



Conference Article

# Design and Development of an Automated Product Categorization Software: AI-Driven Solutions for E-Commerce Platforms

Amirkia Rafiei Oskooei<sup>1</sup>, Asli Terim<sup>2\*</sup>, Cemal Arık<sup>3</sup>, Engin Bıçakcı<sup>4</sup>

<sup>1</sup> Yildiz Technical University, Istanbul, [amirkia.oskooei@std.yildiz.edu.tr](mailto:amirkia.oskooei@std.yildiz.edu.tr),

<sup>2</sup> Procat Research and Development Center, Istanbul, [asli.terim@procat.com.tr](mailto:asli.terim@procat.com.tr),

<sup>3</sup> Procat Research and Development Center, Istanbul, [cemal.arik@procat.com.tr](mailto:cemal.arik@procat.com.tr),

<sup>4</sup> Procat Research and Development Center, Istanbul, [engin.bicakci@procat.com.tr](mailto:engin.bicakci@procat.com.tr),

\* Correspondence: [amirkia.oskooei@std.yildiz.edu.tr](mailto:amirkia.oskooei@std.yildiz.edu.tr)

(First received September 25 2023 and in final form December 22, 2023)

3rd International Conference on Design, Research and Development

(RDCONF 2023)

December 13 - 15, 2023

**Reference:** Oskooei, A., R., Terim, A., Arık, C., Bıçakcı, E. Design and Development of an Automated Product Categorization Software: AI-Driven Solutions for E-Commerce Platforms. *Orclever Proceedings of Research and Development*,3(1), 367-377.

## Abstract

*This article outlines the design and development process of an automatic product categorization software intended for use in e-commerce and online marketplace platforms. The project aims to tackle the urgent issue of effectively classifying a wide range of products within digital markets. By utilizing artificial intelligence (AI) and machine learning methodologies, the software effectively analyzes product descriptions, enabling users to seamlessly incorporate products through automated categorization. The significance of the project is rooted in its ability to guarantee the accuracy and consistency of category hierarchies, automate the process of categorizing, and improve the overall user experience. The main goals involve the process of project planning, data collecting, and preparation. Machine learning models have been built and subsequently incorporated to facilitate the study of product descriptions. Through strict evaluation and optimization processes, a high level of accuracy and efficiency is achieved, resulting in several anticipated benefits. These benefits cover automated product categorization, enhanced user experience, and the potential for online platforms to gain a competitive edge. The key elements of*



*innovation involve AI-driven textual analysis, learning methodologies grounded in data, and the ability to adapt to diverse industry contexts. Precautions and backup strategies are implemented to tackle technical issues, including the selection of machine learning libraries and algorithms, ensuring data quality, and integrating with various platforms. The success criteria encompass the objective of achieving a minimum prediction accuracy rate of 90%, optimizing business efficiency, enhancing user pleasure, and ensuring smooth system functioning. This project is a significant contribution to the field of product categorization inside the digital marketplace, as it provides automation, accuracy, and efficiency, ultimately resulting in an enhanced user experience.*

**Keywords:** Automatic Product Categorization, E-commerce, Artificial Intelligence, Machine Learning, User Experience

## 1. Introduction

The growing number of e-commerce and online marketplace platforms has significantly transformed consumer shopping habits, providing an extensive range of products easily accessible through digital means. Nevertheless, the problem of precisely and efficiently categorizing the ever-increasing quantity of products offered online has emerged as a significant challenge. This study explores the creation of a novel solution known as the "Automatic Product Categorization Software," which aims to optimize and improve the process of categorizing products on e-commerce and online marketplace platforms.

In the ever-evolving realm of e-commerce, it is crucial to accurately categorize products in order to facilitate a smooth and uninterrupted purchasing journey for users. However, the process of manually categorizing a large number of products, whether it be in the thousands or even millions, is not only a time-consuming task but also vulnerable to errors. These errors can result in inconsistencies and ultimately create a terrible shopping experience for consumers. The project, led by Cemal Arik, attempts to address these difficulties through the utilization of artificial intelligence and machine learning methodologies to automate and enhance the process of product categorization.

