NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES
(KARACHI CAMPUS)
FAST School of Computing
Spring 2024

Course Instructor: Mr. Basit Jasani
Project Name: Product Title Classification

Team Members:
21K-3279 Insha Javed
21K-4503 Muhammad Tahir
21K-4606 Sabika Shameel

The basic understanding of our project:

The project titled "Bagging Model for Product Title Quality with Noise" aims to enhance the quality of online product titles. Over the years, the availability and utilization of computers in educational institutions has significantly increased. These powerful machines serve various purposes, including administration, instructional delivery (through exercises, tutorials, and simulations), and problem solving. However, despite their widespread adoption, there remains a lack of research on the effectiveness of computer-assisted instruction compared to traditional classroom methods. This study seeks to address this gap by specifically examining the effectiveness of computer-assisted instruction in a class setting. By addressing noisy titles, this project contributes to better search results, user understanding, and overall satisfaction during online shopping. The research focuses on data processing, model development, and evaluation to achieve these objectives. Ultimately, it seeks to optimize product titles for efficient communication and improved user engagement.

Workflow of SVM-based Product Category Classification:

1. **Preprocessing**:

   Data Cleaning: Text in titles and descriptions is cleaned by removing punctuation, stop words (common words like "the", "a"), and potentially applying stemming to reduce words to their root form.

2. **Feature Extraction:**

   TF-IDF: Alternatively, consider TF-IDF (Term Frequency-Inverse Document Frequency) that accounts for both word frequency and its rarity across the entire dataset, downplaying the weight of common words.

   Combine Features:A single feature vector is created for each product by concatenating the BoW (or TF-IDF) vectors of its title and description. This captures information from both features.

   Label Encoding: Product categories are converted into numerical labels (e.g., "Electronics" = 1, "Clothing" = 2).

   Train the SVM Model: An SVM model is trained using a training dataset containing the combined feature vectors and corresponding category labels. The model learns the optimal hyperplane in the feature space to separate different product categories. Kernel functions are used if necessary for non-linear relationships between features and categories. Hyperparameters are tuned to optimize the model's performance.

3. **User Input and Classification :**

   User Input: The user enters a new product title.

   Preprocess New Title: The title is cleaned and converted into a feature vector using the same method as in preprocessing (TF-IDF).

   Prediction: The trained SVM model takes the new product's feature vector as input.

   Category Prediction:Based on the model's learned hyperplane, the model predicts the most likely category label (e.g., "Electronics") for the new product.

4. **Output:**

The system displays the predicted product category for the user.

This process happens in real-time when a user enters a new title.

The accuracy of the prediction depends on the quality of the training data, feature selection techniques, and model hyperparameter tuning.

Additional Considerations:

Feature Engineering: Explore alternative feature extraction techniques beyond TF-IDF for potentially better accuracy.

Model Evaluation:Regularly evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score to ensure its effectiveness.

MODEL 1

```
Accuracy: 93.05%
Precision: 93.13%
Recall: 93.05%
F1 Score: 93.05%
```

