Sentiment Analysis & Product Recommendation For Women's Clothing Business(E-Commerce)

SULAKKHANA HERATH

GRADUATE DATA SCIENTIST

Bio

Education

- Master of Information Technology
- Postgraduate Diploma in Computer Science
- Bachelor Degree in Computer Science and Mathematics

Projects

- Air Quality Analysis for Australia
- Employee Turnover Prediction and Analysis
- Sentiment Analysis for Airline using Twitter Data

Agenda

Define

- Why Sentiment?
- Industry and Stakeholders
- Problem Definition
- Business Question
- Data Question and Dataset

Design

- EDA and Visualization

Deliver

- Feature Engineering
- Model Evaluation
- Deployment

Summary

Next Steps

Define

Why Sentiment Analysis?

Sentiment Analysis is a process of identifying opinions and provide a quantifiable results using AI and Natural Language Processing techniques to process raw data. Studies have shown that customers are likely to spend 140% more after having a positive experience with brand. (https://www.revechat.com)

Value of Sentiment Analysis for Business:

- Useful tool for businesses to understand their customers and maximize customer satisfaction.
- To maintain brand reputation.
- Optimize marketing campaigns
- Fine tuning of new product launches.
- Minimize customer churn over and attract new customers.
- Minimize abandon purchases.
- Minimize stock return rate.
- Increase revenue.

Industry and Stakeholders

Targeted Industry is Fashion retailers. Fashion retailers are discovering new ways to understand customer information and opinions to provide personalized experience to existing customers and attracting new customers.

Potential Stakeholders :

CUE

Sportsgirl

REVIEW

FOREVER NEW

WITCHERY

sussan

JACQUI·E

Problem Definition

The goal is to develop a supervised machine learning model to predict customer sentiment whether positive, negative, and neutral and to predict whether the product is recommended to other users.

Business Question

Sentiment Analysis can help businesses and services to understand their customer's experience and make necessary changes to improve the quality of their services.

Business Question:

- Understand customer's opinion whether positive, negative, or neutral.
- Identify if the product is recommended by customer.
- Identify customer likes and dislikes.
- Identify Target audience for the shop.
- Identify least recommended Items.

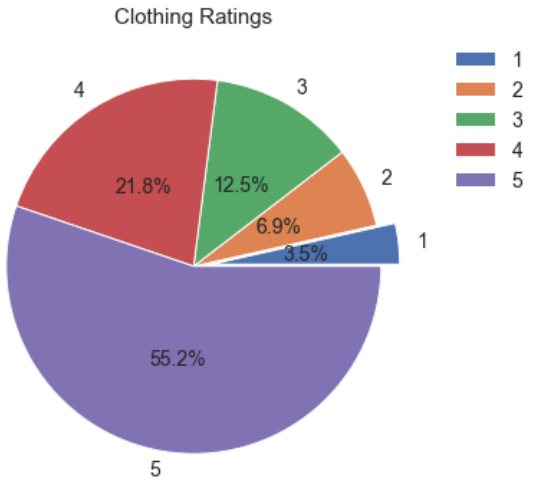
Data Question and Dataset

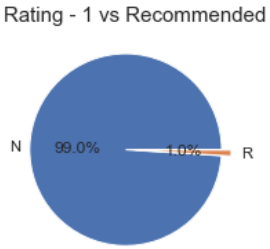
- Data Question
  - What models can be used to analyze customer sentiment and product recommendation?
  - What are target features?

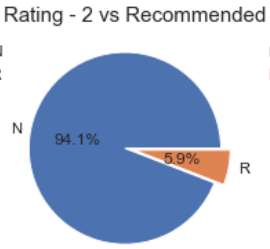
- Data set
  - Dataset downloaded from https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews
  - This dataset includes 23486 rows and 10 feature variables. Each row corresponds to a customer review, and includes the variables:
    - Clothing ID: Integer Categorical variable that refers to the specific piece being reviewed.
    - Age: Positive Integer variable of the reviewer's age.
    - Title: String variable for the title of the review.
    - Review Text: String variable for the review body.
    - Rating: Positive Ordinal Integer variable for the product score granted by the customer from 1 Worst to 5 Best.
    - Recommended IND: Binary variable stating where the customer recommends the product where 1 is recommended, 0 is not recommended.
    - Positive Feedback Count: Positive Integer documenting the number of other customers who found this review positive.
    - Division Name: Categorical name of the product high level division.
    - Department Name: Categorical name of the product department name.
    - Class Name: Categorical name of the product class name.

