Hierarchical E-commerce Product Categorization using Ensemble Machine Learning Techniques

Atharva Patil 11866
School of Computing
Dublin City University
Dublin, Ireland
atharvarajkumar.patil2@mail.dcu.ie

***Abstract*—Proper product classification is beneficial in the e-commerce industry to improve search, recommendation websites, and overall customer experience across online marketplaces. This is to produce a strong machine learning pipeline for product category classification on a real-world dataset. The dataset consists 400,000 product listings with accompanying data. The classification procedure consists of performing two distinct stages for either prediction at the top-level category or at the more specific bottom-level category. The preprocessing steps applied substantial methods to resolve data problems that included the treatment of data gaps and the use of TF-IDF vectorization and SMOTE for text model and class balance. Various supervised learning techniques were used to perform both classification operations. The combination of hard-voted Logistic Regression Random Forest and LightGBM delivered an optimum weighted F1-score of 0.91 for the top-level category classification task. The Ridge Classifier demonstrated the most predictable and understandable performance during bottomlevel category classification beyond what other algorithms could achieve with a weighted F1-score of 0.75. The project selected lightweight scalable models along with dataset downsampling to achieve training efficiency because of computational limits. The selected models achieved high validation accuracy which proved that both standard and ensemble learning methods result in effective solutions. The final predictions generated were saved for deployment purposes after performing predictions on test data.**

I. Introduction

The world wide growth of e-commerce has made automated product classification into an indispensable machine learning duty which controls massive product databases. Manual classification lacks viability for the e-commerce domain because it tries to categorize thousands of product types from millions of different inventory listings. User satisfaction along with improved search results and recommendations happen in digital commerce because text descriptions and metadata allow accurate product classification.

The identification of product categories exists as a multi-class classification challenge that covers multiple hierarchical levels because products belong to nested categories. The use of Optuna-based hyperparameter optimization techniques led to higher classification accuracy in SVMs and Decision Trees and Random Forests. The product categorization process proves difficult to execute because it contains three main problems regarding overwhelming class bias alongside unclear description texts and combined regular and nonstandard features [1]. A robust classification pipeline needs the combination of Machine Learning and natural language processing (NLP) alongside feature engineering and automated model selection for its development. The combination of hierarchical models with multimodal learning methods extends classification boundaries for available product image data and metadata [8], [7].

Multiple approaches to handle this challenge appear in recent writings. The extraction methods which combined NLP preprocessing with TF-IDF vectorization coupled with SVM yielded superior F1 scores for e-commerce product classification [6]. BERT shows better performance in extracting semantic details from product descriptions through collaboration with LightGBM boosting algorithms [9].

For this assignment, a private dataset from an e-commerce business has been utilized. The dataset includes product listings that are annotated with both top-level and bottom-level category IDs, representing a hierarchical category hierarchy. Each record has a combination of structured fields (e.g., product type, material, recipient, and occasion) and unstructured fields like product titles, descriptions, and tags. This diverse combination of data fields offers a realistic and challenging multi-class classification task, especially due to class imbalance and noisy or contradictory text data.

The work builds a complete machine learning framework from the data preprocessing and afterwards TF-IDF and embedding feature extraction along with SMOTE and ADASYN resampling methods and ensemble classifiers such as Random Forest, Logistic Regression, LighGBM. The goal is to classify products into their respective category levels correctly with the help of both structured and textual data features efficiently.

II. Related Work

Product categorization in the e-commerce field requires complex multiclass classification through structured metadata analysis with unstructured text evaluation of product names and descriptions. Rapid market growth in online platforms made machine learning techniques necessary for e-commerce operations because they enable automated large-scale product classification. Research experts currently examine product categorization through direct comparison between conventional classifiers and modern artificial deep learning network models which combine recursive neural technologies and transformer systems and unified models. Modern techniques focus on both precision enhancement of categorization and handling unorganized datasets across multiple category levels. Scientists conduct research investigations on product classification through various machine learning techniques and optimized data sources within this academic section.

