



UNDERGRADUATE THESIS

PREDICTION OF BRONCHITIS DISEASE INDICATIONS USING CATBOOST ALGORITHM AND SHAPLEY ADDITIVE EXPLANATIONS WITH A RULE-BASED SYSTEM FOR DRUG RECOMMENDATIONS

LUKMAN HAKIM

NPM 21081010118

THESIS ADVISORS

Yisti Vita Via, S.ST., M.Kom.

Eva Yulia Puspaningrum, S.Kom., M.Kom.

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026**

APPROVAL SHEET

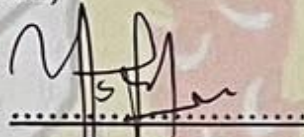
PREDICTION OF BRONCHITIS DISEASE INDICATIONS USING CATBOOST ALGORITHM AND SHAPLEY ADDITIVE EXPLANATIONS WITH A RULE-BASED SYSTEM FOR DRUG RECOMMENDATIONS

By:
LUKMAN HAKIM
NPM. 21081010118

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination
of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan
Nasional Veteran Jawa Timur, on April 17, 2026:

Approved,

Yisti Vita Via, S.ST., M.Kom.
NIP. 19860425 2021212 001



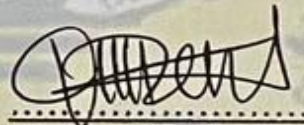
(Advisor I)

Eva Yulia Puspaningrum, S.Kom.,
M.Kom.
NIP. 19890705 2021212 002



(Advisor II)

Henni Endah Wahanani, ST. M.Kom.
NIP. 19780922 2021212 005



(Head Assessor)

Retno Mumpuni, S.Kom., M.Sc
NIP. 19870716 2025212 045



(Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science



Prof. Dr. Ir. Novirina Hendrasarie, MT..
NIP. 19681126 199403 2 001

APPROVAL SHEET

**PREDICTION OF BRONCHITIS DISEASE INDICATIONS USING CATBOOST
ALGORITHM AND SHAPLEY ADDITIVE EXPLANATIONS WITH A RULE-
BASED SYSTEM FOR DRUG RECOMMENDATIONS**

By:
LUKMAN HAKIM
NPM. 21081010118

Approved to proceed to the Thesis Examination

Approved by,

Coordinator of Informatics Study Program
Faculty of Computer Science



Dr. Intan Yuniar Purbasari, S.Kom. MSc.
NIP. 19800602 202521 2 029

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Lukman Hakim
NPM : 21081010118
Degree Program : Bachelor (S1)
Study Program : Informatics
Faculty : Faculty of Computer Science

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher education institution. Furthermore, it does not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.



Surabaya, June 13, 2026

Declarant,



LUKMAN HAKIM
NPM. 21081010118

ABSTRACT

Student Name / NPM: Lukman Hakim/21081010118
Thesis Title: Prediction of Bronchitis Disease Indications Using Catboost Algorithm And Shapley Additive Explanations With A Rule-Based System For Drug Recommendations
Supervisor: 1. Yisti Vita Via, S.ST., M.Kom
2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

Bronchitis is one of the most common respiratory diseases that requires accurate and timely medical treatment. However, the process of diagnosis and medication recommendation often relies heavily on manual analysis by medical personnel. Therefore, this study aims to develop a predictive model for bronchitis indications using the CatBoost algorithm and implement a rule-based drug recommendation system to support medical decision-making.

The research methodology follows the stages of machine learning implementation, including data collection from patient symptoms at Apotek RH Farma Surabaya January 2025 period until April 2025, data preprocessing, feature engineering, labeling, hyperparameter tuning, and model evaluation using Accuracy, Precision, Recall, and F1-Score metrics. The model was tested using three training-to-testing data ratios (60:40, 70:30, and 80:20).

The results show that the model with a 80:20 ratio achieved the best performance, with Accuracy 83%, Precision 75%, Recall 84%, dan F1-Score 79%. The SHAP (Shapley Additive Explanations) analysis revealed that symptoms such as coughing period and wheezing contributed the most to bronchitis prediction. Furthermore, the rule-based recommendation system successfully mapped patient symptoms to relevant Over-the-Counter (OTC) drug lists and provided consultation suggestions when bronchitis indications were detected. Based on these findings, the integration of the CatBoost prediction model and rule-based system proved effective in improving diagnostic accuracy and the relevance of drug recommendations.

Keywords: Machine Learning, CatBoost, SHAP, Bronchitis, Recommendation System, Rule-Based System

TABLE OF CONTENTS

COVER PAGE	1
APPROVAL SHEET.....	2
APPROVAL SHEET.....	3
STATEMENT OF ORIGINALITY	4
ABSTRACT.....	5
TABLE OF CONTENTS	1
IMAGE LIST	4
TABLE LIST.....	5
CHAPTER I INTRODUCTION	6
1.1 Background.....	6
1.2. Problem Formulation	8
1.3. Purpose	9
1.4. Benefits.....	9
1.5. Problem Limitations	9
CHAPTER II LITERATURE REVIEW	11
2.1 Previous Research.....	11
2.2 Data Mining	12
2.2.1 Data Selection	13
2.2.2 Data Cleaning.....	14
2.2.3 Transformation	14
2.2.4 Data Mining	14
2.2.5 Evaluation	14
2.3 ISPA Disease.....	15
2.4 Bronchitis.....	15
2.5 Drug Categories.....	17
2.6 OTC Drug Rules (Over The Counter).....	18
2.7 Prescription Drugs	20
2.8 Doctor Validation	22
2.9 Bronchitis Disease Content	22

