



HARMONIZED SYSTEM CODE RECOMMENDATION: A MULTI-CLASS CLASSIFICATION MODEL

Diyouva Christa Novith

Directorate General of Customs and Excise, Jakarta, Indonesia, Email: diyouva.christa@customs.go.id (penulis berkorespondensi)

ARTICLE INFORMATION

ARTICLE HISTORY

Received
02 September 2024

Accepted to be published
31 Desember 2024

KEYWORDS:

HS Code
Multi-Class Classification
Customs
Text Mining

ABSTRACT

Menilai keakuratan kode Harmonized System (HS) untuk barang secara manual rentan terhadap kesalahan dan memakan waktu. Keterbatasan petugas bea cukai yang mampu dengan pengetahuan yang memadai dalam mengklasifikasikan barang, serta peningkatan transaksi lintas batas karena konsep baru perdagangan elektronik membuat petugas bea cukai menghadapi tantangan besar untuk memenuhi tugasnya. Artikel ini mengeksplorasi pendekatan otomatis yang menggunakan algoritma pembelajaran mesin untuk membangun model yang paling sesuai untuk memverifikasi kebenaran deskripsi sesuai dengan Kode HS-nya. Proyek ini menggunakan data historis dari dokumen impor yang dideklarasikan dan penilaian resmi dari petugas bea cukai. Di antara model yang dieksplorasi, terbukti bahwa *Linear Support Vector Classification* memiliki akurasi tertinggi untuk mengklasifikasikan kode HS berdasarkan deskripsinya. Penemuan ini berkontribusi pada pengetahuan dengan memberikan solusi praktis yang memanfaatkan pembelajaran mesin untuk klasifikasi kode HS. Implikasi studi ini meluas ke peningkatan operasi bea cukai dan otomatisasi.

Evaluating the accuracy of Harmonized System (HS) code for goods manually is both error-prone and time-consuming. The limitation of capable customs officers with adequate knowledge in classifying goods, also the increase of cross-border transactions due to the new concept of e-commerce make customs officials face great challenges to meet its duty. This paper explored an automated approach employing machine learning algorithms to build the most suitable model to verify the correctness of description according to its HS Code. This project used historical data from declared import documents and official assessments from customs officers. Amongst the explored model, it is proved that Linear Support Vector Classification held the highest accuracy to classify HS Code based on its description. This findings contribute to the body of knowledge by providing a practical solution that leverages machine learning for HS code classification. The study's implications extend to enhancing customs operations and automation.

1. INTRODUCTION

The past decade has witnessed a remarkable transformation in how businesses operate, and consumers engage with the marketplace. The rapid growth of online transactions shifts the concept of consignment goods, which used to be non-commercial, to become dominated by profit-oriented activities. Digitalization has removed international borders and introduced electronic commerce, enabling easy trade in the electronic environment. The removal of borders in international trade has led to lower costs compared to traditional trade, and more importantly, it has made it easier to compete in global markets. This has introduced the concept of cross-border e-commerce. Cross-border e-commerce allows businesses to reach and establish a broader market presence by offering more affordable and efficient distribution channels for their products or services (Asian Development Bank, 2023). The introduction of Electronic Trade Customs Declarations within the scope of low-value export has further enhanced the competitiveness of enterprises in cross-border markets.

In 2022, 5 of the 10 fastest-growing countries in e-commerce transactions are in Southeast Asia, with Indonesia among the countries with online transactions increasing by more than 30% while average world growth is nearly 10% (Statista, 2022). Cross-border e-commerce is predicted to better meet the diversified needs of consumers (Liu & Wang, 2019) therefore it needs to effectively be promoted in the development of its regulation to support transaction fairness. While it holds opportunities in a prosperous economy, various risks are possibly emerging under the increase of cross-border e-commerce activities. The main problem businesses require is the need for more competent staff in customs procedures, especially related to customs duty and taxation (Saydam & Civelek, 2022; Yuan, 2023). On the other hand, customs authorities have the same issue, the need for more personnel to monitor and physically examine the consignment goods raises issues in security and revenue collection. To regulate, customs authorities use Harmonized System (HS) codes, which are widely used in the classification of goods during import and export processes. The Harmonized System is a standardized numerical method for categorizing traded products, enabling countries worldwide to uniformly identify and describe products for duties assessment and statistical purposes. However, it requires adequate knowledge for businesses to fulfill their duty. Beyond these

weaknesses, there are additional challenges in applying the HS Code to achieve desirable accuracy. These challenges encompass the complexity of the HS system, terminology gaps, and the evolving nature of the HS Code, among others.

