = 8	1.2879	0.005226
Epochs = 9	1.2875	0.005217
Epochs = 10	1.2872	0.005216

The researchers found that the scope of the study was reduced to only the data with which the algorithm was trained. Therefore, if new data were to be introduced these results would not be expected. The scope of this research will be discussed in the Discussion section of this research article.

6 Discussion

6.1 Key Thoughts

The researchers chose this topic for the study due to personal interests. The authors involved have an interest in e-commerce, fashion, and deep learning, particularly applied to this field of study. The primary audience for this research is vendors, and ultimately, consumers in the e-commerce fashion industry market space. The researchers hope to solve a critical business problem and address such a gap in the fashion market space. According to the results, the goals the researchers laid out were successfully and objectively met, based on the statistical research found in previously in this research article. The results of this article are significant because of their novelty. To the researchers' knowledge, this approach has not been documented by the open-source community. However, the researchers had limited access to private ventures outside of the public domain. The researchers hope that the findings in this article may contribute to the fascination and contributions to the deep learning community at large to solve more challenging and humanitarian issues facing the world today.

When examining the results from the stylist module, the training and validation mean squared error losses were very small, which indicates that the model performed relatively well. Since the H&M dataset was used as a source to train and validate the model, the performance was limited to the values within the dataset. The training mean squared error loss had the lowest value during the third epoch, with a value of 0.0363, and the second lowest value during the eighth epoch, with a value of 0.0344. Since the lowest loss was not present at the final epoch, the learning parameter did not adapt as the number of parameters increased. However, the validation squared error loss continued to decrease as the number of epochs increased, which means that the learning parameter was adapting. However, if a larger, more comprehensive data set was used, then the results for the training and the validation losses could have been vastly different.

For the recommender, a mean cross entropy loss was also used on the H&M dataset, specifically on the transactions train file that was included in the open-source data. The transactions train file contained a history of customers, the respective purchased articles of clothing, the price of the clothing, and the sales channel associated with the sale. The number of epochs was the same for the recommender module. The mean cross entropy loss on this subsection of the dataset slowly decreased as the number of epochs increased. The final value for the mean cross entropy loss was 0.0833 at the last epoch. This meant that the learning parameter performed better as the number of epochs increased. This also suggests that the model is adapting with the influx of data. When looking at the mean average precision score, the lowest value was 0.005216 at the tenth epoch. When the mean average precision

score is at 0.005, this means that the model is well-balanced. When examining the results from the recommender module, the lowest binary cross entropy with logits loss occurred during the tenth epoch, with a value of 1.2872. Depending on the source data, the values for the recommender could also be different than as produced above.

6.2 Ethical Considerations

Regarding ethics and privacy concerns as it relates to this research. The researchers believe that there is always an inherent risk in using Computer Vision algorithms, including Convolutional Neural Networks and Convolutional Autoencoders, to target any group or entities from a corpus of data. Such examples include discrimination, alienation, or isolation for the end goal of taking advantage of the party or entity involved for personal gain. This unfairness is the crux of the danger that Computer Vision algorithms offer to the e-commerce fashion market space. In combination with a recommendation algorithm, this novel solution might pose a danger to consumers with shopping addictions, making it easier for online realtors to harass.

6.3 Limitations/Future Research

To the researchers' surprise, this problem was challenging in an unusual way. The engineering of pipelines for the data to flow from module to module was quite exciting to the researchers, almost as exciting as applying the statistical deep learning algorithms. This is also one of the more obvious rooms for improvement in the methodologies of this research article. Further research may be found in the fortification, or substitution, of the Convolutional Autoencoder algorithm used in the "Stylist" module. For example, substituting the Convolutional Autoencoder with a Generative Adversarial Network to recognize similar articles of clothing. Generative Adversarial Networks are known for their robustness to future, unforeseen data, and ability to thwart overfitting.

The researchers could have chosen alternative methodologies for obtaining the end objective. Firstly, a Generative Adversarial Network (GAN) pair could be used in substitution for the Stylist module. GANs might be utilized to find similar images within a corpus of image data, where each vector in the feature space might be mapped to similar vectors in other feature spaces for inference. Generative Adversarial Networks are known for their high accuracy and their ability to generate synthetic data like real data. The current Stylist utilizes Convolutional Autoencoders. Compared to Generative Adversarial Networks, which train in unsupervised learning, Autoencoders solve their objective via semi supervised learning. Transforming the Stylist into an unsupervised learning objective can eliminate potential costs in obtaining labelled data.

One particularly difficult area to solve was applying the above methodologies to creating DataSet class objects in the famous PyTorch library. Another difficult area, centering around these DataSet class objects, was understanding the organization of the class objects and their indices. The researchers, once these class objects were successfully created, gained a fandom of PyTorch, seeing as how the library is vast and efficient in its Pythonic abstractions.

The researchers hope that this solution can and will be implemented and adopted by the e-commerce fashion industry. It is of particular importance that this novel solution's user's central to the user's experience. The design philosophy of this ensemble of algorithms is with the end-user in mind, and at its core. The researchers believe that this ensemble of algorithms might be applied to other industries and areas of research as well. The applications of which might be limited to any industry concerning computer vision, and recommendation of due actions and diligence, of course.

