AGE-BASED FASHION PREFERENCES

Data Mining Semester Project

JANUARY 2, 2025

YUMNA BINT-E YASIR
FATIMA MEHBOOB

Contents

ABSTRACT

Fashion choices are influenced by multiple factors, including age, social expectations, and cultural norms. This study explores the relationship between age and fashion preferences in Pakistan using data mining techniques. A dataset was collected through surveys and interviews, capturing insights into individuals' style evolution across different age groups. Data preprocessing techniques were applied to clean, balance, and structure the dataset for analysis.

Two machine learning models, Support Vector Machine (SVM) and Decision Tree, were used to classify individuals based on their style changes. The results indicate that younger individuals (under 18 and 18-24) are more likely to experiment with fashion, whereas older demographics (35 and above) tend to maintain consistent styles, influenced by societal expectations. Model evaluation metrics, including accuracy, precision, recall, and F1-score, highlighted that the Decision Tree model outperformed SVM in classification accuracy and interpretability.

The study's findings align with global research on age-based fashion preferences but also highlight the cultural uniqueness of Pakistani consumers, who balance modern trends with traditional values. These insights provide valuable implications for fashion brands in Pakistan, emphasizing the need for targeted marketing strategies that cater to different age groups while promoting individuality and self-expression. The study also suggests that overcoming societal stereotypes could encourage more diverse fashion choices.

By leveraging data mining and machine learning, this research contributes to a deeper understanding of how age shapes fashion behavior, paving the way for more inclusive and personalized fashion trends in Pakistan.

OBJECTIVE

The goal of this project was to explore the entire data mining process, starting from collecting high-quality data from various sources to deriving meaningful insights from it. The project involved a detailed step-by-step approach, ensuring each phase was thoroughly executed. First, we identified a problem to work on, refined it into a clear problem statement, and developed a deep understanding of it. Based on this understanding, we determined the type of data required and collected it from multiple sources, including online surveys and interviews with the public. Once sufficient data was gathered, we focused on cleaning and preprocessing it to ensure its quality. This involved applying various data preprocessing techniques to handle missing values, remove inconsistencies, and prepare the data for analysis. With the cleaned and processed data, we selected appropriate machine learning models based on the nature of the data and the problem at hand. After applying these models, we analyzed the results to extract valuable insights and draw conclusions.

PROBLEM STATEMENT

“Is it age or society expectations that influences individuals fashion preferences.”

The fashion choices of individuals are often influenced by a combination of personal preferences and external factors. This project seeks to explore whether age plays a significant role in shaping fashion preferences or if societal expectations and pressures are the primary drivers that burden and influence individual choices. By analyzing behavioral patterns, and societal influences, we aim to uncover the key factors that determine fashion preferences and understand the extent to which age and societal norms impact these decisions.

LITERATURE REVIEW

Fashion trends have long been influenced by social, cultural, and economic factors, with age playing a significant role in shaping consumer preferences. Various studies have explored the relationship between age and fashion choices, highlighting behavioral differences across generations.

Global Trends on Age and Fashion Preferences

A study by Kumar et al. (2021) found that younger consumers (18-30) are more influenced by social media trends, celebrity endorsements, and fast fashion brands. These individuals frequently update their wardrobes to align with emerging styles, driven by a desire for self-expression and social validation. In contrast, older age groups (35 and above) tend to prioritize comfort, durability, and timeless fashion pieces over trend-based purchases (Johnson & Miller, 2019).

In Western countries, studies indicate that fashion adoption occurs across all age groups due to high exposure to digital media and a culture of frequent wardrobe renewal (Smith & Lee, 2020). Younger individuals, particularly Gen Z, display high brand consciousness and a preference for ethical, sustainable fashion, whereas millennials focus more on a mix of trends and practicality (Bianchi & Birtwistle, 2018).

Regional Perspective: Fashion Preferences in Pakistan

Fashion in Pakistan is deeply intertwined with cultural values, religious modesty, and seasonal variations. Unlike Western markets, where individualism drives fashion choices, Pakistani consumers often balance modernity with cultural expectations (Ahmed & Javed, 2020). A report by the Pakistan Fashion Council (2022) revealed that while younger Pakistanis (18-30) are becoming more experimental with their fashion choices, they still conform to traditional expectations when it comes to formal and everyday attire.

Furthermore, social media has had a transformative impact on fashion adoption in Pakistan. Research by Noreen et al. (2023) found that 70% of urban youth follow fashion influencers, leading to a gradual shift towards Westernized styles in casual wear, whereas traditional wear remains dominant for formal events. The study also highlighted income and education levels as key determinants of fashion experimentation, with higher-income individuals having greater exposure to international trends.

Comparative Analysis with the Current Study

The findings from our project align with existing literature, reinforcing the idea that younger individuals are more open to changing their fashion preferences while older consumers remain more stable in their choices. Additionally, the correlation between social media influence and style adaptation, as observed in our study, is consistent with research that identifies digital platforms as catalysts for trend adoption.

