

UNLOCKING REMANUFACTURING COMPONENT
DEMAND POTENTIAL THROUGH BIG DATA ANALYTICS

By

Eko Ichwan Ibrahim

22452017

MASTER'S DEGREE

in

MASTER PROGRAM OF MECHANICAL ENGINEERING
ENGINEERING MANAGEMENT



SWISS GERMAN UNIVERSITY

The Prominence Tower

Jalan Jalur Sutera Barat No. 15, Alam Sutera

Tangerang, Banten 15143 - Indonesia

Revision after Thesis Defense on August 19th, 2025

Eko Ichwan Ibrahim

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.

Eko Ichwan Ibrahim

Student

Date : 03 – 09 - 2025

Approved by:

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

Thesis Advisor

Date : 03 – 09 - 2025

Dr. rer. nat. Filiana Santoso

Thesis Co-Advisor

Date : 03 – 09 - 2025

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

Dean

Date : 03 – 09 - 2025

Eko Ichwan Ibrahim

ABSTRACT

UNLOCKING REMANUFACTURING COMPONENT DEMAND POTENTIAL THROUGH BIG DATA ANALYTICS

By

Eko Ichwan Ibrahim

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

Dr. rer. nat. Filiana Santoso, Co-Advisor

SWISS GERMAN UNIVERSITY

This study aimed to unlock remanufacturing component demand potential at United Tractors (UT) through big data analytics, specifically to predict demand for Komatsu and Scania remanufactured parts and improve demand visibility. A two-stage XGBoost model (classification and regression) was developed using comprehensive internal data (customer transactions, machine operations) and external data (weather, coal price, calendar). Model performance was evaluated using RMSE, MAE, and accuracy, comparing XGBoost against Naïve Bayes. Post-processing integrated RFM (Recency, Frequency, Monetary) analysis to prioritize sales opportunities. XGBoost consistently outperformed Naïve Bayes (RMSE 83.672 vs. 111.098; MAE 34.961 vs. 51.429). The model achieved high accuracy in 2024 (up to 98.04%) with average 70,44% but declined to ~44.5% in S1 2025, likely due to uncaptured external market shifts like sharp coal price drops. Key predictors included historical demand, agroclimate factors, and operational/commercial metrics. Adoption of analytical opportunities by UT's team significantly increased to 45-48% by May-June 2025. The research successfully applied big data analytics for remanufactured component demand prediction, with future improvements requiring dynamic adaptation to external market variables

Keywords: Remanufacturing Component Business, Demand Forecasting, Big Data Analysis, Machine Learning, Heavy Equipment.

Eko Ichwan Ibrahim



DEDICATION

With gratitude, I dedicate this work to my parents for their unwavering prayers, love, and support throughout my life and academic journey. To my beloved family—my wife, Tri Puji Astuti, thank you for being a source of encouragement and peace of mind with every step I take. Your patience and understanding are invaluable.

To all the MME lecturers, I'm deeply grateful for the knowledge and insights they have shared, which have greatly shaped my perspective to this day. To my fellow students and colleagues at the company where I work who helped me complete this thesis, thank you for the support and opportunity given me to realize one of my dreams and make it a reality.



ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to my Thesis Advisors, Dr. Maulahikmah Galinium, S.Kom, M.Sc, and Dr. rer. nat. Filiana Santoso, from Swiss German University, for their invaluable guidance, constructive feedback, and unwavering support throughout the entire research process. Their profound expertise was instrumental in shaping this Master's thesis.

I would like to sincerely thank Dr. Ir. Gembong Baskoro, M.Sc., IPU, ACPE, APEC Eng., as the Head of the Master of Mechanical Engineering Program, for his exemplary leadership and continuous commitment in fostering an academically enriching environment. My gratitude also extends to Dr. Maulahikmah Galinium, S.Kom., M.Sc., as the Dean of the Faculty of Engineering and Information Technology, for the valuable support and encouragement provided throughout my academic journey.

Finally, I express my sincere appreciation to my family and friends for their continuous encouragement and understanding. This achievement would not have been possible without their unwavering support.

