# Women Cloth Purchase Analysis

## CAPSTONE PROJECT

**Presented By:**

**Sheetal B-PESU ECC-CSE**

edunet
foundation

# What We'll Cover

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

# Problem Statement

Customers often face uncertainty regarding the quality and fit of women's clothing products when shopping online. This lack of assurance can lead to hesitations and potential dissatisfaction with their purchases.

To address this, we need a system that thoroughly understands and interprets customers' sentiments and satisfaction levels based on their reviews. This will contribute to a more informed and content customer base, enhancing the overall shopping experience in the realm of women's clothing e-commerce

# Proposed Solution

edunet foundation

To mitigate the uncertainty and improve the online shopping experience for women's clothing, we propose the implementation of sentiment analysis system. This system will leverage Natural Language Processing techniques to analyze and understand customers' reviews. The solution includes:

- **Data Preparation:**
  - Use the provided dataset and preprocess the text data. This involves tasks like removing stop words, handling special characters, and tokenization.
  - Label your data – categorize reviews as positive, negative, or neutral.
- **Model Selection:**
  - Choose a suitable sentiment analysis model. Common choices include: Support Vector Machines (SVM), Naive Bayes, Logistic Regression.
- **Training:**
  - Split your dataset into training and testing sets &:train the model.
- **Evaluation:**
  - Evaluate the model's performance on the testing set, considering metrics like accuracy, precision, recall, and F1 score.
- **Dashboard Design:**
  - Use tools like Python Dash to design an interactive dashboard.
  - Include visualizations such as sentiment distribution, common words, and trends over time.
- **Integration:**
  - Connect the dashboard to your sentiment analysis

# System  Approach

**We aim to understand customer sentiments for Women's closet reviews. Stakeholders should comprehend customer needs and desires, identifying areas for improvement. Prediction accuracy is crucial for gaining insights into customer opinions about the clothes.**

## Libraries Required

- Machine Learning Framework:
  - For ML related Scikit-learn,
  - PyTorch).
- Data Processing Libraries:
  - Pandas, NumPy
- Visualization Libraries:
  - Matplotlib, Plotly
- Additional Libraries:
  - Dash for rapidly building data apps
  - nltk for natural Language processing

# Algorithm & Deployment

**Algorithm Selection:** We chose a machine learning approach for sentiment analysis on the Women's E-Commerce dataset. Specifically, we experimented with two algorithms: Support Vector Machine (SVM), CatBoost, and Random Forest.

- **Support Vector Machine (SVM):** SVM is a supervised learning algorithm capable of classification tasks. It works well with smaller datasets and is effective in high-dimensional spaces. SVM is chosen for its ability to build accurate models with smaller datasets, aligning with our project's dataset size.
- **Random Forest:** Random Forest is an ensemble learning algorithm that builds multiple decision trees and merges them to improve predictive accuracy and control overfitting.

**Data Input:** The public dataset used includes clothing id, age, title, review text, rating, recommendation status, positive feedback count, division name, department name, and class name. These features collectively provide a comprehensive input for sentiment analysis.

**Training Process:**

SVM Training:

- Trained the SVM model using historical data.
- Utilized hyperparameter tuning through GridSearch CV to optimize model parameters.