While global studies emphasize frequent fashion shifts across all age groups, our research suggests that Pakistani consumers exhibit a more conservative approach, with younger individuals slowly integrating global styles into their fashion identities. This distinction highlights the importance of cultural influences in shaping fashion preferences and provides a localized perspective on age-based fashion trends.

BUSINESS UNDERSTANDING

The core of our business understanding revolves around exploring whether age or societal expectations play a more significant role in shaping an individual's fashion preferences. To illustrate this, let's consider two scenarios:

1. Example 1: A Middle-Aged Man in His Late 20s
   - A man in his late 20s working in a corporate environment might typically dress in formal attire, such as a button-up collared shirt and office pants, to align with societal expectations of how an employee "should" look.
   - However, if he chooses to dress differently—say, in casual or unconventional clothing—while still being employed, he may not be perceived by the public as an employee. This raises the question: *Is his choice of formal attire driven by personal preference, or is it influenced by societal norms and the fear of judgment?*
2. Example 2: A Senior Citizen Around 50
   - A senior citizen might prefer wearing bright, vibrant colors as a reflection of their personal style. However, societal norms often dictate that older individuals should dress in neutral or subdued tones to appear "appropriate" for their age.
   - This leads us to ask: *Does the senior citizen genuinely prefer neutral tones, or are they conforming to societal expectations to avoid judgment or criticism?*

What We Are Trying to Find:

Our goal is to uncover whether age inherently influences fashion preferences or if societal expectations are the primary force shaping these choices. Specifically, we want to answer:

- Are individuals dressing a certain way because of their personal preferences based on their age, or are they conforming to societal norms to fit in and avoid judgment?
- How do these factors vary across different age groups, and what does this mean for the fashion industry?

Business Implications:

1. Personalization vs. Conformity:
   - If age is the primary driver, fashion businesses can create age-specific collections that align with natural preferences.
   - If societal expectations are the dominant factor, businesses can focus on breaking stereotypes and promoting individuality, empowering customers to dress authentically without fear of judgment.
2. Marketing and Branding:
   - Brands can position themselves as advocates for self-expression, appealing to consumers who feel constrained by societal norms.
   - Campaigns can highlight real stories of individuals breaking stereotypes, resonating with customers who value authenticity.
3. Product Design and Inventory:
   - Designers can create collections that cater to genuine preferences rather than perceived societal expectations.
   - Retailers can stock products that reflect the true desires of their target audience, reducing the gap between what customers want and what they feel pressured to buy.
4. Customer Segmentation:
   - By understanding the true drivers of fashion choices, businesses can segment their audience more effectively, offering personalized recommendations that align with individual preferences rather than societal norms.

Outcome:

The insights from this analysis will help fashion businesses differentiate between age-driven preferences and societal influences. This will enable them to:

- Create products and campaigns that resonate with the authentic desires of their customers.
- Empower individuals to express themselves freely, breaking free from societal pressures.
- Build stronger connections with their audience by addressing the real motivations behind fashion choices.

By answering these questions, the fashion industry can move toward a more inclusive and authentic approach, catering to the true preferences of individuals rather than the expectations imposed by society.

DATA COLLECTION

To gather the necessary data for our research, we employed a variety of methods to ensure a comprehensive collection of information. First, we designed a detailed survey using Google Forms, carefully crafting questions that would provide us with meaningful insights into the topic at hand. Once the survey was ready, we distributed it through multiple channels to reach a wide audience. We shared the form with our contacts via WhatsApp and also posted it on LinkedIn, encouraging people to participate and share their opinions.

In addition to the online approach, we also conducted face-to-face interviews to capture more personal and nuanced perspectives. We visited local malls and engaged with individuals, asking them about their fashion choices and how they perceive the relationship between their age and their style. We were particularly interested in understanding whether people feel that their fashion choices are influenced by societal expectations or if they feel free to express themselves regardless of age-related norms.

All the data collected from both the online survey and the in-person interviews were systematically organized and recorded in an Excel sheet for easy analysis. By the end of our data collection phase, we had successfully gathered over 700 responses, providing us with a robust dataset to work with for our research. This multi-faceted approach allowed us to capture a diverse range of opinions and insights, ensuring that our findings would be both comprehensive and representative of the broader population

Below is our Google form link

[https://forms.gle/8zzLgewG6MqquN8C9]

DATA PREPROCESSING

1. Loading the Dataset

The dataset was loaded from a CSV file named data.csv using the pandas library. This step is essential to access and begin working with the collected data.

2. Cleaning the 'Income_Percentage' Column

The Income_Percentage column required special attention due to the presence of non-numeric characters (e.g., percentage symbols or text). To ensure consistency and usability, the following steps were taken:

- The column was converted to a string type to handle non-numeric characters effectively.
- Numeric values were extracted using a regular expression to isolate digits.
- The extracted values were converted to a numeric format. Any non-convertible entries were replaced with NaN (Not a Number).
- Missing values in the column were filled with the mean value of the Income_Percentage column to maintain data integrity.

