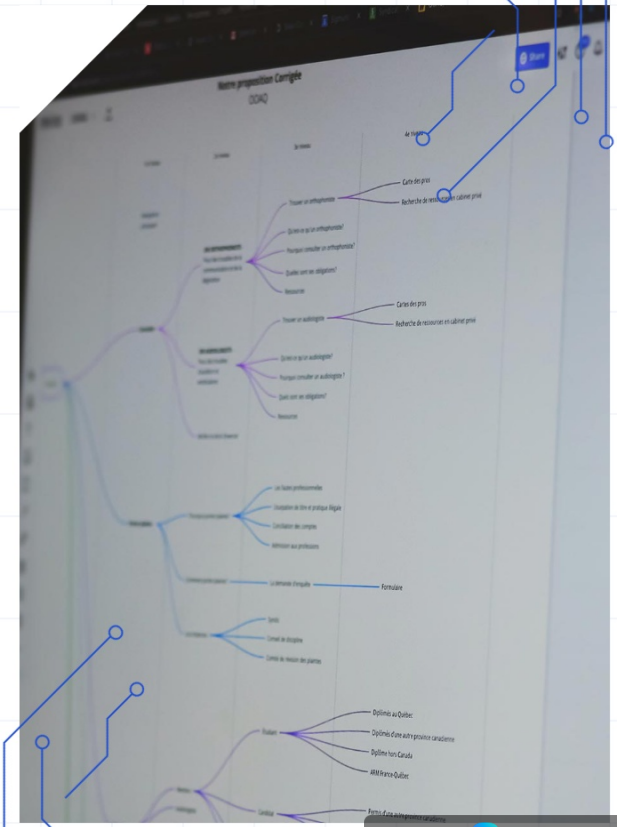
MULTI-CLASS TEXT CLASSIFICATION FOR PRODUCT CATEGORIATION

Hackthon Solution for Improved ecommerce Product Categorization

YEDUGURU JAGANMOHAN REDDY



PROBLEM OVERVIEW

Challenges and Objectives in Product Categorization

IMPORTANCE OF PRODUCT CATEGORIZATION

Accurate product categorization is essential for enhancing customer experience, improving product discoverability, and boosting sales in eCommerce.

CHALLENGES IN CURRENT SYSTEMS

Existing systems face difficulties due to ambiguous product descriptions, inconsistent naming conventions, and multilingual data handling.

IMPORTANCE OF ACCURATE CATEGORIZATION

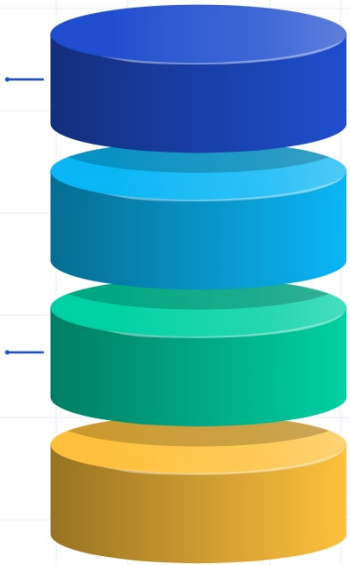
Boost Business Success with Efficient Product Organization

CUSTOMER EXPERIENCE

Accurate categorization helps customers find products faster, improving their shopping experience and satisfaction.

SCALABILITY

Automated categorization supports seamless scaling as new products are added, maintaining efficiency and organization.



SALES & REVENUE

Increased product visibility through proper categorization leads to better conversions and boosts sales and revenue.

MARKETING EFFICIENCY

Accurate categorization enables targeted marketing campaigns, enhancing marketing efficiency and effectiveness.

PROBLEM SOLVING APPROACH

Detailed Steps for Effective Machine Learning Solutions

DATA PREPROCESSING

Clean and prepare product descriptions using tokenization, stopword removal, and stemming.

FEATURE ENGINEERING

Convert text into numerical features using techniques like TF-IDF and Word2Vec.

MODEL SELECTION & TRAINING

Train models such as Logistic Regression, SVM, and Neural Networks.

EVALUATION & OPTIMIZATION

Tune models using hyperparameter optimization and cross-validation.

