Automating Product Classification in E-Commerce through Machine Learning and Natural Language Processing

P. Jayadharshini
*Department of Artificial Intelligence
Kongu Engineering College, Perundurai,
Erode, Tamil Nadu*
jayadharshini.ai@kongu.edu

Lalitha Krishnasamy
*Department of Artificial Intelligence and
Data Science, Nandha Engineering
College, Erode, Tamil Nadu*
vrklalitha24@gmail.com

Harshini.M
*Department of Artificial Intelligence
Kongu Engineering College, Perundurai,
Erode, Tamil Nadu*
harshinim.22aid@kongu.edu

Arunachalam.M
*Department of Artificial Intelligence
Kongu Engineering College, Perundurai,
Erode, Tamil Nadu*
arunachalamm.22aid@kongu.edu

Ajaykumar. N
*Department of Artificial Intelligence
Kongu Engineering College, Perundurai,
Erode, Tamil Nadu*
ajaykumarn.22aid@kongu.edu

Hari Sri Uvan.M
*Department of Artificial Intelligence
Kongu Engineering College, Perundurai,
Erode, Tamil Nadu*
harisriuvanm.22aid@kongu.edu

Abstract—An efficient product category classification has now emerged as a vital element of the rapidly developing world of e-commerce and online retail, providing better control over inventory and enabling customized presentations for each customer. This paper presents the investigation into using machine learning techniques to automatically classify products based on textual descriptions. Using a large, annotated dataset, authors collect information by extracting relevant information from product descriptions and use advanced methods in both machine learning and natural language processing in showing well the classification of categories of products. The study also includes the critically important data gathering, preprocessing, and feature extraction phases that are vital in harvesting subtle semantic-related information hidden in descriptions of products. The process involves careful selection and implementation of complex algorithms, such as Naive Bayes, Random Forest, and Linear SVM, best suited for dealing with product categorization-specific problems. Thus, training and validation procedures play a vital role in optimizing the model parameters so that the model can perform robustly on unseen data and hence generalize effectively to new products. The model, deployed on an e-commerce platform, provides real-time category prediction as new products are added into the system. It regularly updates and maintains itself so that any variations in the product catalog or changes in the structures of language used over time in describing products are incorporated within the model. It has addressed issues such as data incompleteness and inconsistency and therefore offers integrated methods that may lead to improving the accuracy of predictions. The system is flexible and adaptive and capable of learning and evolving over time, according to new categories that may arise and different contexts of operations. The results do reflect significant strides in creating large-scale, efficient machine learning solutions for e-commerce, automating the process of product category prediction and the further efficiency enhancement of online commerce platforms.

Keywords—*Product categorization, Machine Learning (ML) Algorithms, Naive Bayes (NB), Random Forest (RF), Linear Support Vector Classifier (SVC)*

I. INTRODUCTION

Online businesses spend most of their time, budgets, and energies arranging the inventory, understanding the behaviour of customers so that the marketing can be enhanced, and which products to sell. This study like to apply machine learning methods that help us categorize product categories, say "Electronics," as well as potential subcategories, such as "Printers." This would come in very handy when someone had a list of new products they wanted to sell and could automatically categorize them using some training data about previous products the companies owned and the corresponding classifications. The training data-for example, this study might want to create a more detailed structure if a company only sold electronic equipment. Based on discussions with a few small- to medium-sized businesses, assuming once the system gains adequate amounts of historical training data and an enriched set of classifications, then this study will require an accuracy range of 95%. Product category prediction is another machine learning task. This task is one of predictive modelling where a system learns how to assign products to a predefined set of products based on their features and descriptions. This is very helpful in many industries, such as e-commerce, to manage inventory, boost search and improve user experience. Product category prediction is a classification machine learning task where it predicts the product category to which a product belongs based on its name and description. Predictive modelling is about creating models based on the existing data and predicting consequences to better manage the problem or situation. To implement product category prediction, consider the following steps: Set up the data: In this first stage, prepare the data that you will use to build and test the machine learning model. You need to obtain the dataset (a list of items with numerous carefully categorised products). Use this information to train and evaluate the ML algorithm. Then comes the first stage of processing, which is called preliminary data where a lot of optimisation and homogenisation of data reigns in the form of data cleaning, missing values, and language modelling. One of the most fundamental of the steps in this process is model selection, i.e. deciding from a set of available methods between one (e.g. logistic regression in older niches, decision trees(DT), RF, SVC).The chosen model is then tuned and trained in the most ruthless possible way in a region of available data, i.e. adjusting the parameters of the model in such a way as to maximise the predictive capability and the efficiency of the model. 2015 seemed to be a pure number when it was created, but it now has a meaning. The next step in this chain of events is to find out how well the model developed