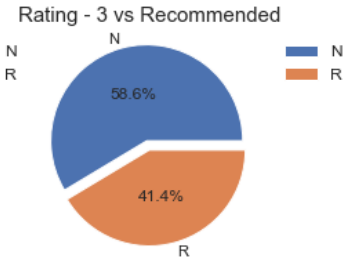
Design

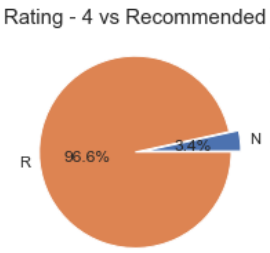
Exploratory Data Analysis(EDA)

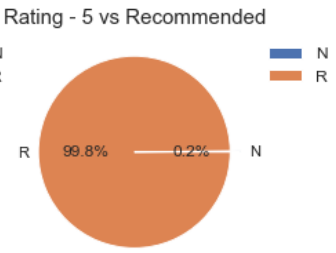


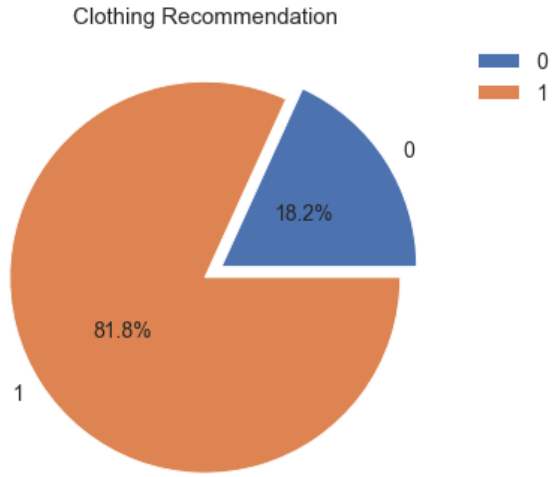