The limitation of dedicated personnel and the huge number of transactions force customs authorities to explore new approaches. Utilizing technological advancement to give recommendations for HS Codes will be one of the most favorable actions. Developing a model specified to HS Code using artificial intelligence can enhance accuracy, reduce processing time, minimize errors, and be less costly in classifying commodities. The historical data can be used to train machine learning models to classify goods and maintain their consistency for statistical purposes, risk management, and other advanced analyses (Rukanova et al., 2019). Several previous studies have explored artificial intelligence for automating various aspects of customs operations, such as optimizing tariff classifications (Cuaya-Simbro et al., 2022), detecting fraud (Sembiring et al., 2023), and improving the overall efficiency of trade processes. These studies have demonstrated the potential of AI to handle large volumes of data and complex decision-making tasks, making it a promising tool for addressing the challenges faced by modern customs authorities. By building on these insights, this research aims to develop an AI-based model that not only streamlines the HS code classification process but also provides a scalable solution that can adapt to the challenge in classifying goods based on its description.

The purpose of this research is to address the complexities and gaps inherent in the current Harmonized System (HS) code classification process by developing a HS Code prediction using a multi-class classification model. This study is particularly focused on the context of Indonesia Customs, where the rapid increase in cross-border e-commerce has intensified the need for accurate and efficient customs procedures. By integrating Artificial Intelligence (AI) into the HS code system, the research aims to improve the accuracy of classifications, reduce processing times, and minimize errors associated with manual code assignments. The importance of this research lies in its potential to significantly enhance the capabilities of customs operations, thereby supporting the growing demands of global trade. By automating the classification process, the model is expected to alleviate the burden on customs personnel, who are often overwhelmed by the high volume of

transactions and provide a more reliable and consistent approach to managing international trade. Ultimately, this research could serve as a foundational step toward modernizing customs procedures, making them more resilient and adaptable to the evolving landscape of global commerce.

2. LITERATURE REVIEW

Harmonized System, introduced in 1988, is a standardized system of names and numbers used to classify traded products. It is managed by the World Customs Organization (WCO) and is adopted globally to facilitate international trade, collect trade statistics, and to apply tariff and non-tariff barriers. Since then, it has been adopted as an international code of cross-border trade by more than 200 countries around the world, which means 98% of international merchandise trade. The nature of HS Code is structured hierarchically, starting from broad categories down to more specific subcategories, which allows for detailed classification of goods.

Figure 1. Harmonized System Code Structure SECTION XVI

Machinery and mechanical appliances; Electrical Equipment; Parts Thereof; Sound Recorders and Reproducers, Television Image and Sound Recorders and Reproducers, and Parts and Accessories of Such Articles;

CHAPTER 85

Electrical machinery and equipment and parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles;

8519

Sound recording or reproducing apparatus.

851981

- - Using magnetic, optical or semiconductor media :

85198162

- - - Telephone answering machines

The HS code is vital for several reasons. It provides a common language for international trade, ensures uniform application of tariffs, and facilitates the collection of trade statistics. The system also plays a crucial role in strategic trade control and policy implementation, enabling countries to monitor and control the flow of goods effectively. Moreover, it is reviewed and renewed to accommodate new products and changes in global trade practices regularly, at least once every 5 years. However, this adaptability also contributes to its complexity, making accurate classification a challenging task

for customs officials and traders. The complexity of the HS, with its detailed and evolving nature, poses significant challenges. Misclassification of goods can lead to incorrect duty payments, trade disputes, and even legal consequences.

Utilizing machine learning for HS code prediction aims to automate and enhance the accuracy of the classification process. The traditional methods of HS Code classification rely heavily on human expertise and manual processes, which are prone to errors and inefficiencies. Machine learning models, by contrast, can process large volumes of data, recognize patterns, and make accurate predictions much faster than humans. This is particularly useful in customs operations, where timely and accurate classification is critical. The application of machine learning in customs operations, particularly in HS classification, is still in its early stages but shows great potential. By automating the classification process, machine learning can reduce the likelihood of human error, streamline customs operations, and ensure more consistent application of tariffs. Moreover, machine learning models can adapt to changes in the HS system more quickly than traditional.

