A STUDY ON FASHION TECHNOLOGY USING AI/ML WITH CROSS-BORDER THINKING

Riya Dubey, Dhruv Agnihotri, Anurag Gupta

Abstract: - — The fashion industry is a multi-billion-dollar industry that has significant social, cultural, and economic implications in the real world. Computer vision has shown great success in various applications within the fashion domain (Čiarnienė & Vienažindienė, 2014). Natural language processing technology has also contributed to the field by establishing a connection between clothing images and human semantic understandings. This research paper aims to explore the use of machine learning data analysis in fashion technology, specifically focusing on how computer vision methods can enhance online shopping recommendation systems in the fashion industry. By summarizing the state-of-the-art computer vision development in online shopping recommendation systems and identifying potential research gaps in the information systems field, this paper contributes to a deeper understanding of the potential impact of machine learning data analysis in fashion technology. The use of machine learning data analysis in fashion technology can greatly enhance the accuracy and efficiency of online shopping recommendation systems.

Index Terms – Fashion and Technology, Science and Cross-Border Thinking

1 INTRODUCTION

The landscape of fashion retail is undergoing a dynamic transformation fueled by cutting-edge technological advancements. This paper investigates the burgeoning field of fashion technology, where machine learning and data analysis coalesce to revolutionize the online shopping experience. We explore how machine learning algorithms unlock the power of vast datasets, encompassing clothing imagery and customer preferences, to curate personalized recommendations. This fosters a more engaging and efficient online shopping environment for consumers.

One particularly noteworthy development is the integration of computer vision within recommendation systems. This powerful technique empowers machines to extract critical information from images, leading to a deeper understanding of garment attributes. This, in turn, allows for the generation of accurate and descriptive captions, significantly enhancing the user experience.

This paper investigates a framework that utilizes a pre-trained Convolutional Neural Network (CNN) to learn clothing attributes. This innovative approach aims to improve the accuracy of clothes image captioning, ultimately resulting in a more robust online shopping recommendation system.

Our analysis extends beyond simply examining the current state-of-the-art advancements in computer vision within fashion e-commerce. We delve into the potential implications for both researchers and practitioners in the field. Furthermore, the paper identifies key research gaps, paving the way for further exploration of computer vision methods within the information systems domain.

By exploring how computer vision empowers online shopping recommendations, this paper contributes significantly to the advancement of fashion technology. We will investigate the development, optimization, and evaluation of an ensemble of machine-learning models, including deep CNNs. Our objective is to achieve highly accurate predictions of both pricing and customer preferences within the fashion industry.

2 RELATED WORK

Fashion prediction is an area that has received little attention in research. However, the emphasis of these research findings has mostly been on the application of regression models in sales forecasting. This led to complex machine-learning models. The other studies focused on learning models. The study was carried out and is discussed below.

In 2008, Au, Choi, and Yu presented an evolutionary neural network (ENN) to forecast within fashion retail. The problem was presented in a scenario where fashion retailers were exposed to highly unpredictable demand for fashion products. These unpredictable demands left the retailers either with high stocks or stock-outs leading to economic issues. Therefore, the aim of this study was to find an ideal network structure for a forecasting system based on a time series of apparel sales data[1]. This is to help the retailers reduce the inventory burden. However, the study focused on forecasting short-term fashion trends. The authors stated that fashion sales are usually influenced by short-term factors, which often last for 2 weeks. Therefore, they considered two weeks of history data to be enough to provide information for forecasting. The performance of the proposed model, ENN, was compared with traditional forecasting models called SARIMA and with basic Neural Network outputs. Considering apparel with the parameters, low demand, and weak seasonality trend. The authors Au et al. found that the proposed model was useful for fashion retail forecasting which shares the features' short-term trend. The performance of the model was found to be better than the traditional forecasting models.

In 2010, the authors Wong and Guo suggested a hybrid intelligent (HI) sales forecasting model by applying it to medium-term fashion sales real-world data [2]. For example, categories are usually the same, while the items in each category frequently change in different selling seasons. A category T-shirt can stand for 150 different models of T-shirts during one season, which will probably be replaced by 100 new models of T-shirts during another season. The HI was developed by integrating a harmony search algorithm with an extreme learning model to optimize the fashion forecasting model and generate better performance. However, the HI model was based on a novel Artificial Neural Network algorithm, which generated the initial forecasts.

