

**INTEGRATING KNOWLEDGE GRAPHS AND LLM FOR INDONESIA HS
CODE CLASSIFICATION: A FOCUS ON HEAVY EQUIPMENT PARTS**

By

**Syepta Hendrawan
22351018**

MASTER'S DEGREE
in

MASTER OF INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

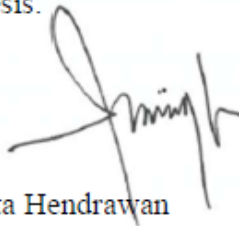


SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

Revision after the Thesis Defense on 06 February 2025

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.



Syepta Hendrawan

20 January 2025

Date

Student

Approved by:



Dr. Ir. Heru Purnomo Ipung, M.Eng

20 January 2025

Date

Thesis Advisor

Dr. Ir. Eka Budiarto, S.T., M.Sc

20 January 2025

Date

Thesis Co-Advisor

Dr. Maulahikmah Galinium, S.Kom, M.Sc

20 January 2025

Date

Dean

Date

Dean



ABSTRACT**INTEGRATING KNOWLEDGE GRAPHS AND LLM FOR INDONESIA HS
CODE CLASSIFICATION: A FOCUS ON HEAVY EQUIPMENT PARTS**

By

Syepta Hendrawan

Dr. Ir. Heru Purnomo Ipung, M.Eng, Advisor

Dr. Ir. Eka Budiarto, S.T., M.Sc, Co-Advisor

SWISS GERMAN UNIVERSITY

Accurate classification of heavy equipment parts under Indonesia's Harmonized System (HS) codes is vital for regulatory compliance, cost mitigation, and operational efficiency. This paper presents a proof-of-concept framework that integrates Knowledge Graphs (KGs), Large Language Models (LLMs), and Machine Learning (ML) techniques to address the persistent challenges of HS code misclassification. The framework leverages historical customs data to train ML models for HS code prediction and incorporates a Knowledge Graph to encode domain-specific rules and hierarchical relationships. Complementing these methods, LLMs provide contextual understanding and generate explainable justifications for classification decisions. As a proof-of-concept, the system is demonstrated through a case study at a leading enterprise in Indonesia's heavy equipment sector which exhibits enhanced accuracy, transparency, and operational efficiency in HS code assignment. Accounting for nuanced part descriptions, evolving regulations, and the inherent complexity of hierarchical classification, the proposed approach minimizes reliance on implicit knowledge and reduces error rates. The findings highlight the framework's scalability and adaptability, indicating its potential to benefit not only the heavy equipment industry but also other sectors that confront similar global trade compliance challenges.

Keywords: HS code classification, knowledge graphs, large language models, machine learning, heavy equipment parts, regulatory compliance, RAG.



© Copyright 2025

by Syepta Hendrawan

All rights reserved

DEDICATION

This work dedicated to my family, whose unwavering support, love, and encouragement have shaped my journey every step of the way.



ACKNOWLEDGEMENTS

I wish to begin by expressing my profound gratitude to Dr. Ir. Heru Purnomo Ipung, M.Eng, for his steadfast mentorship and insightful guidance as my advisor. His expertise and dedication have been pivotal in refining this research.

My sincere thanks also go to Dr. Ir. Eka Budiarto, S.T., M.Sc, who has generously contributed his knowledge and perspective as my co-advisor. His thoughtful input has greatly enriched the scholarly depth of this work.

Lastly, I extend my warm appreciation to my friends from Batch 33 MIT SGU. Their unwavering support and camaraderie have been a constant source of motivation, reminding me of the power of collaboration and shared aspiration throughout this academic journey.

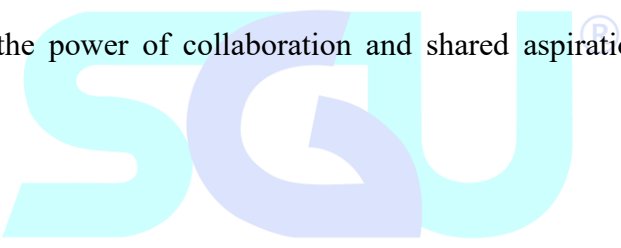


TABLE OF CONTENT

DEDICATION	6
ACKNOWLEDGEMENTS	7
TABLE OF CONTENT	8
LIST OF FIGURES	12
LIST OF TABLES	14
CHAPTER 1 – INTRODUCTION	15
1.1. Background	15
1.1.1 Import Process and HS Code Classification	28
1.1.2 KUM HS and Its Role in Indonesia’s Trade Classification Framework Error! Bookmark not defined.	
1.1.3 Complexity and Challenge in HS Code Classification in PT XYZ	28
1.1.4 Introduction to Machine Learning, Knowledge Graphs and AI Language Models (LLM) as Potential Solutions	Error! Bookmark not defined.
1.2. Research Problem	20
1.3. Research Objective	21
1.4. Research Question	22
1.5. Hypothesis	22
1.6. Research Scope and Limitation	23
1.7. Significance of Study	24
1.8. Thesis Structure	25
CHAPTER 2 - LITERATURE REVIEW	26
2.1 Introduction	26
2.2 Problem Overview	28
2.2.1 Hierarchical Complexity	28
2.2.2 Regulatory Changes Impact on Heavy Equipment Supply Chains	32
2.2.3 Part Numbers and Their Structure	36
2.3 Traditional Methods for HS Code Classification	37
2.3.1 KUHMS	37
2.3.2 Reliance on Tacit Knowledge	32
2.4 Machine Learning for HS Code Classification	44
2.4.1 Text Mining Transforming Unstructured Text into Usable Data	45