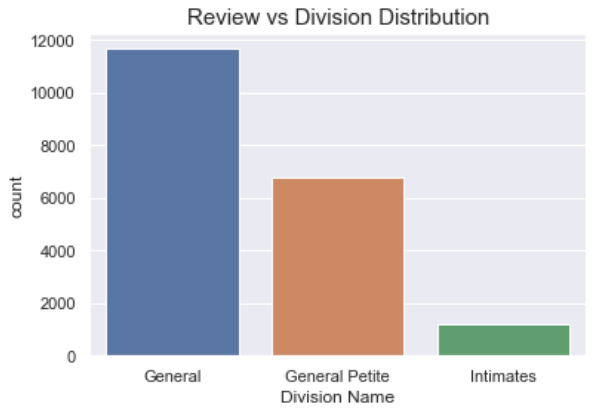


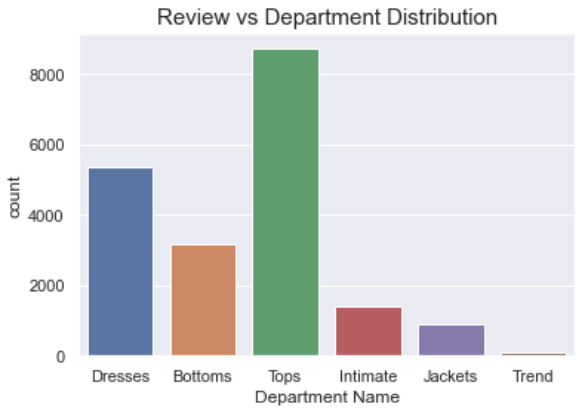


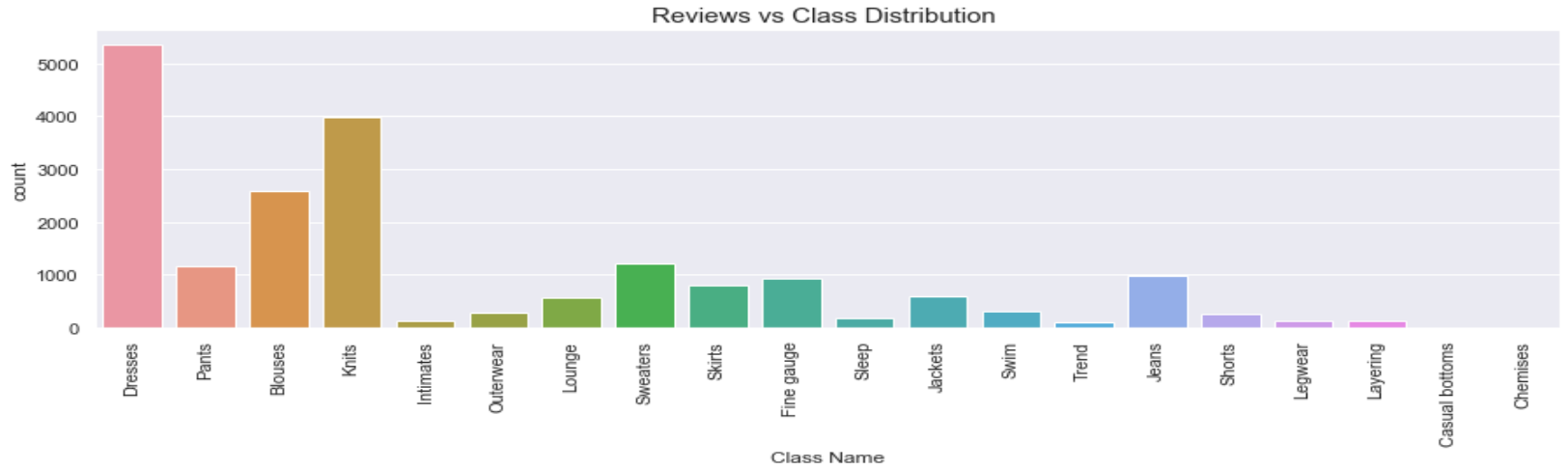




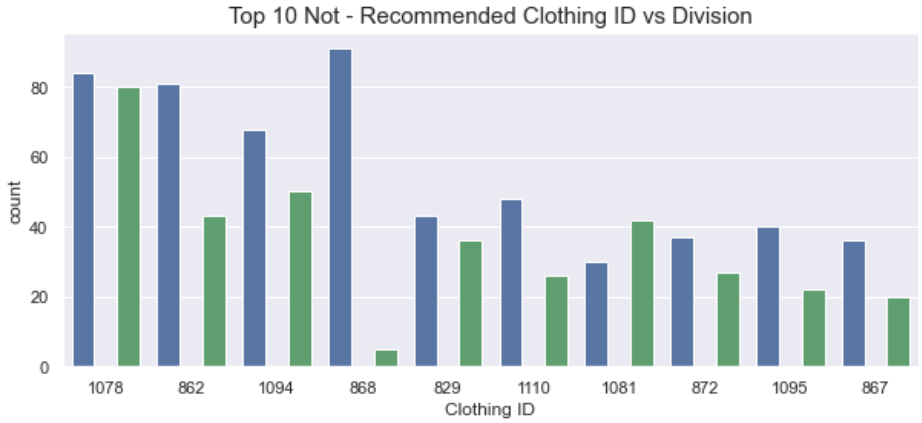
Review Distribution with Division, Department and Clothing Class