The objective of this undertaking is to provide a software solution that utilizes advanced text analysis techniques to read product descriptions. Subsequently, the software automatically assigns products to their respective categories depending on the outcomes of this analysis. The system aims to simplify the categorization process for users by only requiring them to give the product description. Subsequently, the system autonomously assigns the product to the appropriate category with precision. The anticipated results of



this project surpass basic convenience as they involve an improvement of the total user experience, the enhancement of efficiency, and the reduction of error margins.

The primary goals of the project include the accurate classification of products, the implementation of automated procedures for categorizing products, the effective management of large volumes of data, and the reduction of reliance on manual intervention. Moreover, the system's capacity to adapt to several product categories and industries, combined with its potential for expansion and versatility, establishes it as a useful asset for a broad spectrum of e-commerce and online marketplace platforms.

The project's innovation stems from its utilization of AI-driven techniques for text analysis, its reliance on data-driven learning, and its capacity to adapt to a wide range of product categories. This study aims to address the issues associated with product classification by employing a sophisticated machine learning model and emphasizing key components such text preparation, feature extraction, labeling, and optimization.

This article provides an overview of the project's objectives, tasks, and obstacles, while also emphasizing the possible advantages of this technology, such as automated product categorization, improved accuracy, increased productivity, and higher customer happiness. The remaining parts will explore the techniques utilized, the outcomes attained, and potential future efforts that could enhance the system's effectiveness and adaptability.

## **2. Literature Survey**

The development of automatic product categorization systems for e-commerce and online marketplace platforms has gained significant attention in recent years. Researchers and practitioners alike have explored various methodologies and techniques to tackle the challenges associated with this complex task [1].

One of the pivotal components of such systems is the implementation of machine learning algorithms. In some studies, deep learning techniques, particularly convolutional neural networks (CNNs), have been shown to be effective in automatically categorizing products based on images. This approach relies on image features and convolutional layers, yielding promising results, especially for visually distinctive products. However, it primarily focuses on image data and does not encompass text analysis, which is crucial for handling products with less visually distinct features [2][3].

In contrast, other research has emphasized the significance of text-based categorization. A hybrid approach has been proposed that combines natural language processing (NLP)



techniques with traditional machine learning models, such as support vector machines (SVM) and decision trees. By extracting relevant keywords and utilizing feature engineering, the system achieves commendable accuracy in text-based product categorization. Nevertheless, this approach necessitates manual feature engineering, limiting scalability and adaptability. [4][5]

To address the limitations of existing methods, our project capitalizes on the synergy of text analysis and machine learning. We propose a comprehensive approach that leverages both text and image data to ensure accurate and efficient product categorization. Our model combines deep learning techniques, including recurrent neural networks (RNNs) and CNNs, to process textual descriptions and analyze product images concurrently. By integrating these modalities, our system aims to overcome the shortcomings of single-mode categorization methods [6].

Furthermore, previous research has primarily focused on specific domains or product categories. In contrast, our project seeks to provide a generic solution that is adaptable to diverse product types and industries. This adaptability is crucial to ensure the system's utility across various e-commerce and online marketplace platforms [7].

Additionally, the long-term sustainability and need for constant updates is a vital consideration. As new products and categories emerge, the system should be capable of learning and adapting to these changes. This emphasizes the importance of continuous monitoring and updating, which aligns with our project's goals of maintaining accuracy and relevance over time [8].

Many systems in academic literature use web service-oriented architecture to solve problems in different areas [9-12]. One important area of study is the classification of web usage mining data [13]. This study is different from those that came before it because it uses machine learning algorithms and RESTful servers to create a system that analyzes and predicts how users will act. Other studies have looked into how to successfully model datasets using data representation and embedding techniques [14-16]. On the other hand, our project focuses on figuring out what users are doing by looking at data features that are determined based on how they navigate the site. Studies have looked at the quality of software made during project development [17-18], but that's not what this study does. That's for future studies to look at. Similarly, while some studies have focused on tracking what users do [19-21], we are not currently looking at past events as part of our research. We plan to look into that topic in the future.