suits the task. There are criteria here which tell us how well tuned such methods as recall, accuracy, precision or F1 score give the final success of the model. It is also necessary to fine-tune these measures of effectiveness so that the developed model satisfies the given requirements. Finally, the ready-made (after training) model is installed in production and now they are field-tested and present to the world to predict anything that the model was trained on in an instant. The decision of the model becomes its own reality, and we run away with it. The model we obtain here is already integrated into an app or system and creates a product that goes on sale (e.g. Amazon platform). The product is placed in an e-commerce system. And that's it, distribution works. Regular care and maintenance is crucial for it to remain relevant in the long run. This involves evaluating the model's performance in real-life situations and modifying it as necessary to accommodate changes in the product catalog or changes in instructions. Many challenges and decisions, including inconsistent data, cross-referenced data, and changes to new products, increase the complexity of the process. Techniques such as oversampling, undersampling, and examining correlations can help to overcome these reaching. Increase efficiency by centralizing the deployment process, thus reducing manual work and time. Improve user experience with more accurate and personalized recommendations. The scalable system's ability to process dynamic products further enhances the overall success of this predictive model.

II. LITERATURE SURVEY

Applying combined system getting-to-know techniques, Anitha and Vimal Kumar (2023) introduce one possibility that can be used for the e-commerce product category, suggesting the requirement for accuracy and efficiency in the marketing process, specifically regarding the classification of products displayed on e-trading sites. In a study in power augmentation methods of image uploading, Jayadharshini et al. (2023) target improved situational awareness with the stimulation of developing successful methods for the transfer of photos toward a range of applications. Koksal et al. (2022) discuss the presentation of an all-inclusive natural language processing-based model for the categorization of an e-commerce product category. It is based on state-of-the-art techniques to successfully explore the product descriptions and to classify objects. The development learning techniques use the textual information in classifying product categories at a hierarchical level, and Jahanshahi et al. (2021) contribute towards providing a framework that predicts multi-level product categories based on these aspects. It is supportive of the series-to-collection hierarchical class technique by Hasson et al. (2021), which helps in achieving product class that is based on hierarchical structure and therefore, in e-commerce, helps in finding accurate categories. This helps to classify commodities very efficiently by using sequential data, hence creating a push in category recognition in e-commerce by the sequence-to-collection hierarchical class technique.

Kejriwal et al., for example, offer the progressive evaluation and annotation method in categorization matching for electronic commerce as a means of enhancing the precision and dependability of categorization methods on web-based applications. David and Boghossian carry out a detailed discussion on purpose-derived category generation and nesting of product categories to gain better insight into class construction mechanics and implications of such strategies for organizations. Vazquez (2021) analyze how the category of product and type of selected social media platform would impact the perceived quality of e-retail products. So, it would go on to unlock interesting views in relation to how platform selection would influence customer perceptions. Özdemir and Turanlı (2021) provides insight in the comparative analysis upon the performance and application of various algorithms associated with the forecasting of customer behavior. The related insights regard the knowledge by algorithms about a category of devices utilized for this purpose. Kim et al. 2021 found trends in statistics-aware deep learning for product classification that can aid in developing the strategies of artificial intelligence on accurate and efficient product classification in e-commerce activities. In the study of 2021, Abirami et al. focus their attention on the potential in human-computer interaction in public transport and provide important insights into the way in which HCI technology should be implemented into transport with a view to the quality of experience and function for the passengers.