The integration of machine learning into HS Code classification offers several benefits. First, it can significantly reduce the time and effort required for classification by automating the process, thereby improving efficiency. Machine learning models can also enhance the accuracy of classifications by learning from large datasets and identifying patterns that may not be immediately apparent to human classifiers. However, numerous challenges remain. The quality of input as product descriptions provided by users can significantly impact the performance of machine learning models. Poorly described or intentionally misleading data can lead to incorrect classifications, undermining the effectiveness of the model. Furthermore, HS Code contains multi-class, which makes the model complex and prone to be overfitting. Parameter tuning, and other techniques to maintain the model can generalize the data pattern should be taken into consideration to refine a model that is able to classify commodity description based on its HS Code.

Several previous studies have been explored to train model in classifying HS Code. One is conducted by Harsani et al., (2020) explored the use of an N-gram model combined with methods such as Libshorttext, text categorization, and topic modeling for classifying products based on HS codes. The study used description from import declaration and public dataset, consists of 33 classes from all 52 classes

in Chapter 64 and found that the Libshorttext approach with a bigram model achieved the highest accuracy, demonstrating the effectiveness of the N-gram model in dealing with the short text descriptions typically associated with HS codes. In another study by using anonymized dataset of shipments provided by DHL, Chen et al., (2021) employed Neural Network with combining of hierarchical loss. The model managed to classify about 29% of the population of the data with accuracy in 85%. Focusing on chapter 22 and 90 in HS Code, Ding et al., (2015) indicate that while traditional machine learning techniques such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) have been commonly employed, innovations like the B-Net approach offer promising alternatives for enhancing classification accuracy in the context of international trade and customs operations. The integration of topic modeling and the careful selection of features, such as those captured by N-gram models, are crucial for overcoming the inherent challenges posed by the short and often unstructured nature of product descriptions.

A significant portion of previous research on HS codes focuses on developing a knowledge base to facilitate manual code searches based on shipment descriptions or from declared description. In this study, the dataset will include the official assessment from customs authorities to enhance the quality of the data (Baeza-Yates & Liaghat, 2017). Furthermore, the data used in this study consists of various chapters to test the accuracy and ability of the model to generalize data. It is necessary to use various chapters due to HS Code nature that contains multi-class, therefore this study will develop a multi-class classification model using several machine learning algorithms.

3. RESEARCH METHODOLOGY

This study leveraged the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to guide the development of a machine learning model for predicting Harmonized System (HS) codes. Employing several techniques of data cleansing, model selection, and rigorous evaluation, the model was trained to meet the practical needs of customs operations. CRISP-DM, which is a well-established approach in the field of data science and machine learning, provides a structured, iterative process that ensures thorough and effective project execution (Schröer et al., 2021). This methodology was chosen for its flexibility and its ability to accommodate the complex

nature of data mining projects, which involve the prediction of HS codes based on commodity descriptions.

The project began with a clear business understanding phase, where the primary objective was defined: to develop a machine learning model capable of predicting the most probable 8-digit HS codes for a given commodity description. This objective addresses a significant challenge in customs operations, where accurate and efficient classification of goods is critical. Automating this process not only enhances accuracy but also reduces the time and effort required for manual classification (Spichakova & Haav, 2021). Following this, the data understanding phase involved an in-depth analysis of the dataset, which was sourced from historical data of Indonesia's import declaration from 2020 to the midst of 2024. This dataset included a wide variety of commodity descriptions, each linked to its corresponding HS code. To give a comprehensive understanding of the data, an exploratory data analysis was conducted to identify key patterns, depict the distribution of data, and detect any potential issues such as missing values or class imbalances. In the data preparation phase, several preprocessing steps were applied to the raw text data. Initially, the commodity descriptions were cleaned and normalized, which involved converting all text to lowercase, removing punctuation, and eliminating stop words to focus on the most relevant terms. Further refinement was achieved through tokenization and stemming, which broke down the text into individual words and reduced them to their root forms. This approach ensured that the model would focus on the essential elements of the descriptions, enhancing its ability to learn meaningful patterns (Hosseinzadeh et al., 2023).

To transform the text data into a format suitable for machine learning, the cleaned descriptions were vectorized using the Term Frequency-Inverse Document Frequency (TF-IDF) method.

$$TF - IDF(t, d, D) = \frac{f(t, d)}{N_d} \times \log \left(\frac{N}{1 + |\{d \in D : t \in d\}|} \right)$$

Its effectiveness in reflecting the importance of words within the dataset allows the model to differentiate between commonly used terms and those more specific to certain commodities (Xiang, 2022). This vectorization process resulted in a rich feature set that captured the nuances of the text data, providing

a solid foundation for the modeling phase. To predict HS Code accurately, exploring models specified to a multi-class classification problem is a must. Therefore, the modeling phase explored several machine learning algorithms, including Multinomial Naive Bayes, Logistic Regression, Random Forest, and Linear Support Vector Classification (LinearSVC) that have been chosen due to characteristics of the data, specific nature of the issue and the desired outcome.