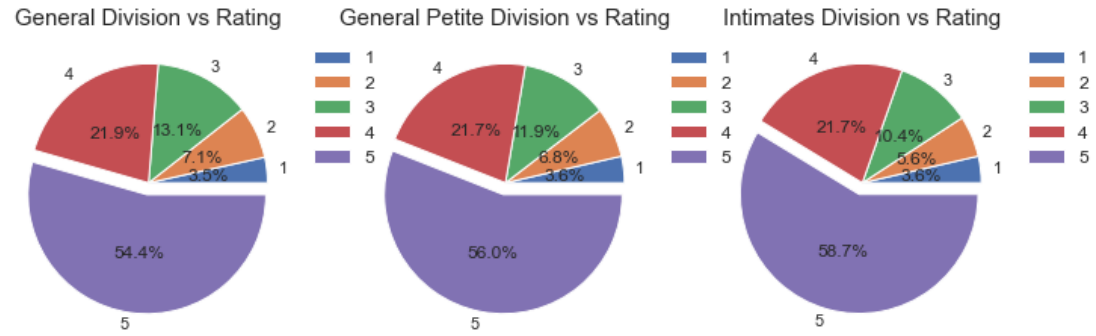


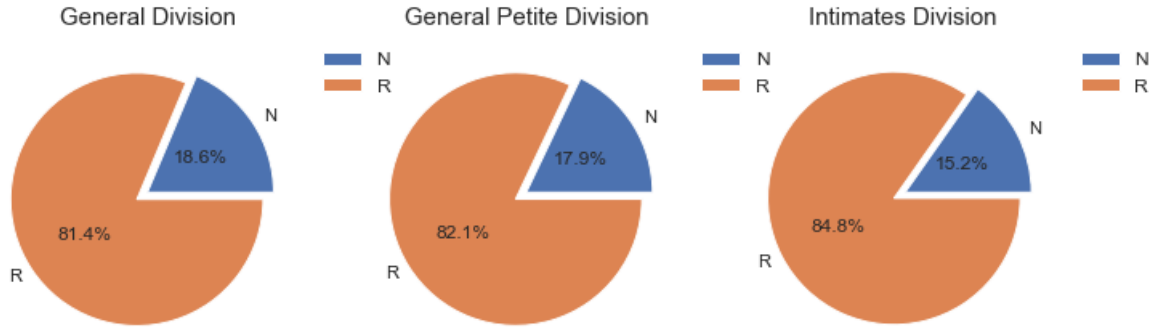




Least Recommended Products

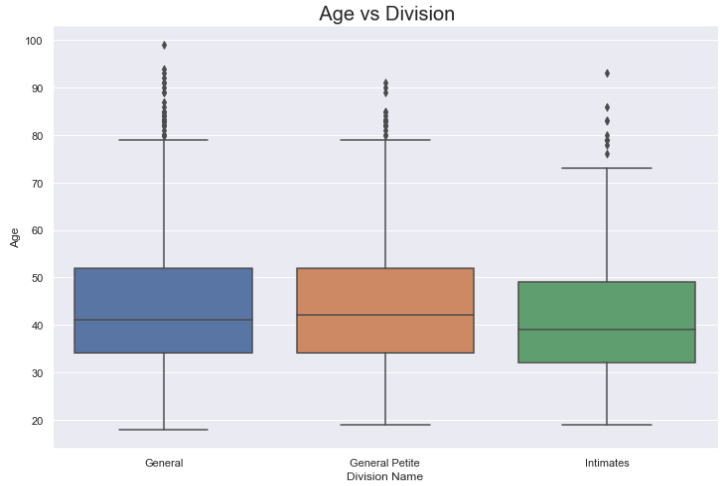


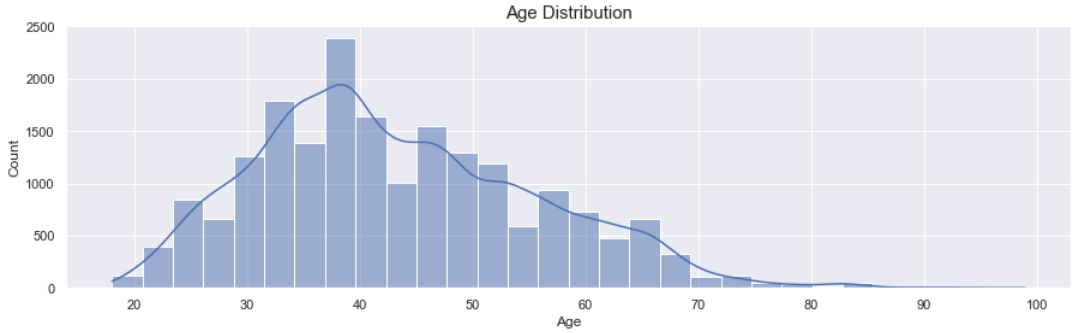




Age


Age vs Clothing Class
Age
Class Name
Dresses
Pants
Blouses
Knits
Intimates
Outerwear
Lounge
Sweaters
Skirts
Fine gauge
Sleep
Jackets
Swim
Trend
Jeans
Shorts
Legwear
Layering
Casual bottoms
Chemises