To know how exceptional product categories influence consumers' perceptions of the provider lovely within the logistics industry, Cheon and Kim (2020) analyzed the impact of the product class on buying decisions into logistic provider quality. Deep hierarchical class, a class prediction strategy with a sophisticated mechanism that accurately classifies products into appropriate categories based on their abilities and qualities, was offered by Gao et al. (2020) for e-commerce structures. Lin et al. (2020) investigates the effects of product classes, brand alliance health, and character developments on customers' logo attitude and buy intents to understand the dynamics of brand perception and customer behavior. They accomplish this by concentrating on the Spotify case. Heistracher et al. (2020) classify methods of acquiring system knowledge into the description of products from the darknet platform; in any case, they add to the known body of knowledge related to how high-level algorithms can help in classification with regard to products in complex and unknown online environments. Allweyer et al. (2020) suggests a way to classify the products in retail by using, in part abbreviated product names most effective indicating that this is a reasonable method to classify products efficiently despite that less statistics might be provided within the product names. Offer a grocery product categorization and recommendation system by Hanumantha Raju and Murthy (2020) using the integration of machine learning and customer profile identification, which can give personalized suggestions based on user preferences. Chen et al. (2019) proposes techniques to classify products with fine granularity for coarse-grained product classification in e-commerce and enhance the browsing and searching experience for electronic buyers. Hang et al. (2019) applies gadget studying to percent type showing superior algorithms capability for categorizing and classifying items towards correct packing features. Kumar S et al. (2024) also advocates a technique for part category using supervised device gaining knowledge, leading to better efficiency and accuracy of parts categorization in manufacturing and engineering fields. Taneja et al. (2024) are of the view that open-world learning methods used in producing the product type provides new techniques to address the difficult situations of classifying products in dynamic and ever-changing environments with e-commerce portals.

In their study published in Scientific Reports, Sehito et al. (2024) investigate various machine learning methods for the extensive classification of products. Their research may investigate the use of machine learning algorithms to efficiently classify different types of objects at scale. The aim of the Muniyandi et al. study in 2024 is to achieve the overriding function of distribution of agricultural products by machine learning-based technology that can enhance the efficiency and accuracy in the agricultural process-sorted through this technique. The research by Abirami et al. (2023) is oriented toward the application of techniques of machine learning to analyze medical images with a view to detection of tuberculosis accurately at appropriate time for early diagnosis and treatment of this infectious disease.In the study of Santhiya et al. (2024) Glove Embedding has been used to identify the Hate and offensive speech whereas BERT has been used for preprocessing the offensive statements which identifies the each and every word individually. Lalitha et al. (2018) endorses the detailed explanation of hearing loss for Teenagers including its surgical and equipment costs. This study also compares the traditional and advanced equipment and its usage level among teenagers. Singh et al. (2024) explains the benchmark effectiveness of IIoT in real world application and in various authentication schemes.

Ravichandran et al. (2024) study the possible use of IoT, combined with data analytics, for creating smart agriculture practices in the field of digital agriculture. The paper points out the importance of acquiring and assessing data in real-time for better crop management, ideal resource allocation, and a more efficient overall farm. This integration of technology aims to solve agricultural problems while streamlining decision-making methods and increasing aspects of sustainability. Nagappan et al. (2023) have proposed a cluster-oriented context-sensitive routing service management framework especially for intelligent autonomous vehicles used for industrial transport systems. The methodologies make dynamic changes in routes through contextual information, thereby helping to reduce the delay time and overall performance efficiency in operational efficiencies. This is about optimizing logistics in forms of industrial and smart network transportation modes. Jayadharshini et al. (2024) propose a thorough machine learning framework designed for the prompt identification and categorization of both hyperthyroidism and hypothyroidism. The study underscores the potential of machine learning methodologies to improve diagnostic precision and facilitate the early recognition of these conditions, thereby enabling timely medical responses. The suggested approach seeks to diminish instances of misdiagnosis and enhance patient results by ensuring accurate classification of thyroid-related ailments.