The authors Joy and Selvan (2022) investigated the effect of hyperparameter tuning on multi-class classification through Optuna framework evaluation. The research team evaluated SVM, Decision Trees and Random Forests on four structured datasets through four optimization methods: Grid Search, Random Search and TPE as well as CMA-ES. The Random Forest model achieved maximum accuracy while Random Search proved as the superior tuning method. Structured and semi-structured classification tasks require model selection and hyperparameter optimization according to the research conducted in [1].

The authors Kanaan et al. (2023) [2] developed a method to classify e-commerce products through GPT model fine-tuning. The research methodology relied on product descriptions and metadata which delivered better accuracy along with F1-scores than standard deep learning and rule-based models. The research demonstrates that transformer models particularly GPT excel at processing text-based multi-class product classification problems.

The research by Karakaya et al. (2023) [3] introduced an ensemble learning-based system that utilizes behavioral website data to forecast online shopping purchase intentions. The model utilized k-Nearest Neighbors (kNN) together with Random Forest and Modlem classifiers designed as a stacking ensemble system that Naive Bayes operated as the meta-classifier. With the Online Shoppers Purchasing Intention Dataset (12,330 sessions) the proposed model underwent training while performing tests using clickstream along with session-based features including bounce rate, exit rate and page value and time spent on product pages. The implemented system demonstrated 87.4% accuracy and 88% F1-score because ensemble techniques prove highly effective for behavior-based classification tasks. This research stresses the need to merge different classifiers to enhance prediction assessment while directly aligning with the current project that aims to classify products into multiple categories through metadata and textual information.

The researchers Parashar and Gupta (2017) [4] developed an Artificial Neural Network (ANN) method to create an evaluation system that ranks e-commerce product quality. The method employed ANN technology to handle massive customer reviews by applying numerical rating algorithms that produced an integrated product quality score spanning from 1 to 100. This automatic platform enabled e-commerce companies to identify superior and inferior products for platform display or elimination through an algorithm that resolved review cascade and biased appraisal problems. This research explores ANN's applications in opinion mining and quality assurance which meet the objectives of metadata-based product classification used by e-commerce platforms.

A deep learning system designed for aspect and entity recognition in e-commerce search at eBay was proposed by Wen et al. (2019) [5]. The end-to-end system implements Bi-directional LSTM along with Conditional Random Fields (BiLSTM-CRF) for named-entity recognition (NER) from billions of user interactions extracted through query log mining. The system extracts product attributes as brand and color and size and other details from noisy short search queries via its predefined design. The system incorporates two key modules which include a query intent prediction module together with a ranking booster powered by aspect bit-vectors to enhance search results. The platform deployed the model to eBay's proprietary AI infrastructure where it outperformed the current production systems. Deep learning integrated with user behavior signals and infrastructure components creates an effective method to improve product understanding systems for e-commerce applications.

Reddy et al. (2024) [6] developed a computing platform based on machine learning to categorize e-commerce products through text-based documentation. Text normalization involved multiple NLP operations beginning with case folding together with stopword removal and stemming that preceded TF-IDF vectorization. Among the variety of evaluated machine learning classifiers SVM (sigmoid kernel) delivered the best F1-score at 95.19% during testing. The authors utilized Optuna for conducting hyperparameter optimization which led to better model results. The research demonstrates how NLP + ML pipelines succeed in product metadata classification tasks and meets all project requirements.

DeepCN represents an end-to-end deep learning framework for large-scale product categorization that operates using multiple RNNs according to Ha et al. (2016) [7]. The system uses separate RNN cells for different metadata attributes (title and brand and image signature) to analyze semantics individually without combining information into unified inputs. The model processed 94 million products from 4,100 categories on NAVER Shopping with superior accuracy compared to single-RNN and traditional bag-of-words models. The specialized RNN architecture brought enhanced performance by dealing efficiently with both lengthy product categories and disordered and unbalanced data types which commonly occur in online retail classification workloads.

The proposed deep learning technique from Yu et al. (2018) [8] built a multi-level classification framework to analyze extensive e-commerce products through their titles. FastText teams up with AbLSTM algorithms to process product titles through hierarchical tree structures that solve class distribution issues and deep product category hierarchies. The researchers applied single-label and multi-level label prediction techniques in their approach to achieve a top performance level of 0.8404 F1-score on the SIGIR eCom Challenge leaderboard. Product classification performance showed significant improvement when semantic word embeddings were used with hierarchical category trees and ensemble models according to research findings.