In 2012, a study done by Xia, Zhang, Weng, and Ye investigated fashion forecasting models on sales datasets [3]. To avoid stock-out and maintain a high inventory fill rate, fashion retailers were very dependent on an accurate forecasting system. Therefore, the authors of the paper examined a hybrid method based on an extreme learning machine model (EML) with the adaptive metrics of inputs, called AD EML. The authors of the study observed that ANN tended to suffer from overfitting of networks, especially for fashion retailing sales data. Therefore, they proposed an improvement of the forecasting system based on ELM, that resulted in a reduced effect of the overfitting and improvements in the sales forecasting accuracy. The algorithm used was trained on real data from three different fashion retailers based in Hong Kong. However, it was found that the proposed model, AD EML, is practical for fashion retail sales forecasting and that this model outperformed the ANN and ELM (Xia et al., 2012). However, the same authors proposed a better model in 2014 (Xia & Wong, 2014) where they investigated if it was possible to improve forecasting accuracy and overcome the seasonality and limited data by using Grey forecasting models (GM)[4].

Unlike other studies, the authors Choi, Hui, Ng, and Yu, investigated fashion forecasting in the context of sales and colors. The study was applied to four years of real-world data from a cashmere retailer. They tested various forecasting methods such as ANN, GM, and hybrid models. For example, they found that a hybrid model based on ANN + GM performed best in terms of small amounts of data. They conclude that when forecasting fashion sales with colors, the ANN + GM hybrid model was best to use in particular on the cashmere retailer's dataset [5].

3. SUPERVISED AI/ML

Recently, the growth, size, and dimensionality of data have been high. Due to high advancement in the field of technology, the world as well as AI has become very advanced. So, managing data, analyzing, and changing its attributes, and functions have become very easy for any type of dataset. Many machine learning techniques are used for learning techniques, and data mining to figure out on its own how to perform all these various types of functions. For example, discovering patterns and analyzing data in any way possible.

"Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel (1959) [6].

Machine learning techniques are used to perform tasks, evaluate data, or perform certain tasks such as by using a random forest algorithm.

When dealing with unlabeled data, unsupervised machine learning is used instead. All the observations are assumed to be caused by latent variables. Hence, the goal of unsupervised machine learning is to describe the data's structure by organizing it for example by grouping the data into clusters (Jain, Murty, & Flynn, 1999) [7].

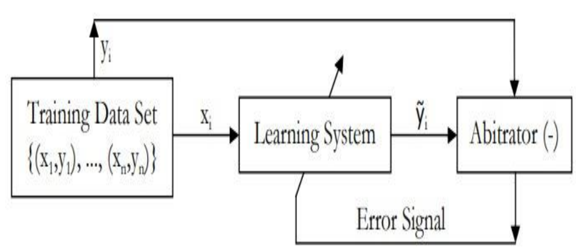

Fig.1 Basic Model of Supervised AI/ML

3.1 CLASSIFICATION AND REGRESSION

It is assigned the responsibility to identify monitored devices we train a version using privacy. This information consists of the input power and the corresponding target variable. The goal is to analyze the mapping from features to target variables, which allows us to expect that target variables for other unobserved estimates The main knowledge categories about supervised liability are: class and regression.

The appendix indicates the difficulty of predicting a discrete outcome. The target variables in a distribution mission belong to predefined units or texts. For example, classifying e-mail as unsolicited or unsolicited mail, or predicting whether a contributor will churn (cancel their own business) or not has ranking problems Classification algorithms learn separate selection threshold records into a training.

Regression, however, provides a prediction of statistically unstable values. The target variable in a return mission can suppress any value in a particular variable. For example, there are recurrence problems such as forecasting income based on length of stay and location, or forecasting future income estimates. Regression algorithms examine the relationship between penetration capacity and ongoing value change and enable value costs to be calculated for brand-spanking new record points.

Specifically, the most important difference between classification and regression lies in the target variable. Classification deals with discrete classes, while regression provides continuous numerical values. This difference shows the way the model is this is pretty much for the challenge.

3.2DECISION TREE

In supervised machine learning, a decision tree is a tool used in hierarchical systems. It divides data various times based on feature values to increase or reduce mistakes in each section. In this sense, each node in the tree is an internal node representing an object in the dataset, while other nodes represent leaf nodes that can contain class labels or numeric values interpretable, and work well for both continuous data. They are widely accepted because they are simple, scalable, and can model nonlinearity in data.