2.4.2	TF-IDF: Extracting Significant Features	46
2.4.3	Data Source for HS Code Classification.....	Error! Bookmark not defined.
2.4.4	Machine Learning Algorithms for HS Code Classification.....	48
2.4.5	Hyperparameter Tuning for Machine Learning Algorithms.....	Error! Bookmark not defined.
2.4.6	Limitations of Machine Learning in HS Code Classification Automation.....	55
2.5	Knowledge Graph in HS Code Classification	62
2.5.1	Introduction to Knowledge Graph	62
2.5.2	Historical Context and Development.....	68
2.5.3	Rationale of Knowledge Graphs.....	70
2.5.4	Building the Knowledge Graph	70
2.6	Large Language Model (LLM) for Contextual Understanding.....	76
2.6.1	Introduction: What is a Large Language Model (LLM)?	76
2.6.2	Historical Overview of LLMs.....	76
2.6.3	Types of Large Language Models (LLMs).....	77
2.6.4	LLMs Capabilities and Limitation.....	79
2.6.5	Retrieval-Augmented Generation (RAG) and LLMs	81
2.7	Integration of Knowledge Graphs with LLMs	85
2.7.1	Roadmap of Unifying LLM and Knowledge Graph.....	85
2.7.2	Synergized KG and LLM Method in HS Code Classification	88
2.7.3	LangChain: Integrated Framework of LLM and Knowledge Graph for HS Code Classification	91
2.7.4	Evaluation Framework for KG-LLM Systems	93
2.8	Risks and Challenges of AI in HS Code Classification	94
2.8.1	Data-Related Risks	97
2.8.2	Systemic Risks.....	98
2.8.3	Operational Risks.....	98
2.8.4	Ethical Risks	99
2.8.5	Role of Human-in-the-Loop (HITL).....	99
2.9	Related Works and Previous Studies.....	Error! Bookmark not defined.
2.9.1	HS Code Classification Using Machine Learning.....	57
2.9.2	ML Data Collection for HS Code Classification	48
2.9.3	Evaluation Methods in HS Code Classification ML	55

2.9.4	Knowledge Graph in Hierarchical System	62
2.9.5	Large Language Models in Knowledge Graph Information Retrieval	94
2.10	Synthesis of Findings	102
2.10.1	Strengths of Existing Approaches.....	Error! Bookmark not defined.
2.10.2	Limitations and Gaps	Error! Bookmark not defined.
2.10.3	Relevance of Proposed Framework	105
2.11	Proposed Framework	102
2.11.1	Overview of the Framework	102
2.11.2	Framework Workflow	106
2.11.3	Validation Framework through Process Modelling Success (PMS)	107
CHAPTER 3 – RESEARCH METHODS		109
3.1	Research Framework	109
3.2	Data Collection and Preprocessing	111
3.3	Model Development	112
3.4	Knowledge Graph Construction	115
3.5	Integration of Machine Learning Model with Knowledge Graph	117
3.6	Controlled Evaluation for AI Risk Exposure	119
3.7	System Deployment and User Interaction via Console-Based Chatbot	120
3.7.1	Console-Based Chatbot Interface	121
3.7.2	Real-Time Query and Explanation	121
3.7.3	User Feedback Integration and KG Updates	122
3.8	Evaluation Method	122
3.8.1	Quantitative and Qualitative Evaluation.....	122
3.8.2	Evaluation Using the RAGAS Framework.....	124
3.9	Validation Method Focus Group Discussion Against Experts	126
3.9.1	Expert Criteria.....	126
3.9.2	Questionnaire Design Using the PMS Framework.....	127
3.9.3	Focus Group Discussions (FGDs)	129
CHAPTER 4 - RESULTS AND DISCUSSIONS		132
4.1	Introduction	132
4.2	Data Preparation and Text Processing.....	132
4.2.1	Data Collection	132

