International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.037 ISSN: 2581-4621

Empowering Sentiment Analysis For Improved Fashion Choices

Kavi Priya S[1], Porkodi J[2] , Kaviya Sri A N[3] , Shweatha K[4]

[1]*Dr .S.Kavi Priya ,Associate Professor, Computer Science and Engineering ,Mepco Schlenk Engineering College , Sivakasi , Taminadu*

[2]*Mrs. J. Porkodi, Head and Assistant Professor, Department of Computer Applications (UG), V.V.Vanniaperumal College for Women, Virudhunagar, Tamilnadu*

[3]*Ms.A.N.Kaviya Sri , Final Year Student , Computer Science and Engineering(UG), Mepco Schlenk Engineering College , Sivakasi , Tamilnadu*

[4]*Ms.K.Shweatha, Final Year Student , Computer Science and Engineering(UG), Mepco Schlenk Engineering College , Sivakasi , Tamilnadu*

Orcid ID : 0000-0003-0173-7955

ABSTRACT

The world of fashion is dynamic, ever-changing, and deeply influenced by individual preferences and collective sentiment. Making confident and informed fashion choices can be a daunting task, especially in an era marked by rapid trends and diverse styles. The fashion industry is ever-evolving, with trends and styles constantly changing. Making informed fashion choices that align with personal preferences and public sentiment can be challenging. "Empowering Sentiment Analysis for improved fashion choices" presents a novel approach to address this challenge. This abstract introduces a comprehensive framework that harnesses the power of sentiment analysis, a sophisticated natural language processing technique, to provide consumers and fashion enthusiasts with invaluable insights into the realm of fashion. Sentiment analysis, commonly used in understanding public opinions and emotional tones in textual data, is adapted here to decode the fashion landscape. By analyzing textual data from fashion reviews, social media posts, and comments, sentiment analysis can discern public sentiment and opinions about specific clothing items, styles, and trends. The primary objective of this initiative is to empower individuals to make better, more informed fashion choices. Through sentiment analysis, individuals can access an in-depth understanding of the prevailing sentiments and opinions surrounding specific clothing items, styles, and trends. This knowledge equips them with the tools to align their choices with current trends, explore niche styles, or even express their uniqueness confidently. By leveraging the power of machine learning models like Logistic Regression, Naïve Bayes, Support Vector Machines, Random Forest , Ada Boosting and Deep Learning algorithms the sentiment classification models are built. Furthermore, this technology fosters inclusivity and diversity in fashion decision-making by highlighting a wide range of sentiments and opinions. It acknowledges that fashion is a highly personal and subjective domain and helps individuals discover styles that resonate with their unique tastes and values.

Keywords— Fashion industry, Sentiment analysis, Natural language processing, Logistic Regression, Naïve Bayes, Support Vector Machines, Random Forest, Ada Boosting

1. Introduction

The fashion industry is an ever-evolving landscape, driven not only by artistic creativity but also by consumer sentiment and preference. Fashion choices are deeply personal, influenced by a myriad of factors, including individual style, cultural trends, and emotional resonance with clothing and accessories. In this era of rapid digital transformation, where fashion is increasingly shaped and consumed online, understanding and harnessing consumer sentiment has become paramount for both fashion brands and consumers themselves. The advent of social media platforms, e-commerce websites, and fashion forums has given rise to an unprecedented wealth of text data, wherein

consumers freely express their opinions, critiques, and emotions about fashion products and trends. These expressions hold invaluable insights into what resonates with consumers, what drives their purchasing decisions, and, ultimately, what empowers them to make informed fashion choices.
The potential applications of sentiment analysis in the realm of fashion are far-reaching. By understanding consumer sentiment, brands can tailor their product offerings, marketing strategies, and customer engagement tactics to align with the desires and emotions of their target audience. Consumers, on the other hand, can benefit from more personalized fashion recommendations that resonate with their unique styles and preferences. Machine learning is a rapidly developing technology that has the potential to revolutionize fashion choices. Model is trained in large datasets and thus machine learning can be used to develop highly precise fashion recommendations.
This research endeavors to explore the intersection of sentiment analysis and fashion choices, seeking to empower both consumers and fashion industry stakeholders. Through the lens of sentiment analysis, we aim to decipher the emotions, opinions, and trends that drive fashion choices. Our study seeks to shed light on the effectiveness of sentiment analysis models, their applications in enhancing fashion recommendations, and their potential to transform the fashion landscape into a more emotionally connected and data-driven ecosystem.

