SMU Data Science Review

Volume 3 | Number 1 Article 6

2020

Advancing Performance of Retail Recommendation Systems

Lisa Leininger
Southern Methodist University, lleininger@smu.edu

Johnny Gipson
Southern Methodist University, jlgipson@smu.edu

Kito Patterson
Southern Methodist University, kitop@smu.edu

Brad Blanchard
Southern Methodist University, bablanchard@smu.edu

Advancing Performance of Retail Recommendation Systems

Johnny Gipson[1], Lisa Leininger[1], Kito Patterson[1], and Brad Blanchard[1]

Master of Science in Data Science, Southern Methodist University, Dallas TX 75275 USA {jlgipson, lleininger, kitop, bablanchard}@smu.edu

Abstract. This paper presents two recommendation models, one traditional and one novel, for a retail men's clothing company. J. Hilburn is a custom-fit, menswear clothing company headquartered in Dallas, Texas. J. Hilburn employs stylists across the United States, who engage directly with customers to assist in selecting clothes that fit their size and style. J. Hilburn tasked the authors of this paper to leverage data science techniques to the given data set to provide stylists with more insight into clients' purchase patterns and increase overall sales. This paper presents two recommendation systems which provide stylists with automatic predictions about possible clothing interests of their clients. The first recommendation system is a commonly used content-based collaborative filtering model and serves as the base model to evaluate the second recommendation system. The second recommendation system is an ensemble model comprised of separate clustering, KNN, and time series models that is a novel approach. These models are then fed into a neural network in order to produce recommendations. These recommendations for J. Hilburn's clients will hopefully lead to expanding their customer base and increasing their revenue as a result of more refined clothing and style recommendations. This paper describes the process of building two recommendation systems. Both models are evaluated using AUC as a metric as well as their potential for scalability. The ensemble model has a slightly higher AUC, 91% versus 86%. However, the ensemble model is computationally more extensive resulting in it requiring more resources to run.

1 Introduction

J. Hilburn is a custom-fit, men's clothing company which was founded in Dallas, Texas, in 2007. It provides a unique shopping experience because customers may browse clothing choices online and through their catalogs; however, to make a purchase the customer must work with a personal stylist. These stylists are responsible for taking accurate measurements of the client as well as serving as a personal fashion consultant. The stylists offer clients guidance in creating their own individualized wardrobe through customization of the clothing. J. Hilburn's main business goal for their stylists is to increase items sold per transaction and also increase overall frequency in transactions. J. Hilburn consistently works towards growing their market share by giving J. Hilburn's stylists superior recommendations to competitively assist their clients in making purchases in a crowded clothing market.

The goal of this paper is to create a recommendation system that will assist J. Hilburn stylists in increasing their overall sales. For background, recommendation systems estimate users' preferences to suggest items users might like to purchase; these systems are usually classified into three categories: collaborative filtering, content-based, and hybrid recommender systems [1]. Today recommendation systems are utilized in a variety of areas. Early recommendation system adopters were Netflix and Pandora, who were pioneers in suggesting movies and songs for users, respectively. Stitch Fix is often recognized for its innovation in using data science to produce recommendations within the retail clothing sphere [2]. While the majority of early recommendation systems made use of collaborative filtering when building models, today there are far more complex approaches in the never-ending race to produce increasingly accurate recommendations designed to entice consumers into making additional purchases. The main contribution of this paper is presenting an open source and novel approach for building an ensemble recommendation model.

Section II of this paper summarizes prior research which is the foundation on which the authors based their work. Section III describes the data preprocessing required to build both models. While, Section IV delves further into the exploratory data analysis required for the models in this paper. Next, Section V details the design of the Collaborative Filtering base model recommendation system. After that, Section VI delineates the novel approach ensemble recommendation system. This section is divided into four sections: clustering, KNN model, time series model, and ensemble model deployment. Section VII defines the suggested business rules to be layered into both models. Section VIII analyzes the results of both models. Section IX addresses the limitations of the data used and models built in this paper. Then, Section X proposes possible future work for the problem outlined in this paper. Section XI outlines ethical concerns about recommendation systems and how they were addressed in this paper. Lastly, Section XII concludes this paper.