The authors Shaikh et al. (2024) developed an analytical model which applies BERT for contextual language understanding alongside LightGBM for classification to examine e-consumer behavior in e-commerce. The researchers applied complex data processing techniques before selecting features via Earth Worm Algorithm (EWA) and trained the model using BERT for linguistic understanding and LightGBM for classification. A BERT-LightGBM combination yielded higher success than independent BERT and LightGBM models by reaching 95.72% accuracy. Deep learning algorithms with boosting techniques demonstrate effective consumer prediction and generate useful information for text-based product categorization systems [9]

Xiao and Tong (2021) [10] put together a predictive model which combines TF-IDF with Logistic Regression for buying behavior predictions on e-commerce platforms by processing data from JD.com's competition records. The research model creates an interaction matrix by applying Chinese word segmentation and keyword extraction to product reviews and metadata followed by cosine similarity operations. The logistic regression model utilized weighted textual features extracted through TF-IDF for predicting user purchase behavior during the forthcoming week. The prediction accuracy reached 98% when traditional text mining techniques operated together with interpretable classification models in behavioral prediction tasks.

Analyzing e-commerce performance through machine learning algorithms involved Micu et al. (2019) [11] studying 1,420 Romanian e-commerce businesses. Through an evaluation using Google Cloud's AutoML Vision for image classification the team measured quality aspects in design including brand logos and banners and product images. Financial success of e-commerce websites demonstrated direct correlations with specific visual design elements in their websites. The application of machine learning for visual content analysis supports brand evaluation and both digital marketing strategy development and usability improvements in online platforms which delivers important insights for systems using product metadata and visual features for categorization.

III. METHODOLOGY

A. Dataset Overview

This project utilized a privately acquired set of product listings which were retrieved from an actual e-commerce platform. The system aims to categorize products using textual metadata into designated top-level as well as bottom-level groups. The product listings include five essential fields of information. The Top Category ID identifies the highest classification element which contains the product such as "home-decor". The Bottom Category ID functions as a specialized subcategory which sits beneath the top category with an example of "home-decor-vases". Products get a unique Title during listing which contains vital keywords describing their

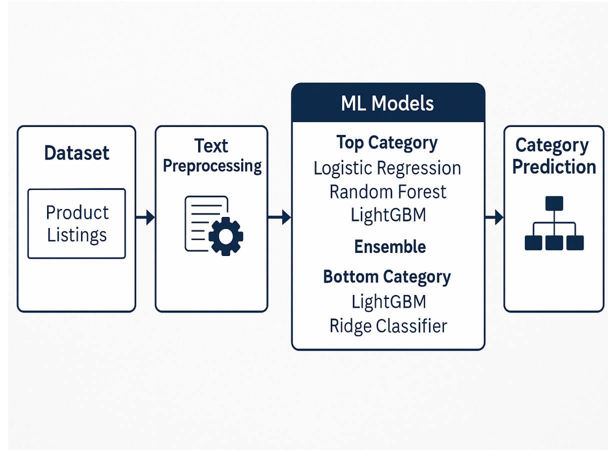

Fig. 1. Workflow of the Machine Learning Pipeline for E-commerce Product Category Classification.

item. The short summary of product attributes and functions exists in the Description field. Products are tagged with brief keywords or phrases named Tags which help users carry out search filters and discovery actions.

A total of 229624 training examples along with 13587 testing examples make up the dataset where each example contains a single product listing. Machine learning predicts suitable top and bottom classifications of new products through title and description analysis in combination with tags evaluation. The machine learning models are parameterized by examining the training data before evaluating the models using the test data to determine their success in classifying new products.

Classification is difficult since it has many options for categorizing and hierarchical label forms and partially unstructured fields of data. Serves as a foundation for multiclass model evaluation and natural language processing usage, making the task valuable.

B. Exploratory Data Analysis

The process began with exploratory data analysis to understand both the quality and structure of the dataset. A set of summary computations in the form of count, mean, standard deviation and value distributions were computed for both numerical and categorical data types. This helps to gain significant information about data distributions together with major statistical classes and outliers in the data. Data type inspection enabled recognition of numerical and categorical and textual variables and boxplots and histograms provided visualization tools that helped detection of numeric column variability and skewness and potential outliers.