Multinomial Naive Bayes is a probabilistic classifier based on Bayes' Theorem, which assumes that features follow a multinomial distribution. It is particularly suited for text classification tasks where features represent word frequencies. The model predicts the class C_k for a given feature vector x using:

$$P(C_k | x) = \frac{P(C_k) \prod_{i=1}^n P(x_i | C_k)}{P(x)}$$

Logistic Regression is a linear model used for binary classification. It estimates the probability that a given input x belongs to a particular class by applying the logistic (sigmoid) function to a linear combination of the input features. The probability of class 1 given x is:

$$P(y = 1 | x) = \frac{1}{1 + \exp(-(w \cdot x + b))}$$

Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the mode of the classes (for classification) or the average prediction (for regression) of the individual trees. Each tree is trained on a bootstrap sample of the data, and at each split in the tree, a random subset of features is considered. The prediction of the Random Forest is:

$$\hat{y} = \text{mode}(\{h_t(x)\}_{t=1}^T)$$

LinearSVC is a linear model for binary classification that attempts to find the hyperplane that best separates the two classes by maximizing the margin between them. The decision function for LinearSVC is:

$$f(x) = w \cdot x + b$$

The model was trained on 75% of the dataset, while the remaining 25% was reserved for testing. To enhance the practical utility of the model, a post-processing step was introduced. This step ensured that the HS code recommendations were unique, addressing the issue of duplicate predictions and improving the

diversity of the model's outputs. The model's performance was evaluated using a range of metrics, including accuracy, precision, recall, and F1-score. The evaluation methods focus on precision and F1-score due to the imbalanced dataset and to make sure the model can generalize data so it can predict new data as inputs. Additionally, confusion matrices were employed to analyze the distribution of predictions, helping to identify common misclassification patterns. This evaluation phase confirmed the model's ability to provide accurate and diverse HS code recommendations, validating its potential for real-world deployment. Finally, the model, along with its preprocessing and post-processing steps, was prepared for deployment. This phase involved integrating the model into a customs environment, where it can operate in real-time to assist with the classification of goods. By automating this critical process, the model is expected to significantly enhance the efficiency and accuracy of customs operations, making it a valuable tool for customs officials and businesses.

4. RESULTS AND FINDINGS

Data Preparation

Having a deep understanding of the data used in the model is a crucial part. The description of the commodity, used to determine which HS Code the commodity it is, is retrieved from text input from the reporter and some of them which underwent physical inspection by customs officer. To make the model work properly, understanding of the data can be insightful to determine how to cleanse the raw data. In the preliminary steps, maintaining the integrity of data such as lowering the description of HS Code, dropping the duplicates and null values and making sure the amount of data is sufficient to train and test the model is necessary. The initial dataset consisted of 999,999 observations and 2 features. After the data underwent several processes in the preliminary steps, the dataset was reduced to 588,438 observations, while maintaining the same number of features.

Furthermore, to make sure the description can be read by the model, cleansing the data into readable form should be done. The cleansing step can remove noises in the description that can mislead the model to uncover the pattern. Several functions were employed with their respective specific tasks start from doing simple stemming, expanding the contractions and removing accented characters, punctuation, irrelevant characters, extra whitespace, and stop words as follows in figure 2.

Figure 2. Data Preparation

Sample of Rows affected by **remove_accented_chars_func**: [246 rows x 2 columns]

Before	After
PVC¿APRON BAIK & BARU	PVCAPRON BAIK & BARU
HAND BAGS WOMEN K-8195¿ BAIK & BARU	HAND BAGS WOMEN K-8195 BAIK & BARU

Sample of Rows affected by **remove_punctuation_func**: [119721 rows x 2 columns]

Before	After
BRASS BUSS KD: BAIK / BARU	BRASS BUSS KD BAIK BARU
TELESCOPE T-EAGLE SR 2X28	TELESCOPE T EAGLE SR 2X28

Sample of Rows affected by **remove_irr_char_func**: [107367 rows x 2 columns]

Before	After
TELESCOPE T EAGLE SR 2X28	TELESCOPE T EAGLE SR X
BAGS 3708263 WITH OUTER SURFACE OF PU POLYE...	BAGS WITH OUTER SURFACE OF PU POLYE...