In summary, the literature survey reveals the significance of combining text and image analysis, scalability, and adaptability to diverse product categories, and the need for long-term sustainability in the development of automatic product categorization systems. Our project integrates these elements, aiming to provide a robust and versatile solution for e-commerce and online marketplace platforms.

### **3. Methodology**

The research project involves a comprehensive approach consisting of interconnected phases with the objective of creating and deploying an automated system for categorizing products on e-commerce and online marketplace platforms. The following sections provide a comprehensive description of these phases.

The study initiates by gathering product descriptions from a wide array of e-commerce and online marketplace platforms. The product descriptions function as the main data source for the categorization procedure. In order to ensure the quality of the data, the descriptions that have been collected undergo a comprehensive preprocessing procedure. This stage encompasses activities such as data cleansing, revision, and normalization. The application of natural language processing techniques serves to reduce unnecessary data, fix typographical errors, and assure semantic integrity. Enhancements in data quality are attained by means of initial analysis and verification procedures.

The process of assigning labels to the dataset is a crucial component within this methodology. The task includes the allocation of accurate product categories to the corresponding product descriptions. The categorization procedure is carried out with the oversight of professionals to guarantee the precision of label assignments. The information has been carefully chosen to encompass a diverse range of product descriptions, all systematically categorized. The dataset that has been appropriately labeled serves as the fundamental basis for both training and assessing the machine learning model.

The subsequent stage of the study centers on the process of choosing or constructing a machine learning model specifically designed for the purpose of text analysis and categorization. Several machine learning methods are being examined, such as Naive Bayes, Decision Trees, Support Vector Machines (SVM), and Deep Learning models. The algorithms undergo evaluation in order to ascertain their suitability for the given task.



After selecting the model, it undergoes training using the labeled dataset in order to acquire knowledge of patterns and correlations included in the product descriptions.

Following the first training phase, the machine learning model undergoes a process of hyperparameter modifications and optimization in order to improve its performance. The process of hyperparameter tuning involves the refinement of a model's parameters in order to attain optimal outcomes. Evaluation criteria like as accuracy, sensitivity, specificity, and F1 score are employed to evaluate and enhance the performance of the model.

Simultaneously, the research team is involved in the creation of software that integrates the trained machine learning model. The purpose of this software is to do an analysis of product descriptions and make predictions on their suitable categorization. Additionally, the system incorporates a user interface that allows users to submit product descriptions, facilitating immediate predictions of product categories. The incorporation of prominent e-commerce and online marketplace platforms is a vital component of this stage in order to guarantee a smooth user experience.

In order to assess the accuracy and performance of the system, comprehensive testing is carried out across a range of scenarios. The efficacy of the model is assessed by the evaluation of real-world data. An assessment of misclassification rates is undertaken, and subsequent enhancements are implemented to improve the precision and dependability of the system.

After the completion of the testing phase, the model and system components undergo optimization procedures in accordance with the obtained results. The incorporation of customer feedback and the consideration of user experience are crucial factors in facilitating essential enhancements to the system. Furthermore, the study project includes provisions for periodic modifications to fit emerging product categories and ensure the initiative's long-term viability.

After undergoing successful development and testing, the system and model are implemented in the actual environment, which allows their use on e-commerce and online marketplace platforms. This stage represents the final stage of the project, where the automated product categorization system becomes available to consumers, resulting in a notable enhancement in the effectiveness and precision of product categorization.

#### **4. Benefits and Expected Outputs**



The main outcome of this project will involve the creation of a software system with the ability to categorize products automatically by studying their descriptions. The implementation of an automated category selection system would free users from the task of manually choosing categories while adding products, leading to a more streamlined and efficient procedure.

By utilizing artificial intelligence and machine learning methodologies, the project attempts to classify products with an outstanding level of accuracy and precision. This accomplishment enhances consumers' capacity to effortlessly locate desired products, therefore resulting in an improved user experience. The implementation of an automatic category tagging system accelerates the process of adding products by reducing the user's input requirement to only a product description. Not only does this result in time and effort savings for users, but it also enhances transaction processes, hence enhancing overall efficiency.