2 Related Work

Within the past twenty years, online shopping has reinvented the clothing shopping experience. This transformation has driven significant changes in how clothing retail companies market to their customer base [3]. For most people, it is now a daily experience to encounter a recommendation system that suggests purchases, allows items to be easily added to each customer's shopping cart, and increases e-commerce's sales all with just the click of a mouse or tap of a finger. Guan, Qin, and Ling state that, "Most existing recommendation systems predict similar products that users may like from buying/like history based on data mining technology" [4]. These data mining technology approaches include many machine learning algorithms and process features such as screen views, customer clicks, historical purchases, and personal demographics. Recommendation systems also compare customers with one another. All of these factors must be considered when building a retail recommendation system, which are then utilized by companies to increase profits and to help consumers make useful purchases. Simply put, recommendation systems determine the similarity between individual products and users, in different combinations, in order to make purchase suggestions to customers. Therefore, user feedback of purchases is also very important in building and evaluating recommendation systems. One study built two recommendation models, one with user feedback as a feature and one without. The model which utilized user feedback as a factor resulted in the number of satisfied users being nine times higher than the model without user feedback [5]. Similarity between customers and items is determined by distances between them which are calculated by different machine learning algorithms. Different metrics to calculate these distances between items or customers can be used such as Euclidean distance, Pearson similarity, Manhattan distance, Minkowski distance, or Cosine similarity. When comparing different models which utilized different distance metrics, a Chinese retail study achieved higher results using cosine similarity over using Pearson similarity [6].

There are two main types of recommendation systems in use today: collaborative filtering systems and content-based systems [7]. Collaborative filtering has been widely used for many years because it has high prediction performance [8]. Collaborative filtering models are also considered advantageous because they have good scalability as users increase and require little computation, which produces results quickly [9]. However, as recommendation systems continue to dominate marketing practices across industries, engineers are continually increasing the complexity of their models in an attempt to grow market share for their businesses by accurate purchase suggestions. For this reason, ensemble models are gaining traction as they allow for integration of analytical insight attainable only through multifaceted approaches. Time series and clustering are two models commonly used when building ensemble recommendation models. Clustering works to create groups of customers, who have similar interests and attributes, which assists the recommendation system in making suggestions [8]. K-means clustering has also greatly improved results over original clustering algorithms, which had many shortcomings [8]. Including time series models allows for the model to adapt to sales trends, which might otherwise be overlooked in other models.

However, creating a novel recommendation system is a challenge because there is not a universal definition stating what precisely would make it novel [10]. Opinions on what determines novelty differ so much that some believe it is the composition of the model that makes it novel while others believe it is the evaluation of the model that leads to novelty [10]. There is also not a universally agreed upon method of evaluating recommendation systems. While accuracy has long been the preferred method of evaluation, it is now believed that it should be used in conjunction with other methods, though there is not agreement as to which methods or how they should be weighted [11].

Additionally, recommendation systems for clothing companies require more nuance than one for a more general retail environment, like a video-streaming company. When streaming videos, the videos do not turn over rapidly, instead they accumulate on top of one another. So, this type of recommendation system can pull from clearer historical data to make suggestions. Clothing, as a product, turns over much more rapidly than other items sold. Many clothing stores will introduce new items and will remove old items multiple times a month. Clothing recommendation systems also must factor in weather, seasonal spending trends, and that an identical item will likely not be sold from month to month.

3 Data Preprocessing

The data set used included over 700,000 individual J. Hilburn sales transaction records from about 80,000 unique customers spanning January 2017 to August 2019. The information for each transaction is stored in one fact table and three-dimension tables. The fact table consisted of all transaction-related variables, including stock keeping units (SKU), order date, first order date, customer ID, gross sales, gross units, product category, order ID, and item descriptions. The remaining three-dimension tables consist of general information about their stylists, customers, and products like location, product price, and color. These tables are joined together in Python using shared unique identifiers to create one comprehensive data set.

Before data exploration, the starting population was reduced by defining the beginning population. The definition used for this paper includes only customers who had their first transaction during the data collection window. First time customers within the observation window were flagged if the first order date column equaled the order date column. This was done to ensure the models only included customers whose complete life cycle with J. Hilburn is available and excludes customers with gaps in their purchase history. For example, as part of the data discovery efforts, a cohort analysis may reveal separation in customer groups based on each customer's acquisition date. Finding meaningful cohorts is another way to cluster customers. This decision resulted in a population total of about 44,000 unique customers, or a 46% reduction in our original data set.

4 Exploratory Data Analysis

Once the beginning population is defined to include first-time customers only, it was determined how the product category purchase activity was distributed. The first realization in evaluating the overall sales units was that custom shirts make up a significant portion, approximately 45%, of J. Hilburn's overall sales. This is even more significant as the second most commonly sold item, custom trousers, only make up approximately 13% of the total items sold. Intuitively, this is logical for two reasons. One is that custom shirts and trousers are less expensive than other items, such as suit jackets, making it more affordable to buy them in larger quantities. Also, it is more likely that a man might wear an item like a sports coat, belt, or pair of shoes many times while the same man is less likely to do that with a shirt or pair of trousers. This requires the man to purchase more shirts and trousers than other items of clothing. These two items together make up almost three-fifths of J. Hilburn's sales, which is vital to understand given recommendation systems tendency to over recommend popular items. While recommending commonly sold items results in higher accuracy evaluation metrics, it does not necessarily indicate a successful recommendation system. This is because a model that only recommends a few popular items is a recommendation system which lacks coverage. Coverage is when a recommendation system is successful in recommending a variety of items and not just a few items. Coverage can result in a sacrifice in overall accuracy, as one study realized when increasing items recommended from 49 to 695 resulted in a drop in accuracy from 82% to 68% [12]. Thus, the J. Hilburn models should suggest less common items that customers would not naturally find on their own in addition to some of the popular selling items.

