**Logo

Description automatically generated**

**Teesside University**

**MSc Artificial Intelligence**

**Artificial Intelligence Ethics & Applications**

**Assignment 2**

**MARKETING BIAS IN PRODUCT**

**RECOMMENDATIONS**

**TABLE OF CONTENTS**

1. **Abstract ……………………………………………………………………………………………………….. 1**
2. **Introduction …………………………………………………………………………………………………. 2**
3. **Literature Review …………………………………………………………………………………………. 3**
4. **Methodology ……………………………………………………………………………………………….. 3**
5. **Data Analysis ……………………………………………………………………………………………….. 4**
6. **Statistical Analysis ……………………………………………………………………………………….. 7**
7. **Findings ……………………………………………………………………………………………………….. 9**
8. **Conclusion ……………………………………………………………………………………………………. 10**
9. **References ……………………………………………………………………………………………………. 11**

**ABSTRACT**

Artificial Intelligence (AI) is a cutting-edge technology that is revolutionising digital strategy. It can take data, analyse it, apply it to a problem, and then learn from the results. One of the most common applications of AI is Recommender systems(RS) in which the system provides personalized recommendations to the users by recognising the patterns from the user-product interaction. The applications of recommender systems include in e-commerce, clothes, news, and entertainment websites. However, these recommendations can be biased depending on the other factors. It includes marketing strategies in which the recommendation results can be manipulated. For case, a female client who would like a sports bicycle items may be less likely to suggested in case the sports bicycles are advanced with male clients.

In this report, we chose two datasets from ModCloth, a women's clothes website, and Amazon's Electronics category. These datasets allow us to consider the showcasing propensity triggered by a human model's body shape determination for clothing items, as well as study the effects of a human model's sex on electronics products. Complete data-analysis is performed on these datasets and the electronic products are more likely gets recommended to male users when compared to female users. On the similar note, clothing products are getting recommended more to female users than male users. This is clear evidence to the underlying bias which causing impact on the other quality products.

**1. INTRODUCTION**

Recommender systems are computer programmes that are designed to interact with huge, complicated information spaces and prioritise items that are likely to be of interest to the user. Many online e-commerce services, such as Amazon, Netflix, and Disney, use personalised recommendations. Researchers have been inspired to extend the reach of recommender systems into new and hard fields by the plethora of actual application expertise. Recommender systems are strategies that provide recommendations for any type of content that may be useful to a user during a decision-making process. Personalized suggestions are presented as ranked lists of goods in their most basic form. Recommender systems attempt to forecast the most appropriate products or services depend on the user’s preferences.

There are two techniques in recommender systems.

1. Collaborative filtering
2. Content-based filtering
   1. **Collaborative filtering:**

In Collaborative filtering, products are recommended based upon interaction between similar users. For example, consider a social media platform like Facebook or Instagram. If there are 3 persons and I & II are friends, II & III are friends then I is recommended to III and vice versa.

Collaborative techniques have the advantage of requiring no data about clients or items, allowing them to be used in a variety of situations. Furthermore, the more clients associated with things, the more accurate the unused suggestions became for a fixed set of clients and things, modern intelligent recorded over time deliver unused data and make the framework increasingly effective.

However, because it only considers past interactions to form proposals, collaborative filtering endure from the “cold start problem”. It is incomprehensible to recommend anything to modern clients or to prescribe an unused thing to any clients and numerous clients or things have too few intuitive to be productively dealt with. This downside can be tended to in numerous way: suggesting arbitrary things to unused clients or modern things to arbitrary clients (arbitrary methodology), prescribing well known things to modern clients or modern things to most dynamic clients (greatest desire methodology), suggesting a set of different things to modern clients or a modern thing to a set of targeted users.

**1.2. Content based filtering:**

In this technique, the product is recommended based on the number of users to the product. For example, in Amazon prime video if there is a new movie added, then it got recommended based on the users watched the movie. If there are a greater number of viewers to the movie, then more it got recommended to the other watchers.

Content-based solutions suffer from the "cold start problem" significantly less than collaborative approaches: new people or things can be represented by their attributes (content), allowing for significant suggestions for these newly created entities This problem will inevitably affect new users or products with previously unknown features, however this is unlikely to happen once the framework has developed.