4.2.2	Data Cleaning and Normalization.....	134
4.2.3	Feature Extraction.....	136
4.3	Model Development Results	140
4.3.1	Exploratory Data Analysis (EDA).....	141
4.3.2	Hyperparameter Tuning.....	141
4.3.3	Model Selection and Training.....	142
4.3.4	Addressing Class Imbalance.....	144
4.4	Knowledge Graph Construction Results	145
4.4.1	Tools to Develop Knowledge Graphs.....	146
4.4.2	Developing Knowledge Graph for Part Number	146
4.5	Integration of LLM and Knowledge Graph in Chatbot Development	148
4.5.1	High-Level Overview of the Integration	148
4.5.2	Technical Design and Workflow	149
4.5.3	Chatbot Development and Flow Process	152
4.6	Evaluation: Retrieval Augmented Generation Assessment (RAGAS)	157
4.6.1	Evaluation Process for RAGAS.....	157
4.6.2	RAGAS Evaluation Result	158
4.6.3	Conclusion of RAGAS Evaluation.....	161
4.7	Validation base on Focus Group Discussion (FGD)	161
4.7.1	Participant Demographics.....	161
4.7.2	Quantitative & Qualitative FGD Results	163
4.7.3	Key Observations During FGD	171
4.8	Summary.....	172
CHAPTER 5 – CONCLUSION AND DISCUSSIONS		174
5.1	Conclusions	174
5.2	Recommendation	175
5.3	Suggestions for Future Research (Feasibility-Focused).....	177
REFERENCES		180

LIST OF FIGURES

Figure 1.1 Harmonized System (HS) code structure showing the division of the first 6 digits standardized globally and the 7th and 8th digits customized for Indonesia Harmonized System Code Structure	16
Figure 1.2 Flow process illustrating key steps in HS code classification during the import-export process in Indonesia.....	29
Figure 1.3 Penalty tariffs imposed due to HS code misclassification, showing the financial impacts of errors in duty assessment.	30
Figure 1.4 Financial impacts of HS code misclassification on PT XYZ, including penalties, overpayment, and increased costs from delays.....	31
Figure 1.5 Fishbone diagram identifying key challenges faced by PT XYZ in classifying HS codes for heavy equipment parts.....	Error! Bookmark not defined.
Figure 2.1 The Hierarchical BTKI Framework	27
Figure 2.2 Hierarchical HS Code for PN Cylinder Assy	35
Figure 2.3 Hierarchical structure of HS Code for chapter 84 Machinery and Mechanical Appliance in BTKI Regulation	36
Figure 2.4 Structure PN Heavy Equipment PT XYZ systematically organized essential detailed.....	37
Figure 2.5 The Process Flow Chart for KUHMS Classification Goods	40
Figure 2.6 Example of a Knowledge Graph Applied to HS Code Classification in the Heavy Equipment Industry.	67
Figure 2.7 Semantic and Hierarchical Representation in Knowledge Graphs for HS Code Classification	75
Figure 2.8 Workflow of a Retrieval-Augmented Generation (RAG) Framework.....	82

Figure 2.9 Outlined through three key frameworks shown road map of unifying LLM and Knowledge Graph proposed (Pan et al.,2023)	85
Figure 2.10 Retrieval-Augmented Generation Application to retrieve additional information from various sources such as Knowledge Graph (medium.com)	90
Figure 2.11 Process Modelling Success (PMS) measurement framework (Sedera Rosemann and Gabe,2002)	108
Figure 3.1 Depicts the overall structure of the research framework, illustrating the seamless integration of ML, KG, and LLMs across different phases of the HS code classification system.	109
Figure 3.2 The Customs Expert Certification (Sertifikasi Ahli Kepabeanan or SAK).	127
Figure 4.1 Distribution of HS Codes shown imbalance data of Heading and Sub Heading.	141
Figure 4.2 Visualization of the Knowledge Graph Showing Relationships Between Part Numbers, Categories, and Purposes compare to relational database table.	147
Figure 4.3 Comprehensive Knowledge Graph Representing Part Numbers, Categories, and Purposes for HS Code Classification	Error! Bookmark not defined.
Figure 4.4 Chatbot providing detailed information on HS codes, including part specifications, hierarchical breakdowns, and actionable insights to address user queries with clarity and precision.	155

LIST OF TABLES

Table 2.1 Historical Updates of BTKI	32
Table 2.2 Six General Rules for Classification Harmonized System (KUHMS)	41
Table 2.3 XYZ Misclassification HS Code	38
Table 2.4 Comparison of Relational Databases and Knowledge Graphs	73
Table 2.5 Categorization of LLM and Comparative.....	78
Table 2.6 Related Works of HS Code Machine Learning	58
Table 2.7 Related Work Data Processing for Machine Learning	50
Table 2.8 Related Work Application on Knowledge Graph and LLM.....	64
Table 2.9 Related Works Integration LLM and Knowledge Graph	95
Table 4.1 Dataset Attributes Overview of PT XYZ.....	133
Table 4.2 Dataset Cleaning Remove Duplication.....	134
Table 4.3 Text Cleaning and Normalization.....	135
Table 4.4 Categorical Encoding Details	136
Table 4.5 Feature extraction.....	138
Table 4.6 TF-IDF Matrix Representing Term Importance Across Documents.....	140
Table 4.7 Comparative Performance of Candidate Models.....	142
Table 4.8 Classification Report.....	144
Table 4.9 Summarized Iteration Result of RAGAS.....	158
Table 4.10 Summary of Participants Demographics	162
Table 4.11 PMS-Model Quality Insight Summary	163
Table 4.12 PMS-Model Used Insight Summary	165
Table 4.13 PMS-User Satisfaction Insight Summary	167
Table 4.14 PMS-Process Impact Insight Summary	168
Table 4.15 PMS-Organizational Impact Insight Summary.....	170