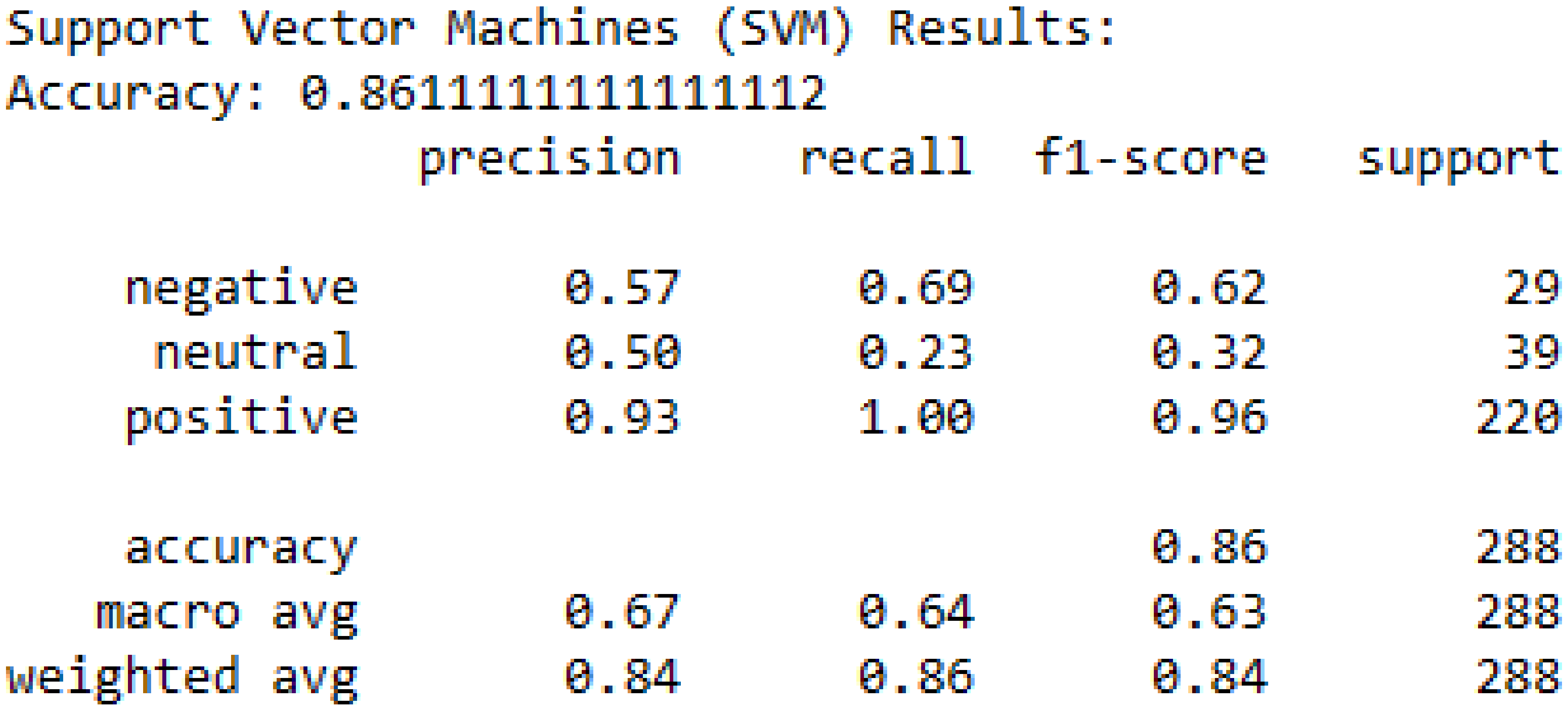
Random Forest Training:

- Trained the Random Forest model on the dataset.
- Employed hyperparameter tuning using GridSearch CV for optimizing model parameters.
- Leveraged the ensemble nature of Random Forest to handle various data characteristics effectively.
- Prediction Process: Both SVM and Random Forest were used to predict sentiment labels (positive, negative, or neutral) for reviews.