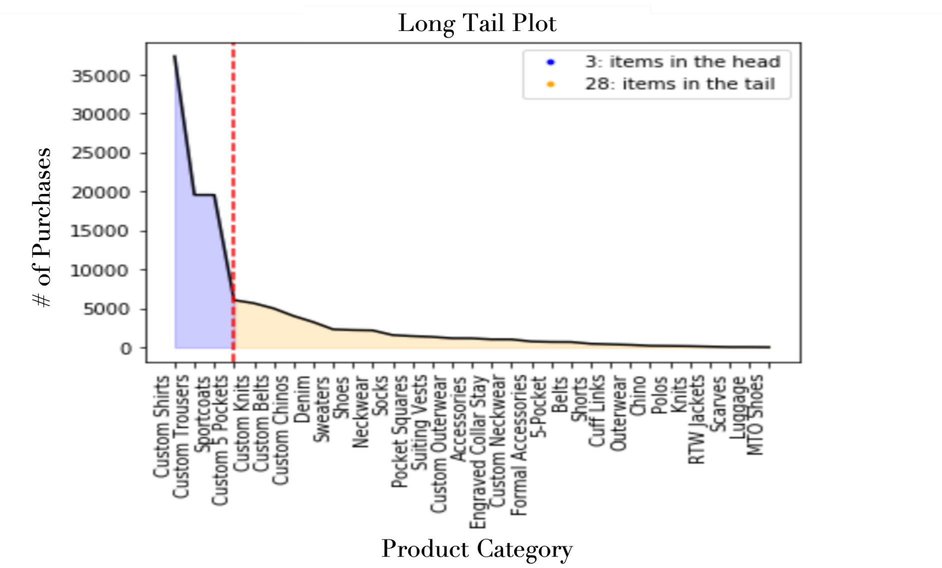

Fig. 1: Long Tail Plot of Product Categories

One major challenge for recommendation systems is the difficulty associated with cold start customers or customers who are making a purchase for the first time. These first-time customers and their lack of interaction data with J. Hilburn's brand makes it difficult for the algorithm to make meaningful item predictions. Conversely, new items that are being sold for the first time also suffer from cold start issues given these items do not have enough historical evidence to recommend items. Alternative engagement measures like clicks, page views, and time spent on different product pages can be used to supplement the lack of historical data points to recommend relevant items to new customers. Other techniques like supervised learning methods can also be used to address and overcome problems associated with cold start customers and items [13]. Access to the metrics mentioned above is not available in this data set, so an alternative method was implemented to determine a customer's interest. This alternative approach interprets a purchase as a signal of interest for an item. Another characteristic of a recommendation system is it typically relies on a customer to make several purchases before it can make accurate predictions. The data is explored to evaluate repeat customers and their subsequent buying patterns. While it would be ideal to build these recommendation systems solely based on customers who made at least four separate purchases, that action would significantly diminish the size of the beginning population. Ultimately, it was determined that striking a balance between excluding cold start customers and having a reasonably large data set is ideal, so the final data set includes first-time customers who made at least two purchases during the data collection time frame. This method results in eliminating about 40% of unique customers while allowing for more accurate prediction models. After filtering the initial data set to only include first-time

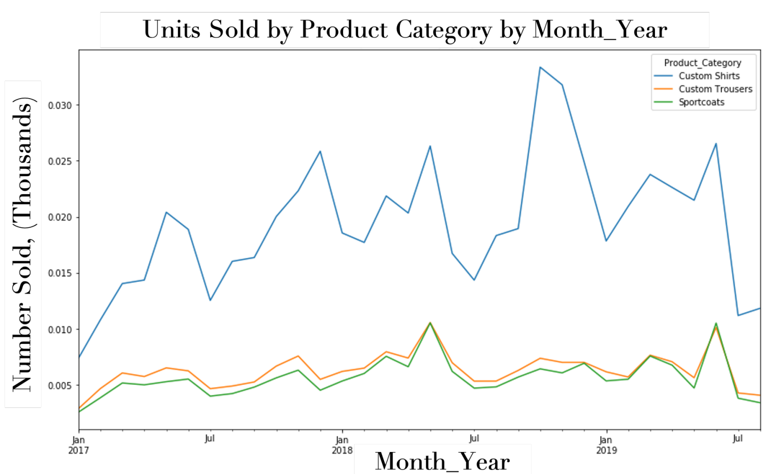

Fig. 2: Time Series Plot of Product Categories

customers with more than one purchase, the exploratory analysis is conducted. A long tail plot of J. Hilburn transactions is shown in Figure 1. It shows that the majority of purchases come from the three categories of custom shirts, custom trousers, and sport coats. These three product categories represent 72% of gross units sold with 45% coming from custom shirts alone. The imbalance of sales data is important to take into consideration when determining the proper approach to creating train and test splits to evaluate model performance. For example, stratification should be considered when trying to maintain proportional product distribution for splitting the data into test and train splits.

