



UNDERGRADUATE THESIS

A COMPARISON OF THE PERFORMANCE OF THE XGBOOST AND CATBOOST ALGORITHMS IN CLASSIFYING THE LEVELS OF DEPRESSION AMONG COLLEGE STUDENTS

KHAIRUNNISA

NPM 22081010325

THESIS ADVISORS

Dr. Faisal Muttaqin, S.Kom, M.T

Chrystia Aji Putra, S.Kom, M.T

MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FACULTY OF COMPUTER SCIENCE

INFORMATICS STUDY PROGRAM

SURABAYA

2026

APPROVAL SHEET

A COMPARISON OF THE PERFORMANCE OF THE XGBOOST AND CATBOOST ALGORITHMS IN CLASSIFYING THE LEVELS OF DEPRESSION AMONG COLLEGE STUDENTS

By:

KHAIRUNNISA
NPM. 22081010325

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 July 9, 2026:

Approved,

Dr. Faisal Muttagin, S.Kom, M.T
NIP. 19851231 2021211 009



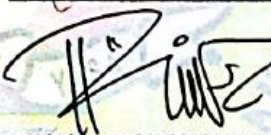
(Advisor I)

Chrystia Aji Putra, S.Kom, M.T
NIP. 19861008 2021211 001



(Advisor II)

Dr. Rizky Parlika, S.Kom, M.Kom.
NIP. 19840518 2021211 003



(Head Assessor)

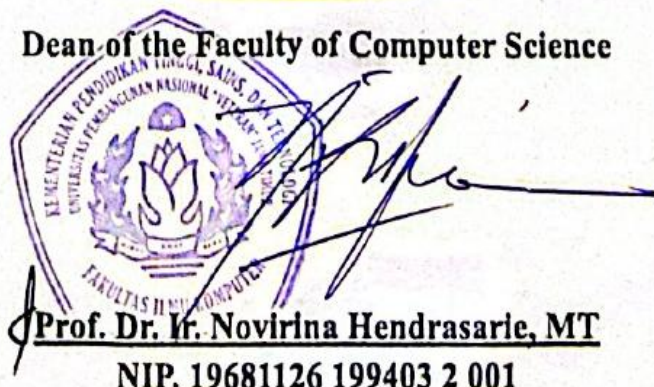
Achmad Junaidi, S.Kom, M.Kom
NIP. 197811102025211048



(Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science



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

APPROVAL PAGE

**A COMPARISON OF THE PERFORMANCE OF THE XGBOOST AND
CATBOOST ALGORITHMS IN CLASSIFYING THE LEVELS OF
DEPRESSION AMONG COLLEGE STUDENTS**

By:

**KHAIRUNNISA
NPM. 22081010325**

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 : Khairunnisa
NPM : 22081010325
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, August 31, 2026
Declarant,



KHAIRUNNISA
NPM. 22081010325

ABSTRACT

Student Name / NPM : Khairunnisa / 22081010325

Thesis Title : A Comparison Of The Performance Of The Xgboost And Catboost Algorithms In Classifying The Levels Of Depression Among College Students

Supervisor : 1. Dr. Faisal Muttaqin, S.Kom, M.T
2. Chrystia Aji Putra, S.Kom, M.T

Students' psychological health and academic performance may be impacted by depression, a mental health condition. This study compares the effectiveness of the XGBoost and CatBoost algorithms and attempts to categorize the depression levels of students. 306 senior undergraduate students made up the data used, and the PHQ-9, GAD-7, and ERQ instruments provided the characteristics. Four categories of depression were identified: normal, mild, moderate, and severe depression. Three data splitting scenarios (80:20, 70:30, and 60:40) were used in the investigation, coupled with SMOTE-NC and GridSearchCV hyperparameter tuning. With accuracy of 0.89, precision of 0.91, recall of 0.81, F1-score of 0.84, and AUC of 0.97, the findings demonstrated that XGBoost performed best at the 80:20 ratio. CatBoost, on the other hand, had an accuracy of 0.85 and an AUC of 0.98 at the 60:40 ratio. The best model for categorizing students' depression levels, according to the total evaluation results, was XGBoost with an 80:20 ratio.

Keywords: College Student Depression, XGBoost, CatBoost, GridSearchCV, SMOTE-NC

ACKNOWLEDGMENTS

The author gives thanks to Allah SWT for all His grace, blessings, and guidance, which have enabled the author to successfully complete this thesis titled **A Comparison of the Performance of the Xgboost and Catboost Algorithms in Classifying Students' Levels of Depression**. This thesis was written as one of the requirements for obtaining a Bachelor's degree in the Computer Science Program.