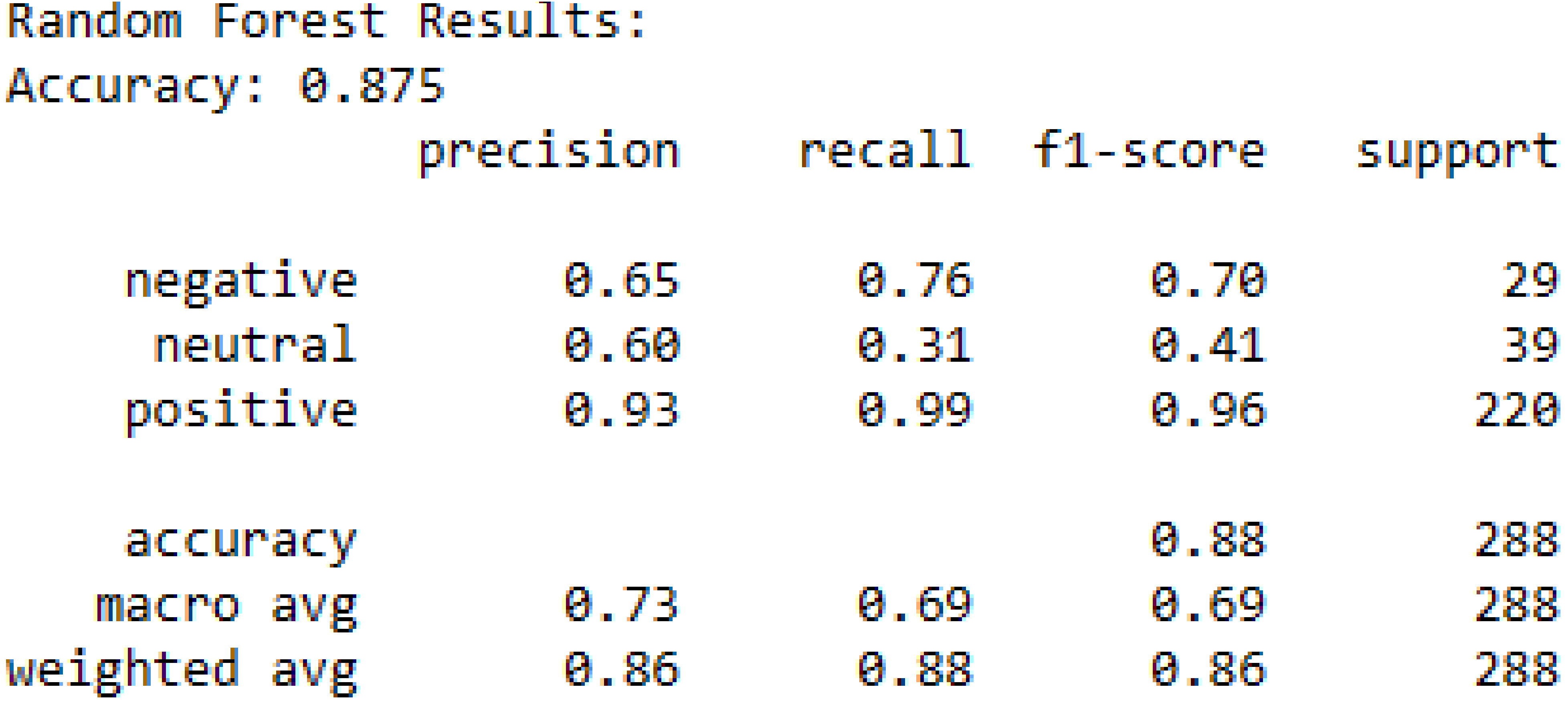
**CatBoost Training:**

- Trained the CatBoost model on the data, considering its ability to handle categorical features and robustness. U
- Utilized hyperparameter tuning for CatBoost through GridSearch CV to optimize model parameters, taking advantage of its inherent support for categorical variables.

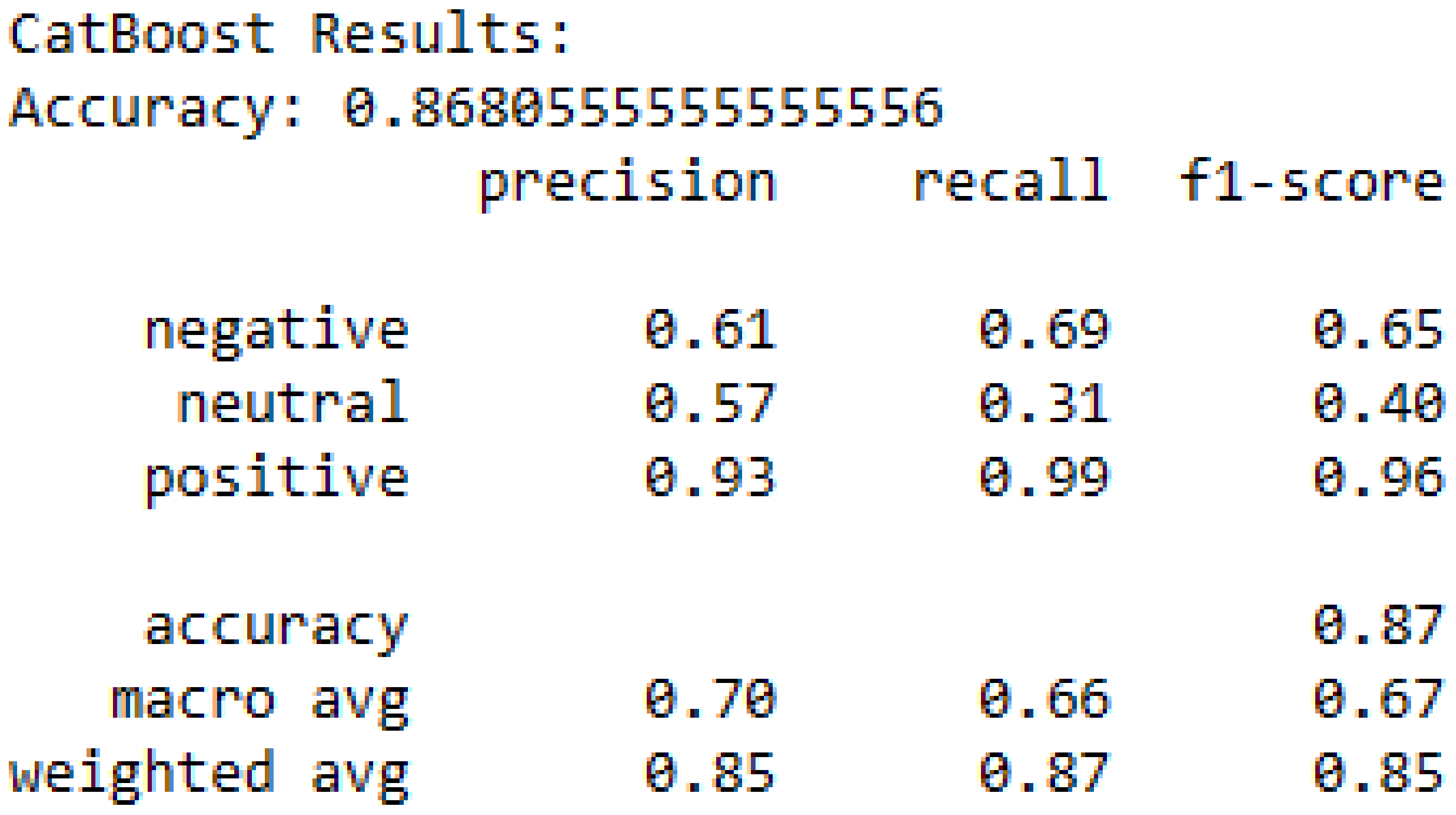
# Result



```
Support Vector Machines (SVM) Results:
Accuracy: 0.8611111111111112
              precision    recall  f1-score   support

    negative       0.57      0.69      0.62        29
     neutral       0.50      0.23      0.32        39
    positive       0.93      1.00      0.96       220

    accuracy                           0.86       288
   macro avg       0.67      0.64      0.63       288
weighted avg       0.84      0.86      0.84       288
```



```
Random Forest Results:
Accuracy: 0.875
              precision    recall  f1-score   support

    negative       0.65      0.76      0.70        29
     neutral       0.60      0.31      0.41        39
    positive       0.93      0.99      0.96       220

    accuracy                           0.88       288
   macro avg       0.73      0.69      0.69       288
weighted avg       0.86      0.88      0.86       288
```