Understanding trends and purchase patterns by time is also important to ensure there are not any significant disruptions in the general historical pattern. Figure 2 shows the units sold by purchase month for the top three Product Categories. The seasonal trends in the data seem to be reasonably consistent. Although not shown, the other items follow a similar historical pattern without significant or unexpected deviations. Understanding this trend informs additional subsequent split methodologies for creating test and train populations.

5 Collaborative Filtering Model

Collaborative filtering attempts to build a prediction model based on users' past behaviors. Although it has many forms, Collaborative Filtering can be reduced down to two main approaches: memory and model-based.

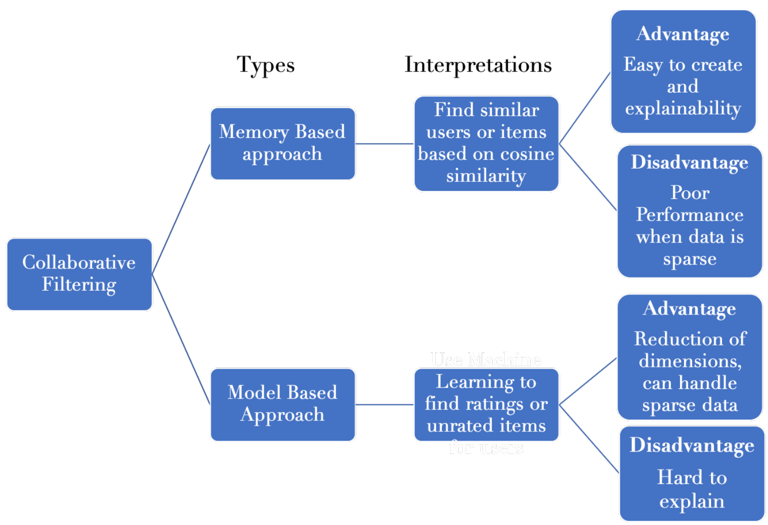

Fig. 3: Types of collaborative filtering approaches.

Model-based collaborative filtering uses machine learning algorithms to predict a user's behavior on unrated items. The model-based approach can be broken down in two ways, as shown in Figure 3. Memory based collaborative filtering can be divided into item and user-based filtering. The user's base collaborative filtering calculates the closest user or item by using cosine similarity or a Nearest Neighbor approach. In the Nearest Neighbor approach, the model looks for clients that have the same rating or buying patterns. That information is used to calculate a prediction for a subsequent customer. In the item-based approach, the model compares the relationships between items. For example, if a customer purchased a pair of trousers, then he typically also purchased a belt to complement his trousers. Cosine similarity is one way to measure similarity between two things. Memory-based models are generally considered the simpler approach because there is no training or optimization. However, memory-based models experience performance degradation when dealing with sparse data sets. The base model in this paper is an item-based collaborative filter model. This model uses customer purchase data as input and calculates the distance to similar items using cosine similarity as the distance function.

6 Ensemble Recommendation Model

After building the item-based, collaborative filter base model, a second recommendation system is built with a focus on having a novel approach. This model consists of individual clustering, K-Nearest Neighbors, and time series models input into a neural network. The motivation is to see if combining unique and focused topical models assisted by unsupervised machine learning could outperform the collaborative filtering base model recommendation system.

6.1 Clustering

J. Hilburn sells clothing to a wide variety of customers who have diverse purchasing power and clothing needs. Therefore, it is important that customer segments are defined for each customer as inputs to the novel ensemble model. The data set does not have many reliable continuous variables to use, so the segmentation strategy is based on recency, frequency, and monetary values (RFM). This assists the ensemble model in refining customer recommendations to items similar to those purchased by other users in the same cluster. Every elbow plot had a natural bend at 3; therefore, three customer clusters were created for recency, frequency, and monetary values.

The grain of the data is not in the proper format to begin clustering, so aggregations for each metric are created. First, customer ID and the max order date are grouped to calculate the number of days since the last purchase for the recency metric. Once completed, the model is initialized and fit using K-means to predict cluster values for each customer ID. These steps are separately repeated for frequency, grouped by customer ID by sum of gross units, and monetary, grouped by customer ID by sum of gross sales. It should be noted the data was not standardized before clustering, given each RFM metric was fit separately. There is no concern of maintaining scale in regard to this approach.

All customers are given an overall score by adding each of their three individual recency, frequency, and monetary cluster scores together. Each individual customer is given an overall score ranging from 0 to 6 based on their RFM clusters. To create a reproducible solution when segmenting scored customers, a quantile-based approach is used. This creates three segments defined as low-value, mid-value and high-value for a more intuitive naming convention. The summary statistics for each cluster can be found in Table 1 and a visualization for the frequency clusters can be seen in Figure 4.