TABLE OF CONTENTS

	Page
Table of Contents	
TABLE OF CONTENTS.....	7
CHAPTER 1- INTRODUCTION	14
1.1. Background.....	14
1.2. Research Problem	19
1.3. Research Objective	19
1.4. Significance of Study	20
1.5. Research Question	20
1.6. Hypothesis	20
1.7. Scope of Study	21
CHAPTER 2- LITERATURE REVIEW	22
2.1. Sales Pipeline Activity.....	22
2.2. Big Data Analytic	24
2.2.1 Big Data Analytic Concept.....	24
2.2.2 Data Source	24
2.2.3 Big Data Analytics Implementation Stages for Demand Prediction	25
2.3. Data Quality Process.....	25
2.3.1 Data Profiling	26
2.3.2 Data Cleansing	26
2.3.3 Handling Missing Value.....	26
2.3.4 Outlier Handling.....	26
2.3.5 Data Integration.....	26
2.3.6 Data Validation and Governance	26
2.4. Machine Learning	27
2.4.1 Linear Regression.....	28
2.4.2 Naïve Bayes.....	30
2.4.3 Random Forrest	31

Eko Ichwan Ibrahim

2.4.4	Long Short-Term Memory	32
2.4.5	XGBoost	34
2.4.6	Model Comparison	35
2.5.	Coal Price Correlation to Heavy Equipment Sales	42
2.6.	Remanufacturing Component Business	43
2.7.	Platform Use	45
2.8.	Previous Study	46
CHAPTER 3- RESEARCH METHODOLOGY		56
3.1.	Research Limitation	21
3.2.	Research Design Framework	56
3.2.1	Define	56
3.2.2	Measure	59
3.2.3	Analyse	59
3.2.4	Improve	60
3.2.5	Control	60
CHAPTER 4- RESULT AND DISCUSSION		61
4.1.	Problem Discussion	61
4.2.	Architecture Design	70
4.3.	Data Collection	74
4.3.1	Data in Brief	74
4.3.2	Transaction Data	75
4.3.3	Machine Population Data	78
4.3.4	Sales Funnel Activity	79
4.3.5	Calendar Data	80
4.3.6	Commodity Data	81
4.3.7	Weather Data	83
4.4.	Big Data	90
4.4.1	Big Data Explanation	90
4.4.2	Big Data Architecture	92

4.4.3	Data Pipeline and Warehouse.....	93
4.4.4	Modelling and Prediction	93
4.5.	Data Pre-Processing	93
4.5.1	Data Cleansing	93
4.5.2	Data Spine	100
4.5.3	Feature Engineering	102
4.6.	Modelling Development	114
4.6.1	Modelling Development in Brief	114
4.6.2	Correlation Analysis	115
4.6.3	Model Comparison	119
4.6.4	Model Development	130
4.7.	Post processing	136
4.7.1	Recency, Frequency, Monitoring	136
4.7.2	Operation information handling	140
4.8.	Implementation	141
4.8.1	Implementation Process	141
4.9.	Result	142
4.9.1	Model Result in 2025	142
4.9.2	Customer Behaviour Analysis.....	149
CHAPTER 5- CONCLUSIONS AND RECOMMENDATIONS		160
5.1	Conclusion.....	160
5.2	Recommendation	162
5.2.1	Enhance Model Robustness and Accuracy:	162
5.2.2	Automate the Data Pipeline and Analytics Process:	162
5.2.3	Refine Customer Segmentation and Profiling:.....	163
5.2.4	Integrate with Supply Chain and Supplier Systems:	163
REFERENCES		165
CURRICULUM VITAE		170