7 Conclusion

The execution of the model created for this study revealed interesting insights into how algorithms operate. When the algorithm was run during the training and validation sets, the error from the loss function decreased. The researchers found that having at least ten epochs, or cycles, proved to be optimal to receive the minimal accuracy necessary for recommendation. Each time the algorithm ran, the learner was able to minimize the overall error and was able to assist in the future predictions.

This research can be utilized in any market space and industry that utilizes Computer Vision and Recommendation algorithms in order to understand the user's needs and enhance their online shopping experience, regardless of location. This problem, and set of algorithms, may be implemented in these market spaces, and include manufacturing, digital product transformation and sales, and logistics.

References

1. Akhlaq, A., & Ahmed, E. (2014). Online Shopping: A Global Perspective. Journal of Basic and Applied Scientific Research.
2. Almousa, M. (2011). LE RISQUE PERÇU DANS DES ACHATS EN LIGNE D'HABILLEMENT : 9.
3. Cheng, W.-H., Song, S., Chen, C.-Y., Hidayati, S. C., & Liu, J. (2021). Fashion Meets Computer Vision: A Survey (arXiv:2003.13988). arXiv. https://doi.org/10.48550/arXiv.2003.13988
4. Fashion Analysis: Current Techniques and Future Directions. (2014). IEEE MultiMedia, 21(2), 72–79. https://doi.org/10.1109/MMUL.2014.25
5. Ferreira, B. Q., Baía, L., Faria, J., & Sousa, R. G. (2018). A Unified Model with Structured Output for Fashion Images Classification (arXiv:1806.09445). arXiv. https://doi.org/10.48550/arXiv.1806.09445
6. Goel, A., Tung, C., Lu, Y.-H., & Thiruvathukal, G. K. (2020). A Survey of Methods for Low-Power Deep Learning and Computer Vision (arXiv:2003.11066). arXiv. http://arxiv.org/abs/2003.11066
7. Goldsmith, R. E., & Flynn, L. R. (2004). Psychological and behavioral drivers of online clothing purchase. Journal of Fashion Marketing and Management: An International Journal, 8(1), 84–95. https://doi.org/10.1108/13612020410518718
8. Lee, E.-J., & Park, J. K. (2009). Online service personalization for apparel shopping. Journal of Retailing and Consumer Services, 16(2), 83–91. https://doi.org/10.1016/j.jretconser.2008.10.003
9. Mahony, N. O., Campbell, S., Carvalho, A., Harapanahalli, S., Velasco-Hernandez, G., Krpalkova, L., Riordan, D., & Walsh, J. (2020). Deep Learning vs. Traditional Computer Vision (Vol. 943). https://doi.org/10.1007/978-3-030-17795-9
10. Moshrefjavadi, M. H., Rezaie Dolatabadi, H., Nourbakhsh, M., Poursaeedi, A., & Asadollahi, A. (2012). An Analysis of Factors Affecting on Online Shopping Behavior of Consumers. International Journal of Marketing Studies, 4(5), p81. https://doi.org/10.5539/ijms.v4n5p81
11. Patel, B., Desai, P., & Panchal, U. (2017). Methods of recommender system: A review (p. 4). https://doi.org/10.1109/ICIIECS.2017.8275856
12. Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x
13. Sekozawa, T., Mitsuhashi, H., & Ozawa, Y. (2011). One-to-one recommendation system in apparel online shopping. Electronics and Communications in Japan, 94(1), 51–60. https://doi.org/10.1002/ecj.10261

14. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience, 2018, e7068349. https://doi.org/10.1155/2018/7068349

15. When AI meets store layout design: A review | SpringerLink. (n.d.). Retrieved June 10, 2022, from https://link.springer.com/article/10.1007/s10462-022-10142-3

16. Yu, F., Zeng, A., Gillard, S., & Medo, M. (2016). Network-based recommendation algorithms: A review. Physica A: Statistical Mechanics and Its Applications, 452, 192–208. https://doi.org/10.1016/j.physa.2016.02.021

Appendix

Table 4: Variable Description for H&M Dataset

Variable Name	Variable	Variable Type
Article ID	article_id	int64
Product Code	product_code	int64
Product Name	prod_name	object
Product Type Number	product_type_no	int64
Product Type Name	product_type_name	object
Product Group Name	product_group_name	object
Graphical Appearance Number	graphical_appearance_no	int64
Graphical Appearance Name	graphical_appearance_name	object
Color Group Code	colour_group_code	int64
Color Group Name	colour_group_name	object
Perceived Color Value ID	perceived_colour_value_id	int64
Perceived Color Value Name	perceived_colour_value_name	object
Perceived Color Master ID	perceived_colour_master_id	int64
Perceived Color Master Name	perceived_colour_master_name	object
Department Number	department_no	int64
Department Name	department_name	object
Index Code	index_code	object
Index Name	index_name	object
Index Group Number	index_group_no	int64
Index Group Name	index_group_name	object
Section Number	section_no	int64
Section Name	section_name	object
Garment Group Number	garment_group_no	int64
Garment Group Name	garment_group_name	object
Detail Description	detail_desc	object