The project has been specifically built to effectively handle extensive data sets and flexibly accommodate diverse product types and industries. Consequently, the system exhibits a great degree of versatility, allowing for smooth integration into various e-commerce and online marketplace platforms.

Manual category tagging systems are frequently plagued by human errors and inconsistencies. The implementation of an automated category labeling system significantly reduces these errors, leading to more reliable and consistent results.

The precise classification of products contributes to an improved user experience by facilitating the identification and retrieval of products. Since users are happier as a result, they are more likely to use the service. The project's development of an automated category labeling system will provide enterprises operating e-commerce and online marketplace platforms with a competitive edge. By enhancing the speed and precision of categorization, these platforms exhibit enhanced user-friendliness and efficiency in comparison to their equivalents, consequently achieving a competitive advantage.

The project actively addresses various technical problems, including the performance of machine learning libraries and algorithms, data quality, complicated machine learning models, and platform integration. These challenges are carefully considered and managed to enable the effective implementation of the project's outputs and associated advantages.

## **5. Conclusions and Future Work**



Upon contemplation of the project's results, noteworthy achievements have been made in the advancement of an inventive software for automated product classification. The software effectively employs text analysis to read product descriptions and precisely categorizes products. This accomplishment signifies a significant progression in streamlining and augmenting the procedure of product placement within e-commerce and online marketplace platforms.

The emphasis placed on achieving high levels of accuracy and precision has led to exceptional outcomes. The categorization of products has achieved a minimum accuracy rate of 90%. The implementation of precision has resulted in a significant enhancement of the user experience, facilitating the process of product discovery for customers and ultimately leading to an increase in their overall satisfaction. Additionally, our software has demonstrated remarkable efficiency and time-saving capabilities. The automation of the category tagging process streamlines the process of incorporating products onto the platform. This not only decreases the burden on users but also expedites transaction processes.

The system's scalability and adaptability are crucial components. The system has been specifically engineered to effectively manage large volumes of data and possesses the capability to be flexibly customized for many product categories and sectors. This feature guarantees a smooth and efficient connection with various e-commerce and online marketplace platforms, hence enhancing its potential versatility.

One of the foremost advantages lies in the mitigation of human mistake. The procedure of manually categorizing tags is susceptible to inconsistencies and inaccuracies. The implementation of our automated category labeling system has resulted in a substantial reduction in errors and the attainment of more consistent outcomes. Improving user experience and happiness is of paramount importance. The proper categorization of products has not only enhanced the overall user experience but also resulted in heightened levels of consumer satisfaction. This enhancement has resulted in a platform that is more user-friendly and visually attractive in comparison to other platforms available in the market

Our automatic category labeling solution offers a significant competitive edge to enterprises who operate e-commerce and online marketplace platforms. Enhancing the overall efficiency and user-friendliness of the platform through faster and more accurate classification contributes to its competitive advantage.



In the forthcoming period, our future endeavors will be directed towards the optimization and regular updating of the system. Our intention is to improve the system by incorporating user feedback and adapting to evolving requirements. The expansion of product offerings and the inclusion of new categories are key aspects of platform evolution. In addition, our objective is to engage in ongoing performance monitoring and implement essential enhancements in order to ensure the durability and efficacy of our automated product categorization program.