**2. LITERATURE REVIEW**

In general, Recommender systems allow us to request controlled tastes from the data firehose that is aimed at us every day of the week, by highlighting a small number of particularly relevant or profitable items from an infinite collection. And, while they're enormously important pieces of innovation, they also have a number of genuine moral disappointment modes, many of which arise as a result of companies' tendency to build recommenders to reflect client criticism rather than considering the broader implications these systems have for society and human civilization. Those proposals are crucial, and they're coming together quickly. Twitter and Google recommender calculations routinely impact open assumption on critical ethical concerns of our time, in some cases intentionally, and in others by accident. So, rather than allowing society to be reshaped in the image of these efficient calculations, perhaps it's time we asked a few big questions about the kind of world we need to live in and then worked backwards to see what our answers would be.

Hence, research questions are as follows.

* What is the marketing bias in product recommendations?
* What are their ethical issues?

**METHODOLOGY**

**Quantitative Study**

**Dataset**

Two real-world e-commerce datasets from ModCloth, a women's clothes website, and Amazon's Electronics category. ModCloth is an online store that sells women's apparel and accessories. Many products have two human models with diverse body shapes and estimations of these models, which is a one-of-a-kind quality of this information. In addition, users can give the item sizes they acquired and fit input ('Just Right’, ‘Slightly Larger, 'Larger, 'Slightly Smaller, ‘Smaller'). Then, in this dataset, we focus on the dimension of human body shape as a source of encouraging propensity. Another survey dataset was collected from Amazon's Electronics category, with Clothing serving as an assistant category. The finest of Amazon's open data has been used to create this collection. This dataset is based on the best of the open Amazon 2018 Dataset and has been tweaked to support the research goals. On this dataset, we consider gender to be the target promoting propensity.

**DATA ANALYSIS**

For the Data analysis, following python libraries are used.

Numpy: NumPy is used to perform mathematical operations.

Pandas: Pandas is used for filtering of data, segmenting, and segregating.

Matplotlib: Matplotlib is used for Data visualization.

Seaborn: Seaborn is used for making statistical graphs.

Scipy.stats: This module is used for correlation functions, masked and statistical tests, probability distributions, and frequency statistics.

**Overview of data**

**Electronics**

**Table

Description automatically generated**

**ModCloth**

**Table

Description automatically generated with low confidence**

**Column Data**

**Table

Description automatically generatedTable

Description automatically generated Electronics ModCloth**

**Check for Null values**

**Table

Description automatically generated with low confidenceA picture containing text

Description automatically generated** **Electronics ModCloth**

**Visualization of data**

**Heatmap**

**Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated Electronics ModCloth**

**User Rating**

Chart, bar chart

Description automatically generated**Chart, bar chart

Description automatically generated Electronics ModCloth**

**Bar Graph**

**A picture containing graphical user interface

Description automatically generatedGraphical user interface

Description automatically generated** **Electronics ModCloth**

**CAT plot**

**STATISTICAL ANALYSIS**

The item predilection of a customer is divided into two dimensions:

1. a user's desire to buy something.

2. the user's assessment of the time spent participating.

Then, using the ModCloth and Electronics datasets, we undertake observational studies to address marketing bias across the two dimensions. To begin, we look to see if a certain promoting influence has a tendency in a consumer's item selection process. In our datasets, we specifically look for a relationship between product image and client character in terms of interaction recurrence. At this point, we consider consumer happiness with acquired things to be a work of item image, client personality, and second-order interactions. Rating scores on ModCloth and Electronics, as well as binarized fit feedback (i.e., if the garment item fits the customer) on ModCloth, are examples of consumer input signals.

**Marketing Bias vs. Product Selection**

We address promoting bias in product choice by analysing the association between product image and user identity in observed information regarding interaction frequency, rather than conducting real-world experiments with irregular assignments, due to the limitations of conducting real-world experiments with irregular assignments. Our erroneous assumption is that the picture of the product and the identity of the user are statistically unrelated. We should expect fewer discrepancies between their measured frequencies and the initially expected values if this assumption is correct. As a result, the Pearson's Chi Squared Test Measurement can be used to determine whether these two elements are related in terms of frequency.