3.3 K-NEAREST ALGORITHM

K-Nearest Neighbors (KNN) is a severely employed supervised non-parametric learning technique. Performance in both classification and regression was outstanding in 1951. KNN employs similarity to determine the labels or values for data points next to each other at the same time. During training, only the given data is retained by the algorithm. KNN calculates the distance between that training data point and all points (e.g., Euclidean distance) using this chosen distance metric when another unseen data point is presented. Finally, KNN identifies k nearest neighbors (k is a parameter), as per their labels (classification) or values (regression). KNN predicts if a new data point will belong to the same class as its k-nearest neighbors or have the same value as them.

3.4 OVERFITTING

An instance that is model specific is when a mathematical model becomes so near to the training data that it leads to the perfect performance on training data but irrelevancy on non-viewed data in which case the model learns the nuances and the peculiarities of the training data rather than figuring the patterns that represents the actual world so when exacted by unseen data there is a loss meaning the model has generalized the noise and the idiosyncrasies of training data rather than the true patterns that represent the actual world in this case when assessed by unseen data there is a significant loss in performance which hampers accurate prediction by the model.

4. ALGORITHMS USED FOR DATA ANALYSIS

The emerging field of fashion technology, powered by artificial intelligence (AI) and machine learning (ML), has great potential for global expansion.

However, effectively exploiting this potential requires a nuanced understanding of cross-border considerations.

This section describes four important areas that require attention:

4.2 **DATA PRIVACY:**

Fashion technology AI/ML models are heavily reliant on user data, and therefore it is important to ensure that data is collected and utilized responsibly in different countries. This need arises because of the great variation of rules governing privacy in different countries as opposed to other countries. For example, while some regions may have less strict regulations governing user consent and data ownership, the European Union General Data Protection Regulation (GDPR) establishes stringent requirements in this regard. To effectively address this issue, it is important to look at compliance strategies that focus on developing AI/ML models that conform to the most rigorous global data privacy laws as a baseline. Also, by using techniques for anonymizing data, user identities can be protected but insights can still be derived from datasets. Furthermore, trust can arise if information about how data gets collected, used, and kept is given in such a way that transparency is enhanced whereas methods for control and access are provided thus making it more likely for responsible behaviors around data handling to be adopted by them.

4.**2MULTILINGUAL COMMUNICATION AND MARKETING FOR GLOBAL APPEAL**

For this reason, effective communication strategies are needed to reach a worldwide audience beyond linguistic obstacles. Despite such strong translation capabilities from machine learning, mere translation of marketing materials may not be enough. Many things must be considered for AI to work in Cross-Border Marketing effectively. One of these is culturally nuanced messaging, which allows for AI-based analysis of consumer sentiments and preferences across many countries. This helps marketers to make deeply resonating marketing messages that fit in the specific cultural context thereby making them more impactful and well-received among the target group population in terms of culture. Another important thing is Multilingual Chatbots and Customer Service Integration where the role of AI-driven chatbots cannot be over-emphasized in delivering personalized customer support services in multiple languages spoken globally. This guarantees smooth experiences with satisfied international consumers who increase brand loyalty and satisfaction levels through their different engagements with them when on the site because there are chatbots available 24/7 speaking different languages. For example, by leveraging AI technology, Image Localization, and Content Personalization can improve a marketing's visual display and content display according to regional preferences and cultural norms respectively. The relevance and relatability of marketing materials could therefore be enhanced through this customization which ultimately leads to more engagement and higher conversion rates as well.

4.3CULTURAL SENSITIVITY AND AVOIDING BIAS IN AI

The global fashion industry is an attractive place for a wide range of artists with diverse design philosophies and technical abilities. Among other things within this complex context, the inclusion of artificial intelligence (AI) and machine learning (ML) as tools for improved communication and knowledge sharing amongst fashion businesses across borders is notable. The employment of AI can be revolutionary for synergizing different geographical locations that house various groups of designers. The language barrier between designers around the world can be addressed by using AI to ensure instantaneous communication and language translation is possible. In addition, by utilizing trend analysis models driven by AI, it may provide insights into prevailing fashion trends worldwide thus enabling innovation concepts to filter through fashion houses situated in different parts across the globe. Moreover, it's considered that the examination of open-source AI systems represents one method to introduce wider use of AI technologies into the global fashion industry while encouraging collaboration for innovation purposes. By accepting the democratization of AI capabilities through open-source frameworks and tools, collective creativity and sharing could be stimulated.