C. Data Pre-processing and Feature Engineering

A preprocessing pipeline must run its complete process to prepare a robust machine learning dataset implementation. Data preprocessing prepares information so that it meets both performance requirements and establishes structure to better optimize model effectiveness. The research utilized exploratory data analysis (EDA) as its first step followed by missing value management and outlier identification after which label encoding took place before employing TF-IDF for text vectorization alongside SMOTE to handle the class imbalance.

1) Handling Missing Data: Text-heavy datasets become challenging when dealing with missing values because they deteriorate both model accuracy and reliability. The first step involved the evaluation of missing (NaN) values and blank strings in all columns. The empty string replaced all empty values present in textual fields which included titles along with descriptions and tags. Text vectorization remains error-free and data structure intact after the implementation of this replacement method. When dealing with numerical fields the median value of each column served as the imputation method because it showed higher resilience to outliers than mean imputation.

2) Label Encoding: The encoding is required since most machine learning algorithms are able to handle numerical input only, and they do not understand categorical strings. After converting category labels into integers, the models can properly learn class-specific patterns and compute evaluation metrics such as accuracy and F1-score for top and bottom category predictions.

3) Outlier Detection: To imporve model stability and avoid distortion in learning, outlier detection was performed using Z-score normalization on numerical columns. With the help of Z-score greater than 3 in any numerical feature were considered outliers and removed. Even though the classification depends mostly on textual data features in this scenario the data cleaning process provided benefits for potential application development through enhanced feature engineering or multi-modal extensions of both top and bottom prediction categories.

4) Text Feature Engineering with TF-IDF: The title field was selected as a primary classification feature because it presented availability and importance in all records. TF-IDF transformed the title field text through its Term Frequency–Inverse Document Frequency vectorization method which generated numerical vectors that maintained important word values. The application of 5,000 most informative terms served as a vocabulary restriction to decrease data dimensions. Separate TF-IDF representations were used to develop different models for detecting top-level and bottom-level category IDs so the models could discover separate yet linked patterns.

5) Handling Class Imbalance in Dataset: The category labels top and bottom had lot of imbalance data that can create bias in proper model prediction. The training data received oversampling through the Synthetic Minority Over-sampling Technique (SMOTE) for underrepresented classes. SMOTE delivered performance enhancement by producing balanced class distributions following TF-IDF processing for the top and bottom category estimates.

6) Dataset Splitting: The dataset was initially provided with a predefined split into training and test sets. During development the preprocessed training data received an additional split into training and validation sets through stratified sampling methods. Both sets maintain their original label distributions for top-level and bottom-level categories which enables models to develop comprehensive knowledge of the whole taxonomy through training and evaluation.

Computational constraints and the large number of bottom categories necessitated reduction of data when analyzing this level. Researchers employed random sampling using the resample function on 10% of training data to cut down training time without altering class distribution balance. The downsampling enabled the model to possess fast training and prediction abilities while supporting regular performance profiles in identifying bottom-level categories.

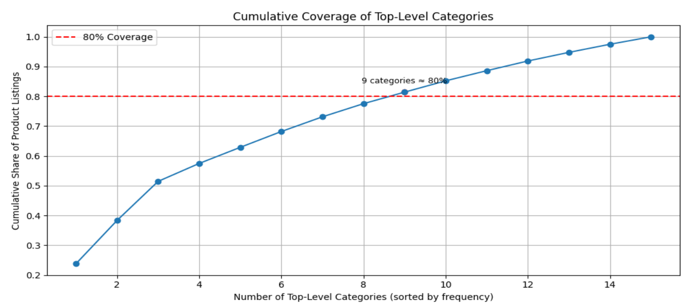

Fig. 2. Cumulative distribution plot of product listings with most frequent categories.

D. Methods Overview

This issue regarding product placement between top and bottom categories required evaluation of multiple machine learning methods. The classification assignment received supervised multi-class classification treatment because it required different models for top-level and bottom-level categories. The models processed TF-IDF text features from product titles which received LabelEncoder for numerical encoding of the corresponding labels.