LIST OF FIGURES

Figures	Page
Figure 1.1 United Tractors Business Revenues Contributor (United Tractors, 2024).	14
Figure 1.2 Coal Price Trend (Jun 2024 to Feb 2025).....	15
Figure 1.3 Komatsu Product Sales YoY Q3 2024 (United Tractors, 2024)	15
Figure 1.4 United Tractors Sales Performance 2024	16
Figure 1.5 Coal Price Projection 2025	17
Figure 1.6 Service Business Contributor	17
Figure 1.7 Revenue Reman Component Trend.....	18
Figure 2.1 - Sales Pipeline	22
Figure 2.2 Linear Regression (Gilbert Tanner, 2020).....	28
Figure 2.3 Naive Bayes (course-net, 2023)	30
Figure 2.4 Random Forest Illustration (Will Koehrsen, 2017)	32
Figure 2.5 LSTM Diagram (Nora Yehia, 2024)	33
Figure 2.6 XG Boost Chart Algorithm Flow Chart ((Hidayaturrohman and Hanada, 2024a)	34
Figure 2.7 Coal Price and Heavy Equipment Sales Correlation (<i>Source : Internal Company Research</i>)	43
Figure 2.8 Remanufacturing Component Business Overview.....	44
Figure 2.9 Big Data Platform.....	45
Figure 3.1 DMAIC Cycle.....	56
Figure 3.2 Empathy Map (hrbartender, 2020)	57
Figure 3.3 Customer Journey Mapping (Ravindra Dolare, 2020)	58
Figure 3.4 How Might We (thegreendotgroup, 2025)	58
Figure 4.1 - Concept Design	70
Figure 4.2 Architecture Design.....	72
Figure 4.3 Big Data Architecture	92
Figure 4.4 Model Development Process	115
Figure 4.5 Feature Correlation Analysis	117
Figure 4.6 XG Boost Feature Important	123
Figure 4.7 Naive Bayes Important Feature	127

Figure 4.8 Concept of Customer Segmentation	136
Figure 4.9 RFM Calculation	137
Figure 4.10 Post Processing Based on Operation Condition	140
Figure 4.11 Opportunity Analytic Implementation	141
Figure 4.12 MAPE & Accuracy Comparison 2024 and 2025	147
Figure 4.13 Coal Price Downtrend since 2024 to 2025	148
Figure 4.14 Reman's Customer Machine Population.....	150
Figure 4.15 Reman Customer's Sector.....	151
Figure 4.16 Customer Profile Based On RFM Category for Reman Transaction	153
Figure 4.17 Customer Profile Based On RFM Category for Parts Transaction	154
Figure 4.18 Combined Customer Profile (Viewed by RFM Parts & Reman Transaction)	156
Figure 5.1 Road Map of Optimizing Remanufacturing Component Business	164



LIST OF TABLES

Table	Page
Table 2.1 Machine Learning Type	28
Table 2.2 Model Comparison	35
Table 2.3 Model Comparison Summary	40
Table 2.4 Previous Study Summary	47
Table 4.1 Ideation	65
Table 4.2 Cluster Idea	66
Table 4.3 Cost & Benefit Formula	67
Table 4.4 Idea Validation	68
Table 4.5 Transaction Data	75
Table 4.6 Machine Population	78
Table 4.7 Sales Funnel Information	80
Table 4.8 Calendar Information	81
Table 4.9 Commodity Data Information	82
Table 4.10 Weather Data Information	84
Table 4.11 Different Unit Model Naming from SAP and Komtrax	95
Table 4.12 Normalized Unit Model Name	95
Table 4.13 Service Product List	96
Table 4.14 Surcharge Material Part Number List	97
Table 4.15 Transaction table before aggregation	102
Table 4.16 Transaction table after aggregation	102
Table 4.17 Base Table	107
Table 4.18 List of Feature in Base Table	107
Table 4.19 Table Summary of Data Variable	118
Table 4.20 Model Comparison	121
Table 4.21 Model Comparison Summary	130
Table 4.22 Feature Importance of XG Boost Modelling Development	132
Table 4.24 Testing Result	134
Table 4.25 RFM Scoring Index	137
Table 4.26 RFM for Reman Customer	139

Table 4.27 Closed Lost Feedback Activity	141
Table 4.28 - Modelling Accuracy in Plant Level	142
Table 4.29 Modeling Result in 2025	145
Table 4.30 Opportunity Source Comparison	145
Table 4.31 Reman Customer Characteristic Comparison	157
Table 4.32 Reman Customer Characteristic based on RFM for Parts Transaction Category	158
Table 4.33 Reman Customer Catagory with Combined RFM from Reman & Parts transaction behaviour	158