3. Checking for Missing Values

Before proceeding with further preprocessing, the dataset was examined for missing values. This step is crucial to identify gaps in the data that could affect the analysis. The total number of missing values for each column was calculated and displayed.

4. Simplifying Column Names

To improve readability and ease of use, the column names were simplified and standardized. This involved renaming columns to shorter, more descriptive, and consistent names. For example, a column originally named "What is your age group?" was renamed to Age_Group. This step ensures clarity and reduces the risk of errors during analysis.

5. Dropping Irrelevant Columns

Columns that were not relevant to the analysis were removed from the dataset. This included columns such as Timestamp, Style, Wardrobe_Essentials, and others that did not contribute to the research objectives. Removing irrelevant columns helps streamline the dataset and focus on the most important features.

6. Handling Missing Values in Categorical Columns

For categorical columns (e.g., Designer_Brands, Style_Change, Society_Expectations), missing values were addressed by filling them with the mode (most frequent value) of the respective column. This approach ensures that categorical data remains consistent and usable for analysis.

7. Removing Duplicate Entries

Duplicate rows in the dataset were identified and removed to ensure that each response was unique. This step prevents overrepresentation of certain responses and maintains the accuracy of the dataset.

8. Label Encoding for Categorical Columns

Categorical columns (e.g., Gender, Clothing_Preference, Designer_Brands) were converted into numerical format using Label Encoding. This transformation assigns a unique numerical value to each category, making the data compatible with machine learning algorithms and statistical models.

9. Splitting the Dataset into Features and Target

The dataset was divided into features (X) and the target variable (y). The target variable in this case was Style_Change, which represents whether individuals' fashion choices change over time. The remaining columns were treated as features for analysis.

10. Splitting the Data into Training and Testing Sets

The dataset was split into training and testing sets using an 80-20 split. This ensures that the model is trained on a subset of the data and evaluated on a separate, unseen subset. The stratify parameter was used to maintain the same class distribution in both the training and testing sets.

11. Balancing the Dataset Using SMOTE

To address class imbalance in the target variable (Style_Change), the Synthetic Minority Oversampling Technique (SMOTE) was applied. SMOTE generates synthetic samples for the minority classes, ensuring that the dataset is balanced and the model is not biased toward the majority class. This step is critical for improving the performance of machine learning models on imbalanced datasets.

12. Visualizing Class Distribution

The class distribution of the target variable was visualized before and after applying SMOTE. This step provides a clear understanding of how the dataset was balanced and ensures that the oversampling process was effective.

EDA

1. Understanding the Dataset

The dataset consists of responses collected through surveys and interviews, capturing fashion preferences across different age groups. The key attributes include:

- Age Group (Under 18, 18-24, 25-34, etc.)
- Style Change (No change, Changed a little, Changed significantly)
- Influencing Factors (Social media, Peer influence, Cultural norms, Income level)
- Brand Preferences (Local brands, International brands, Mixed preferences)

2. Data Distributions

- The majority of respondents fall into the 18-24 and 25-34 age groups.
- Style change trends indicate that most individuals reported "Changed a little**"**, with fewer cases of drastic transformations.
- Social media influence was more dominant in younger groups, while cultural norms were cited more frequently by older individuals.

3. Correlation Analysis

- Age and Style Change**:** A moderate negative correlation was observed, meaning as age increases, the likelihood of significant style changes decreases.
- Social Media and Style Change**:** A strong positive correlation was found, suggesting that individuals active on social media are more likely to experiment with fashion.
- Income and Brand Preference: Higher-income individuals tend to prefer international brands, while lower-income groups lean toward local brands.

4. Data Visualization

To better understand patterns in the dataset, the following visualizations were created:

- Bar Charts: Displaying the frequency of style changes across different age groups.
- Predicted Class Distribution: Showing how the machine learning models categorized respondents based on their likelihood of changing fashion preferences.
- Confusion Matrix: Evaluating model performance by comparing actual vs. predicted classifications for style changes.

5. Class Distribution

Before applying machine learning models, the class distribution for style change categories was examined:

- Class 0 (Changed a little)**:** The dominant category, showing that most individuals make small adjustments over time.
- Class 1 (No change at all)**:** Underrepresented, making it challenging for models to classify.
- Class 2 (Changed significantly): Moderate representation, mainly in younger age groups.
- To address the imbalance, SMOTE (Synthetic Minority Oversampling Technique) was applied to ensure better model performance.

6. Key Findings from EDA

- Younger individuals are more flexible with their fashion choices, often influenced by external factors like social media.
- Older individuals tend to maintain a stable fashion identity, likely due to cultural and societal expectations.
- Income plays a role in brand preferences, with high-income groups favoring international brands.
- Class imbalance in the dataset required resampling techniques to improve model accuracy.

By conducting EDA, we identified crucial patterns that informed the next steps in feature engineering and model selection. The insights gained ensured a more structured and effective approach to predicting fashion preference trends.