The F1-score evaluation metric served as the assessment method for each model since it proves effective in multi-class classification tasks. The F1-score demonstrates superior evaluation capabilities in imbalanced datasets because it analyzes both class precision and recall during assessments.

1) Logistic Regression: The selection of Logistic Regression for baseline modeling stems from its suitable characteristics which include practicality along with universal compatibility with high-dimensional sparse features like TF-IDF. The algorithm runs linear classification through a sigmoid function which enables multi-class classification through its one vs rest scheme. In this project Logistic Regression with SMOTE-balanced dataset is used because it aimed to solve problems with class imbalance when classifying top-level categories. The model's performance was evaluated on the validation set using product titles transformed through TF-IDF, yielding a weighted F1-score of 0.78 Logistic Regression demonstrates reliable performance specially when combined with text vectorization and resampling approaches in the model.

2) Random Forest Classifier: The Random Forest model applies ensemble learning through decision trees aggregation to achieve better generalization rate and minimize overfitting effects. The model contained 100 estimators and used SMOTE to handle class imbalance while processing TF-IDF-transformed product titles. The model achieved a weighted F1-score of 0.79 on top-level category evaluation of the validation set which indicates its ability to effectively deal with high-dimensional sparse data while detecting complex feature associations. The obtained outcome validates Random Forest as an effective method to produce stable and consistent prediction outcomes in a setting with imbalanced multi-class classification.

3) LightGBM Classifier: LightGBM is an open-source gradient boosting framework with engineered features for fast processing that was employed in classifying top-level product categories. The model implemented SMOTE-balanced TF-IDF representations of product titles for training on its large and high-dimensional dataset. The model applies a histogram-based algorithm as its main processing mechanic which enables fast training in conjunction with reduced memory demands suitable for big multi-class applications.

The validation scores showed LightGBM obtained 0.77 weighted F1-score with faster training speed than XGBoost and alternative ensemble models. The performance evaluation indicates that LightGBM provides an appropriate scalable solution which handles text-based classification operations with imbalanced data classes.

4) Ensemble Model for Top-Level Category Classification: A combination learning technique was used for top-level product category prediction because it delivered maximum accuracy rates and overcame single algorithm constraints. Three different classifiers namely Logistic Regression, Random Forest, and LightGBM join together in a hard voting ensemble after training on the same set of features. Due to the majority voting procedure the ensemble model determines its final prediction after receiving results from each of the three component models thus reducing prediction error and improving the robustness of predictions.

The product titles received a TF-IDF transformation before being used to train three models which employed SMOTE for handling class imbalance. SMOTE technique during training guaranteed enough learning of rare categories in the dataset which led to balanced predictions throughout the top-category.

The ensemble received its training data from the validation set which underwent TF-IDF encoding combined with label-encoding for its targets. After evaluation, the model achieved an impressive weighted **F1-score of 0.91**, significantly outperforming individual models such as Logistic Regression (F1 = 0.78), Random Forest (F1 = 0.79) and LightGBM (0.77) individually. The ensemble demonstrates excellent performance by extracting intricate relationships from the textual metadata and handling the uneven distribution of classes.

To make it ready for deployment and ensure reproducibility the model was saved to disk through joblib. The ensemble possesses outstanding capabilities for real-world e-commerce taxonomy classification responsibilities because it performs well and scales properly with multiple classes.

5) LightGBM Classifier for Bottom-Level Category Classification: To address the fine-grained product listing categorization to bottom-level categories, the LightGBM model was employed because of its effectiveness and ability to handle large sparse data. The bottom-up model took the same vectorization procedures from top-level classification with the use of TF-IDF techniques and SMOTE balancing techniques in the training set.

Training at the bottom-level demonstrates higher complexity since there are many categories which frequently present imbalanced distributions. The picture-intensive calculations forced researchers to select 10% of training data for both model training and predictive distribution purposes. This selection provided a real-time solution window. The model managed to deliver good results even though it operated with reduced training data.

The reduction of training data to 10% did not inhibit the model performance since it scored a weighted F1 Score at 0.70.

The hierarchical classification capabilities of LightGBM proved to be effective for large and detailed classification systems along with its ability to handle bottom-category predictions.

