**Customer Review Analysis For Leading Woman Clothing E-Commerce Company**

* All the EDA Process done such as:

1. Cleaning the Review and Calculating the Sentiment Score and Comparing the Sentiment score with Original Review.
2. Checking the Length of Review after Cleaning it.
3. Assigning Threshold to Give Sentiments as Positive, Negative, Neutral.
4. Understanding for each Rating and its Sentiments.
5. Understanding sentiment among customers for different Categories.
6. Understanding sentiment among customers for different Sub-Category1.
7. Understanding Sentiment for different Age-Groups using Line Graph.
8. Understanding Sentiment among different Locations.
9. Word Clouds Created for Positive and Negative Sentiments.

* Ratings assumed as Multinomial Classification and Recommend Flag considered as Binary Classification.
* Vectorization [DTM] done for Cleaned Review using Count Vectorizer.
* Created a Bar Chart for Top 20 Frequent words.
* Review Title and Review DTM’s created using TF-IDF Vectorizer and Both Concatenated as X Variables for Predictive Modelling.
* **Algorithms Used For Multinomial Classification:**

1. **Linear SVC Model**
2. **KNN Model**
3. **Multinomial NB Model**
4. **Logistic Regression**

|  |  |  |
| --- | --- | --- |
| **Models** | **Train Accuracy** | **Test Accuracy** |
| 1] Linear SVC Model | 84% | 64% |
| 2] KNN Model BP  KNN Model OP | 100%  65% | 61%  61% |
| 3] Multinomial NB | 71% | 65% |
| 4] Logistic Regression | 79% | 66% |

For Conclusion:

Most of the Models are Overfitting the only Stable Model is KNN Model OP though it is having less accuracy its quite stable compared to other Models.

Predicting via Categorical variables using XG Boost Model Train and Test accuracy is 55% which is also Stable.

So if want to predict the Ratings we can use KNN Model OP and to predict via Categorical Variables we can use XG Boost!

* **Predictive Analysis for Recommend Flag [Binary Classification]**
* We use same X Variable (DTM of Review Title and Review) for Predicting Recommend Flag.
* **Algorithms used:**

|  |  |  |
| --- | --- | --- |
| **Models** | **Train Accuracy** | **Test Accuracy** |
| 1] Linear SVC Model | 96% | 90% |
| 2] KNN Model BP  KNN Model OP | 100%  90% | 87%  87% |
| 3] Multinomial NB | 92% | 90% |
| 4] Logistic Regression | 94% | 90% |

**To conclude:**

We can clearly see that Multinomial NB Model is highly stable model with 90% Accuracy, so to predict the Recommend Flag we can use this model.

Predicting via Categorical variables using XG Boost Model Train and Test accuracy is 82% which is also Stable.

So if want to predict the Recommend Flag we can use Multinomial NB model and to predict via Categorical Variables we can use XG Boost!

* **Text Segmentation done with 20 Clusters and found out Top 20 words for each segment.**
* **Each Review Title and Review concatenated with its Segment.**
* **Topic Mining**

1. **Cleaned the Tokens and Created unique tokens Using Corpora Dictionary.**
2. **Created DTM using doc2bow function [Bag of Words].**
3. **Created LDA [Latent Dirichlet Allocation] Model using GENSIM Library.**
4. **Created 50 Topics with Top 25 words.**
5. **50 Topics chosen as its Perplexity score is less compare to other Number of Topics.**
6. **For Example: Topic 1 consists of words such as Dress, Online Purchase, Happy, Altered, Dark, Expensive which means the Reviewer using this words is a Neutral Reviewer.**