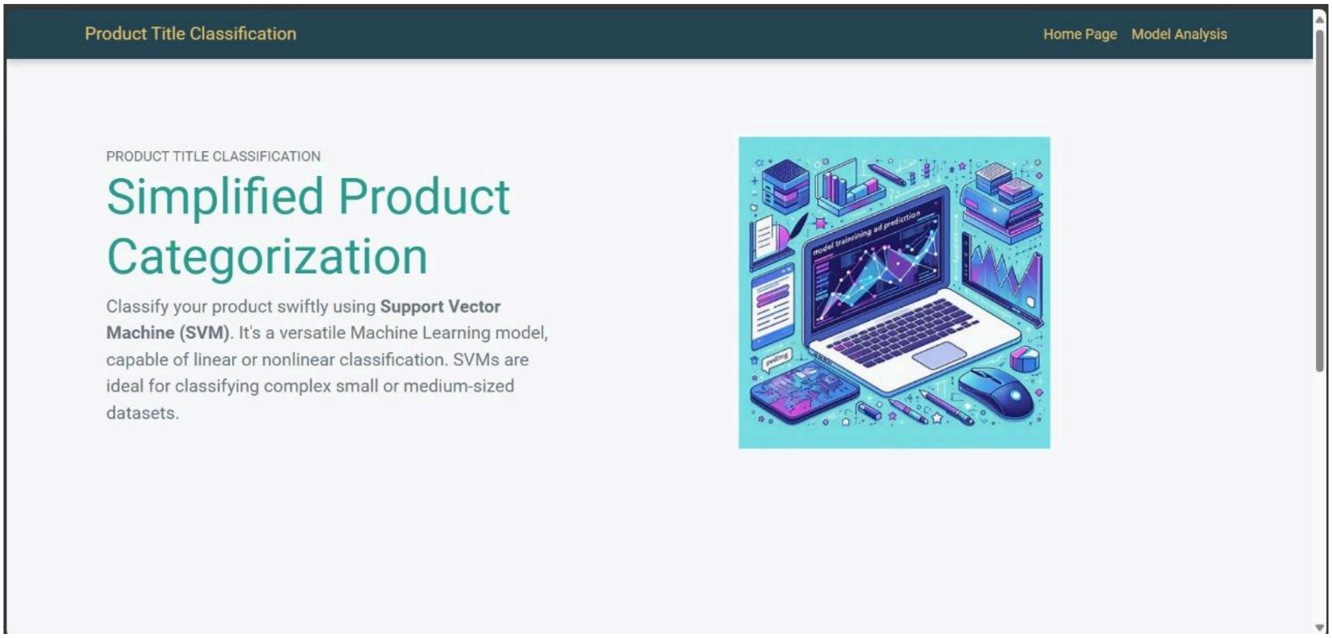
MODEL 2

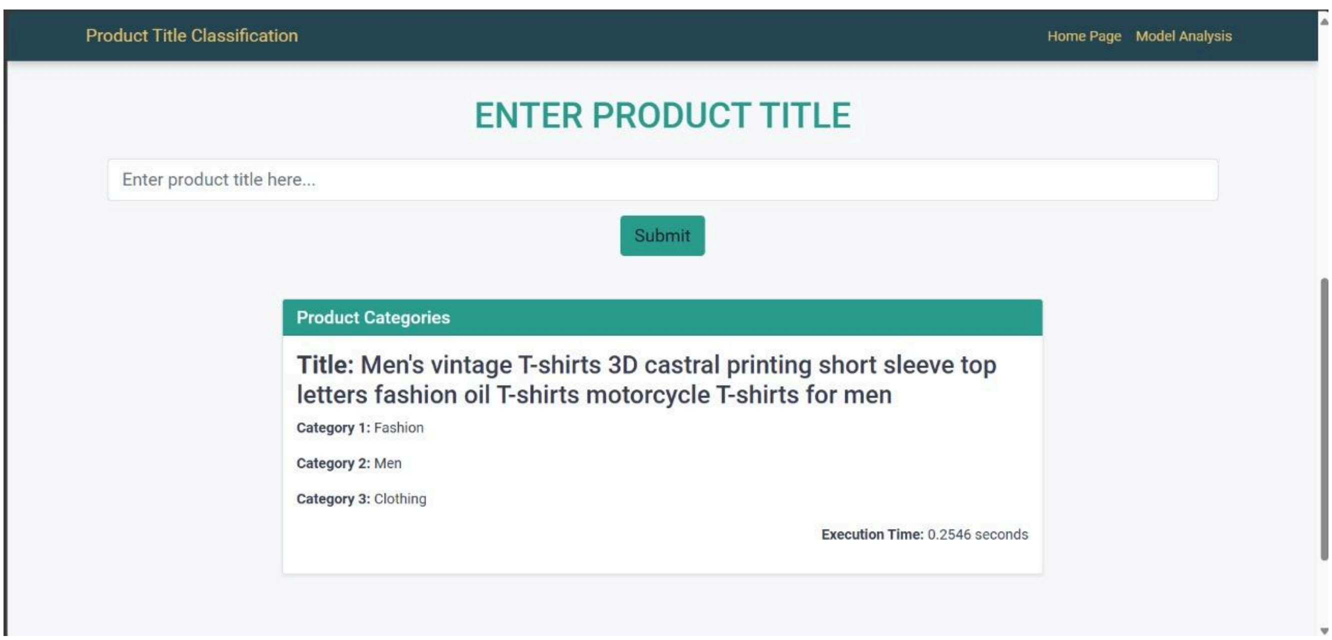
```
Accuracy: 86.42%
Precision: 86.74%
Recall: 86.42%
F1 Score: 86.27%
```

MODEL 3

```
Accuracy: 76.78%
Precision: 80.55%
Recall: 76.78%
F1 Score: 77.70%
```

Project Screenshots





Product Title Classification   Home Page   Model Analysis

Model Performance Visualization

Model 1

Evaluation Metrics

```
Accuracy: 93.05%
Precision: 93.13%
Recall: 93.05%
F1 Score: 93.05%
```

Predicted Results

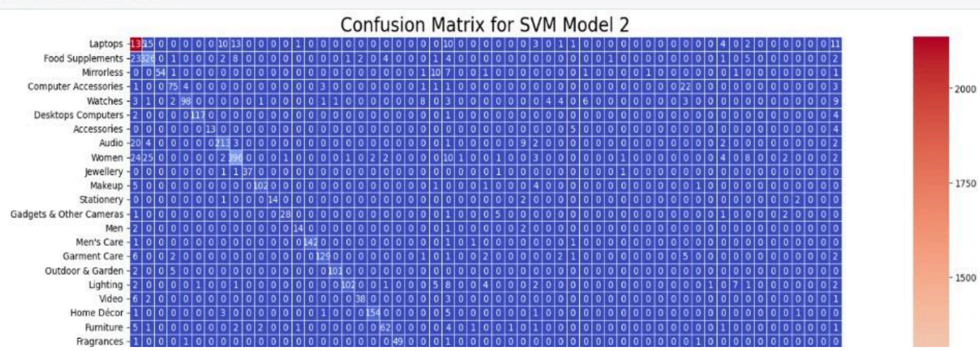
Product Title Classification   Home Page   Model Analysis

Model 2

Evaluation Metrics

```
Accuracy: 86.42%
Precision: 86.74%
Recall: 86.42%
F1 Score: 86.27%
```

Predicted Results



Product Title Classification

Home Page Model Analysis

Model 3

Evaluation Metrics

```
Accuracy: 76.78%
Precision: 80.55%
Recall: 76.78%
F1 Score: 77.70%
```

Predicted Results