PREDICTION & RESULTS

Predict product categories for new data.

EXPLORATORY DATA ANALYSIS (EDA)

Analyzing Product Distribution, Keywords, and Sales Trends

DISTRIBUTION OF PRODUCTS ACROSS CATEGORIES

Analyzing how products are distributed among various categories helps identify market gaps and opportunities. Visuals like graphs or pie charts can effectively illustrate these distributions.

COMMON KEYWORDS IN PRODUCT DESCRIPTIONS

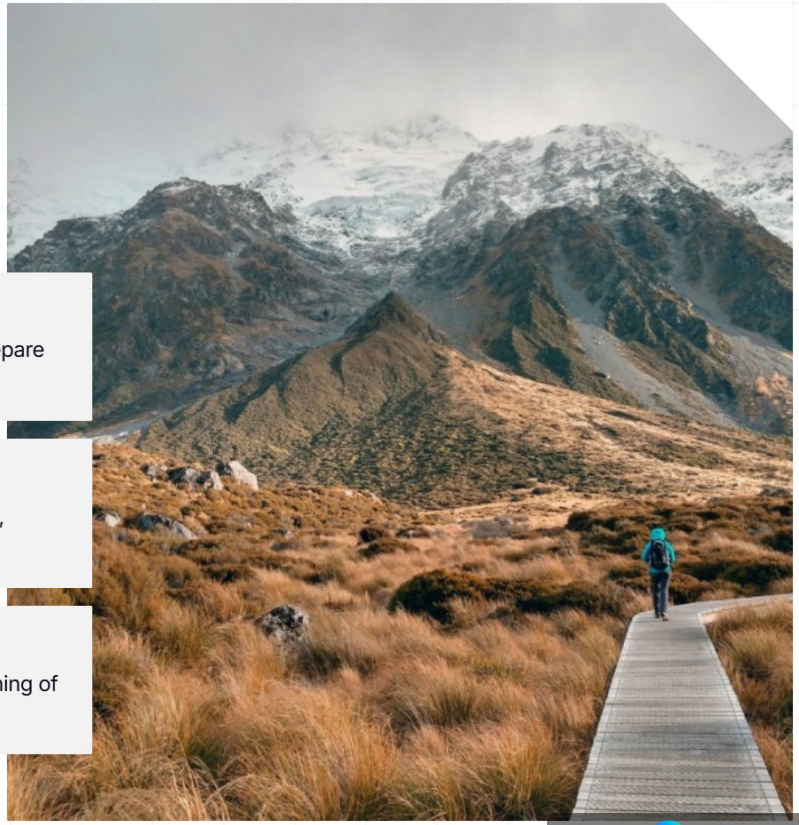
Understanding frequently used keywords in product descriptions can provide insights into consumer interests and trends. This analysis aids in optimizing product listings and marketing strategies.

SALES TRENDS PER CATEGORY

Identifying sales trends within each product category allows businesses to forecast demand and adjust inventory accordingly. Graphs depicting these trends can provide a clear visual understanding.

DATA PREPROCESSING & FEATURE ENGINEERING

Essential Techniques for Data Processing and Analysis



TEXT PREPROCESSING

Tokenize text, remove stopwords, and apply stemming or lemmatization to prepare data for analysis.

FEATURE EXTRACTION

Utilize TF-IDF to identify and quantify important words in product descriptions, converting them into numerical data.

TEXT REPRESENTATION

Implement models like Word2Vec to capture and represent the semantic meaning of words in a vector space.

MODEL SELECTION AND EVALUATION

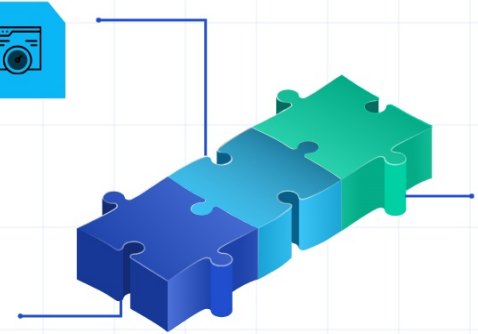
Analyzing Performance of Various Machine Learning Models

EVALUATION METRICS

Key metrics used for model evaluation were Accuracy, Precision, Recall, and F1-Score. Additionally, a Confusion Matrix was utilized to identify misclassifications and potential areas for improvement.