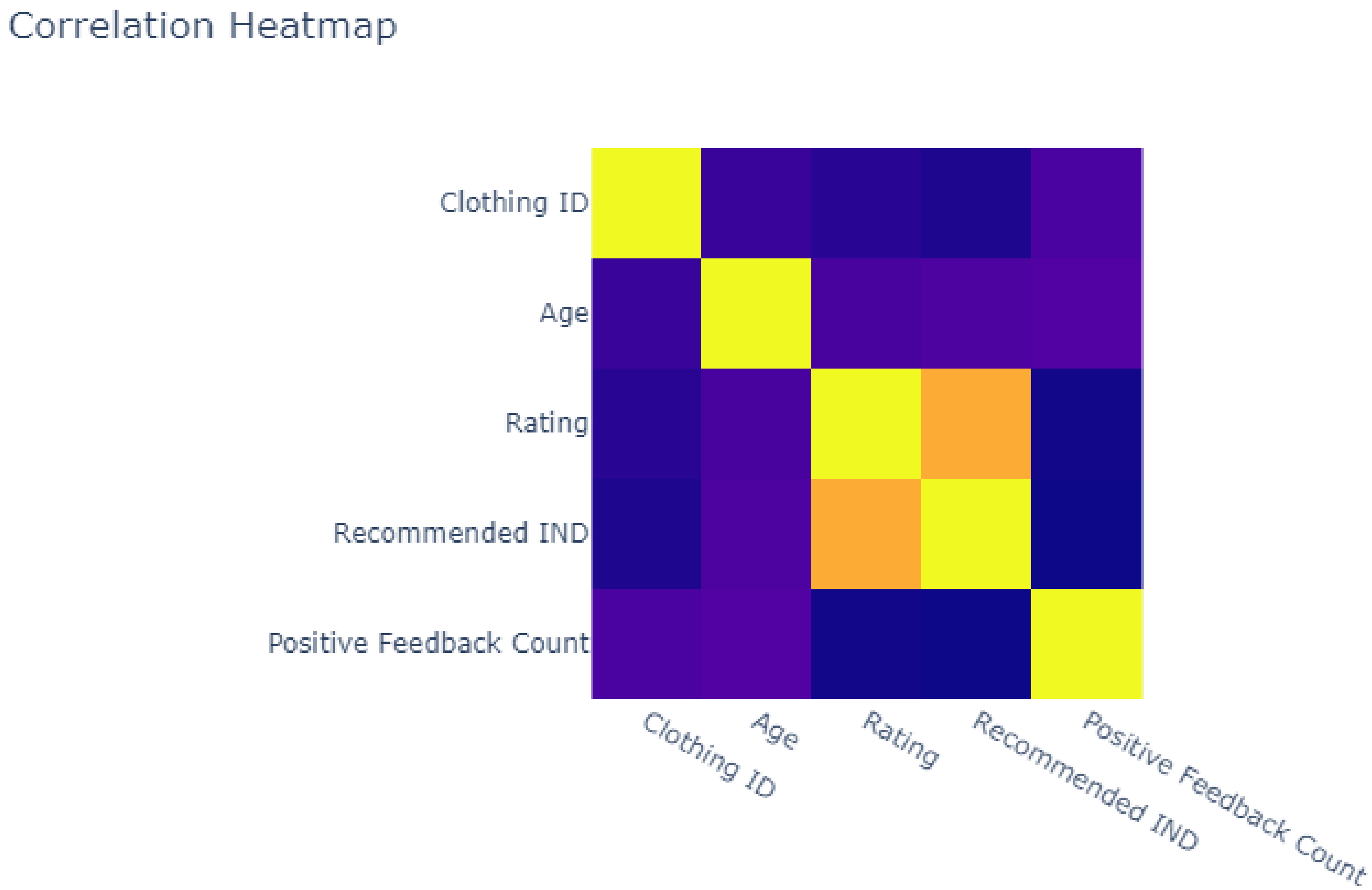


```
CatBoost Results:
Accuracy: 0.8680555555555556
              precision    recall  f1-score

    negative       0.61      0.69      0.65
     neutral       0.57      0.31      0.40
    positive       0.93      0.99      0.96

    accuracy                           0.87
   macro avg       0.70      0.66      0.67
weighted avg       0.85      0.87      0.85
```