Table 1: Customer Segmentation Analysis

Cluster Summary				
Cluster Name	Size	Recency(mean)	Frequency(mean)	Revenue(per item)
High-value	20,334	191 days	10	226
Mid-value	11,090	480 days	3	217
Low-value	12,602	80 days	2	212

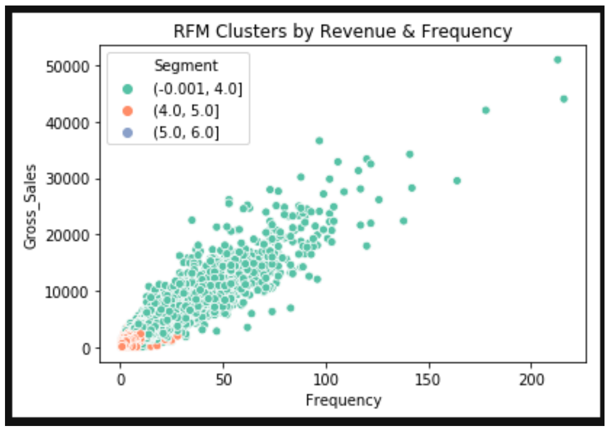

Fig. 4: Scatterplot Visualizing Clusters by Gross Sales and Frequency

6.2 KNN Model

The K-Nearest Neighbor (KNN) model is the portion of the ensemble model that provides recommendations. The KNN algorithm has many advantages over other classification algorithms such as producing competitive results, providing more scalability, yielding lower error rate, and requiring less computational power [14]. While KNN models still suffer from common recommendation system obstacles such as cold start and lack of coverage in items [15], this model was still an essential component to providing recommendations in the ensemble model.

The first step of a KNN model for a recommendation system requires the data to be held within an array containing both the number of items or categories and the number of users. This enables the model to utilize linear algebra algorithms to calculate the cosine similarity between closely related items of clothing. This model's method works similarly to the collaborative filtering base model.

6.3 Time Series Model

User preferences and item profiles change drastically over time [16]. When working with clothing data, seasonality and different periods of items being sold should be considered. For example, a pair of shorts should not be recommended in the winter months for the customer and so this must be accounted for in the model. Another paper documented that using a time series model can improve prediction [17]. In this model, a seasonal Autoregressive Integrated Moving Average (ARIMA) is used to model the seasonality of product categories using the item transaction data [18]. Time series factor tables are also incorporated to determine the appropriate seasonal factor [19]. The resulting seasonal factor is included as a new feature to input into the ensemble model.

6.4 Ensemble Model and Deployment

Ensemble modeling is a process where multiple modeling algorithms are used in tandem to predict outcomes. They combine the decisions from multiple models to improve the overall performance of the system [20]. As long as the base models are diverse and independent, results should have better accuracy and less error. The power of ensemble modeling is that it reduces the error rate of predictions [21]. For this ensemble model, the clustering results and the seasonality of the time series are added as features into the data set. Then weights from a user-item collaborative filtering model using KNN and a time series model were used to create a matrix which was used as the embedding matrix for a neural network. This leads to a higher accuracy and lower error in the test set. The base and ensemble models had an AUC of 85% and 91%, respectively.

7 Business Rules

Building a recommendation system that solely takes input and provides output is not enough to meet the complex needs of a clothing retail company. It is also necessary to consider the unique intricacies of the data set and design the model to produce recommendations which align with the business's goals. These adjustments to the model are referred to as business rules. Business rules by definition are the additional code added to the model to address this need and are commonly used in recommendation system design.

Three business rules were deemed necessary for these two recommendation systems to best fulfill J. Hilburn's corporate goals as can be seen in Figure 5. The first business rule implemented is to remove historical purchases from the list of recommendations. For example, if the customer has already purchased a specific blue dress shirt, the exact same shirt will not be listed on his recommendation list. The second business rule is to eliminate item SKUs that are sold out or out of season. And the final business rule serves to diversify the recommendation sort order. J. Hilburn predominately sells shirts and thus any recommendation system would tend to recommend shirts over other items of clothing. This does not allow customers to see much variety in recommended items. Therefore, once a category of clothing has been recommended to a customer the system will skip over subsequent recommendations of the same category in order to provide more diversity in products recommended.

