

# ECOMMERCE PRODUCT CATEGORIZATION

**Problem Statement:** In The Rapidly Evolving World Of Ecommerce, Accurate Product Categorization Is Crucial For Ensuring Seamless Customer Experiences, Reducing Search Friction, And Increasing Product Discoverability. However, The Sheer Volume Of Diverse Products Poses A Significant Challenge. Current Classification Systems Struggle To Handle Ambiguities, Unconventional Naming Conventions, And Multi-language Data. This Project Aims To Address These Challenges By Inviting Participants To Create Innovative Solutions That Enhance Product Categorization Efficiency, Accuracy, And Scalability.

---

**Objective:** Develop an accurate and efficient machine learning model for eCommerce product categorization using text data from product descriptions and names.

# METHODOLOGY

- **Data Exploration and Preparation:**

- Analyzed key features, detected missing data.
  - Preprocessed data for consistency and noise reduction.
- 

- **Descriptive Analysis:**

- Identified data patterns and category distributions.
- Visualized insights using word clouds.

- **Feature Engineering:**

- Used TF-IDF for text data transformation.

- **Predictive Modeling:**

- Implemented Random Forest Classifier and SVC.

- **Fine-Tuning:**

- Used Grid Search for hyperparameter tuning.

- **Ensemble Modeling:**

- Combined models using Voting Classifier.

# PERFORMANCE METRICS AND BEST PARAMETERS

- **Grid Search Best Parameters:**

- Random Forest Classifier: max\_depth=None, n\_estimators=300
- SVC: C=1, kernel='linear'

- **Performance Metrics:**

- **Accuracy:** 0.966
- **F1 Score:** 0.9645
- **Precision:** 0.9654
- **Recall:** 0.966
- **Ensemble Accuracy:** 0.9643

# RECOMMENDATIONS AND CONCLUSION

---

## •Recommendations:

1. Further Tuning: Explore additional hyperparameters.
2. Feature Engineering: Use advanced methods like word embeddings.
3. Handling Imbalance: Use techniques like SMOTE.
4. Ensemble Methods: Test stacking or boosting.
5. Interpretability: Use SHAP for model interpretation.
6. Scalability: Test performance on larger datasets.
7. Continuous Learning: Implement periodic retraining.

## •Conclusion:

1. Achieved high accuracy and robust performance.
2. Improved product categorization can enhance user experience and operational efficiency.