2. Experimental Methods or Methodology

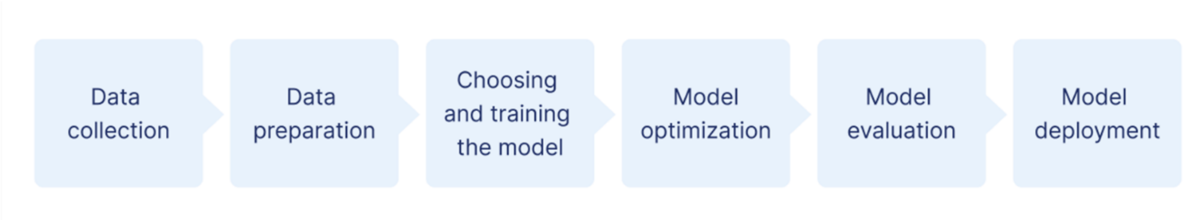

Fig 1. Machine Learning process flow

This article addresses the better fashion recommendation for the customers. Sentiment analysis has the potential to play a pivotal role in personalized fashion recommendations. By factoring in sentiment-related features, recommendation systems can provide consumers with products that align not only with their style but also with their emotional inclinations. We used Machine learning algorithms like Logistic regression , Support Vector Machine , Random Forest Classifier , Ada Boosting , Naïve Bayes and Deep learning algorithm like Recurrent Neural Network to build the model .
Logistic regression, a fundamental statistical and machine learning technique, is a supervised learning model used for binary or multi-class classification tasks. Unlike linear regression, which predicts continuous values, logistic regression models the probability that a given input belongs to one of the predefined classes, typically using the logistic sigmoid function. It's particularly valuable when the relationship between input features and the outcome variable is non-linear, and it provides interpretable results, as the coefficients can be analyzed to understand feature importance. Random forest is a type of ensemble learning algorithm that combines multiple decision trees to improve the accuracy of predictions. It is a popular machine learning algorithm that is used forclassification. Random forests work by building a large number of decision trees on different subsets of the training data. Each decision tree is trained to make predictions independently of the other trees. When a new data point is presented to the random forest, each tree makes a prediction. The final prediction is made by averaging the predictions of all the trees. Random forests are effective at reducing overfitting, which is a problem that can occur when machine learning models are trained on too much data. Overfitting occurs when the model learns the training data too well and is unable to generalize to new data. Random forests are often more accurate than other machine learning algorithms, especially for complex problems. This is because random forests are able to reduce overfitting by training a large number of trees on different subsets of the data. Hyperparameter tuning, also known as hyperparameter optimization or hyperparameter search, is the process of

selecting the best combination of hyperparameters for a machine learning model. Hyperparameter tuning is done by grid search. Involves defining a grid of possible hyperparameter values and exhaustively trying all possible combinations.

Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. It works by finding a hyperplane that best separates data points into distinct classes while maximizing the margin between them. SVM is effective in handling high-dimensional data and can work well even with limited training samples. It employs the concept of support vectors, which are the data points closest to the decision boundary, to create robust and generalized models. SVM can handle both linear and non-linear classification through the use of kernel functions like the radial basis function (RBF). Naive Bayes is a simple yet powerful probabilistic machine learning algorithm often used for classification and text categorization tasks. It's based on Bayes' theorem and the "naive" assumption of feature independence, which simplifies calculations but may not hold true in all cases. Naive Bayes is computationally efficient, particularly well-suited for high-dimensional datasets, and requires relatively small amounts of training data to make predictions.

AdaBoost (Adaptive Boosting) is an ensemble learning technique that combines the predictions of multiple weak learners to create a strong, robust model. It works by assigning different weights to training instances and iteratively training weak learners on these weighted instances, focusing on the ones that are misclassified in previous rounds. This sequential learning process allows AdaBoost to adapt and give more emphasis to the challenging data points, ultimately improving classification accuracy. AdaBoost is effective in various domains, particularly in binary classification problems, and can handle both categorical and numerical features. It's less prone to overfitting, simple to implement, and works well with a variety of base learners.

Recurrent Neural Networks (RNNs) are a class of neural network architectures designed for sequential data modeling. What sets RNNs apart is their ability to maintain hidden states, allowing them to capture temporal dependencies within sequences. RNNs process data step by step, making them suitable for tasks like time series forecasting, natural language processing, and speech recognition. However, traditional RNNs suffer from vanishing gradient problems that hinder learning long-range dependencies. More advanced variants, such as LSTMs and GRUs, mitigate these issues and are widely used.