III. METHODOLOGY

A. Dataset Description

The data collection module is a very basic module that provides an entry point into the whole project. Basically, it has to do with gathering the correct set of data that should be used for the operation. The dataset used in order to predict the categories of the product needs to be filtered according to several criteria. Another point is the data collection process enhanced the dataset through extra external data. There are nearly twenty thousand data records within the dataset. The features that occur in this collection have a distinctive id, product label, category of product, and the short description of the products. It also requires suitable preprocessing for the information, which involves missing value handling and maintaining equal distribution of categories to avoid biases in the model. We would be using the product sales information from the online marketplace - Flipkart. To use the Kaggle dataset, we would apply the model provided to us with information to predict properly.

A. Data Preprocessing

Preparing a dataset for use in training machine learning models is a very important preprocessing step. The main aim of preprocessing is to process the raw data to make it workable for training models, which can enhance the performance of the latter. One has also got to be quite careful while working with data and try out different pre-processing techniques on different data sets so that one can identify the best technique which best suits the category of products that will be predicted in the product category prediction. For this project, the prepare data function has been used, which cleaned both the labels as well as the textual description. Besides negation words, which are optional, email addresses, newline characters, annoying single quotes, numbers, punctuations, single characters, words with accents, and multiple spaces are also taken off from stop words.

Despite the fact that the removal of lemmatization and stop words has empirically been proven to improve performance, it has been performed. The dataset has been divided into training and testing sets, features of all the sets are being extracted using different feature engineering approaches.

C. Feature Engineering

Indeed, it's one of the factors that feed into machine learning, basically processing unprocessed data into a form which captures much more of the underlying patterns in the model. Feature engineering (FE) basically is crucial for machine learning; this includes generating new features or altering old ones toward optimizing the model's performance. Feature engineering is the quality required to achieve maximal output of any model in machine learning. The ML models do not work with textual data. Hence, there is textual data to be formulated into a number. In FE, there are seven methods for converting text-based specifications into numerical attributes. The methods consist of: 1. Unigrams with simply word frequency representation. 2. Unigrams and bigrams shown using word frequency. 3. The depiction of unigrams, bigrams, and trigrams using word frequency. 4. Words represented solely as unigrams using TF-IDF. 5. Word TF-IDF-based bigram and unigram representation. 6. Word representation of bigrams, trigrams, and unigrams using TF-IDF. 7. Character TF-IDF-based representation of bigrams and trigrams. We are utilizing TF-IDF, bigrams, and unigrams in this research.

IV. MODEL DESCRIPTION

A. Random Forest

When employing classification by category of products, it is reasonable to use a Random Forest model. The algorithm is ideally created during the learning process by creating many decision trees and merging their predictions to come up with a more reliable and accurate distribution. Items and information are identified. In such data that has

annotations, various techniques can be used in annotations, including tokenization and vectorization. Then, training process as well as testing process data are generated, and the random forest distribution is trained over the previous process. Such a procedure makes the process more efficient, and one can explore hyperparameter tuning to optimize the performance of the random forests classified according to the uniqueness of the data nature. The score for accuracy gives an overall measure of the model accuracy. The Table I describes the performance evaluation of a RF model with CV, BCV, and NCV as feature extraction techniques. All these feature extraction techniques have been evaluated against three key metrics - Precision, Recall, and F1-Score.