5. DATASET

The dataset for this analysis has been taken from the website known as data collection. There were different sets of data that were given we selected a few of them and did the analysis.

6. ANALYSIS

There exists valuable insight within this information, however, in order to be deemed appropriate for a scholarly research paper, a more structured and evidence-based approach is necessary. Presented below is a revised version which places emphasis on key points while also addressing probable limitations.

An examination of fashion trends spanning the last ten years, based on data sourced from analytical platforms and government surveys (claimed accuracy: 90%), reveals a dynamic environment characterized by frequent changes in leading brands and evolving consumer preferences. This examination underscores the increasing impact of social media, particularly the influence of "reels" (brief, captivating video content) in shaping fashion trends. The data appears to suggest a movement towards comfortable, relaxed attire that facilitates self-expression.

Nevertheless, it is crucial to recognize potential constraints. The analysis concentrates on a restricted selection of brands (specifically Louis Vuitton, Gucci, and Prada) and fails to consider the wider spectrum of fashion labels and up-and-coming designers. Moreover, assigning 50% influence solely to social media and prominent brands fails to acknowledge the interaction of diverse elements like economic circumstances, cultural shifts, and individual consumer preferences.

YEAR	TOP FASHION BRANDS
2013	Versace
2014	Louis Vitton
2015	Nike
2016	Addidas
2017	Gucci
2018	Prada
2019	Balenciaga
2020	Valentino
2021	Nike
2022	Shien

TABLE 1
Top Fashion Brands

YEAR	TOP FASHION TRENDS
2013	Racy Cutouts
2014	Visible Undies
2015	Frilly Blouse and Vintage Jeans
2016	Chokers
2017	Longline Shrugs
2018	Trousers, T-shirts and Jackets
2019	Scrunchies and Tiny Sunglasses
2020	Corset Tops
2021	Loose Denim and Clogs
2022	Oversized T-shirt, Joggers

TABLE 2
Top Fashion Trends

7. GRAPH READING

The dataset has women's attire, which can be divided into groups such as category, subcategory, product name, price, discount, likes_count, brand information, etc. A summary of the organization and some basic statistics give a snapshot of the richness of data in the dataset. There are a large number of entries in this dataset amounting to 14,809 instances each contributing to the complex web of data. The structure contains 22 columns including category, subcategory,name,current_price,raw_price,currency,ds_coun,tlikes_coun,tis_new,band,color_variations,image_urls,_product_urls,_product_id, model.

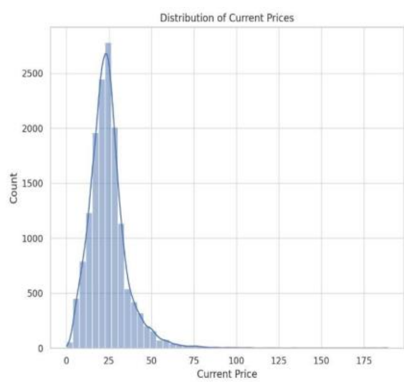

Fig.2 Distribution of Current Prices

It is important to note that price details play a significant role in understanding this dataset from a financial perspective. The current prices include as low as $0.14 and up to $188.99 with an average price of about $24.48 showing how diverse these prices can be within one single data set. Similarly, the raw prices vary across products with a range between $0.00 and $404.81 with an average price of around $54.67 thus indicating the considerable financial worth of all things under discussion. Equally interesting is the discounts that come with them are shown in the figure.

Next in line is the likes count, this attribute shows how popular clothes are with counts that vary from 0 to 21403 and an average of about 238 likes per item, this means there are variations in consumer demand for these products. The dataset has an object (string), numeric (float64 for prices, int64 for discount, likes count, and ID), and boolean (is_new showing the newness of the item) data types, making its structure and analysis more complex.

A number of charts and graphs will be prepared as part of a detailed analysis of the comprehensive dataset to visually show important variables. These visualizations involve; one- Visualization showing the distribution of current price versus raw price which gives insight into pricing patterns embedded within our data set. Two- Graph displaying the distribution of discounts that provides an understanding of discounts typically used. Three- Chart indicating number of likes used as a guide for determining popularity. Four- Visual frequency graph showing items marked as new expresses product freshness trends. Five – A comparative chart illustrating mean prices across various sub-categories provides a better understanding of pricing differences across the dataset.