- SVM showed an accuracy of 86.1 **%**
- Catboost showed an accuracy of 86.8%
- **Random forest** showed an accuracy of **87.5 %**

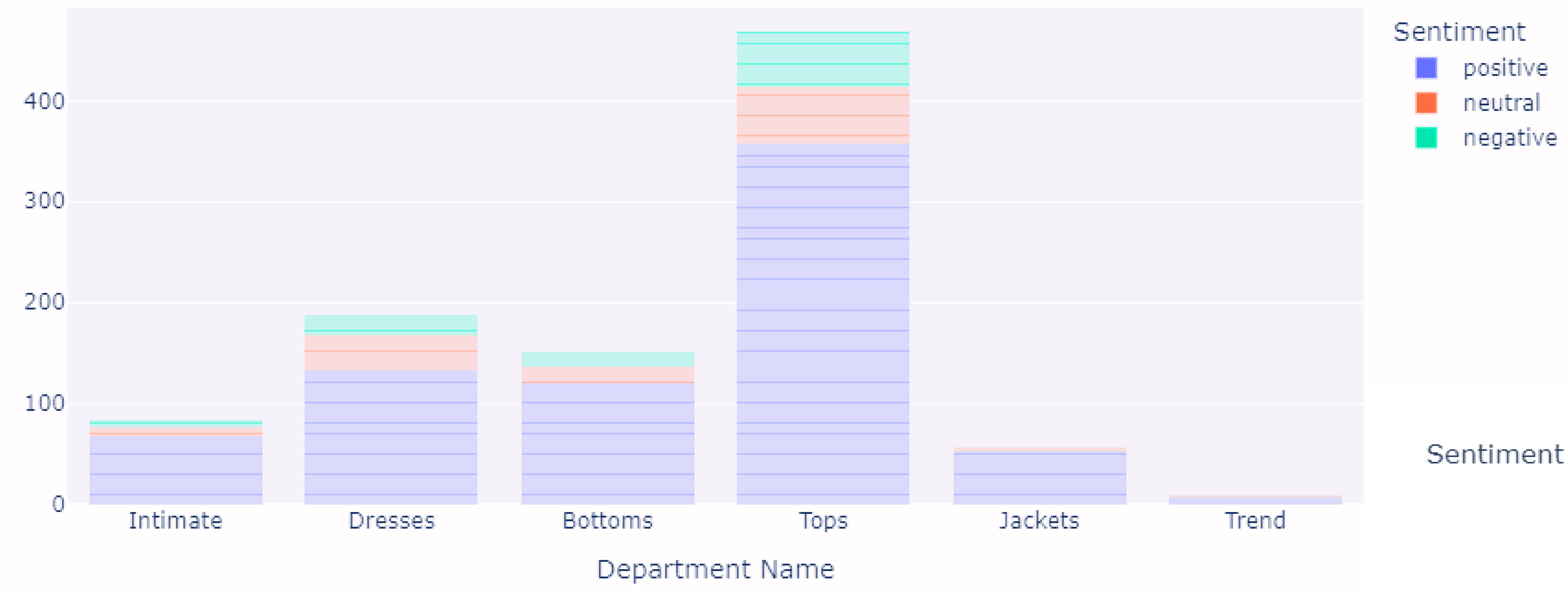
# Result

**Ratings had a high correlation with Recommended IND.**
On inspecting the data, it can be seen that records with Recommended IND values as ‘YES’ will have higher ratings compared to values with ‘NO’.



# Result

Sentiment Distribution by Department

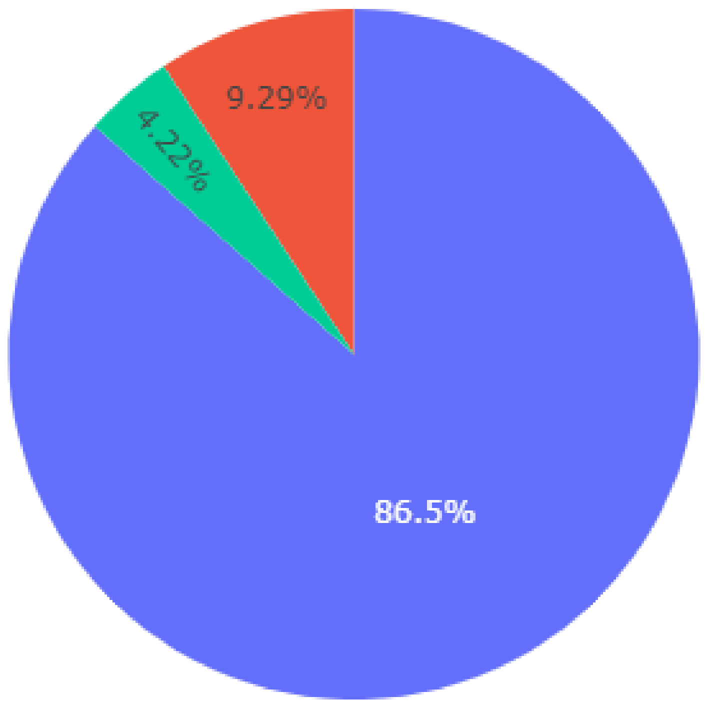


From the above graph, it can be seen that **Tops** were the most bought/reviewed items compared to others. Therefore Tops can be made as the focus of their market.