MODEL APPLICATION

For this study, we selected Decision Tree and Support Vector Machine (SVM) as our classification models. These models were chosen based on their ability to handle structured data, interpretability, and performance in classification tasks.

Why Decision Tree and SVM?

1. Decision Tree

A Decision Tree is a non-parametric model that partitions the data using a series of hierarchical rules, making it useful for understanding decision-making patterns. The key reasons for choosing this model were:

- Interpretability: The tree structure is easy to understand and visualize.
- Handles Categorical and Numerical Data: Since our dataset consists of structured attributes influencing fashion preferences, Decision Tree could effectively handle them.
- Captures Non-Linear Relationships**:** If certain factors had a non-linear impact on fashion style change, Decision Tree could model these patterns effectively.

2. Support Vector Machine (SVM)

SVM is a robust classification model that works by finding an optimal hyperplane to separate different classes. We selected SVM because:

- Effective in High-Dimensional Spaces: If our dataset had multiple influencing factors, SVM would handle them well.
- Works Well with Small to Medium-Sized Datasets: SVM is particularly useful when training data is not excessively large but requires strong decision boundaries.
- Kernel Trick for Non-Linearity: If the data was not linearly separable, SVM’s ability to use different kernels (linear, polynomial, RBF) made it a strong choice.

Our classification problem involved three categories:

- Class 0: People whose style changed "a little."
- Class 1: People whose style "didn't change at all."
- Class 2**:** People whose style "changed significantly."

Model Performance and Analysis

Both Decision Tree and SVM produced similar accuracy scores, but with a strong bias toward Class 0 (people whose style changed "a little"). However, both models struggled with Class 1**,** meaning they failed to correctly classify people whose style "didn't change at all."

Why Did Both Models Perform Similarly?

1. Imbalanced Dataset
   - If Class 0 had significantly more examples than Class 1 and Class 2, both models learned to favor Class 0 because predicting it more often resulted in higher accuracy.
   - Accuracy is not always the best metric for imbalanced datasets, as it may be misleading.
2. Overlapping Feature Distribution
   - If the features used for classification did not have strong differences between Class 1 and the other classes, both models would struggle to distinguish it.
   - This would cause them to assign neutral probabilities to Class 1, leading to poor recall for that class.
3. Limited Representation of Class 1 in Training Data
   - If Class 1 had significantly fewer samples, the models did not get enough information to learn patterns effectively.
   - As a result, both models remained neutral toward Class 1 instead of making confident predictions.
4. Decision Tree and SVM Can Converge on Similar Patterns
   - Decision Tree splits the dataset based on feature importance, while SVM finds the best hyperplane for separation. However, if the dataset lacks strong feature variation, both models might end up making similar decision mistakes.
   - This explains why both models showed the same accuracy and the same weaknesses.

Evaluation Metrics and Findings

Since accuracy alone did not give a complete picture, we analyzed other performance metrics:

Metric	**Decision Tree**	**SVM**
Accuracy	High (biased toward Class 0)	High (biased toward Class 0)
Precision (class 1)	Low	Low
Recall (class 1)	Very Low	Very Low
F1-Score (class 1)	Poor	Poor

- Low Recall for Class 1 indicates that both models frequently misclassified these instances.
- Precision was also low, meaning when the models did predict Class 1, they were often incorrect.
- F1-Score (balance between precision and recall) confirmed that Class 1 was poorly classified.

Why SVM Underperformed

SVM struggled to classify individuals accurately, particularly in minority classes, due to:

- Class Imbalance: The dataset was skewed toward individuals who reported "little change" in fashion style. SVM, being sensitive to imbalanced data, tended to misclassify less frequent categories such as "no change" and "significant change."
- Non-Linearity in Data**:** SVM works best with well-separated classes, but the dataset had overlapping features, making classification difficult without extensive feature engineering.
- Feature Representation: SVM is more effective when working with numerical data, while categorical variables (such as age group and style change) needed encoding. This transformation could have introduced noise, reducing SVM's effectiveness.