The completion of this thesis would not have been possible without the help, guidance, support, and prayers of many people. Therefore, the author would like to take this opportunity to express his deepest gratitude to:

1. Allah SWT, who has consistently granted me health, strength, ease, and smooth progress at every stage of the preparation of this thesis.
2. Dr. Ir. Novirina Hendrasarie, M.T., Dean of the Faculty of Computer Science, UPN "Veteran" East Java.
3. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., Coordinator of the Informatics Study Program, Faculty of Computer Science, UPN "Veteran" East Java.
4. M. Muharrom Al Haromainy, S.Kom., M.Kom., the author's academic advisor, who provided direction, guidance, and support throughout the author's academic journey until the completion of his studies.
5. Dr. Rizky Parlika, S.Kom., M.Kom., and Achmad Junaidi, S.Kom., M.Kom., as the author's thesis advisors, who provided constructive and valuable criticism, suggestions, and feedback to improve this thesis. Thank you for your time, attention, and knowledge shared with the author.
6. Dr. Faisal Muttaqin, S.Kom., M.T., and Chrystia Aji Putra, S.Kom., M.T., as the author's academic advisors, who provided guidance, direction, input, and support throughout the research process until the completion of this thesis. Thank you for the knowledge, time, and patience you have shared with the author.
7. All faculty members at UPN "Veteran" East Java, who have provided knowledge, guidance, and valuable experience throughout the author's

academic journey. Thank you for your dedication and sincerity in supporting the author throughout this academic journey to this point.

8. To the author's first love, my beloved father, the late Mr. Suyono. Thank you for all the dedication, love, sacrifice, and prayers you gave the author during your lifetime. Although you were unable to accompany me to the end of this journey, every piece of advice, every expression of love, and every example you left behind will always be a source of strength and motivation for me. I dedicate this thesis as an expression of love and respect for you. May all your good deeds and acts of worship be accepted by Allah SWT and may you be placed in the best of places. Al-Fatihah.
9. To the gateway to my heaven, my beloved mother, Siti Maemunah. Thank you for all the prayers, love, sacrifices, and unwavering support you have given me. Thank you for being my safe haven, my source of strength, and the person who has always believed in me and prayed for every step I take. Every achievement the author has attained is inseparable from Mother's prayers and sincerity. May Mother always be granted good health, happiness, a long life, and blessings in every step she takes. The author dedicates this thesis as a token of boundless love and gratitude to Mother.
10. To the author's two older brothers, my beloved brothers Abdul Fariz and Rafi Muadz. Thank you for always giving me love, care, support, and encouragement. Thank you for being there to share stories, offer advice, and accompany me through both good and difficult times. Your presence has been a source of strength throughout this journey. May we always be blessed with happiness and good health, and continue to support one another in every step of life.
11. To someone special to the author, M. Rizqi Raharjo. Thank you for being there and accompanying me from the start of college until now. Thank you for all the support, encouragement, patience, and love you've always given me, through good times and bad. Your presence has made my journey in this foreign land feel lighter and more meaningful. Thank you for being a home to me in the midst of this long journey. May we always be able to strengthen one another and move forward together toward a better future.

12. To the author's friends: Hana, Eca, Pia, Shanty, and Chatrine. Thank you for every laugh, story, word of support, and moment of togetherness that has enriched the author's college journey. Thank you for being there through every joy and sorrow and for being a place to share stories during our college years. Your presence has made this journey feel lighter and more memorable. May this friendship endure, even as we eventually go our separate ways.
13. To the author's friends from elementary and middle school. Thank you for being a part of the author's long life journey. Although time and our busy lives have led us down our own paths, every story, laugh, and moment we shared will always be a cherished memory. May our bonds and friendship endure forever.
14. To all my friends whom I cannot name one by one. Thank you for every moment we spent together, for your help, support, and the stories that have been part of my college journey. Every moment we've shared will always be a cherished memory. May all the kindness you've shown be repaid with good things, and may we all achieve success on our respective paths.
15. Last but not least, to myself. Thank you for persevering through all the exhaustion, doubts, tears, and challenges. Thank you for choosing to keep going, even when giving up felt easier. I'm proud to have reached a point that was once only a hope. May I always have the courage to dream, the strength to fight, and a sincere heart in embracing every process of life. Thank you for fighting—you're amazing for making it this far.

In closing, the author hopes that this thesis will be of benefit to readers, future researchers, and all stakeholders in the advancement of science, particularly in the fields of digital image processing, computer vision, and deep learning.

Surabaya, Agustus 31th 2026

Author

TABLE OF CONTENTS

COVER PAGE	i
APPROVAL SHEET	ii
APPROVAL PAGE	iii
STATEMENT OF ORIGINALITY	iv
ACKNOWLEDGMENTS	vi
TABLE OF CONTENTS	ix
LIST OF FIGURES	xii
LIST OF TABLES	xiv
CHAPTER I INTRODUCTION.....	1
1.1 Background	1
1.2 Problem Statement	4
1.3 Research Objectives	5
1.4 Research Benefits.....	5
1.5 Scope of the Study	6
CHAPTER II LITERATURE REVIEW	7
2.1 Previous Research	7
2.2 Theoretical Framework	13
2.2.1 Depression.....	13
2.2.2 The Relationship Between Anxiety and Emotional Regulation and Depression in College Students	15
2.2.3 Patient Health Questionnaire-9 (PHQ-9)	16
2.2.4 Generalized Anxiety Disorder-7 (GAD-7).....	19
2.2.5 Emotion Regulation Questionnaire (ERQ)	21
2.2.6 Data Preprocessing.....	23
2.2.7 Machine Learning	23
2.2.8 Classification.....	24
2.2.9 SMOTE	25
2.2.10 Encoding	26
2.2.11 Boosting	27