Sample of Rows affected by **remove_extra_whitespaces_func**: [128258 rows x 2 columns]

Before	After
BRAKE LEVER KD BAIK BARU	BRAKE LEVER KD BAIK BARU
CLUTCH DISC KD BAIK BARU	CLUTCH DISC KD BAIK BARU

Sample of Rows affected by **expand_contractions**: [774 rows x 2 columns]

Before	After
AUTOPARTS DIFF CASE SET	AUTOPARTS DIFFERENT CASE SET

Sample of Rows affected by **simple_stemming**: [157810 rows x 2 columns]

Before	After
CLUTCH DISCS KD BAIK BARU	clutch disc kd baik baru
BAGS WITH OUTER SURFACE OF PU POLYESTER	bag with outer surface of pu polyester

Sample of Rows affected by **remove_stopwords**: [48302 rows x 2 columns]

Before	After
motorcycle part for gift shoe	motorcycle part gift shoe
plastic ornament to baik baru	plastic ornament baik baru

In the exploratory data analysis phase, one of the key techniques employed was Term Frequency-Inverse Document Frequency (TF-IDF), a widely recognized method in natural language processing for transforming textual data into numerical representations. TF-IDF is instrumental in determining the importance of a word within a document relative to a collection of documents, which, in this case, are commodity descriptions used for Harmonized System (HS) code classification.

TF-IDF operates by balancing two components: term frequency (TF) and inverse document frequency (IDF). The term frequency reflects how often a word appears in a specific document, while the inverse document frequency accounts for how common or rare a word is across the entire dataset. The resulting TF-IDF score allows the model to weigh the significance of words, distinguishing between terms that are merely frequent and those that are particularly informative for classification. This approach was crucial in handling the dataset's large vocabulary, where common words such as "and," "new," or "part" might appear frequently across many descriptions but carry little discriminative power. Conversely, more specific terms like "PVC" or "apron" are critical for identifying the correct HS code. By assigning higher TF-IDF scores to these specific terms, the vectorization process effectively captures the nuances in the commodity descriptions, enabling the model to make more accurate predictions.

The richness of the feature set generated through TF-IDF also played a pivotal role in the modeling phase. The transformation of raw text into a structured format allowed for the

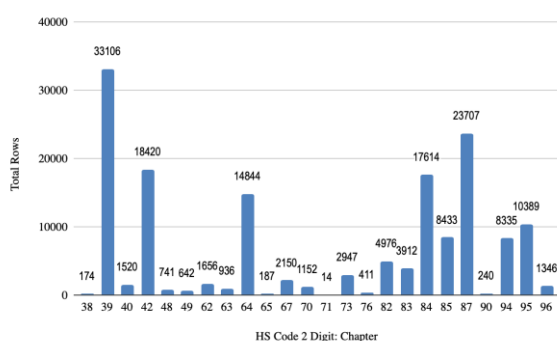
application of various machine learning algorithms, including Multinomial Naive Bayes, Logistic Regression, and Linear SVC, each benefiting from the detailed representation of the text data. This vectorized data provided a solid foundation for building models that could learn and generalize from the complex patterns present in the commodity descriptions. Moreover, the use of TF-IDF in this study not only improved the model's ability to classify HS codes but also contributed to the reduction of noise in the data. By focusing on the most relevant terms, the model's predictions became more robust, especially when dealing with ambiguous or short text inputs. This refinement was essential in ensuring that the machine learning model could effectively handle the diversity of commodity descriptions, which often vary significantly in length and complexity.

In conclusion, the implementation of TF-IDF in the exploratory data analysis phase significantly enhanced the model's capacity to differentiate between important and trivial words in the dataset. This enhancement laid the groundwork for successful model training and ultimately led to more precise HS code predictions, supporting the overall objective of improving customs operations through advanced machine learning techniques.

Model Comparison

Understanding the distribution of Harmonized System (HS) codes across different chapters is crucial for grasping the challenges involved in developing an effective classification model. The dataset reveals significant disparities in the frequency of HS codes, which directly impacts model performance.