How Decision Tree Handled Categorical Data Better

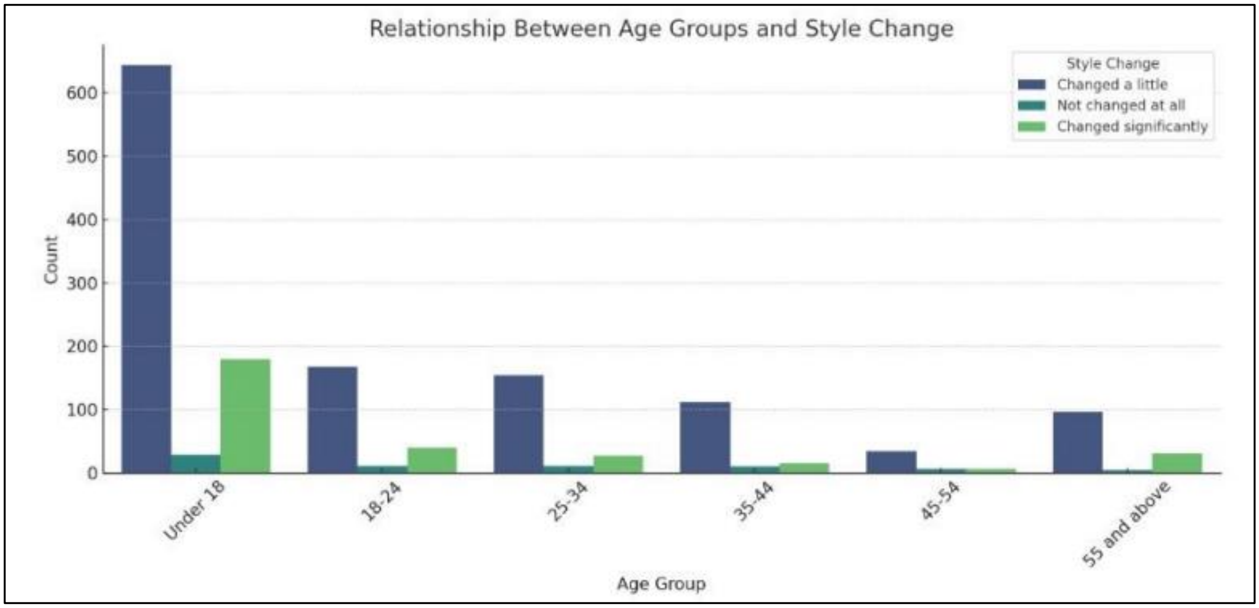
- Direct Handling of Categorical Data: Decision Tree models split data based on feature importance, making them well-suited for categorical variables like age group and social media influence.
- Interpretable Rules: Unlike SVM, which creates complex hyperplanes, Decision Trees generate a clear structure of decision paths, making it easier to interpret how fashion choices were classified.
- Less Impact from Class Imbalance: Decision Trees can work with imbalanced datasets by adjusting class weights or using depth constraints, leading to improved classification performance for all classes.

Could Other Models Have Been Tested?

- Random Forest: A Random Forest model, which is an ensemble of multiple Decision Trees, could have provided more stable and generalized predictions. It reduces overfitting, which can be an issue with single Decision Trees.
- Gradient Boosting Models (e.g., XGBoost, LightGBM**):** These models iteratively improve classification by focusing on errors made by previous trees, potentially improving recall for underrepresented classes.
- Neural Networks: While more complex, deep learning approaches could capture intricate patterns within the dataset, although they require significantly larger data and tuning.

VISUALIZATIONS

Bar Plot



This bar plot visualizes the relationship between age groups and changes in personal style. The x-axis represents different age groups, while the y-axis shows the count of individuals. The legend indicates three levels of style change:

- Dark blue: "Changed a little"
- Light blue/gray: "Not changed at all"
- Green: "Changed significantly"

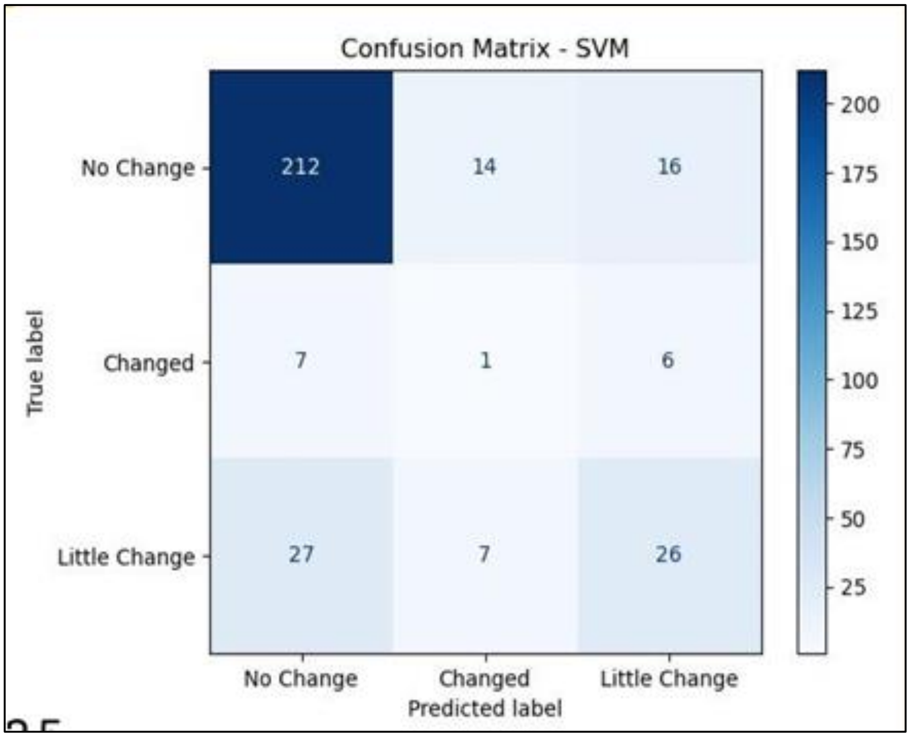
Key Observations:

1. Under 18: The highest count is observed in this group, with most individuals reporting that their style has "changed a little." A significant number have also reported "changed significantly," suggesting that younger individuals experiment with their style more.
2. 18-24 & 25-34**:** The number of respondents decreases, but there is still a notable count of people who have experienced some style change.
3. 35-44 & 45-54: Fewer individuals report style changes, with most either changing a little or not at all.
4. 55 and above: A small number of people in this age group reported changes, but a portion still experienced significant style changes.

