🎤 Presentation Speech (12 Minutes)

---

1. Introduction (1.5 minutes)

Good [morning/afternoon], everyone.
Today, our group is going to present our project on **Consumer Review Analysis of Clothing Products**.

The main idea of our project is to analyze online customer reviews of clothing items using **machine learning techniques**.

We used a dataset from Kaggle, which contains information such as:

- review titles,
- full review text,
- ratings,
- product category,
- material type,
- construction quality,
- finishing, and durability.

Our objective was to build a **classification model** that can predict customer ratings, based on these features.

We mainly focused on two machine learning algorithms:

- **Support Vector Machines (SVM)**
- **K-Nearest Neighbours (KNN)**

By doing this, we aimed to discover useful patterns in consumer behavior and show how machine learning can improve the fashion retail industry.

---

2. Literature Survey (1.5 minutes)

Online reviews are now one of the most important sources of information for both customers and companies. Customers rely on reviews before making purchasing decisions, and retailers use them to improve product design and quality.

However, the number of reviews is often very large. It is almost impossible for humans to read them all, so **machine learning methods are widely used to process and analyze them automatically**.

This type of analysis is often called **sentiment analysis** or **opinion mining**, which classifies reviews as positive, negative, or neutral.

From previous research:

- **SVM** has been proven to work well for text classification, especially when the data is high-dimensional.
- **KNN** is simple, yet powerful, because it works by comparing similarities between reviews.

Most existing studies only use the review text. But some researchers suggest that including extra product-related features, such as material or durability, can give better predictions.

Our project addresses this gap by analyzing reviews with **both text and product features**. This makes the prediction more accurate and valuable for fashion retailers.

---

3. Methodology (2 minutes)

Let me now explain the methodology we followed.

First, we performed **data collection**. The dataset was downloaded from Kaggle.

Next, we went through **data preprocessing**:

- We cleaned the review texts by removing special characters, numbers, and stop words.
- We applied text normalization techniques like lowercasing and tokenization.
- For the text data, we used **TF-IDF Vectorization** to convert the text into numerical features.

After that, we prepared the dataset for machine learning:

- We split the data into **training and testing sets**.
- Both review text and product features were combined as input.

For the machine learning models:

- We trained and tested **Support Vector Machines (SVM)**.
- We also trained and tested **K-Nearest Neighbours (KNN)**.

Finally, we compared their performance using accuracy and other evaluation metrics.

---

4. Implementation (1.5 minutes)

In the implementation phase, we used **Python and Jupyter Notebook**.

The main libraries we used were:

- Pandas and NumPy for data handling,
- Scikit-learn for model training and evaluation,
- Matplotlib and Seaborn for visualization.

We implemented preprocessing steps such as:

- Removing missing data,
- Encoding categorical features,
- Applying TF-IDF on review text.

Once the dataset was ready, we built our models:

- For **SVM**, we experimented with different kernels, such as linear and RBF.
- For **KNN**, we tested different values of K to find the best number of neighbors.

We then trained the models on the training set and tested them on the unseen test set.

---

5. Results & Evaluation (2 minutes)

Now, let us move to the results.

We compared the accuracy of **SVM** and **KNN**.

The results showed that **SVM performed better** than KNN in predicting customer ratings. This is mainly because SVM handles high-dimensional data, like text, more effectively.

We also generated accuracy graphs for both models, which clearly showed that SVM had higher performance.

From evaluation metrics, we looked at:

- Accuracy,
- Precision,
- Recall, and
- F1-score.

Overall, **SVM achieved the best balance** across these metrics.

This shows that SVM is more suitable for analyzing customer reviews when both text and product-related features are included.

---

6. Discussion (1.5 minutes)

From these results, we can discuss a few important points:

- First, adding product-related features improved the performance of the models compared to using text alone.
- Second, SVM was more effective than KNN because it can separate complex decision boundaries in high-dimensional space.
- However, KNN is still useful when the dataset is small or when we want a simple model.

In real-world applications, such models can help clothing companies automatically analyze thousands of reviews and quickly understand customer satisfaction.

---

7. Conclusion (1 minute)

To conclude, our project successfully demonstrated how **machine learning can be applied to consumer review analysis in the fashion industry**.

We used a dataset of clothing product reviews and applied **SVM and KNN** to predict customer ratings.

Our findings show that **SVM outperforms KNN** in terms of accuracy and reliability.

This project highlights the importance of combining **text features** with **product-related features** to improve predictions.

In the future, we can extend this work by:

- Using deep learning models such as LSTMs or Transformers,
- Applying more advanced feature engineering techniques,
- Or analyzing reviews in multiple languages.

---

8. Acknowledgment & Closing (30 seconds)

We would like to thank our lecturers and mentors for their guidance, and also Kaggle for providing the dataset.

Thank you for listening to our presentation. We are now ready to take any questions.