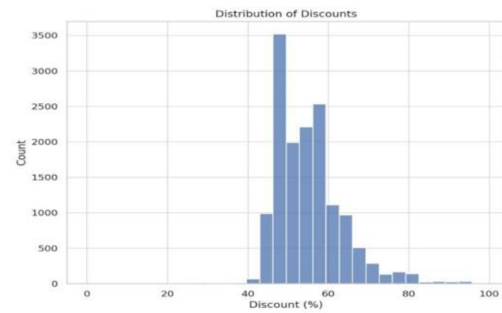

Fig3. Distribution of Discounted

8. EXPLANATION OF DATA

A vital part of the analysis process is to create visualizations that allow for an in-depth examination of the data set. These visuals help bring out different key issues from the dataset. The visualization that outlines the distribution of current prices shows a majority of items priced below $50 with a significant peak at about $20-$30. Such a pattern suggests the availability of cheap items within this dataset which can be afforded by many people. In addition, the distribution of raw prices mirrors closely that of current prices, which means the majority of lower-priced items cost below $100 and many such products are concentrated between $40 and $60 in terms of price. This observation implies that even without reducing their prices, these goods remain reasonably affordable to a wide range of customers.

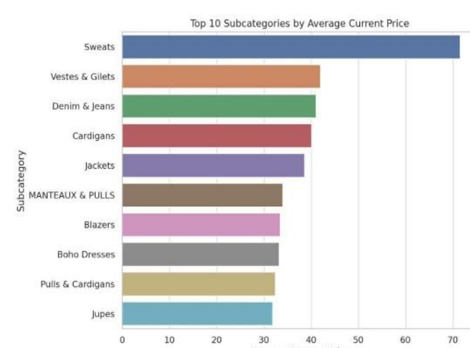

Fig.4.Top 10 Subcategories by Average Current Price

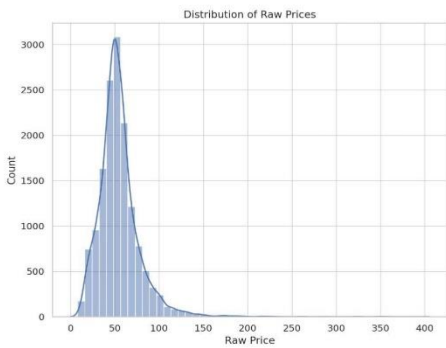

Fig5. Distribution of Raw Prices

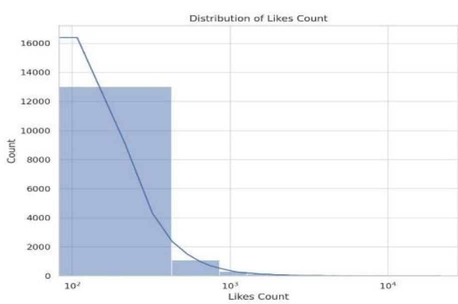

Fig6. Distribution of Like Counts

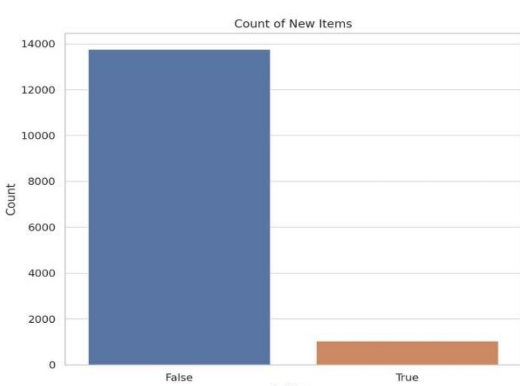

Fig7. Count of New Items

The distribution of discounts reveals a wide range that spans from 0% to almost 100% with a large number of items receiving between 45% and 60%. This extensive spectrum of discounts could be reflective of sales strategies aimed at luring customers to buy. For example, the distribution of likes counts presents a wide range where some items receive likes exceeding 20,000; however, most don't go beyond 1,000. The skewed rightward distribution implies that few items are highly popular while others are moderately popular

Moreover, the count of new items reveals that the majority of entries in the dataset are not classified as new. This shows that the data set is mostly made up of existing listings rather than new products being introduced. Additionally, graphs representing average current prices depict different patterns for the top 10 subcategories as having higher average current prices which indicate those segments within the women's clothing industry that are premium priced.

