**Data Science Hackathon**

**Hackathon Topic**

**:– 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 hackathon aims to address these challenges by inviting participants to create innovative solutions that enhance product categorization efficiency, accuracy, and scalability.

Develop a text classification model that categorizes products with maximum accuracy based on description of the product.

**Objectives:**

This hackathon focuses on creating accurate and efficient solutions for eCommerce product categorization. The key objectives are:

**1. Data Exploration and Preparation:**

- Explore and analyze the dataset to understand key features, detect missing data, and identify ambiguities.

- Preprocess the dataset for consistency, noise reduction, and missing value handling.

**2. Descriptive Analysis:**

- Perform descriptive analysis to identify data patterns, category distributions, and inconsistencies, especially for text data.

- Visualize data insights using word cloud and other text viz techniques.

**3. Feature Engineering/Text to Features:**

- Transform raw product data into informative features suitable for machine learning models.

- Convert text descriptions into numerical features using techniques such as TF-IDF, word embeddings, or custom methods.

**4. Predictive Modeling:**

- Design and develop machine learning models using text data for accurate product categorization.

- Consider classification, clustering, or hybrid models based on problem requirements.

- Ensure the model can handle ambiguous products

**5. Fine Tuning:**

- Optimize models by tuning hyperparameters and selecting relevant features.

- Evaluate and compare models using metrics such as accuracy, F1 score, precision, and recall.

- Validate models through cross-validation or a separate test dataset.

**6. Enhance Categorization Accuracy:**

- Improve model accuracy by incorporating domain-specific knowledge or ensemble methods.

- Ensure solutions can accurately categorize unconventional product names and handle ambiguities.

- Develop models that support scalable, real-time categorization.

**Expected Outcomes:** By the end of the hackathon, participants are expected to deliver the following outcomes:

**1. Predictive Model:**

- A text based predictive machine learning model capable of accurate product categorization across multiple categories.

- The model will be evaluated using accuracy, F1 score, precision, and recall, ensuring high classification performance.

**2. Visualizations:**

- A comprehensive set of visualizations will accompany the model, providing insights into data patterns, category distributions, word clouds etc.

**3. Well-Documented Jupyter Notebook:**

- Participants will deliver a well-documented Jupyter Notebook, including all steps of data exploration, feature engineering, model development, and fine-tuning.

- The notebook will serve as a reproducible guide for others to follow and understand the solution.

**4. Compelling Presentation in a Video Walkthrough Format:**

- A compelling video walkthrough presentation will highlight the model's features, predictive accuracy, and business impact.

- The presentation will include:

- An overview of the problem and approach.

- Key insights from data analysis and visualizations.

- A demonstration of the predictive model.

- An explanation of the business value provided by the solution.

The hackathon aims to deliver innovative product categorization models that enhance customer experience, product discoverability, and business revenue while providing deployment-ready solutions for immediate impact.