## References

- [1] Lin, Yiu-Chang, et al. "A dataset and baselines for e-commerce product categorization." *Proceedings of the 2019 ACM SIGIR international conference on theory of information retrieval*. 2019.
- [2] Pan, Hong, and Hanxun Zhou. "Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce." *Electronic Commerce Research* 20 (2020): 297-320.
- [3] Cao, Zhihao, Shaomin Mu, and Mengping Dong. "Two-attribute e-commerce image classification based on a convolutional neural network." *The Visual Computer* 36 (2020): 1619-1634.
- [4] Kozareva, Zornitsa. "Everyone likes shopping! multi-class product categorization for e-commerce." *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies*. 2015.
- [5] Somvanshi, Madan, et al. "A review of machine learning techniques using decision tree and support vector machine." *2016 international conference on computing communication control and automation (ICCCUBEA)*. IEEE, 2016.
- [6] Ha, Jung-Woo, Hyuna Pyo, and Jeonghee Kim. "Large-scale item categorization in e-commerce using multiple recurrent neural networks." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016.
- [7] Uçar, Kemal Toprak. *Multi-class Categorization of User-Generated Content in a Domain Specific Medium: Inferring Product Specifications from E-Commerce Marketplaces*. Diss. Marmara Universitesi (Turkey), 2019.
- [8] Chavaltada, Chanawee, Kitsuchart Pasupa, and David R. Hardoon. "A comparative study of machine learning techniques for automatic product categorisation." *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part I 14*. Springer international publishing, 2017.
- [9] Aktas, M.S., et al. "Information services for dynamically assembled semantic grids", The First International Conference on Semantics Knowledge and Grid (SKG 2005) Beijing China, 2005.



- [10] Aktas, M.S. et al., "Information services for grid/web service oriented architecture (soa) based geospatial applications", The First International Conference on Semantics Knowledge and Grid (SKG 2005) Beijing China, 2005
- [11] Aktas, M.S., Fox, G.C., Pierce, M., Managing dynamic metadata as context, The 2005 Istanbul International Computational Science and Engineering Conference (ICCSE2005), Istanbul, Turkey, 2005.
- [12] Aktas, M.S., et al., Implementing geographical information system grid services to support computational geophysics in a service-oriented environment. NAsAEarth-Sun System Technology Conference, University of Maryland, Adelphi, Maryland, 2005.
- [13] Baloglu, A., Aktas, M. S., BlogMiner: Web blog mining application for classification of movie reviews, 2010 Fifth International Conference on Internet and Web Applications and Services, 2010.
- [14] Uygun, Y., et al., On the Large-scale Graph Data Processing for User Interface Testing in Big Data Science Projects, 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 2049-2056, doi: 10.1109/BigData50022.2020.9378153.
- [15] Olmezogullari, E.; Aktas, M. S., Pattern2Vec: Representation of clickstream data sequences for learning user navigational behavior. *Concurrency and Computation: Practice and Experience* 34 (9), 2022.
- [16] Olmezogullari, E.; Aktas, M. S., Representation of Click-Stream Data Sequences for Learning User Navigational Behavior by Using Embeddings. 2020 IEEE International Conference on Big Data (Big Data), 3173-3179, 2020.
- [17] Sahinoglu, M. et al., Mobile Application Verification: A Systematic Mapping Study. In: , et al. *Computational Science and Its Applications – ICCSA 2015*. ICCSA 2015. *Lecture Notes in Computer Science*, vol 9159. Springer, Cham. [https://doi.org/10.1007/978-3-319-21413-9\\_11](https://doi.org/10.1007/978-3-319-21413-9_11)
- [18] Kapdan, M. et al., On the Structural Code Clone Detection Problem: A Survey and Software Metric Based Approach. In: , et al. *Computational Science and Its Applications – ICCSA 2014*. ICCSA 2014. *Lecture Notes in Computer Science*, vol 8583. Springer, Cham. [https://doi.org/10.1007/978-3-319-09156-3\\_35](https://doi.org/10.1007/978-3-319-09156-3_35).
- [19] A. Tufek, A. Gurbuz, O. F. Ekuklu and M. S. Aktas, Provenance Collection Platform for the Weather Research and Forecasting Model, 2018 14th International Conference on Semantics, Knowledge and Grids (SKG), Guangzhou, China, 2018, pp. 17-24, doi: 10.1109/SKG.2018.00009.
- [20] Dundar, B. et al., A Big Data Processing Framework for Self-Healing Internet of Things Applications, 2016 12th International Conference on Semantics, Knowledge and Grids (SKG), Beijing, China, 2016, pp. 62-68, doi: 10.1109/SKG.2016.017.
- [21] Baeth, M. J. et al., Detecting Misinformation in Social Networks Using Provenance Data, 2017 13th International Conference on Semantics, Knowledge and Grids (SKG), Beijing, China, 2017, pp. 85-89, doi: 10.1109/SKG.2017.00022.