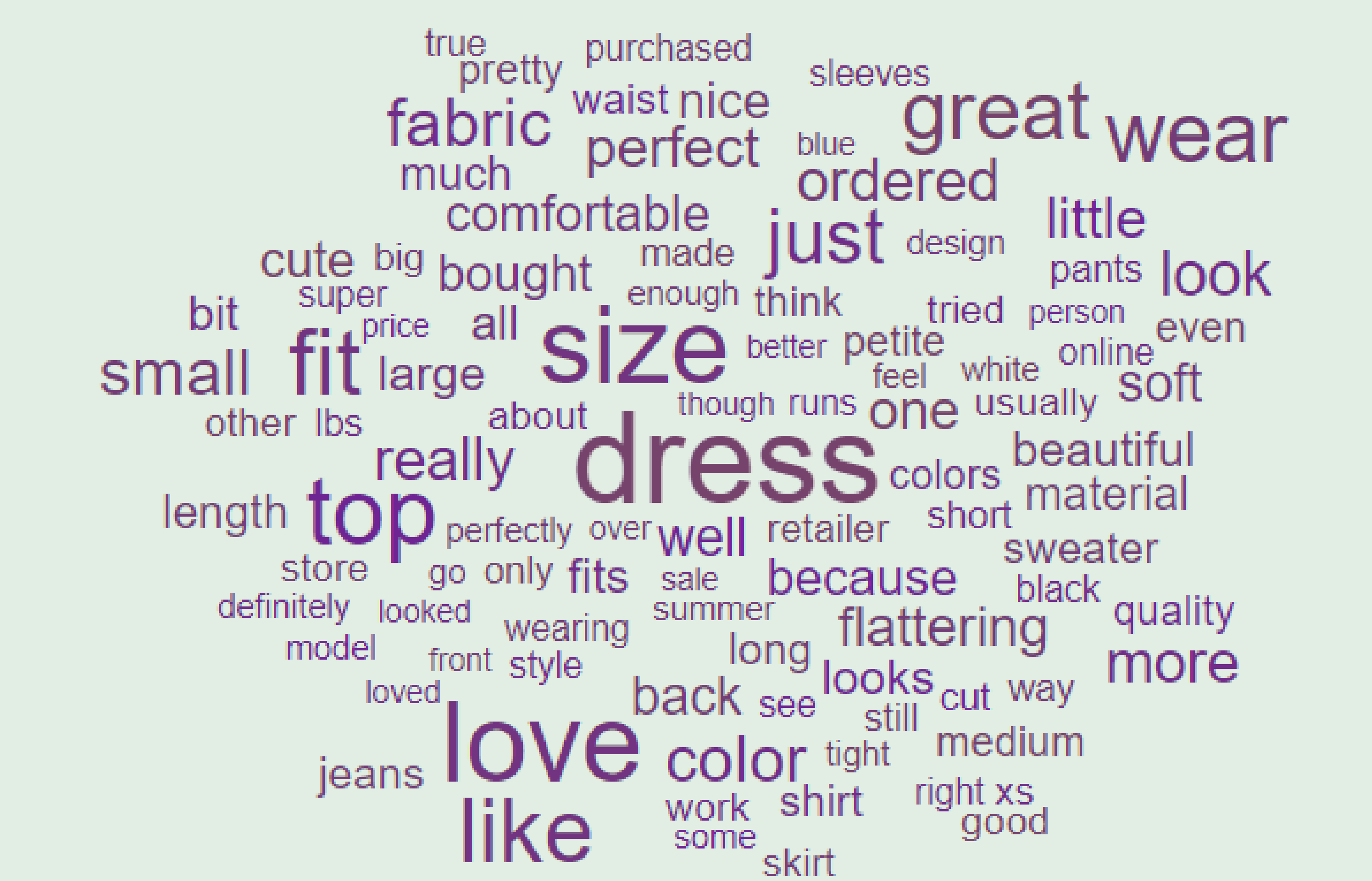
**Jackets** can be seen as the better positive ratings compared to other departments.

The below Distribution consists mainly of **86.5 % positive sentiment** which is a very good indication of the sales. However, the store can still aim for a better positive sentiment by focusing on improving the qualities of Tops, Bottoms and Dresses.