TABLE I. PERFORMANCE EVALUATION OF RANDOM FOREST

Features	Precision	Recall	F1-Score
Count Vectors(CV)	0.87	0.78	0.82
Bigram Count Vectors(BCV)	0.99	0.89	0.93
Ngram Count Vectors (NCV)	0.98	0.98	0.99
Average(Avg)			0.97
Macro Avg	0.94	0.88	0.91

B. Naive Bayes

In machine learning for product classification, The Naïve Bayes algorithm has proven to be a helpful tool, especially appropriate for information-based tasks such as product descriptions. Regular use must be taken with careful arranging of the file into "Text" and "Category" columns; "Text" would represent the description of items and "category" represents their category. Changing the number of features vectors to enable more effective naïve Bayes classifiers. The multinomial naïve Bayes classifier was chosen because it is suitable for performing classification tasks.

The testing phase is one for making predictions based on the patterns learned from the training data. Accuracy, precision, and recall are all measures of a model's performance during its testing phase.

This iterative process ensures the classifier will generalize well to unseen data, making it reliable for real-world product categorization tasks. Model evaluation includes important indicators such as data accuracy and distribution, as shown in Table II.

C. Linear SVC

SVM is a supervised learning algorithm in the type of machine, especially for classification and regression tasks. In linear classification, SVM finds the best possible hyperplane which maximally separates points of other classes.This means they look for the line (in 2D) or plane (in higher dimensions) that best divides the data into distinct groups.

Linear SVC is a specific type of SVM that uses a linear kernel. It is especially useful for classifying items into classes and finding patterns for data that is highly dimensional.

TABLE II. PERFORMANCE EVALUATION OF NAÏVE BAYES

Features	Precision	Recall	F1-Score
CV	1.00	0.97	0.98
BCV	0.98	0.97	0.97
NCV	0.94	0.97	0.95
Avg			0.97
Macro Avg	0.97	0.97	0.96

The classifier is trained on a dataset, and then later, it makes predictions on new data. In the evaluation of the model performance, the use of measurements like accuracy has occurred.The accuracy score provides a broad evaluation of the model's accuracy, as indicated in Table III.

TABLE III. PERFORMANCE EVALUATION OF LINEAR SVC

Features	Precision	Recall	F1-score
CV	0.99	0.99	0.99
BCV	1.00	0.97	0.98
NCV	0.99	1.00	0.81
Avg			0.99
Macro Avg	0.99	0.99	0.99

D. Decision Tree

The most common method of implementing data classification for classifiers is the decision tree classifier. Many scholars from all walks of life, such as statistics, machine learning, and pattern recognition, have challenged themselves to develop a decision tree from accessible data. Probably, the most widely used method of data classification is a decision tree (DT) classifier. It's simple, very interpretable, and very effective in a wide range of domains. A DT is a method whose purpose is to break down any dataset into smaller subsets at the same time incrementally developing an associated tree. Any decision tree has internal nodes which are supposed to represent the decision that is made based on some given feature, and leaf nodes are supposed to represent an outcome or category. Many applications have been proposed for it, including classification of texts, recognition of images, mobile user categorization, medical disease evaluation, and even more. The experimental evaluation of decision trees presented in the paper is elaborated in Table IV.

E. Logistic Regreesion

For analyzing and classifying binary as well as proportionate response data sets, one of the most important statistical and datamining methods which statisticians and researchers have used for years comes in the form of logistic

regression (LR). The model extends to multi-class classification issues and has very important advantages in producing probabilistic output. Other benefits are that, as Table V explains, most model analysis techniques follow the principles of linear regression.