Fig 1 explains initially the data is collected and exploratory data analysis is done.Feature Selection and data cleaning which includes handling missing values is done.Text mining which comprises of tokenization , noise removal and lexicon normalization is carried out.Word Cloud is generated.Then train test split is carried out to split the data into training and testing.Vectorization which includes count vectorization and TF – IDF vectorization is done.Then machine learning as well as deep learning models are built with various algorithms.

3. Results and Discussion

3.1 Dataset

The dataset consist of 11 variables out of which 4 are numerical variables and 7 are categorical variables. Total observations are 23486 and missing cells are 4697.The total size in memory is 2.0 MB.The dependent variables in the dataset are unnamed , Clothing Id , Age , Title, Review text , Rating , Recommended Id , Positive feedback count, Division Name , Department Name ,Class Name. The sentiment analysis is concerned only with the review text and recommended Id so the rest of the columns are dropped.

3.2 Performance metrics

Performance metrics are used to evaluate the performance of machine learning models. They are used to determine how well a model is able to make predictions on new data. The performance metrics used in this paper is accuracy , precision , recall and f1 score .

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.037 ISSN: 2581-4621

Accuracy measures the overall correctness of the model's predictions. Formula: (TP + TN) / (TP + TN + FP + FN). Precision measures the proportion of true positive predictions among all positive predictions. It is particularly useful when false positives are costly. TP / (TP + FP).
Recall measures the proportion of true positives among all actual positive instances. It is valuable when false negatives are costly. TP / (TP + FN) and The F1-Score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance. 2 * (Precision * Recall) / (Precision + Recall).

Fig. 2. Customer rating distribution

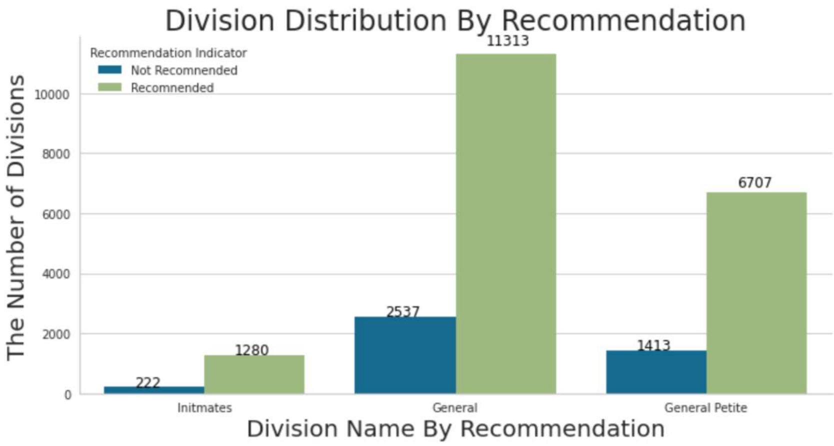

Fig. 3. Division Name distribution

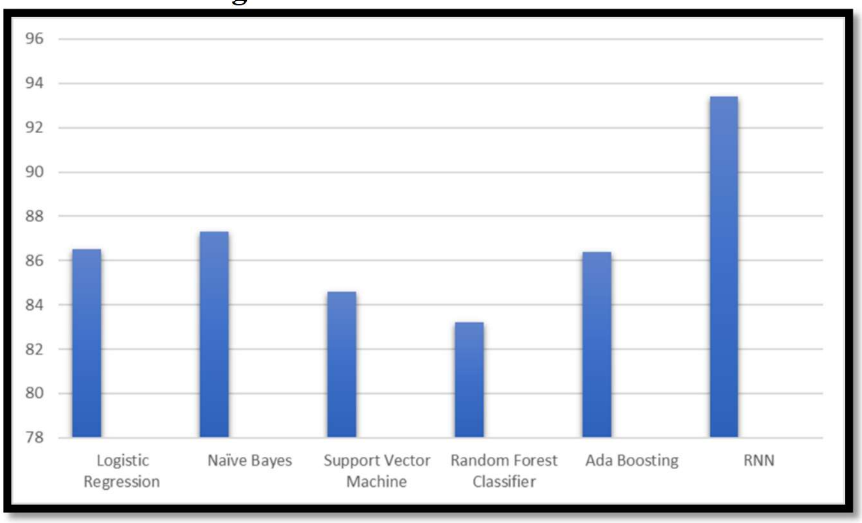

Fig. 4. Accuracy of the machine learning model

DOI:10.46647/ijetms.2023.v07i05.037 ISSN: 2581-4621

The Logistic Regression has the accuracy of 86.3% , Naïve Bayes has 87.8% , Support Vector Machine has 84.3% , Random Forest Classifier has the accuracy of 83.2% , Ada Boosting has the accuracy of 87.2% and RNN has the accuracy of 93.4%. The classification report is analysed and the performance metrics for each machine learning algorithm is compared and finally for this dataset RNN algorithm performs well compared to other machine learning algorithms.