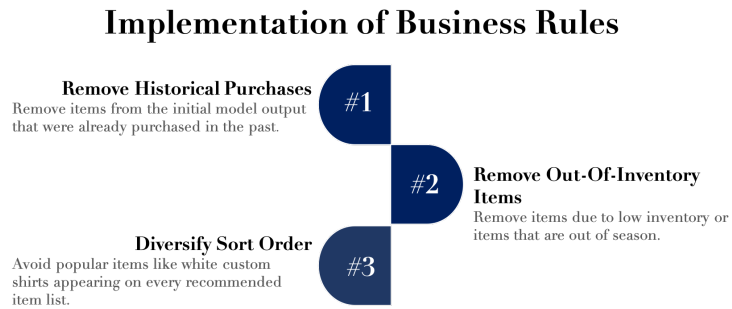

Fig. 5: Business Rules for Recommendation Model

8 Results

For this paper, two recommendation models were generated with the function of suggesting possible clothing items for J. Hilburn's customers to purchase. The first model, a collaborative filter, item-item model serves as a baseline. The second model is a significantly more complex neural network model composed of clustering, KNN, and time series models as input. Area under the curve, or AUC, was used as the metric to evaluate the performance of these models. AUC reports the measure of separability in the model and a higher AUC indicates a better performing model.

The base model had an AUC score of 86% and the neural network had an AUC score of 91%. For recommendation systems these AUC scores are extremely high; usually, the AUC metric is around one-tenth of these results. The likely reason behind the high AUC for these two models is that product categories were predicted instead of individual items or SKU numbers. For example, the product category of white shirts has many different permutations after allowing the customer to choose different collars, buttons, and sleeve options. This decision was made to increase explainability and transparency for both models. Either model could be adjusted to accept more specific input, such as a specific color of item or even a SKU number. However in this paper, utilizing product categories in the models reduced the number of features in the data set and thus allowed for unusually high AUC results when measuring the success of the models' predictions.

In Figure 6, how the recommendations differ for the input of a white shirt between the two models can be seen. Both models recommend a custom shirt first. Recommending the majority class is common among recommendation systems and this model is no different. Recommending a custom shirt, the item that makes up a significant portion of J. Hilburn's sales, is a likely guess for the model. Such a recommendation would also increase the accuracy of the model's predictions. Both models also predict pocket squares and a type of pants. These items are likely similar in cosine similarity to a shirt because they would be bought in tandem to complete an outfit. Thus, it is clear that the models perform similarly in regard to output, just in a slightly different order. The vision for a deployment strategy for both models is outlined in Figure 7. This model illustrates that the extensive preprocessing and machine learning to create the models in this paper should be deployed so that all customer recommendations are provided to the stylist via a mobile device for easy use.

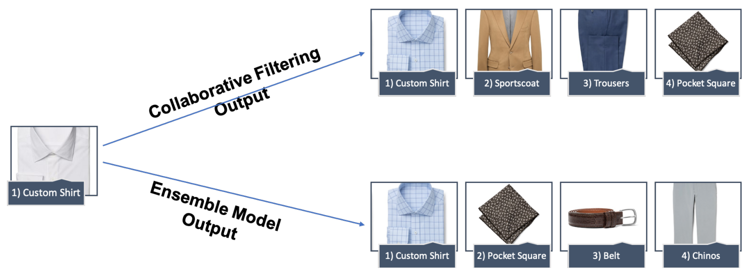

Fig. 6: Model Results based on White Shirt

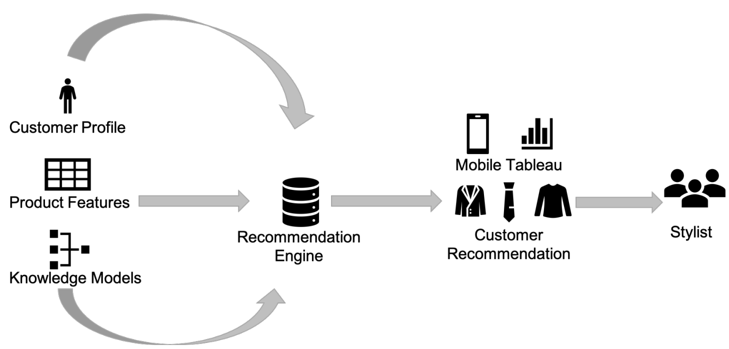

Fig. 7: Deployment Vision

While the AUC results for both models in this paper were significantly inflated over expected results due to the structure of the models, the neural network model did achieve the higher AUC result over the base model. However, there are still shortcomings to the neural network approach despite its slightly better AUC performance. The novel, ensemble model is difficult to explain to stakeholders due to the neural network component functioning as a black box. Inputting multiple models into a neural network is also much more computationally extensive than the base model, which lacked this complexity. The novel approach also required significantly more prepossessing and algorithmic generation, such as clustering and time series, before being input into the neural network. The item-item, base model is more easily explained because it utilizes a simple collaborative filtering algorithm. The data preprocessing required for the base model is also simpler and more easily understood. Since there is a trade-off between the resources required and the performance outcome, there is not a perceptible recommendation between the collaborative filter base model and the novel approach of the ensemble neural network model.