Age vs Division
Age
General
General Petite
Intimates
Division Name

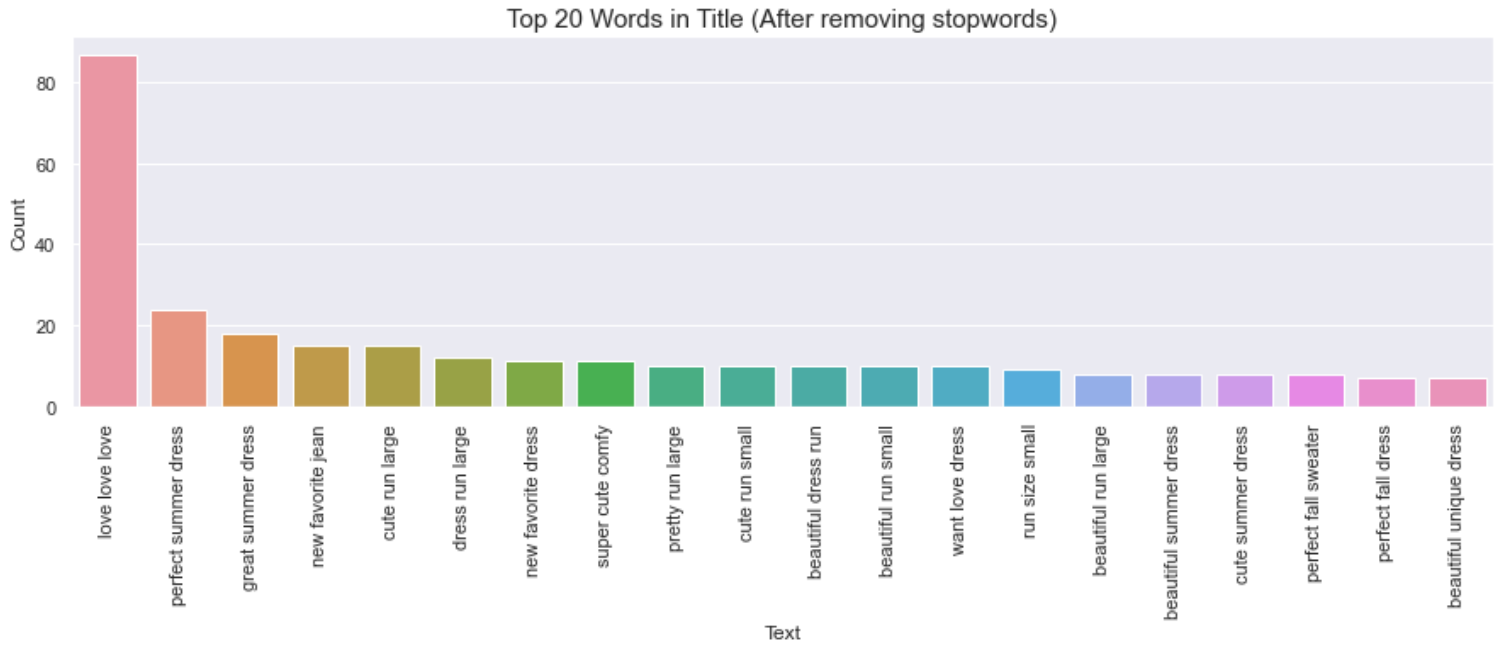

Age Distribution
Count
Age

Word Cloud for Recommended Product with Rating

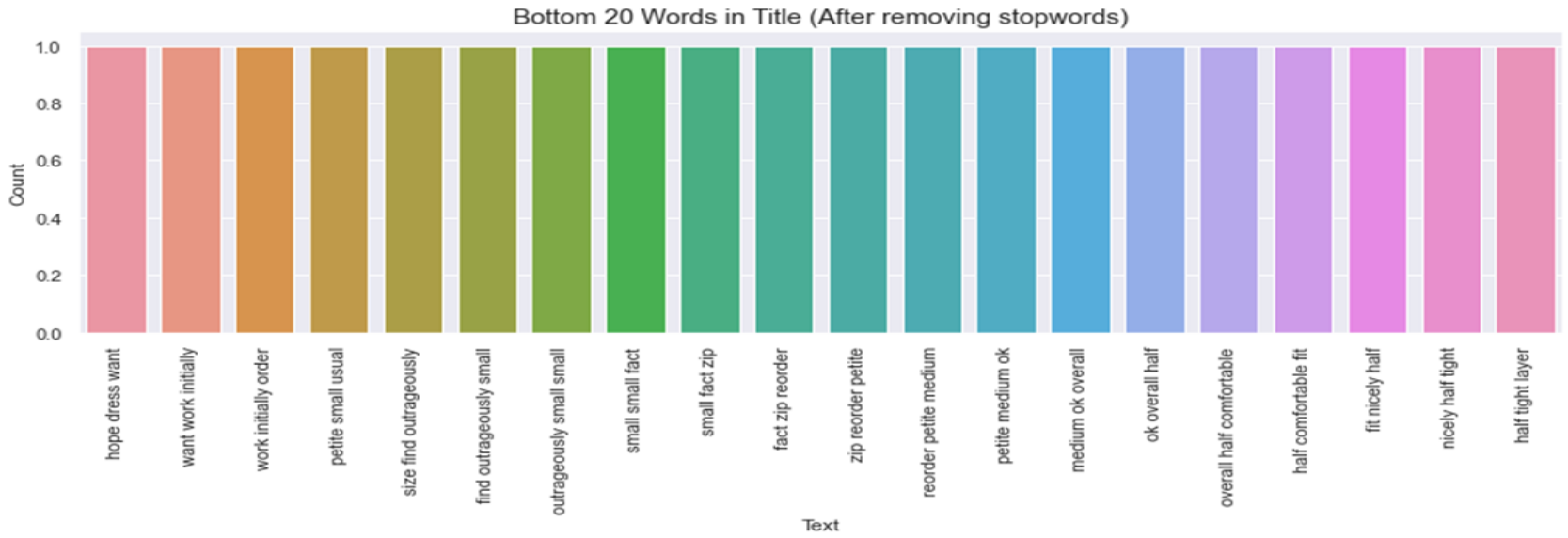
Product Recommended with Rating 5

Product Not - Recommended with Rating 1

Top 20 words(tri-grams)



Bottom 20 Words(tri-grams)



Deliver

Feature Engineering

- Feature Engineering
  - Combined Review and Title and created new Review text
  - Polarity as a feature
  - Counter vectors as features
  - TF-IDF as features
  - Text/NLP as features

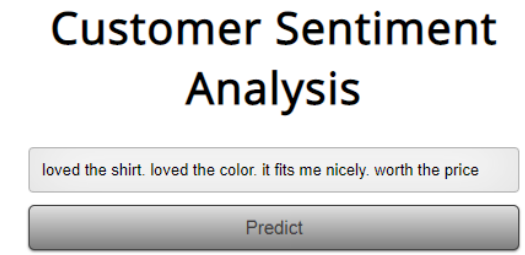
- Model Fitting
  - Data were split into two groups - 80% for training the model and 20% for testing the model.

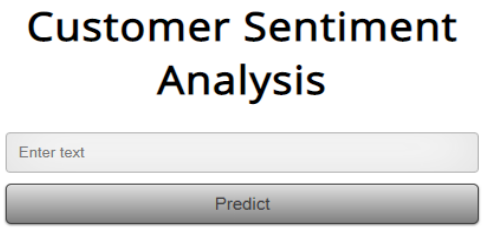
Model Evaluation

Model (Sentiment Classification)	Accuracy	Recall
Logistic Regression	76%	90%
Naïve Bayes	67%	77%
Random Forest Classifier	95%	97%
Gradient Boosting Classifier	78%	93%
Stacking Classifier	81%	52%

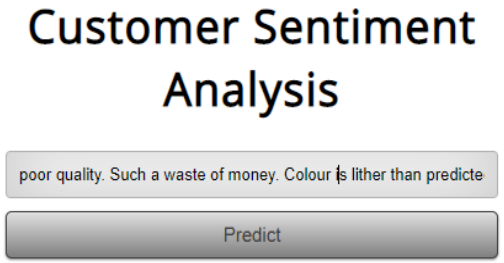
Model (Recommendation Classification)	Accuracy	Recall
Logistic Regression	90%	91%
Naïve Bayes	85%	81%