Figure 3. Total Observations After Cleansing



As seen in figure 3, chapter 39, which covers plastics and articles thereof, stands out with over 33,000 rows, making it the most represented category. This high concentration could lead to a bias in the model, potentially resulting in better performance when predicting these specific

codes. Similarly, Chapters 87 and 40, which deal with vehicles and rubber articles, also have a substantial presence with 23,707 and 18,420 rows, respectively. The abundance of data in these categories likely enhances the model's predictive accuracy, as it has more examples to learn from. On the other hand, some chapters are significantly underrepresented. For instance, Chapter 38, which includes miscellaneous chemical products, and Chapter 94, covering furniture and bedding, have only 174 and 240 rows, respectively. The scarcity of data in these categories poses a significant challenge for the classification models, which tend to struggle with accurately predicting codes from underrepresented groups due to the limited number of training examples. This imbalance in the dataset necessitates a careful evaluation of model performance, as it influences the accuracy and reliability of the classification outputs.

In assessing the performance of the machine learning models used for HS code classification, the first figure provides insight into each model's mean accuracies across five-fold cross-validation (Figure 4). The Random Forest Classifier shows the lowest mean accuracy at 0.24, suggesting that its decision trees may not effectively capture the complex relationships required for this multiclass classification task. This could be due to overfitting or insufficient hyperparameter tuning. In contrast, the Linear Support Vector Classification (LinearSVC) model achieves the highest mean accuracy at 0.74, indicating its ability to identify a hyperplane that effectively separates different classes, even in high-dimensional space which is a crucial capability when dealing with the nuanced textual data found in HS code descriptions.

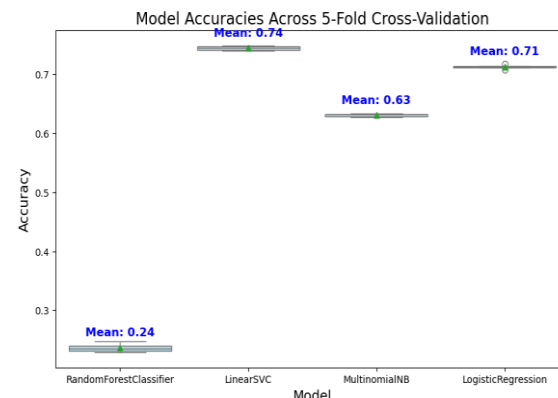


Figure 4. Model Performance

The Multinomial Naive Bayes (MultinomialNB) model, with a mean accuracy of 0.63, performs reasonably well. However, its assumption of feature independence might have limited its ability to fully capture the

complexities of the dataset. Logistic Regression, with a mean accuracy of 0.71, performs nearly as well as the LinearSVC. Its ability to model the probability of class membership in a linear fashion contributes to its solid performance in this multiclass scenario.

Given the LinearSVC model's superior performance, a deeper analysis is conducted using its confusion matrix, as depicted in the third figure. The matrix provides a detailed view of the model's ability to distinguish between different HS codes. The matrix shows strong diagonal dominance, indicating that the model correctly predicts most HS codes, particularly those with ample training data, such as those in Chapters 39, 87, and 40. However, some misclassifications are present in the off-diagonal elements, particularly in categories with less data or where certain HS codes have similar descriptions. These misclassifications highlight areas where the model could be improved, possibly through more advanced feature engineering or the application of data augmentation techniques to underrepresented classes.

Overall, while the LinearSVC model demonstrates strong performance, the confusion matrix reveals specific areas for improvement, underscoring the importance of ongoing model refinement and dataset balancing to achieve optimal classification accuracy across all HS codes.

5. CONCLUSIONS

This study has successfully demonstrated the feasibility and effectiveness of using a multiclass classification model to predict Harmonized System (HS) codes based on commodity descriptions. The research explored the core issue of improving the accuracy and efficiency of HS code classification that contains multi-class data, a critical challenge in international trade and customs operations. The implementation of TF-IDF in the exploratory data analysis phase significantly enhanced the model's capacity to differentiate between important and trivial words in the dataset. LinearSVC as the proposed model, trained on a dataset of commodity descriptions, showed significant potential in automating the classification process, thereby reducing manual errors and speeding up the overall classification task. Though data used in training the model did not include all classes in HS Code, the result of model's performance in predicting numerous class in HS Code reach 74% at the highest.