9 Limitations

Working with real life data comes with inherent limitations. The data used in this project was collected by J. Hilburn over the course of three years. Because the data was not collected specifically to build a recommendation system, it did not include some features which are commonly used for building recommendation systems. For example, many recommendation systems rely on users to like items, which is a key component in the output of other items a user may like, and this data set did not include that information. The alternative approach used to address this was to interpret a purchase as a like. This results in the recommendation systems created being classified as item-item recommendation systems.

There are also shortcomings with ensemble models, especially if they include a neural network. These issues include interpretability, data requirements, and hyperparameter tuning. Despite the successful results of ensemble models, it is difficult to understand what is going on within the neural network piece of the model. It is also challenging to explain to stakeholders why the model chooses the interactions and relationships it does. For most models to have successful operation and deployment, the person using it must understand the input and output to know if the model is working properly or not. This deficit in explainable recommendations results in these models not being deployed in the real world [22]. At the time of publishing, it was not yet determined whether or not J. Hilburn will implement the deployment of either recommendation system.

10 Future Work

Since the base model and ensemble models perform comparably, it defaults to J. Hilburn to choose whether they prefer a simpler, explainable model, or a less explainable but slightly more accurate model. J. Hilburn should then perform A/B testing with their stylists to determine if the recommendation system does in fact increase sales, either in recency between sales or in items sold per sale. This method is the approach most frequently taken by retail companies to evaluate recommendation systems and is most appropriate in this case even though the authors of this paper were not able to do so because they did not have the ability to deploy either model. Another approach to consider would be to completely remove shirts from the recommendation system results. Ultimately, the goal is to help a customer find items he would not normally find on his own. Shirt transactions make up almost 50% of J. Hilburn's portfolio, so removing shirts from the output would increase the prominence of other items. Also, the models in this paper produce recommendations of product category. This decision was made to enable a clean and clear presentation of final product. When this model is deployed for actual use, it would be important to adjust it to recommend SKU numbers instead of product category.

J. Hilburn also collects additional data on their customers' sizing and body measurements, which was not included in the data set provided to the authors of this paper. Adding this information would likely further improve the models as it is reasonable that a man's size would influence his clothing preferences. J. Hilburn could also initiate collecting additional data to indicate a customer's sentiment such as customer likes, ratings, clicks, or time spent looking at an item. Data like this is commonly used in building recommendation systems. This data would also allow for a more accurate model as ratings data is more reliable than working on the assumption that a purchase indicates satisfaction with the item.

Additionally, the K-means clustering algorithm can suffer from a lack of reproducibility. Every time it is executed, the centroid for each cluster is randomly chosen before computing local minimums for each group. In other words, K-means will create slightly different clusters every time it is executed. This issue can be addressed by employing a method that initializes from the same point every time regardless of how many times K-means is executed [23].

Lastly, recommendation systems learn and improve over time as companies collect more data. This allows for further testing of the models and for the engineers to continue to fine tune the recommendation system and continue to increase performance. The recommendation systems in this paper will also benefit from continuous oversight, additional data, and frequent updates.

11 Ethics

It is important to consider possible ethical repercussions before building any data science model. Recommendation systems are known to make customers wary in many ways, including concerns such as companies hoarding data and even cell phone applications eavesdropping on conversations [24]. While there are many examples of recommendation systems overstepping their bounds, perhaps the most notorious of recommendation system mishaps occurred when Target accurately diagnosed a teenager's pregnancy before her own father knew [25]. After this publicity nightmare, companies began to build more subtle and nuanced recommendation systems. After all, increasing accuracy of predictions is not worth scaring off paying customers.

Since consumers already feel uneasy about the personal data corporations possess, the recommendation systems built in this project were designed to find a balance between building an accurate recommendation system for J. Hilburn stylists while still protecting customer data. One way this was accomplished was to use anonymized data from the very beginning. The data set contained unique customer identification numbers as a feature but did not have any features which could be used to personally identify any specific customers. An example of this anonymity is that while the data set contained city, state, and size for each individual customer, it did not contain addresses or unique measurements, such as height and weight. Approaching the recommendation system in this way respects the confidential relationship between stylists and customers, while also protecting the customer's personal data from third parties.

12 Conclusion

In conclusion, there are many possible approaches for companies to increase sales by providing recommendations to their customers. This paper explored both a collaborative filtering model as a base model and an ensemble model as a novel approach. The novel ensemble model slightly outperformed the more traditional collaborative filtering model when using AUC as a metric. It is possible this is due to complexity of the different individual models comprising the ensemble model and the computational power of the neural network processing it. Though the novel approach model performed higher, it also required significantly more time to produce and run. Thus, in the interest of conserving man and computing power, the Collaborative Filtering base model is a more economical solution to the problem outlined in this paper. Additionally, both models were only as a good as the data available. Additional customer transactions and data features would have likely improved the performance of both models. As J. Hilburn continues the development and deployment of a retail recommendation system, it is appropriate to utilize these specific models and to consider adding additional data features as well as implementing real world testing.