TABLE IV. PERFORMANCE EVALUATION OF DECISION TREE

Features	Precision	Recall	F1-score
CV	0.98	0.87	0.92
BCV	0.94	0.98	0.96
NCV	0.95	1.00	0.97
Avg			0.99
Macro Avg	0.95	0.95	0.95

TABLE V: PERFORMANCE EVALUATION OF LOGISTIC REGRESSION

Features	Precision	Recall	F1-Score
Count Vectors	0.88	0.95	0.96
Bigram Count Vectors	1.00	0.97	0.98
Ngram Count Vectors	0.92	0.97	0.81
Average			0.99
Macro Avg	0.93	0.96	0.91

TABLE VI. COMPARISON OF PROPOSED MODELS

Algorithm	Accuracy
Random Forest	97
Naïve Bayes	96
Linear SVC	99
Decision Tree	95
Logistic Regression	94

V. RESULT AND DISCUSSION

The powerful method that is used for the classification of products is by leveraging the machine learning algorithms in order to classify products through their unique feature identification. This approach not only simplifies the organization and management of multiple kinds of products but also significantly enhances operational efficiency and improves user experience overall. Efficient categorization of products helps manage large-scale inventories regarding an e-commerce platform while making sure that the customers view the right products quickly. Some machine learning models have been used in this project, including Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), Linear Support Vector Classification (Linear SVC), and deterministic approach. Each of the models above has its own good performance, though the best performance is achieved with the Linear SVC model. In this project, Linear SVC came out to be the most efficient model for predicting

a product category, being a variation of SVM. It's good with high-dimensional data and can very well differentiate between different product categories by separating them using a hyperplane. This ability to deal with large feature sets and work efficiently with text-based data propels Linear SVC as an appropriate model for the task at hand. The Linear SVC model has already been used very successfully in most the classification problem domains and, within this project, high-accuracy predictions were based on high-frequency components of product descriptions. This meant that the model could catch nuances that other models miss. In the evaluation of the best performance achieved, accuracy, precision, recall, and the F1 score were used. All categories in this case were correctly classified by Linear SVC as it had a great performance across all of them. All the product categories were nearly perfect for those metrics, which may be interpreted as that the model made an extremely confident prediction for the whole product dataset. For precision, it is the measure of how many of the predicted categories of product are correct. Recall measures how many of the actual categories of products are correctly identified, and also the F1 score is a balance between precision and recall. Hence, it also turns out to be a perfect device for massifying the products' classification in the extensive e-commerce systems to make this process bring user experience to an optimal and smooth state. Table VI shows the comparison of the proposed models.

VI. CONCLUSION

Probably, the case with machine learning-based product prediction in e-commerce and retail is the application of the correct algorithms and methods that minimize errors. The use of several well-known models such as NB, RF, and Linear SVC proves to be effective under the right influence of proper data processing, like handling the product description. Naïve Bayes, especially the polynomial version, performs well on text classification tasks with the aid of vectorization techniques like CV, TF-IDF and n-grams, which makes it robust for product descriptions classification. Also, Random Forest performs well as a method of ensemble learning with all data mixed in, complex relationships, strong accuracy, and resistance to overfitting. The accuracy of linear SVC-high dimensional space performance has been acknowledged, as it classifies and offers reliable metrics in particular, accuracy, precision, recall, and F1 score. Deployment reports guide decisions by revealing performance patterns across products and help fine-tune the model. Maintenance, as well as improvement, of the model is of prime necessity, which will further ensure accuracy and efficiency so that business can make data-driven decisions to better forecast and optimize products for sale.Besides the choice of a suitable algorithm, models also require frequent updating and refinement as new data and market trends come on the surface. As product offerings and customer behaviors change, they require models to update by retraining on new data to keep them still highly accurate. This will, among other things, be facilitated by techniques such as hyperparameter tuning, feature engineering, and cross-validation to improve model performance. These models can also be integrated in scalable systems that can deliver real-time product recommendations and demand forecasting for businesses. Companies will, therefore, be best positioned to face better competition from the facts indicated above where customer demands can be met better,

133

and the inventory and sales strategies can be optimized on the whole.

REFERENCES

[1] R. Anitha and D. V. Kumar, "E-commerce product classification using fused machine learning," SJIS-P, vol. 35, no. 1, pp. 795-808, 2023.