CONCLUSION

In conclusion, the research on "Empowering Sentiment Analysis for Improved Fashion Choices" represents a pivotal step toward bridging the gap between the world of fashion and the realm of data-driven decision-making. Through the exploration of sentiment analysis, this study has shed light on the intricate relationship between consumer emotions, preferences, and fashion choices. By harnessing the power of sentiment analysis, we have uncovered the potential to revolutionize the fashion industry, benefiting both consumers and fashion stakeholders. Our research has demonstrated the effectiveness of sentiment analysis models in deciphering the emotions and opinions embedded within fashion-related data. These models have shown the capability to accurately categorize sentiments, allowing for a deeper understanding of consumer behaviour and preferences.

Furthermore, the application of sentiment analysis extends beyond mere insights. It carries substantial implications for the future of fashion, paving the way for more personalized, emotionally resonant fashion recommendations and experiences. Brands and retailers can leverage sentiment analysis to craft tailored marketing campaigns, optimize product offerings, and engage with consumers on a more profound level. However, we acknowledge that sentiment analysis is not without its challenges. The complexity of human emotions, the nuances of language, and the evolving nature of fashion trends pose inherent difficulties in analysis. Ongoing research and development efforts should focus on refining sentiment analysis models to address these complexities.

References

1. Federico Becattini, Lavinia De Divitiis, P.Y. Mok "Fashion recommendation based on style and social events" Multimedia Tools and Applications (2023)

2. L. Javier Cabeza-Ramírez , Sandra M. Sánchez-Cañizares," Exploring the effectiveness of fashion recommendations made by social media influencers in the centennial generation, Textile Research Journal, Volume 93, Issue 13-14,February 16, 2023.

3. Maryam Mahdikhani," Exploring commonly used terms from online reviews in the fashion field to predict review helpfulness",International Journal of Information Management Data Insights, Volume 3, Issue 1, April 2023.

4. Hung-Jen Chen, Hong-Han Shuai," A Survey of Artificial Intelligence in Fashion", IEEE Signal Processing Magazine ,Volume: 40, Issue: 3, May 2023.

5. Harmanjit Singh, Somnath Chakrabarti," How do gratifications to read reviews and perceived reviewers' credibility impact behavioural intentions in fashion e-commerce? A mediating-moderating perspective." Computers in Human Behavior, Volume 143, June 2023.

6. Pierfrancesco Bellini, Luciano Alessandro Ipsaro Palesi, Paolo Nesi," Multi Clustering Recommendation System for Fashion Retail", Sentient Multimedia Systems and Universal Visual Languages,January 13 2022.

7. Artur M. Pereira, J. Antão B. Moura, Evandro De B. Costa," Customer models for artificial intelligence-based decision support in fashion online retail supply chains", Decision Support Systems, Volume 158, July 2022

8. A Di Vaio, R Hassan, G D'Amore, R Tiscini," Responsible innovation and ethical corporate behavior in the Asian fashion industry: A systematic literature review and avenues ahead",Asia Pacific Journal of Management ,Springer 2022.

9. Eunsoo Baek, Shelley Haines, Omar H. Fares,” Defining digital fashion: Reshaping the field via a systematic review”,Computers in Human Behavior , Volume 137, December 2022.

10. N Liu, J Lin, S Guo, Fashion platform operations in the sharing economy with digital technologies: recent development and real case studies, Information- Transparent Supply Chains, Annals of Operations Research ,Springer(2022).

11. S Sysko-Romańczuk, P Zaborek, A Wróblewska,” Data modalities, consumer attributes and recommendation performance in the fashion industry”,Electronic Markets,Springer 2022.

12. Weiqian Li, Bugao Xu,” Aspect-Based Fashion Recommendation With Attention Mechanism”,IEEE Access , Volume 8 , 2020.

13. Wei Zhou, P.Y. Mok,” Fashion recommendations through cross-media information retrieval”,Journal of Visual Communication and Image Representation , Volume 61, May 2019.

14. Cairong Yan, Yizhou Chen” Differentiated Fashion Recommendation Using Knowledge Graph and Data Augmentation”,Volume 7,23 July 2019.

15. Ana Paula Faria , Joana Cunha,” Fashion communication in the digital age: findings from interviews with industry professionals and design recommendations”, Procedia CIRP, Volume 84, 2019.