References

1. Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G.: *Recommender Systems: An Introduction*. First Edition, September 2010.
2. Lake, K.: *Stitch Fix's CEO on Selling Personal Style to the Mass Market*. Harvard Business Review On AI, Analytics and the New Machine Age. June 2018.
3. Banerjee, S., Ghali, N. I., Roy, A.and Hassanein, A. E.: *A bio-inspired perspective towards retail recommender system: Investigating optimization in retail inventory*. IEEE, 2012.
4. Guan, C., Qin, S., Ling, W., and Ding, G.: *Apparel Recommendation System Evolution: An Empirical Review*. International Journal of Clothing Science and Technology, November 2016.
5. Yu-Chu, L., Kawakita, Y., Suzuki, E., and Ichikawa, H.: *Personalized Clothing-Recommendation System Based on a Modified Bayesian Network*. IEEE, 2012.
6. Sun, C., Gao, R., and Xi, H.: *Big data based retail recommender system of non E-commerce*. Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE, November 2014.
7. Hwangbo, H., Kim, Y.S., and Cha, K.J.: *Recommendation system development for fashion retail e-commerce*. Electronic Commerce Research and Applications, Volume 28, March-April 2018, Pages 94-101.
8. Logesh, R., Subramaniyaswamy, V., Malathi, D., Sivaramakrishnan, N., and Vijayakumar, V.: *Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble model*. Neural Computing and Applications, November 2018.

9. Su, X. and Khoshgoftaar, T.M.: *A Survey of Collaborative Filtering Techniques*. Advances in Artificial Intelligence, Volume 2009, October 2009.
10. Zhang, L., Peng, L.F., and Phelan, C.A.: *Novel recommendation of user-based collaborative filtering*. Journal of Digital Information Management, Volume 12, Issue 3, June 2014.
11. Sohrabi, B., Toloo, M., Moeini, A., and Nalchigar,S.: *Evaluation of recommender systems: A multi-criteria decision making approach*. Iranian Journal of Management Studies, Volume 8, Issue 4, October 2015.
12. Zhang, H., Ge,D., and Zhang, S.: *Hybrid recommendation system based on semantic interest community and trusted neighbors*. Multimedia Tools and Applications, 77, March 2017.
13. Park, S.T. and Chu, W.: *Pairwise Preference Regression for Cold-start Recommendation*. RecSys '09: Proceedings of the third ACM conference on Recommender, October 2009.
14. Adeniyi, D.A., Wei, Z., and Yongquan, Y.: *Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method*. Applied Computing and Informatics, Volume 12, Issue 1, January 2016, Pages 90-108.
15. Li, G. and Zhang, J.: *Music personalized recommendation system based on improved KNN algorithm*. IEEE 3rd Advanced Information Technology, Electronic, and Automation Control Conference (IAEAC), December 2018.
16. Koren, Y.: *Collaborative Filtering with Temporal Dynamics*. KDD, pages 447–455, 2009.
17. Zhang, Y.,Zhang, M.,Zhang, Y.,Lai, G.,Liu, Y.,Zhang, H., and Ma, S.: *Daily-Aware Personalized Recommendation based on Feature-Level Time Series Analysis*. In Proceedings of the 24th International Conference on World Wide Web (WWW '15). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1373–1383. DOI:https://doi.org/10.1145/2736277.2741087
18. Shumway, R.H. and Stoffer, D.S.: *Time Series Analysis and Its Application*. Springer, 2010.
19. Woodward, W.A., Gray, H.L, and Elliott, A.L.: *Applied Time Series Analysis with R*. CRC Press, 2017.
20. Ghai, B., Dhar, J., and Shukla, A.: *Multi-Level Ensemble Learning based Recommender System*. (2018).
21. Bar, A., Rokach, L., Shani, G., Shapira, B., Schclar, A.: *Improving Simple Collaborative Filtering Models Using Ensemble Methods*. 2013. Multiple Classifier Systems. MCS 2013. Lecture Notes in Computer Science, vol 7872. Springer, Berlin, Heidelberg
22. Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, Tat-Seng Chua *Attentional Factorization Machines: Learning the Weight of Feature Interactions via Attention Networks*. Twenty-Sixth International Joint Conference on Artificial Intelligence, Auguat 2017.
23. Khan, S.S. and Ahmad,A.: *Cluster Center Initialization Algorithm for K-means Clustering*. Pattern Recognition Letter, Volume 25, Issue 11, May 2003.
24. Cumbley, R. and Church, P.: *Is "Big Data" creepy?* Computer Law & Security Review Volume 29, Issue 5, October 2013, Pages 601-609
25. Corrigan, H.B., Craciun, G., and Powell, A.M.: *How Does Target Know So Much About Its Customers? Utilizing Customer Analytics to Make Marketing Decisions*. Marketing Education Review, Volume 24(2), July 2014, Pages 159-166.