[2] P. Jayadharshini, T. Madhuvanthi, T. Abirami, and P. Thangaraj, "Energy enhancement of image sharing for situation awareness," *in AIP Conference Proceedings*, vol. 2901, no. 1, AIP Publishing, 2023.

[3] D. Koksal, M. M. Alacan, E. Olgun, and C. O. Sakar, "Natural language processing-based product category classification model for e-commerce," *in 2022 30th Signal Processing and Communications Applications Conference (SIU)*, 2022, pp. 1-4, IEEE.

[4] [4] H. Jahanshahi, O. Ozyegen, M. Cevik, B. Bulut, D. Yigit, F. F. Gonen, and A. Başar, "Text classification for predicting multilevel product categories," *arXiv preprint arXiv:2109.01084*, 2021.

[5] [5] I. Hasson, S. Novgorodov, G. Fuchs, and Y. Acriche, "Category recognition in e-commerce using sequence-to-sequence hierarchical classification," *in Proc. 14th ACM International Conference on Web Search and Data Mining*, pp. 902-905,2021.

[6] M. Kejriwal, K. Shen, Ni C.-C., and N. Torzec, "An evaluation and annotation methodology for product category matching in e-commerce," *Computers in Industry*, vol. 131, p. 103497, 2021.

[7] [7] J. Boghossian and David R.J.,"Under the umbrella: Goal-derived category construction and product category nesting," *Administrative science quarterly*, vol. 66, no. 4, pp. 1084-1129, 2021.

[8] [8] Vazquez, E.E. "Effect of an e-retailer's product category and social media platform selection on perceived quality of e-retail products," *Electronic Markets*, vol. 31, no. 1, pp. 139-157, 2021.

[9] [9] R. Özdemir and M. Turanlı, "Comparison of machine learning classification algorithms for purchasing forecast," *Journal of Life Economics*, vol. 8, no. 1, pp. 59-68, 2021.

[10] [10] Y. Kim, H. J. Lee, and J. Shim, "Developing dataconscious deep learning models for product classification," *Applied Sciences*, vol. 11, no. 12, p. 5694, 2021.

[11] [11] T. Abirami, K. Lalitha, P. Jayadharshini, and T. Madhuvanthi, "Future: HCI in public conveyances," *in AIP Conference Proceedings*, vol. 2387, no. 1, AIP Publishing, 2021.

[12] [12] O. Kim and H. J. Cheon, "Product category and shopping options of logistic service quality," *Journal of Distribution Science*, vol. 18, no. 8, pp. 113-125, 2020.

[13] [13] D. Gao, W. Yang, H. Zhou, Y. Wei, Y. Hu, and H. Wang, "Deep hierarchical classification for category prediction in e-commerce system," *arXiv preprint arXiv:2005.06692*, 2020.

[14] [14] Y. C. Lin, H. N. Tsai, and Y. C. Lee, "The effects of product categories, brand alliance fitness and personality traits on customer's brand attitude and purchase intentions: A case of Spotify," *Journal of Statistics and Management Systems*, vol. 23, no. 3, pp. 677-693, 2020.

[15] [15] C. Heistracher, F. Mignet, and S. Schlarb, "Machine learning techniques for the classification of product descriptions from darknet marketplaces," in *ICAI*, pp. 128-137,2020.

[16] [16] O. Allweyer, C. Schorr, R. Krieger, and A. Mohr, "Classification of products in retail using partially abbreviated product names only," *in DATA*, pp. 67–77,2020.

[17] [17] R. HanumanthaRaju and T. N. Murthy, "Grocery product classification and recommendation system based on machine learning and customer profile identity," *in ICDSMLA 2019: Proceedings of the 1st International Conference on Data Science, Machine Learning and Applications*, Springer Singapore, pp. 199–210,2020.

[18] [18] H. Chen, J. Zhao, and D. Yin, "Fine-grained product categorization in e-commerce," *in Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2349–2352,2019.