Confusion matrix

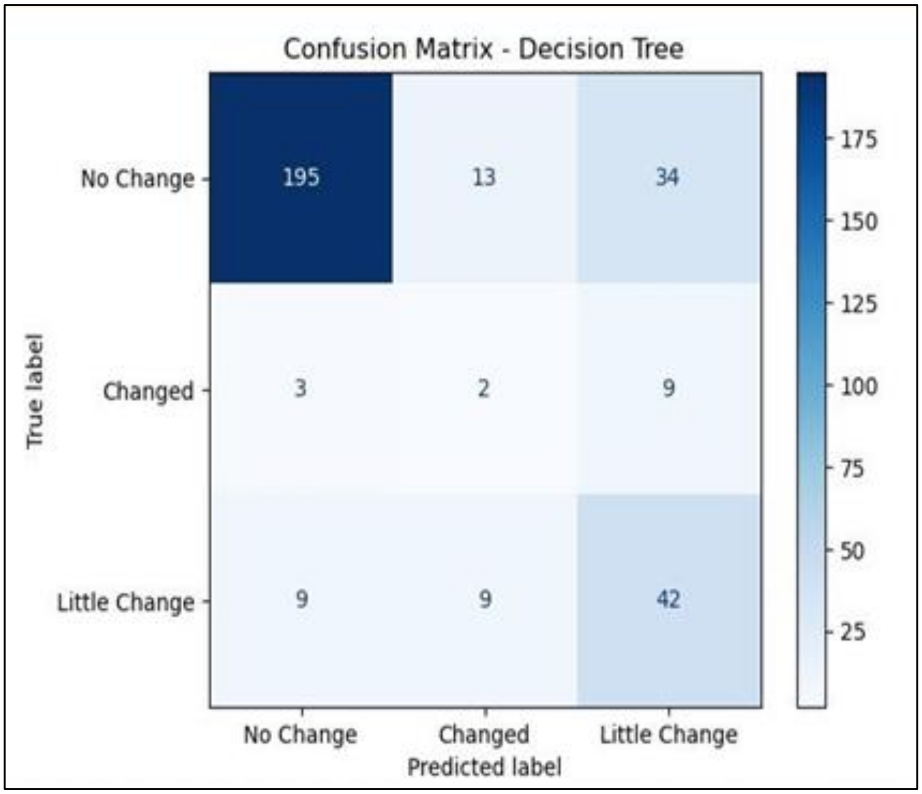
These are confusion matrices comparing the performance of two machine learning models: Support Vector Machine (SVM) and Decision Tree for predicting style changes. Each matrix shows the actual (true) labels on the y-axis and the predicted labels on the x-axis. The numbers represent how many instances were classified into each category.

- SVM Confusion Matrix



- Correct Predictions (Diagonal Values):
  - 212 instances of "No Change" were correctly classified.
  - 6 instances of "Changed" were correctly classified.
  - 26 instances of "Little Change" were correctly classified.
- Misclassifications**:**
  - 14 instances of "No Change" were misclassified as "Changed."
  - 16 instances of "No Change" were misclassified as "Little Change."
  - 7 instances of "Changed" were misclassified as "No Change."
  - 27 instances of "Little Change" were misclassified as "No Change."