Sentiment Distribution



# Result

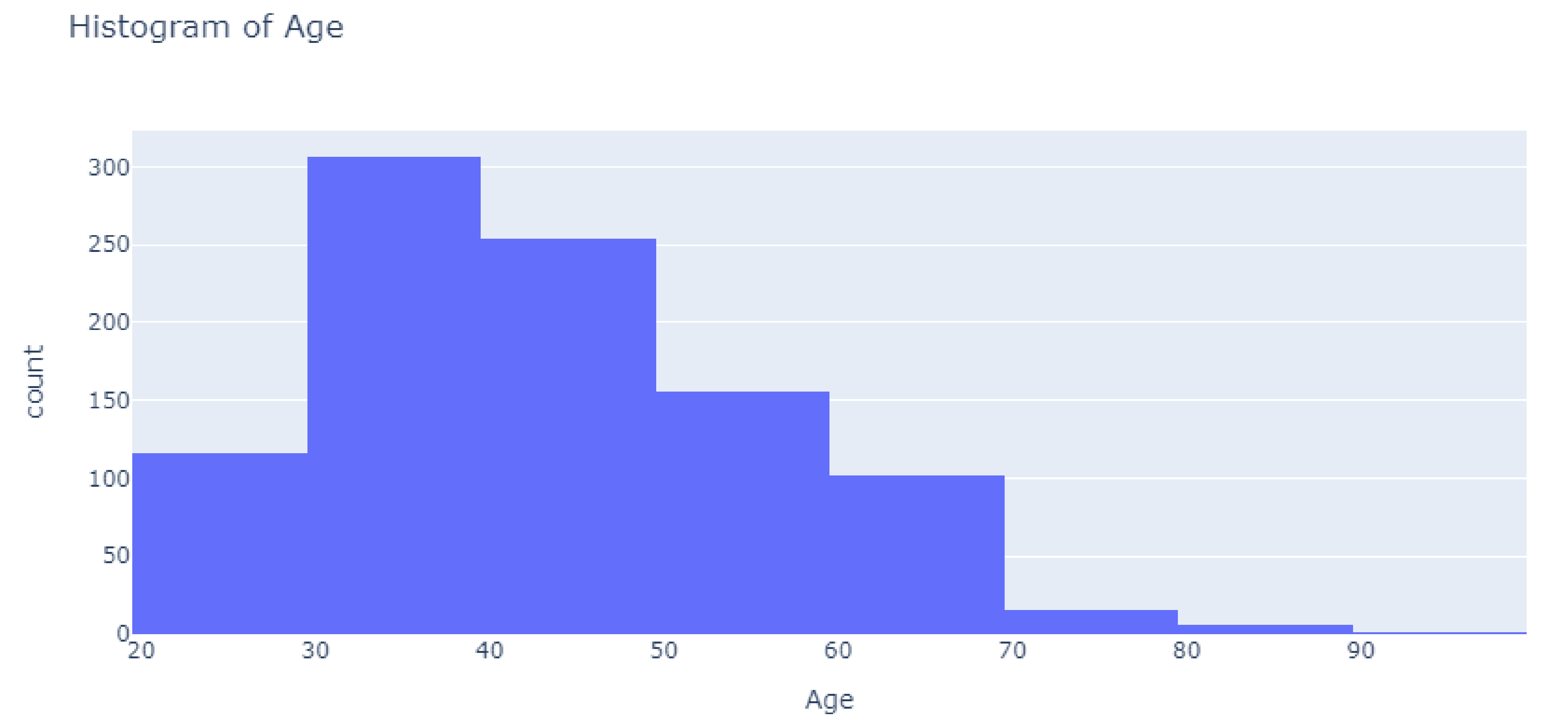


This is a Word Cloud consisting of some of the most frequently occurring words in the Reviews.