2.10	Machine Learning.....	23
2.10.1	Supervised Learning	23
2.10.2	Unsupervised Learning	24
2.10.3	Semi-Supervised Learning and Reinforcement Learning	24
2.11	Label Encoding.....	25
2.12	Hyperparameter Tuning	25
2.13	GridSearchCV	26
2.14	Algoritma CatBoost	26
2.15	AUC-ROC.....	30
2.16	Confussion Matrix	30
2.17	Shapley Additive exPlanations (SHAP).....	32
2.18	Rule-Based System.....	33
2.19	Recommendation System.....	33
CHAPTER III SYSTEM DESIGN AND IMPLEMENTATION.....		35
3.1	Research Stages.....	35
3.2	Literature Studies.....	36
3.3	Data Collection.....	37
3.4	Praproses Data.....	39
3.4.1	Data Cleaning.....	39
3.4.2	Feature Engineering	40
3.4.3	Labeling	41
3.5	Model Implementation	42
3.5.1	Data Sharing.....	42
3.5.2	Hyperparameter Tuning	42
3.5.3	Catboost Algorithm Initiation.....	43
3.5.4	Model Evaluation	53
3.5.5	SHAP Interpretation	54
3.5.6	Medication Recommendations	56
3.6	Testing Scenarios.....	60
CHAPTER IV TESTING AND ANALYSIS.....		62
4.1	Data Collection.....	62
4.2	Data Cleaning.....	62
4.2.1	Data Duplication.....	63
4.2.2	Empty Value Cleaning.....	63

4.2.3	Removal of Unused Columns.....	64
4.3	Feature Engineering	65
4.4	Labeling	66
4.5	Data Sharing.....	67
4.6	Hyperparameter Tuning	71
4.7	Modeling.....	72
4.8	Model Evaluation	76
4.9	ROC-AUC.....	78
4.10	SHAP Interpretation	81
4.11	Medication Recommendations	85
4.12	Test Scenario Results.....	90
CHAPTER V CONCLUSION.....		94
5.1	Conclusion	94
5.1	Suggestions	95
BIBLIOGRAPHY		96

IMAGE LIST

Figure 2. 1 Data Mining Process	14
Figure 2. 2 Diagram Alur Algoritma CatBoost	28
Figure 3. 1 Research Stages	36
Figure 3. 2 Pre-Processing Data	40
Figure 3. 3 Model Implementation	43
Figure 3. 4 Decision tree CatBoost.....	53
Figure 4. 1 Dataset	63
Figure 4. 2 Data Duplication.....	64
Figure 4. 3 Empty Value Cleaning	65
Figure 4. 4 Dataframe After Deletion	66
Figure 4. 6 Data After Feature Engineering	67
Figure 4. 5 Labeling Target.....	68
Figure 4. 7 Data Sharing Ratio 60:40	70
Figure 4. 8 Data Sharing Ratio 70:30	71
Figure 4. 9 80:20 Ratio Data Sharing	72
Figure 4. 10 Best Hyperaparameters.....	73
Figure 4. 11 Models with 60:40 data split.....	75
Figure 4. 12 Models with 70:30 split data.....	75
Figure 4. 13 Models with 80:20 data split.....	76
Figure 4. 14 Probability Algorithm	77
Figure 4. 15 Model Evaluation.....	79
Figure 4. 16 Kurva ROC-AUC.....	81
Figure 4. 17 SHAP Interpretation.....	83
Figure 4. 18 Value Mean SHAP for Profit.....	84
Figure 4. 19 Graphics Summary Plot SHAP	85
Figure 4. 20 Graphics Initjs SHAP	86

TABLE LIST

Table 2. 1 Confusion Matrix	30
Table 3. 1 Data Attributes	37
Table 3. 2 Data Duplication	40
Table 3. 3 Blank Values.....	40
Table 3. 4 Data Consistency	40
Table 3. 5 RH Farma Pharmacy Patient Prescription Data.....	41
Table 3. 6 Hasil Encoding Label	41
Table 3. 7 Sample data of RH Farma Pharmacy Patients.....	44
Table 3. 8 Residual Value Results.....	45
Table 3. 9 Sequence of patients by Label and Symptoms	46
Table 3. 10 TS Ordered Results	48
Table 3. 11 Table distribution by gender	49
Table 3. 12 Table distribution by cough period.....	50
Table 3. 13 Table distribution by Shortness of Breath.....	50
Table 3. 14 Table distribution by Fever	51
Table 3. 15 Confusion matrix.....	54
Table 3. 16 Example of SHAP Absolute Value Result Simulation.....	55
Table 3. 17 Symptom Mapping.....	57
Table 3. 18 Examples of Drug Recommendation Results	58
Table 3. 19 Split Data Based Test Scenario Table.....	60
Table 3. 20 Drug Recommendation Scenario Table	61
Table 4. 1 Confusion Matrix	79
Table 4. 2 Test Scenarios Based on Data Split.....	91
Table 4. 3 Drug Recommendation Scenario.....	92