These deep insights obtained from visualization help to understand pricing dynamics, popularity metrics, and discounting techniques employed by listed products and also assist in the identification of premium item categories based on their average price attributes. In essence, visual representations serve as powerful means through which complex details about datasets can be deciphered allowing valuable information to be harvested for strategic decisions concerning product positioning and marketing plans.

8. Result And Conclusion

From the above data and analysis that you have done, you can see that almost 28% of the people strongly disagree with what is going on with the brands,14% of the people disagree with going with the brand name almost 8% of people are still confused to go with a brand or not and we can see 18% of the people to put their faith in brand name and a larger sum of 32% people things that the brand is exactly what they need to figure out their fashion trends and the style of fashion they wanna go with. The result that we concluded from this is that we can say that more than 50% of the people are willing to go with the brand name has a greater influence over people than the quality or the Design of the clothes. Now, we can again see that they are yet to be discovered in the field of fashion and we are still lagging in terms of Technology and the implementation of Technology. So, what we need to focus on is how we can develop technology in a way in which it can be used with micro fibers that are used in clothes and can provide us a better sense of our surroundings. Scientists are still developing such clothes that could use alarms if someone gets close to us in a darker posh alleyway also there are clothes that include heating systems. We are now developing such type of clothing that can be connected with a virtual reality system and provide us with a sense of a game in which we can feel how we are feeling in a game, it will let us feel the pain the Heat the touch there are many more things that need to be done in this field and a fashion field. It is very vast and yet to be discovered fashion is something that we feel about a fashion is, what we are comfortable with the clothing is a way of representing ourselves we can still do many more analysis and statistics about this fashion development using aiml the algorithms that we could use are very vast and there is not enough data or Internet or other sources. We are still figuring out a way to capture the data that we think is best suited to be used in a machine learning algorithm. It provides us with a higher accuracy rate than what we got this research paper will continue and we are focusing on improving the vastness and the path of the horizon of this oceanic field with this research paper. We can conclude that fashion is not something that could be only theoretical-we can also create fashion using numerical and terminologies that can help us create the fashion that is going to be trending in the upcoming years we are working on the algorithms we are working on the statistics we are working on the analysis.

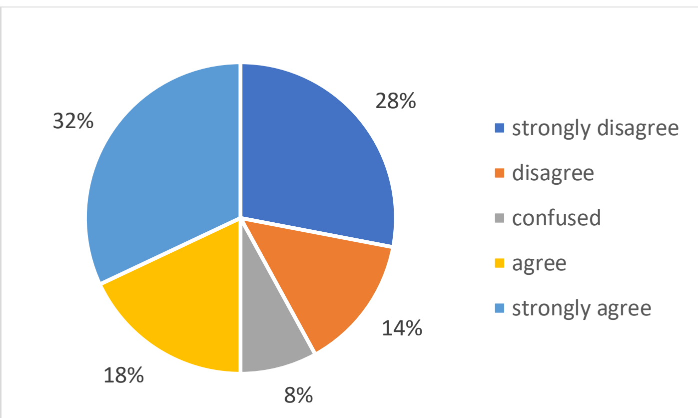

Fig.5.Graph

9. References

[1]Au, K.-F., Choi, T.-M., & Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. International Journal of Production Economics, 114 (2), 615–630.
[2]Wong, W., & Guo, Z. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. International Journal of Production Economics, 128 (2), 614–624.
[3]Xia, M., Zhang, Y., Weng, L., & Ye, X. (2012). Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. Knowledge-Based Systems, 36 , 253–259.
[4] Xia, M., & Wong, W. (2014). A seasonal discrete grey forecasting model for fashion retailing. Knowledge-Based Systems, 57 , 119–126.
[5] Choi, T.-M., Hui, C.-L., Ng, S.-F., & Yu, Y. (2012). Color trend forecasting of fashionable products with very few historical data. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42 (6), 1003–1010.
[6] DEGREE PROJECT IN COMPUTER SCIENCE AND ENGINEERING, SECOND CYCLE, 30 CREDITS STOCKHOLM, SWEDEN 2017
[7]Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a .review. ACM computing surveys (CSUR), 31 (3), 264–323.