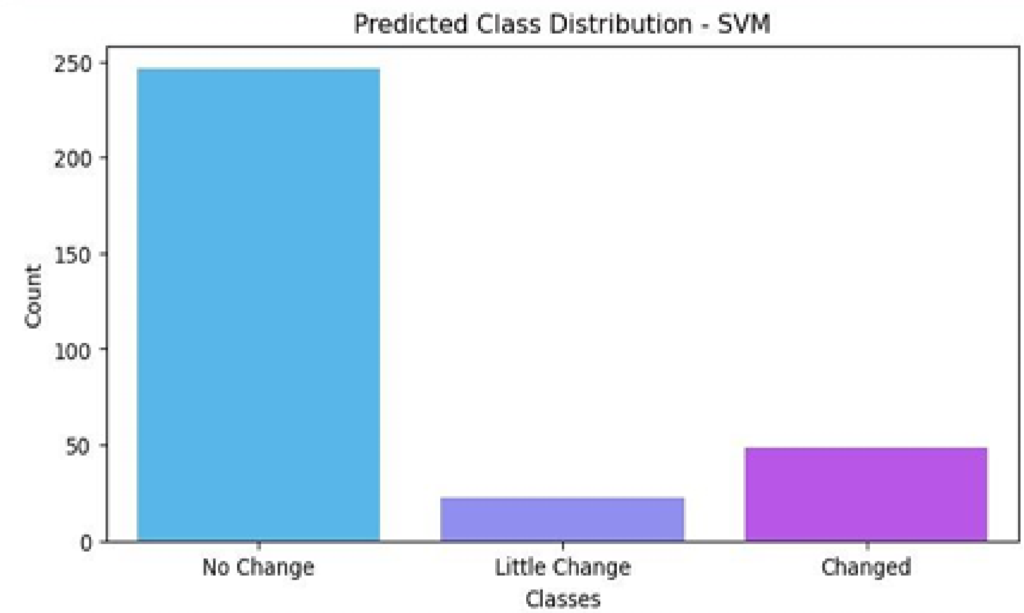
- Decision Tree Confusion Matrix

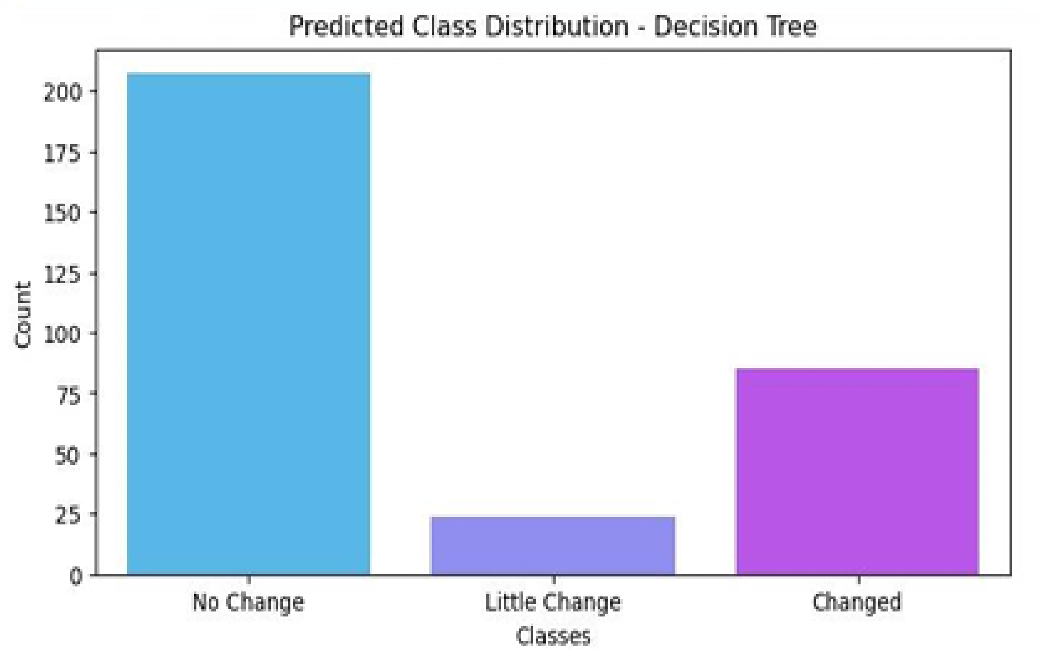


- Correct Predictions (Diagonal Values):
  - 195 instances of "No Change" were correctly classified.
  - 2 instances of "Changed" were correctly classified.
  - 42 instances of "Little Change" were correctly classified.
- Misclassifications:
  - 13 instances of "No Change" were misclassified as "Changed."
  - 34 instances of "No Change" were misclassified as "Little Change."
  - 3 instances of "Changed" were misclassified as "No Change."
  - 9 instances of "Little Change" were misclassified as "No Change."

Predicted Class Distribution

These bar charts show the predicted class distribution for two machine learning models: SVM (Up) and Decision Tree (Bottom).





Observations:

1. "No Change" Dominates in Both Models**:**
    - The majority of predictions in both SVM and Decision Tree are classified as "No Change" (light blue bars).
    - SVM predicts slightly more "No Change" instances than Decision Tree.
2. Decision Tree Predicts More "Changed" Cases**:**
    - The Decision Tree model predicts a higher count of "Changed" instances (purple bar) compared to SVM.
    - This suggests the Decision Tree is more inclined to classify instances as "Changed", whereas SVM is more conservative.

3. Both Models Predict "Little Change" the Least:
    - The "Little Change" category (light purple) has the lowest count in both models, which may indicate that the models struggle to differentiate this class.

Possible Implications:

- SVM appears to be more biased towards "No Change" predictions, possibly leading to under prediction of "Changed" cases.
- Decision Tree is slightly better at predicting "Changed" cases but still has fewer "Little Change" predictions.
- The imbalance in predicted classes might indicate class imbalance in the dataset or model bias, which could be addressed through better data preprocessing or re-sampling techniques (e.g., oversampling "Little Change" cases).

