

Report

This project aims to classify products into subcategories using their description as given in Assignment. We will be using the Logistic Regression model for classification. The dataset contains product descriptions and relevant subcategories. We'll preprocess the dataset, vectorize the descriptions using TF-IDF, and then train a machine learning model.

Step 1: Load and Preprocess Dataset

We begin by loading the dataset using pandas. Then we clean data by removing any rows with missing values in the Description or SubCategory columns. Once clean, we then select the columns. The columns used will be the Description, SubCategory, and ProductName.

Step 2: TF-IDF Vectorization

To convert the text data into a format that can be used for machine learning, we use TF-IDF vectorization. We apply TfidfVectorizer on the Description column. This vectorizer removes English stop words and limits the features to 5000.

Step 3: Model Training

We split the data into training and test sets. We are using The Logistic Regression model, which is suitable for multi-class classification, is then trained on the training data. The hyperparameters of the model are selected to enhance accuracy and overall performance. data set is too large so training process will take long time .

Step 4: Model Evaluation and Displaying Results

After training, we evaluate the model's performance using the test set. We use confusion matrix to evaluate performance. y-axis contains True-Sub Category and x-axis contains Predicted Sub Category. Evaluation helps us to see where model performs well, and where it doesn't.

Result

| | ProductName | Predicted_SubCategory | True_SubCategory |
|------|---|-----------------------|---------------------|
| 4614 | "Nylon Swim Shorts" | Active Pants | Shorts |
| 397 | "Men's Basic Fit Chino Pants" | Chinos | Chinos |
| 1176 | "Men's Paige Brysen Regular Fit Crewneck Sweatshirt" | Sweatshirts/Hoodies | Sweatshirts/Hoodies |
| 4364 | "J.Crew Stretch-Cotton Twill Chinos" | Chinos | Chinos |
| 739 | "padded gilet" | Vests | Vests |
| 3143 | "Gucci Printed Silk-Twill Pocket Square" | Pocket Squares | Pocket Squares |
| 3162 | "Waxed Leather Lace-up Boots" | Boots | Boots |
| 1161 | "Dries Van Noten Yellow Rib Cardigan" | Cardigans | Cardigans |
| 151 | "Logo Icon Cotton Jersey Boxer Briefs" | Boxers & Briefs | Boxers & Briefs |
| 477 | "Iphone XR wings cover" | Tech Accessories | Tech Accessories |
| 2999 | "Gucci Men's GG Wool Slipper With Horsebit - Blue Red " | Slides/Slipper | Slides/Slipper |
| 2765 | "Brunello Cucinelli Men's Reversible Pocket Square - Green" | Pocket Squares | Pocket Squares |
| 1762 | "Embossed Cotton Jersey Jogging Pants" | Active Pants | Active Pants |

Confusion Matrix

