A COMPARISON ANALYSIS OF KNN AS A PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS TO RANDOM FOREST, AND MLR

FINAL THESIS

DINDA FATIMAH KAUTSARINA

41521010114

DEPARTMENT OF INFORMATICS ENGINEERING

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS MERCU BUANA

JAKARTA

2025

A COMPARISON ANALYSIS OF KNN AS A PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS TO RANDOM FOREST, AND MLR

FINAL THESIS

DINDA FATIMAH KAUTSARINA

41521010114

Submitted as one of the requirements to receive a bachelor's degree

DEPARTMENT OF INFORMATICS ENGINEERING

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS MERCU BUANA

JAKARTA

2025

STATEMENT OF ORIGINALITY

I, the undersigned:

Name	:	DINDA FATIMAH KAUTSARINA
NIM	:	41521010114
Department	:	Teknik Informatika
Research Proposal Title	:	PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS USING KNN

I declare that this Final Thesis is my own work and not plagiarism and that I have correctly stated all sources cited and referenced. If it is found that there are elements of plagiarism in my Final Thesis, I am prepared to face the academic sanctions applicable at Mercu Buana University.

Jakarta, 8 Juli 2025

Dinda Fatimah Kautsarina.

APPROVAL PAGE

The undersigned hereby declares that the final thesis report from the student:

Name	:	DINDA FATIMAH KAUTSARINA
NIM	:	41521010114
Department	:	Teknik Informatika
Research Proposal Title	:	PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS USING KNN

Has been successfully defended at the hearing before the Board of Examiners and accepted as part of the requirements needed to obtain a Bachelor's degree in the Informatics Engineering Study Program, Faculty of Computer Science, Universitas Mercu Buana.

Approved by:

Supervisor	:	Ir. Emil R. Kaburuan, Ph.D., IPM., PMP, CISM, ASEAN Eng., SMIEEE
NIDN	:	0429058004
Head Examiner	:	Ilham Nugraha, S.Kom., M.Sc
NIDN	:	0307098904
Examiner 1	:	Dr. Hadi Santoso, S.Kom., M.Kom.
NIDN	:	0225067701
Examiner 2	:	Dr. Ir. Eliyani, M.Kom.
NIDN	:	0321026901

Jakarta,

Acknowledged by,

Dean	Head of Study Program
Dr. Bambang Jokonowo, S.Si., MTI	Dr. Hadi Santoso, S.Kom., M.Kom
NIDN : 0320037002	NIDN : 0225067701

ACKNOWLEDGMENTS

Praise be to God Almighty for His grace and blessings, enabling the author to complete this research proposal. This proposal is a requisite for the completion of the Undergraduate Program (S1) in the Department of Information Technology at Mercu Buana University.

The author realizes that this research proposal is far from perfect, as true perfection belongs only to God Almighty. Therefore, constructive suggestions and feedback are always welcomed with open arms. It is with the support, motivation, assistance, guidance, and prayers from many parties that this proposal has been completed. The author extends heartfelt gratitude to:

1. Prof. Dr. Andi Adriansyah, M.Eng., Rector of Mercu Buana University.
2. Dr. Bambang Jokonowo, S.Si., MTI, Dean of the Faculty of Computer Science.
3. Dr. Hadi Santoso, S.Kom., M.Kom., Head of the Information Technology Study Program at Mercu Buana University.
4. Emil Robert Kaburuan, S.T., M.A., Ph.D., MPTI supervisor, who has provided guidance, motivation, time, effort, and thought, ensuring the research proposal was well-scheduled.
5. My parents and family, who have always supported and encouraged me throughout my studies at Mercu Buana University.
6. All my friends at UMB, Calosa Ballet, and my IISMA dorm-mates at UNISI, who have supported and motivated me during the making of this thesis.

Lastly, the author hopes that God Almighty will repay the kindness and always bestow His grace, guidance, and long life upon us all. Amen. Thank you.

Jakarta, 8 Juli 2025

Dinda Fatimah Kautsarina

STATEMENT OF FINAL PROJECT PUBLICATION CONSENT FOR ACADEMIC PURPOSES

As an academic community member of Universitas Mercu Buana, I, the undersigned:

Name	:	DINDA FATIMAH KAUTSARINA
NIM	:	41521010114
Department	:	Teknik Informatika
Research Proposal Title	:	A COMPARISON ANALYSIS OF KNN AS A PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS TO RANDOM FOREST, AND MLR

Hereby declare that, in the interest of knowledge, I grant and approve Universitas Mercu Buana to have a **Non-exclusive Royalty-Free Right** to my final project report, for use as scientific or non-scientific study material by Universitas Mercu Buana.
Universitas Mercu Buana has the right to store, convert/transfer formats, manage in the form of a database, distribute, and display it in softcopy form for academic purposes within Universitas Mercu Buana, such as in Theses/Papers/Dissertations, as well as in other media, without the need to request my permission, provided that my name remains stated as the author and copyright holder.
This statement is made truthfully.

Jakarta, 8 July 2025
Declarant,

Dinda Fatimah Kautsarina

ABSTRACT

Nama	:	DINDA FATIMAH KAUTSARINA
NIM	:	41521010114
Program Studi	:	Teknik Informatika
Judul Proposal Penelitian	:	A COMPARISON ANALYSIS OF KNN AS A PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS TO RANDOM FOREST, AND MLR
Dosen Pembimbing	:	Emil Robert Kaburuan, S.T., M.A., Ph.D.