The findings contribute to the body of knowledge by providing a practical solution that leverages machine learning for HS code classification. The study's implications extend to enhancing customs operations, improving trade compliance, and supporting businesses in navigating complex international trade regulations. However, the research is not without limitations. The model's performance is inherently dependent on the quality and diversity of the training data (Baeza-Yates & Liaghat, 2017), for example the usage of two languages, namely Indonesian and English, in product descriptions will make the model more difficult to classify the commodity. Additionally, the model may require further refinement to handle ambiguous or highly complex commodity descriptions accurately. Furthermore, the variety in the input needs to be considered. To make the model generalized the pattern in data and can classify various commodity for each HS Code, the model needs to be trained with data from all classes in HS Code especially when the imbalanced data for each code will impact the performance of model.

Building on the conclusions, several recommendations can be made for future research and practical application. Operationally, it is suggested that customs authorities and businesses involved in international trade consider integrating machine learning models into their HS code classification systems used. This integration could streamline operations, reduce errors, and improve compliance with international trade laws. From a policy perspective, governments and international trade organizations should consider developing standardized datasets for training and validating such models, ensuring that they can be widely adopted across different regions and industries. Furthermore, there should be continuous updates to the model to reflect changes in trade patterns and the introduction of new commodities. For future research, it is recommended to explore the incorporation of additional features such as product images and textual data from trade documents to enhance the model's predictive accuracy.