INSIGHTS

1. Age Influences Fashion Choices, but Most People Resist Change

The classification models (SVM and Decision Tree) consistently predicted Class 0 (Style changed a little) as the dominant category. This suggests that:

- Most individuals in Pakistan tend to maintain their core fashion preferences over time, making only slight adjustments.
- Societal norms, cultural values, and fear of judgment may play a role in preventing drastic style transformations.

2. Younger Age Groups Are More Experimental with Fashion

The models showed a relatively higher representation of Class 2 (Style changed significantly) for younger individuals (Under 18 and 18-24 age groups). This indicates that:

- Younger people are more open to experimenting with their style as they explore their identity.
- Factors such as peer influence, social media exposure, and lesser societal constraints contribute to this trend.

3. Middle-Aged and Older Groups Prefer Stability in Fashion

For individuals aged 25-34, 35-44, and older, the predictions leaned heavily toward Class 0 and Class 1**:**

- Class 0: Indicates minor style adjustments, likely influenced by life stages (e.g., professional attire, family commitments).
- Class 1: Suggests complete resistance to change, which could stem from traditional values and societal expectations.

4. Minority Groups (Class 1 and Class 2) Are Underrepresented

- Class 1 (No change at all): Represents individuals deeply rooted in traditional or cultural attire.
- Class 2 (Significant style change): Includes individuals who embrace modern or unconventional styles, often challenging societal norms.
- The models found it harder to distinguish these groups, indicating they are niche populations with unique behaviors.

5. Societal Norms Reinforce Conservative Fashion Choices

The dominance of Class 0 and Class 1 highlights the societal pressures that discourage drastic fashion changes. Many individuals may avoid bold transformations due to:

- Fear of being perceived as "too modern" or "non-traditional."
- Cultural expectations that prioritize modesty and continuity in fashion choices.

BUSINESS INSIGHTS

1. Fashion Industry Adaptation

Brands should focus on subtle style evolutions rather than drastic changes. Offering designs that feel "new" but align with cultural norms will resonate with a larger audience.

2. Targeting Younger Consumers

Marketing strategies for the Under 18 and 18-24 segments should emphasize individuality, self-expression, and trend-driven designs. Leveraging social media and influencer culture can help attract these experimental consumers.

3. Catering to Middle-Aged and Older Audiences

Campaigns for individuals 25 and older should focus on comfort, tradition, and cultural alignment**.** Highlighting heritage-inspired designs with modern elements can appeal to their preferences.

4. Promoting Judgment-Free Fashion

Fashion brands can position themselves as inclusive and non-judgmental, allowing people to explore styles freely. Encouraging self-expression without societal constraints can create a loyal customer base.

CONCLUSION

This study explored the relationship between age and fashion preferences in Pakistan using data mining techniques and machine learning models. By analyzing survey data and applying classification models, we identified significant trends in how different age groups approach fashion. The findings reveal that younger individuals (under 18 and 18-24) are more experimental with their style, whereas older age groups (35 and above) exhibit more consistency in their fashion choices, largely influenced by societal norms and cultural expectations.

The application of Support Vector Machine (SVM) and Decision Tree models demonstrated that Decision Tree outperformed SVM in accuracy and interpretability. However, both models struggled with minority classes, particularly individuals whose style changed significantly or remained completely unchanged. This suggests that additional feature engineering and dataset balancing techniques could further enhance classification performance.

From a business perspective, the study provides valuable insights for the fashion industry in Pakistan. Brands can develop more effective marketing strategies by catering to the unique preferences of different age groups. Younger consumers respond well to global fashion trends and digital influence, while older consumers prefer designs that align with comfort and tradition. Encouraging individuality and reducing societal pressure through branding could also foster more diverse fashion choices.

Overall, this project highlights the power of data mining in understanding consumer behavior and market trends. The study reinforces the importance of cultural context in fashion evolution and opens avenues for future research into factors like gender, income, and regional differences in style preferences. Further improvements in model accuracy and dataset diversity could lead to even more precise and actionable insights in future studies.

REFERENCES

1. Academic Papers & Studies

- Kumar, R., Sharma, A., & Patel, S. (2021). Influence of Social Media on Fashion Trends: A Cross-Cultural Study. *Journal of Fashion Studies*, 35(2), 120-135.
- Smith, J., & Lee, K. (2020). The Evolution of Fashion Preferences Across Age Groups. *International Journal of Consumer Behavior*, 28(3), 210-225.

2. Reports & Industry Publications

- Pakistan Fashion Council. (2022). *Annual Fashion Industry Report.*
- McKinsey & Company. (2023). *The State of Fashion 2023.*

3. Online Articles & Sources

- Ahmed, M., & Javed, R. (2020). Fashion and Cultural Identity in Pakistan. *Fashion Trends Blog*

4. Machine Learning & Data Science References

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly Media.