**ML Major Project**

E-Commerce Women’s Clothing Reviews

Q. Analysis:

1. Describe the data
   1. Descriptive statistics, data type, etc.
2. Analyze the text comment/ review and share the findings
3. Convert the ratings into 2 classes
   1. Class: Bad when Rating <=3
   2. Class: Good otherwise
4. Develop a model to predict the Rating class (created above)
   1. Focus on steps to build a model
   2. Which algorithm can be used and why
5. Share the findings of the model.

Ans. The following is the analysis we drew from the given data:

1. Out of **22,628 products, 17,435 (77%) products** have a ‘**Good’ rating** and **5,193** **(23%) products** have a ‘**Bad’ rating.**
2. **12,527 (55.3%) products** have **5 rating**, **4,908 (21.7%) products** have **4 rating,** and **2,823 (12.5%) products** have **3 rating**.
3. **18,527 (81.8%) reviews recommended** the product, whereas **4101 (18.2%) reviews did not recommend** the product.
4. **Age groups (30- 40)** and **(40- 50)** have the **highest count of ‘Good’ reviews** (Rating > 3). Further, **age group (10-20)** and **(70 and above)** gave the **least amount of reviews.**
5. **Age groups (0-20)** gave the **most** amount of **recommendations**, while also suggesting the **most** amount of **‘No Recommendation’.**
6. **Tops**, **Dresses**, and **Bottoms** have the highest Review counts among Departments, respectively, with **most Reviews** from **age group (30-40)** and **(40-50). And Trend has the less Review count among Departments.**
7. **Clothing IDs 1078 (871 counts) , 862 (658 counts) and 1094 (651 counts)** have the **highest number of reviews (both ‘Good’ and ‘Bad’)**, respectively.
8. We see that there is a **positive correlation (0.79)** between **Rating** and **Recommend\_IND (Recommend Index).** The better the rating, the more the recommendation.
9. The **sentiment** from the **Review Text** data indicates that they are **more** **neutral** and **positive** reviews than negative reviews as seen from the polarity graph.
10. The **most frequently occuring words** in Review Text are- dress, size, love, fit, top, like, wear, great and would.
11. Support Vector Machines(SVM) algorithm is giving