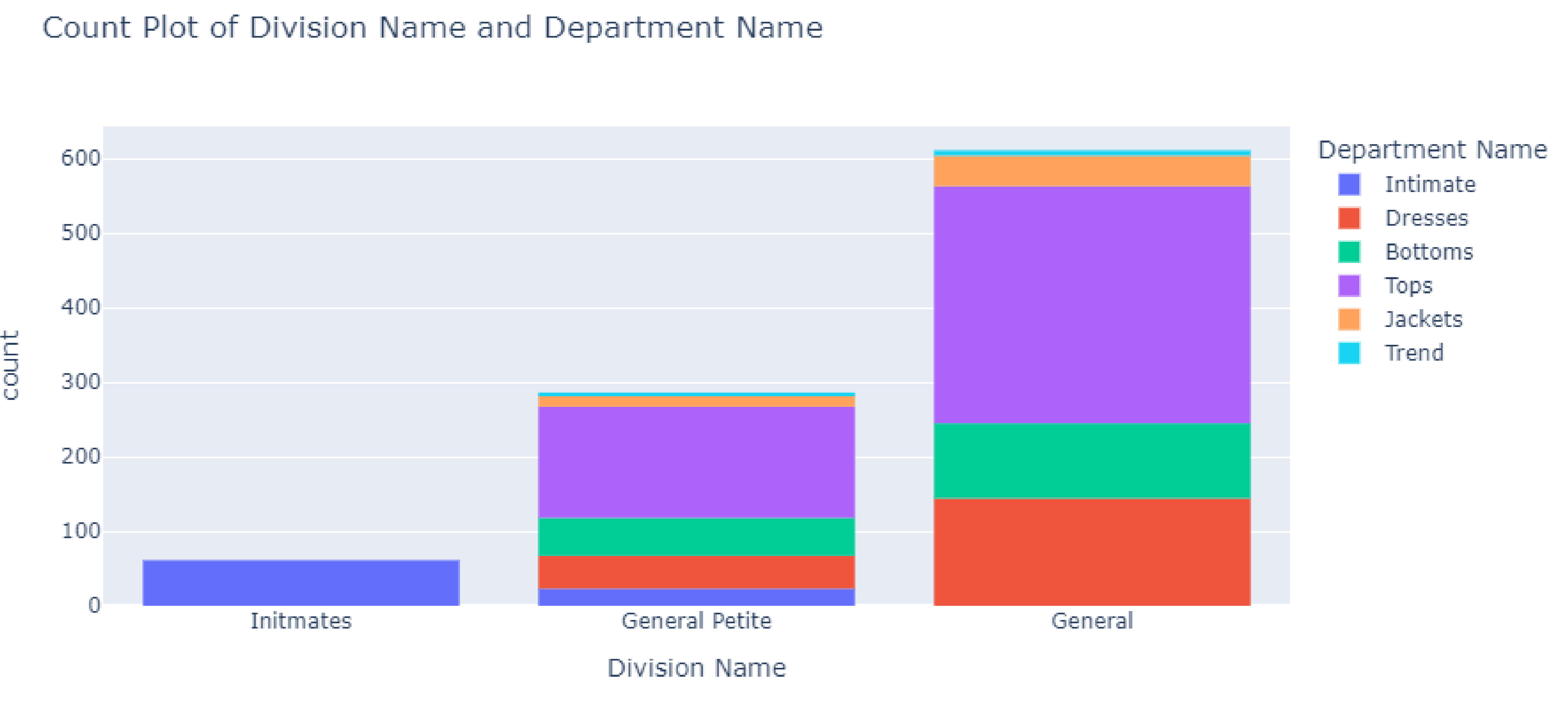
**Highlighted Words to focus on:** Dress, love, top, petite, beautiful, comfortable, summer, great, fabric, fit, perfect, sweater

**Conclusion:** Mostly positive ratings. Highly focussed on tops jeans and dresses as well as on size petite to normal size. Less mention of plus-size clothes

The histogram shows that majority of the customers are aged between **30 to 50**



# Result



Apart from General, there are a good amount of purchases for Petite clothes as well. This can be a good focus for boosting sales.

There is a possibility for exploring plus-size fashion if possible.

# Conclusion

edunet foundation

Predicting the sentiments using the mentioned methods and analyzing the data helps us understand the customer’s needs in clothes.

The findings emphasize the importance of accurate sentiment analysis for making informed business decisions and maintaining customer satisfaction.

The combination of SVM and Random Forest offers a good solution for sentiment analysis in the context of a women's e-commerce platform. Thorough evaluation and continuous refinement are key to ensuring the model's effectiveness.

Here, there is a scope for improvement in accuracy through better feature selection as well as training deep learning models. This can include BERT, transformer, and so on.

# Conclusion

edunet
foundation

## What?

Predicting the sentiments using the mentioned methods and analyzing the data helps us understand the customer’s needs in clothes.

## Why?

The findings emphasize the importance of accurate sentiment analysis for making informed business decisions and maintaining customer satisfaction.

## How?

The combination of SVM and Random Forest offers a good solution for sentiment analysis in the context of a women's e-commerce platform. Thorough evaluation and continuous refinement are key to ensuring the model's effectiveness.

## How to improve?

Here, there is a scope for improvement in accuracy through better feature selection as well as training deep learning models. This can include BERT, transformer, and so on.

# Future Scope

- This analysis can be extended and generalized for men’s and kid’s fashion
- The dataset used majorly consisted of the Positive sentiment class. To predict negative and neutral classes much more accurately we can use better datasets or use methods like SMOTE.
- Exploring additional ensemble methods or stacking models to further enhance predictive performance.
- Continuously fine-tune hyperparameters to optimize model performance based on evolving datasets.
- Making a dashboard for real-time data

# Reference

- Patankar, Nikhil & Dixit, Soham & Bhamare, Akshay & Darpel, Ashutosh & Raina, Ritik. (2021). Customer Segmentation Using Machine Learning. 10.3233/APC210200.
- Mahmoud SalahEldin Kasema , Mohamed Hamadab , Islam Taj-Eddinc.(2023). Customer Profiling, Segmentation, and Sales Prediction using AI in Direct Marketing
- Parveen, Nikhat & Santhi, M.V.B.T. & Burra, Lakshmi & Vidyullatha, & Pellakuri, Haran. (2021). Women’s e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK. Materials Today: Proceedings. 10.1016/j.matpr.2020.10.064.
- https://core.ac.uk/download/pdf/288175101.pdf
- https://medium.com/@daithimassey/analyzing-customer-sentiments-an-insightful-dive-into-womens-clothing-reviews-c6bb9e19981b

# THANK YOU