Despite being a significant contributor to the nation's GDP and expected to grow, the cement industry poses substantial environmental challenges, contributing significantly to GHG emissions and climate change. This study is aimed at developing and comparing three predictive models, namely the K-Nearest Neighbor (KNN), Random Forest, and Multivariable Linear Regression (MLR) to create Scope-1 CO_2 emissions projection from Indonesian cement manufacturers. Using five years of historical data from cement plants across Indonesia, this study has identified key emission factors and the most suitable machine learning model to predict Scope-1 CO_2 emissions. Each model was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R2 to assess their accuracy. Furthermore, this study also proposed an integration of these machine learning models into conventional emission calculation methods to provide a robust emissions projection framework that supports the Indonesian Net Zero goal.

Key words: GHG Emission, K-Nearest Neighbors (KNN), Machine Learning, Emission Prediction

ABSTRACT

Name	:	DINDA FATIMAH KAUTSARINA
NIM	:	41521010114
Study Program	:	Teknik Informatika
Research Proposal Title	:	A COMPARISON ANALYSIS OF KNN AS A PREDICTIVE MODEL FOR SCOPE-1 GHG EMISSIONS FROM INDONESIAN CEMENT MANUFACTURERS TO RANDOM FOREST, AND MLR
Supervisor	:	Emil Robert Kaburuan, S.T., M.A., Ph.D.

Despite being a significant contributor to the nation's GDP and expected to grow, the cement industry poses substantial environmental challenges, contributing significantly to GHG emissions and climate change. This study is aimed at developing and comparing three predictive models, namely the K-Nearest Neighbor (KNN), Random Forest, and Multivariable Linear Regression (MLR) to create Scope-1 CO_2 emissions projection from Indonesian cement manufacturers. Using five years of historical data from cement plants across Indonesia, this study has identified key emission factors and the most suitable machine learning model to predict Scope-1 CO_2 emissions. Each model was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R2 to assess their accuracy. Furthermore, this study also proposed an integration of these machine learning models into conventional emission calculation methods to provide a robust emissions projection framework that supports the Indonesian Net Zero goal.

Keywords: GHG Emission, K-Nearest Neighbors (KNN), Machine Learning, Emission Prediction

TABLE OF CONTENTS

TITLE PAGE i
STATEMENT OF ORIGINALITY ii
APPROVAL SHEET iii
ACKNOWLEDGMENTS iv
STATEMENT OF APPROVAL FOR THE PUBLICATION v
ABSTRACT vi
ABSTRACT vii
TABLE OF CONTENTS viii
LIST OF TABLES x
LIST OF FIGURES xi
LIST OF ATTACHMENTS xii
CHAPTER I INTRODUCTION 1
1.1 Background 1
1.2 Problem Statement 3
1.3 Research Objectives 3
1.4 Research Benefits 4
1.5 Research Limitations 5
CHAPTER II LITERATURE REVIEW 6
2.1 Main Theory 6
2.1.1 CO_2 from Raw Material Calcination 6
2.1.2 CO_2 from Kiln Fuels 7
2.2 Supporting Theory 8
2.2.1 K – Nearest Neighbor 8
2.2.2 Random Forest 9
2.2.3 Multivariable Linear Regression 9
2.3 Previous Research 10
2.4 Research Gap 15
CHAPTER III RESEARCH METHODS 16
3.1 Research Approach 16
3.2 Research Design 16

3.3 Research Subjects 17
3.4 Research Instruments 18
3.5 Data Collection Method 18
3.6 Data Analysis 19
3.7 Research Procedure 20
3.8 Results Evaluation 22
CHAPTER IV DISCUSSIONS 23
4.1 Data Analysis and Identification of Key Emission Factors 23
4.1.1 Exploratory Data Analysis (EDA) Findings 23
4.1.2 Feature Engineering 27
4.1.3 Key Features in Specific Net CO_2 Emissions 29
4.2 Performance Evaluation of K-Nearest Neighbor (KNN) in Emission Prediction 29
4.2.1 KNN Model Training and Validation Results 29
4.2.2 Comparison with Other Machine Learning Algorithms 30
4.3 Proposed Integration with Conventional Emissions Calculation 35
CHAPTER V CONCLUSIONS & SUGGESTION 37
5.1 Conclusions 37
5.2 Suggestions 38
References 41
ATTACHMENT 44

LIST OF TABLES

Table 2.1 Typical components of raw materials in the calcination process [14].... 6
Table 2.2 Related Research..... 11
Table 4.1 Scope 1 CO_2 Emission Factors Descriptive Analysis..... 23
Table 4.2 KNN Model Evaluation Results 30
Table 4.3 Algorithm Performance Comparison..... 31

LIST OF FIGURES

Figure 3.1 Simple cement manufacturing processes showing the CO_2 generation systems [14]. 17
Figure 3.2 Research flow for a predictive model using machine learning algorithms [14] 20
Figure 4.1 Distribution of data for Scope-1 CO_2 Emissions 24
Figure 4.2 Correlation Matrix of the Features and the Target Variable, Specific Net CO_2 Emissions 25
Figure 4.3 ACF and PACF of each feature in the input dataset 26
Figure 4.4 Correlation matrix post further pre-processing using ACF and PACF. 26
Figure 4.5 Correlation Matrix of the Engineered Features with the Target Variable 28
Figure 4.6 A comparison of the training and testing R^2 values by model 31
Figure 4.7 Specific CO_2 Projection compared to baseline 2010 - 2030 at 75% CCR and TSR 5% [7] 33
Figure 4.8 Specific Net CO_2 Projection using KNN 33
Figure 4.9 Specific Net CO_2 Projection using Random Forest 34
Figure 4.10 Specific Net CO2 Projection using Multivariable Linear Regression 35

LIST OF ATTACHMENTS

Attachment 1 Kartu Asistensi 44
Attachment 2 Curriculum Vitae 45
Attachment 3 BNSP Certificate 46
Attachment 4 Turnitin Results 47