[19] [19] W. Hang and T. Banks, "Machine learning applied to pack classification," *International Journal of Market Research*, vol. 61, no. 6, pp. 601–620, 2019.

[20] [20] S. A. Kumar, M. K. Muchahari, S. Poonkuntran, L. S. Kumar, R. K. Dhanaraj, and P. Karthikeyan, "Application of hybrid capsule network model for malaria parasite detection on microscopic blood smear images," *In Multimedia Tools Applications*, pp. 1–27, 2024.

[21] [21] A. Taneja, S. Rani, R. K. Dhanaraj, and L. Nkenyereye, "GCIRM: Towards green communication with intelligent resource management scheme for radio access networks," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 3, pp.1018-1025, 2024, doi: 10.1109/TGCN. 2024. 3384542.

[22] [22] Sehito, N., Shouyi, Y., Alshahrani, H.M., Alamgeer, M., Dutta, A.K., Alsubai, S., Nkenyereye, L. and Dhanaraj, R.K., "Optimizing User Association, Power Control and Beamforming for 6G Multi-IRS Multi-UAV NOMA Communications in Smart Cities", *IEEE Transactions on Consumer Electronics*, vol. 70, no. 3, pp. 5702-5710, 2024, doi: 10.1109/TCE.2024.3388596.

[23] [23] A. P. Muniyandi, B. Balusamy, R. K. Dhanaraj, V. Ellappan, S. Murali, M. Sathyamoorthy, and L. Nkenyereye, "Privacy preserved reinforcement learning model using generative AI for personalized e-learning," *IEEE Transactions on Consumer Electronics*, 2024, https://doi.org/10.1109/tce.2024.3398824. 2024.

[24] [24] T. Abirami, S. Mapari, P. Jayadharshini, L. Krishnasamy, and R. R. Vigneshwaran, "Streamlined deployment and monitoring of cloud-native applications on AWS with Kubernetes, Prometheus, Grafana," in *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, pp. 1149–1155,2023.

[25] [25] S. Santhiya, S. Uma, N. Abinaya, P. Jayadharshini, S. Priyanka, and M. N. Dharshini, "A comparative exploration in text classification for hate speech and offensive language detection using BERT-based and GloVe embeddings," *in 2024 2nd International Conference on Disruptive Technologies (ICDT)*, pp. 1506–1509,2024.

[26] [26] K. Lalitha, S. Varadhaganapathy, S. Santhoshi, and D. R. Kumar, "A review on possibilities of hearing loss and implantable hearing devices for teenagers," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1–4,2018.

[27] [27] A. Singh, R. K. Dhanaraj, and A. K. Sharma, "Personalized device authentication scheme using Q-learning-based decision-making with the aid of transfer fuzzy learning for IIoT devices in zero trust network (PDA-QLTFL)," *Computers and Electrical Engineering*, vol. 118, p. 109435, 2024, doi:10. 1051/e3sconf/202454702003.

[28] [28] D. Ravichandran, S. Dhanabalan, A. Santhanakrishnan, S. Sarveshwaran, and R. Yogesh, "Digital agriculture: Harnessing IoT and data analytics for smart farming solutions," E3S Web of Conferences, vol. 547, p. 02003, 2024, doi:10.1051/e3sconf/ 202454702003.

[29] [29] G. Nagappan, K. G. Maheswari, C. Siva, and M. Shobana, "Cluster-based context-aware route service management for smart intelligent autonomous vehicles with industrial transport system," *International Journal of Communication Systems*, vol. 37, no. 5, p. e5682, 2024.

[30] [30] P. Jayadharshini, A. S. Kumar, R. Sivasankar, K. U. Harsine, P. Geetha, and B. Logu, "Comprehensive machine learning approach for early detection and classification of hyperthyroidism and hypothyroidism," in *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies*, pp. 1–7, 2024, DOI: 10.1109/TQCEBT59414.2024.10545151, 2024.