Random Forest Classifier	94%	94%
Gradient Boosting Classifier	91%	93%
Stacking Classifier	82%	67%

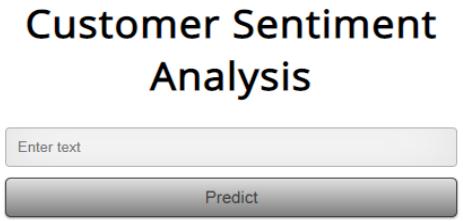
Deployment





Sentiment is positive and Product is recommended.





Sentiment is negative and Product is not recommended.

Summary

- Sentiment can be predicted with 95 % accuracy and Recommendation can be predicted with 94% with Random Forest Classifier.
- Target age group for the shop identified as 30 to 55.
- Highest rated clothing classes are Dresses, Knits, Blouses, sweaters, and pants. Trend, legwear, sleep, Layering, and Intimate classes have received least number of rating of five.
- Negative reviews are related to issues with size, color, fabric, quality, dresses, sweaters, and shirts.
- Trend cloths are only recommended even though it has lowest rating of 5.

Next Steps

- Apply Deep Learning Model for the Analysis.
- Combine social media website to retrieve more reviews.
- Expand Model to predict the customer churn over.
- Combine the model with stock data for stock management.
- Develop a chatbot to collect customer reviews and perform chatbot sentiment analysis.

Question & Answers