6) Ridge Classifier for Bottom-Level Category Classification: A Ridge Classifier served as the method to handle the high-dimensional sparse data extracted from TF-IDF representations for bottom-level category prediction. The Ridge Classifier uses L2 regularization in its linear structure to prevent overfitting and manage multicollinearity features that natural text classification typically contains.

Class-weight was used to adjust training weights properly for minority bottom categories since the data contained an uneven class distribution. This system processed the TF-IDF-transformed titles within the SMOTE-balanced dataset through training until it obtained results based on weighted F1-score evaluation.

The Ridge Classifier achieved high performance in fine-grained product categorization through its weighted F1 Score value of 0.75 that indicated success in dealing with imbalanced and multi-class tasks. The Ridge Classifier performed so well that researchers use it as an effective and understandable baseline model to evaluate complex ensemble models.

IV. EVALUATION

The classification task was divided into two hierarchical levels top-level and bottom-level categories. Multiple machine learning models are evaluated using the weighted F1-score to deal with unequal distribution of classes in both situations. The

TF-IDF vectorization algorithm processed product titles before the training datasets received balance processing through SMOTE.

The top category prediction models acquired evaluation from three techniques including Logistic Regression, Random Forest and LightGBM. All models were trained on TF-IDF-transformed data and evaluated on a stratified validation set. The Voting Ensemble model consisting of Logistic Regression and Random Forest and LightGBM reached the best detection results with a weighted F1-score at 0.91. This combination of classifiers applied linear and non-linear characteristics effectively protected against mistakes in dominant as well as minority class predictions. The individual models reached slightly lower scores than the ensemble yet maintained satisfactory baseline metrics among each other. LightGBM and Logistic Regression showed such performance levels.

The bottom-level category classification task required increased complexity because of its large-scale number of classes combined with severe class unbalance issues. The LightGBM model operated with a 10% reduced downsampled portion of training data that was SMOTE-balanced. The LightGBM model reached a weighted F1-score of 0.70 while providing fast execution time to generalize its predictions effectively. A reduction in the training dataset constrained the model from properly recognizing subtle differences between infrequent categories. Execution of the Ridge Classifier relied on training it using SMOTE-balanced data while downsampling the input was omitted. The linear model succeeded in handling high-dimensional sparse data to deliver consistent results across all experiments with a weighted F1-score reaching 0.75. The model demonstrated reliable predictions for each class together with effective performance in the detailed classification of the bottom category. Based on the results the final models for each classification task were selected with careful consideration of both predictive performance and high F1-score. For top-level category prediction, the best results were obtained using a Voting Ensemble that combined Logistic Regression, Random Forest, and LightGBM classifiers. Along with for bottom-level category prediction, the Ridge Classifier gave great results.

TABLE I
EVALUATION RESULTS FOR TOP AND BOTTOM CATEGORY CLASSIFICATION

Model	F1-Score (Weighted)
Top-Level Category	
Logistic Regression	0.78
Random Forest	0.79
LightGBM	0.77
Ensemble (LR + RF + LGBM)	**0.91**
Bottom-Level Category	
LightGBM	0.70
Ridge Classifier	0.75

V. LIMITATIONS AND FUTURE WORK

The positive outcomes of this research were faced with a number of challenges in the implementation of research. Workplace environment network issues were among the major challenges to smooth data processing. Inadequate resources during Google Colab sessions impeded modeling of XGBoost and CatBoost and deep learning-based designs using the full dataset Downsampling operations were used to preserve session stability yet these steps introduced a model performance risk by potentially causing a reduced generalization ability when applied to less common categories. The dataset included broken and fragmented text segments that could affect both feature analysis capabilities and classifying performance. Hyperparameter tuning itself was also constrained by system resources as well as run time. Although an attempt was made to balance the between accuracy and feasibility, some advanced techniques tuning techniques and ensemble techniques could not achieve. The future research should tackle these difficulties through high-performance computing capabilities for enabling comprehensive training of advanced algorithms. Using BERT transformer models along with text modification techniques and AutoML libraries would achieve maximum model outcomes. Combining seller reputation metadata and product tags into the feature set would give more accurate and specific classification outcomes. System enhancements would incorporate scalability as well as make it interpretable with production-ready features for e-commerce websites.