MODELS TESTED

We evaluated several models including Logistic Regression, SVM (Support Vector Machine), and Neural Networks such as LSTM and CNN.



BEST PERFORMING MODEL

The model with the highest performance in terms of accuracy or F1-score was highlighted as the best performing. This model demonstrated superior capability in accurately classifying inputs.

HYPERPARAMETER TUNING

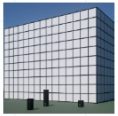
Approaches to Optimize Model Performance

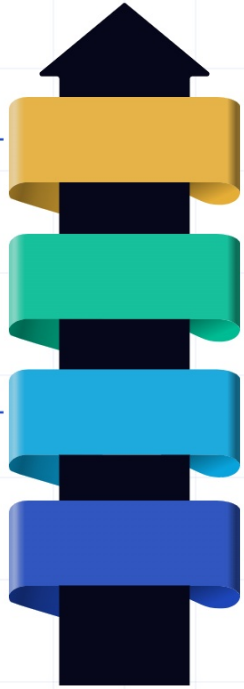
METHODS USED: CROSS-VALIDATION

Use k-fold cross-validation to ensure robust performance across different data splits.

METHODS USED: GRID SEARCH

Exhaustively search through predefined hyperparameters.



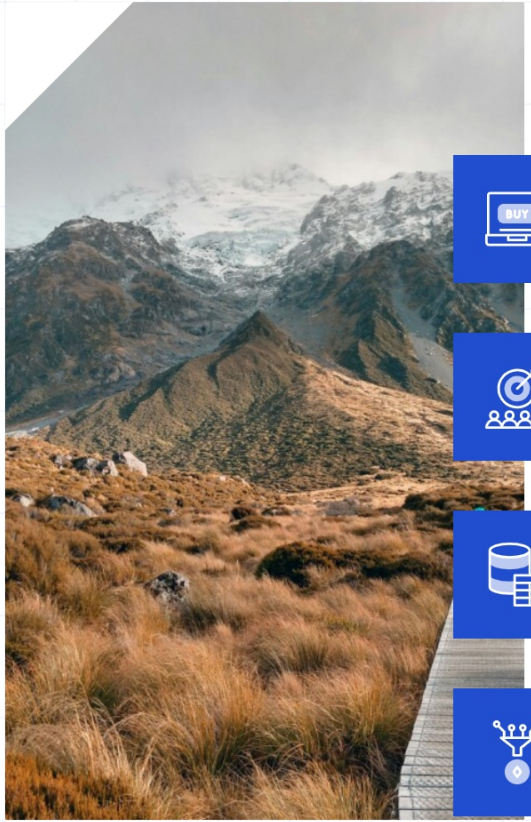


METHODS USED: RANDOM SEARCH

Randomly sample hyperparameters to find optimal settings.

OBJECTIVE

Optimize the model for better performance by adjusting hyperparameters.



BUSINESS IMPACT AND VALUE

Understanding the Strategic Benefits

IMPROVED CUSTOMER EXPERIENCE

Accurate product categorization enhances search results and navigation.

ENHANCED SALES & MARKETING

Enables more effective targeting of marketing campaigns.

DATA CONSISTENCY

Automated categorization reduces errors and ensures uniformity across product catalogs.

REVENUE GROWTH

Easier product discovery leads to higher conversion rates and reduced return rates.

CONCLUSION AND NEXT STEPS

Strategic Path Forward

01

CONCLUSION

The developed NLP-based solution provides accurate product categorization, which significantly enhances customer experience, boosts sales, and optimizes marketing strategies.

02

DEPLOY THE MODEL

Deploy the NLP model for live predictions to facilitate real-time product categorization and improve operational efficiency.

03

MODEL UPDATES

Continuously update the model with new data to maintain accuracy and adapt to changing market trends.

04

MULTI-LANGUAGE SUPPORT

Explore multi-language support to expand the solution's applicability for international markets, enhancing global reach and customer engagement.