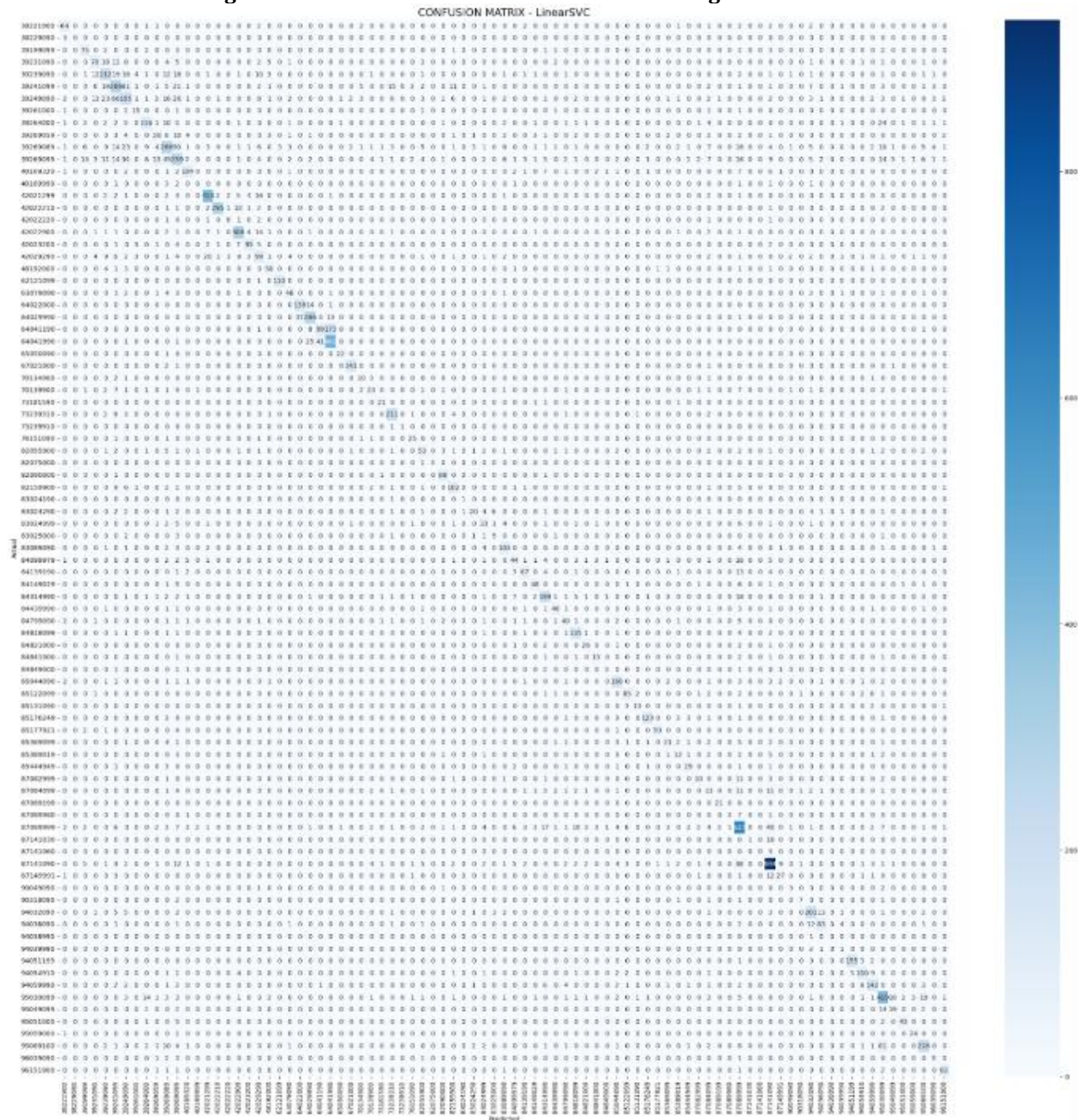
REFERENCES

Asian Development Bank. (2023). *E-Commerce Evolution in Asia and the Pacific: Opportunities and Challenges*. <https://doi.org/http://dx.doi.org/10.22617/TCS230473-2>

- Baeza-Yates, R., & Liaghat, Z. (2017). Quality-efficiency trade-offs in machine learning for text processing. *2017 International Conference on Big Data*, 897–904. <https://doi.org/10.1109/BigData.2017.8258006>
- Chen, X., Bromuri, S., & Van Eekelen, M. (2021). Neural machine translation for harmonized system codes prediction. *ACM International Conference Proceeding Series*, 158–163. <https://doi.org/10.1145/3468891.3468915>
- Cuaya-Simbro, G., Hernández-Vera, I., Ruiz, E., & Gutiérrez-Fragoso, K. (2022). Automatic Tariff Classification System using Deep Learning. *IJACSA International Journal of Advanced Computer Science and Applications*, 13(7), 904–911. www.ijacsa.thesai.org
- Ding, L., Fan, Z. Z., & Chen, D. L. (2015). Auto-categorization of HS code using background net approach. *Procedia Computer Science*, 60(1), 1462–1471. <https://doi.org/10.1016/j.procs.2015.08.224>
- Harsani, P., Suhendra, A., Wulandari, L., & Wibowo, W. C. (2020). A study using machine learning with Ngram model in harmonized system classification. *Journal of Advanced Research in Dynamical and Control Systems*, 12(6 Special Issue), 145–153. <https://doi.org/10.5373/JARDCS/V12SP6/SP20201018>
- Hosseinzadeh, M., Azhir, E., Ahmed, O. H., Ghafour, M. Y., Ahmed, S. H., Rahmani, A. M., & Vo, B. (2023). Data cleansing mechanisms and approaches for big data analytics: a systematic study. *Journal of Ambient Intelligence and Humanized Computing*, 14(1), 99–111. <https://doi.org/10.1007/s12652-021-03590-2>
- Liu, T., & Wang, Z. (2019). Practical Problems in the Customs Supervision on Cross-Border E-Commerce Goods and Its Solutions. *Advances in Economics, Business and Management Research*, 110, 1013–1020. <http://creativecommons.org/licenses/by-nc/4.0/>
- Rukanova, B., Tan, Y. H., Slegt, M., Molenhuis, M., van Rijnsoever, B., Plecko, K., Caglayan, B., & Shorten, G. (2019). Value of Big Data Analytics for Customs Supervision in e-Commerce. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11685 LNCS, 288–300. https://doi.org/10.1007/978-3-030-27325-5_22
- Saydam, S. G., & Civelek, M. E. (2022). Problems in Cross-Border E-Commerce and Development of Cross-Border E-Commerce Performance Scale. *Journal of International Trade, Logistics and Law*, 8, 133–140.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Sembiring, A. R., Novith, D. C., & Ridho, M. H. (2023). Customs Fraud Detection Using Extremely Boosted Neural Network (XBNET). *Jurnal BPPK*, 16, 67–77. <https://doi.org/10.48108/jurnalbppk.v16i1.816>
- Spichakova, M., & Haav, H. M. (2021). Application of machine learning for assessment of HS code correctness. *Baltic Journal of Modern Computing*, 8(4), 698–718. <https://doi.org/10.22364/BJMC.2020.8.4.13>
- Statista. (2022). *Key Figures of E-commerce*.
- Xiang, L. (2022). Application of an Improved TF-IDF Method in Literary Text Classification. *Advances in Multimedia*, 2022. <https://doi.org/10.1155/2022/9285324>
- Yuan, Y. (2023). Cross-Border E-Commerce Logistics Service Challenges and Development. *Frontiers in Business, Economics and Management*, 8(1), 262–265.

APPENDIX

Figure 5. Detailed Confusion Matrix for Each 8-Digit HS Code



This article is licensed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Any derivative works must be distributed under the same license as the original.