VI. CONCLUSION

In this study, here a multilevel machine learning pipeline was created to classify products into their corresponding top and bottom level categories according to the dataset. In this approach, different text preprocessing techniques like TF-IDF vectorization and SMOTE based class balancing were utilized to preprocess high-dimensional and imbalanced text data for effective classification. There are various machine learning models employed for both category prediction tasks. For top level category classification, a Voting Ensemble of Logistic Regression, Random Forest, and LightGBM achieved highest weighted F1-score of 0.91 showing better generalization on frequent as well as rare classes. For bottom level category classification, which was a difficult task but the Ridge Classifier performed exceptionally well with a F1-score of 0.75. The experiment confirms that ensemble algorithms work very well on coarse-grained classification, but linear models with regularization like Ridge are suitable for fine-grained and high-dimensional tasks. This work demonstrates the impact of meticulous preprocessing, model selection, and ensemble modeling on hierarchical text classification problems and lays the ground for further improvements with transformer-based or deep learning models in future research.

REFERENCES

[1] Joy, J., & Selvan, M. P. (2022). A comprehensive study on the performance of different multi-class classification algorithms and hyperparameter tuning techniques using Optuna. 2022 International Conference on

Computing, Communication, Security and Intelligent Systems (IC3SIS). IEEE.

[2] Kanaan, A. G., Wan Hamzah, W. M. A. F., Wahsheh, F. R., Pandey, B., El-Ebiary, Y. A. B., & Stenin, N. P. (2023). An Evaluation and Annotation Methodology for Product Category Matching in E-Commerce Using GPT. 2023 International Conference on Computer Science and Emerging Technologies (CSET). IEEE.

[3] Karakaya, A., Karakaya, İ., & Temizceri, T. (2023). An Online Shoppers Purchasing Intention Model Based on Ensemble Learning. 2023 4th International Informatics and Software Engineering Conference (IISEC), IEEE.

[4] Parashar, A., & Gupta, E. (2017). ANN Based Ranking Algorithm for Products on E-Commerce Website. In 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB) (pp. 420–425).

[5] Wen, M., Vasthimal, D. K., Lu, A., Wang, T., & Guo, A. (2019). Building Large-Scale Deep Learning System for Entity Recognition in E-Commerce Search. Proceedings of the IEEE/ACM 6th International Conference on Big Data Computing, Applications and Technologies (BDCAT), Auckland, New Zealand.

[6] Reddy, N. S. S., Rohith, V. V. A., Abhiram, P. S., Saran, M. D. S. R., & Rebecca, S. (2024). Enhancing Product Categorization in E-commerce using NLP and Machine Learning. In Proceedings of the 7th International Conference on Inventive Computation Technologies (ICICT), IEEE.

[7] Ha, J., Pyo, H., & Kim, J. (2016). Large-Scale Item Categorization in E-Commerce Using Multiple Recurrent Neural Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 107–115).

[8] Yu, W., Sun, Z., Liu, H., Li, Z., & Zheng, Z. (2018). Multi-level Deep Learning Based E-commerce Product Categorization. Proceedings of the SIGIR 2018 eCom Workshop, Ann Arbor, Michigan, USA.

[9] Shaikh, I. A. K., Akram, N., Gamini, P., Haris, A. A. A., Pilli, D., & Bhoopathy, V. (2024). Understanding E-consumer Online Behavior: Establishing E-commerce Performance Metrics through BERT-LightGBM Approach. 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), IEEE.

[10] Xiao, S., & Tong, W. (2021). Prediction of User Consumption Behavior Data Based on the Combined Model of TF-IDF and Logistic Regression. Journal of Physics: Conference Series, 1757(1), 012089.

[11] Micu, A., Geru, M., Capatina, A., Avram, C., Rusu, R., & Panait, A. A. (2019). Leveraging e-commerce performance through machine learning algorithms. Annals of “Dunarea de Jos” University of Galati, Fascicle I. Economics and Applied Informatics, 25(2)

[12] Uwais, Aiman Muhammad, and Hamidreza Khaleghzadeh. ”Loan default prediction using spark machine learning algorithms.” AIAI 29th Irish Conference on Artificial Intelligence and Cognitive Science. CEUR Workshop Proceedings, 2022.