**Table

Description automatically generatedMarketing Bias vs. Consumer Satisfaction**

Now we explore consumer satisfaction as a work of product image and user identity through a standard statistical technique: two-way analysis of variance (ANOVA). We utilize rating scores to speak to users’ satisfaction with respect to the overall quality of their consuming experience on both ModCloth and Electronics. For ModCloth, we study consumer satisfactions regarding their fit feedback. The null hypotheses of our tests includes

* Average consumer happiness is equal across product image groups
* Average consumer satisfaction is equal across consumer identity groups
* Table

  Description automatically generated There is no interaction impact on satisfaction between product groups and consumer groups.

**FINDINGS**

From the analysis, we can say that the user data is co-related to marketing strategies used by the retailers. We focused on recommendation systems that were trained on explicit feedback in this study. The affiliation between user identity and product image is consistently unique in terms of recurrence distribution, implies the presence of bias within the collected datasets. The ‘self-congruity’ design is consumers may generally tend to be associated with products with similar impressions as their identities. Such an affiliation eminently causes underrepresentation of market segments. The relationship between marketing factors and customer satisfaction is rather complicated. We watch disparities across product groups and user groups, whereas the presence of their interaction impact depends on the type of product and the type of satisfaction measure.

**Improving market fairness in recommendation systems.**

Although there is market bias, there are certain practices which can avoid this problem. Following are the some of the techniques which increases the market fairness in recommendation systems.

* Calibrating the prediction errors across different markets segments eventually leads to better recommendation results.
* Better data collection is also one the remedy for the market bias.
* Matrix factorization
* Error Correlation loss

**CONCLUSION**

In this report, on two real-world e-commerce datasets, we investigated a potential source of bias marketing in the form of the relationship between interaction input, product image, and client character. The inter-correlations between these components can be confirmed through observational thinking, and 'self-congruity' patterns can be discerned within the item selection process, resulting in the underrepresentation of a few highlight segments. We focused on showcasing reasonableness and investigated how traditional collaborative filtering calculations respond to one-sided input data. We discovered that this predisposition can be transferred to recommendation outcomes in varying degrees. We created a blunder relationship system, which explicitly calibrates the value of forecast mistakes over diverse market segments. Exploratory comes about illustrate that a common accuracy-fairness trade-off may be achieved by using this relationship misfortune.

Personalized experiences are appealing to both online consumers and internet users. Most people prefer to save time by using recommendations provided by a separate website since they do not want to lose time searching for and becoming lost in information. More businesses are employing different recommender systems to customise their business transactions for a better customer experience. Implementing a recommender system can be costly, but the benefits of highly tailored content will undoubtedly outweigh the costs. However, marketers should be use Artificial Intelligence wisely to prevent losing the trust of customers Transparency in data gathering and utilisation, as well as the use of appropriate marketing strategies, are all part of this. Despite the hazards, artificial intelligence is here to stay, and it will undoubtedly continue to influence marketing in the future.

**REFERENCES**

[1] “Introduction to recommender systems.” https://tryolabs.com/blog/introduction-to-recommender-systems (accessed Apr.26 2022)

[2] “Tutorial on Fairness of Machine Learning in Recommender Systems”

https://fairness-tutorial.github.io/ (accessed Apr.26 2022)

[3] "Various Implementations of Collaborative Filtering." https://dl.acm.org/doi/10.1145/3130348.3130372 (accessed Apr.27 2022)

[4] "Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey". https://github.com/nzhinusoftcm/review-on-collaborative-filtering/blob/master/7.ExplainableMatrixFactorization.ipynb (accessed Apr.27 2022)

[5] "Your First Machine Learning Project in Python Step-By-Step". https://microsoft.github.io/ML-For-Beginners/#/ (accessed Apr.27 2022)

[6] Stacy Landreth Grau and Yorgos C Zotos. 2016. Gender stereotypes in advertising:

a review of current research. International Journal of Advertising 35, 5 (2016),

761–770

[7] "ANOVA using Python". <https://www.reneshbedre.com/blog/anova.html>

[8] "Marketing bias in recommender systems". <https://github.com/MengtingWan/marketBias>