2.2.12	GridSearch CV	28
2.2.13	Extreme Gradient Boosting (XGBoost)	30
2.2.14	CatBoost	34
2.2.15	K-Fold Cross Validation	38
2.2.16	ROC-AUC	38
2.2.17	SHAP Analysis	39
2.2.18	Confusion Matrix.....	39
2.2.19	Streamlit	41
CHAPTER III RESEARCH METHODOLOGY.....		43
3.1	Research Phases.....	43
3.2	Literature Review	43
3.3	Data Collection	44
3.4	Data Preprocessing	46
3.4.1	Data Cleaning	47
3.4.2	Data Transformation	47
3.4.3	Label Assignment	48
3.5	Data Distribution	49
3.6	Modeling.....	50
3.6.1	SMOTE.....	50
3.6.2	Hyperparameter Tuning Initialization	50
3.6.3	K-fold Cross Validation	51
3.6.4	XGBoost Modeling	51
3.6.5	CatBoost Modeling.....	57
3.7	Evaluation.....	64
3.7.1	Testing Scenarios	66
3.7.2	Selection of the Best Model	67
3.8	Website Implementation	68
CHAPTER IV RESULTS AND DISCUSSION		70
4.1	Data Collection	70
4.2	Data Cleaning	70
4.2.1	Handling Missing Values.....	70
4.2.2	Removal of Unused Columns.....	71

4.3	Separating Feature and Target Data	71
4.4	Exploratory Data Analysis (EDA)	72
4.4	Dataset Transformation	73
4.5	Dataset Splitting	75
4.7	Balancing Data	76
4.8	Modeling	78
4.8.1	XGboost	78
4.8.2	CatBoost	80
4.9	Selection of the Best Parameters and Model	81
4.9.1	XGBoost	81
4.9.2	CatBoost	83
4.10	Model Evaluation	84
4.10.1	XGBoost	84
4.10.2	CatBoost	86
4.10.3	Test Scenario Results	90
4.10.4	Best Model Selection and SHAP Analysis	91
4.11	Website Implementation	93
CHAPTER V CONCLUSIONS AND RECOMMENDATIONS		95
5.1	Conclusion	95
5.2	Recommendations	96
REFERENCES		97
APPENDIX		104

LIST OF FIGURES

Figure 2. 1 Machine learning.....	24
Figure 2. 2 Classification.....	25
Figure 2. 3 Boosting	28
Figure 2. 4 The XGBoost Algorithm Block Diagram.....	31
Figure 2. 5 CatBoost Workflow	35
Figure 3. 1 Research Phases	43
Figure 3. 2 Data Preprocessing	46
Figure 3. 3 Modeling	50
Figure 3. 4 XGBoost Model	51
Figure 3. 5 The Catboost Method	57
Figure 3. 6 Decision Tree CatBoost.....	62
Figure 3. 7 Evaluation Stages	64
Figure 3. 8 Wireframe of the Website for Classifying Students' Depression Levels	69
Figure 4. 1 Data after column deletion is irrelevant	71
Figure 4. 2 Feature and Target Separation	72
Figure 4. 3 Class Distribution of Students' Depression Levels	73
Figure 4. 4 Encoding Results for Categorical Features	74
Figure 4. 5 Encoding Results for the Target Feature	75
Figure 4. 6 Data Results Before and After SMOTE 80:20	77
Figure 4.7 Data Results Before and After SMOTE 70:30	77
Figure 4. 8 Data Results Before and After SMOTE 60:40	78
Figure 4. 9 Optimal XGBoost parameters for a 60:40 data split	82
Figure 4. 10 Optimal XGBoost parameters for a 70:30 data split	82
Figure 4. 11 Optimal XGBoost parameters for an 80:20 data split	82
Figure 4. 12 Optimal CatBoost parameters for a 60:40 data split	83
Figure 4. 13 The optimal CatBoost parameters for a 70:30 data split.....	83
Figure 4. 14 Optimal CatBoost parameters for an 80:20 data split	83
Figure 4. 15 Confusion matrix XGBoost 80:20.....	85

Figure 4. 16 Classification report XGBoost 80:20.....	86
Figure 4. 17 Confusion matrix CatBoost 80:20	87
Figure 4. 18 Classification report CatBoost 80:20	88
Figure 4. 19 Feature Importance XGBoost	89
Figure 4. 20 SHAP Visualization	92
Figure 4. 21 Website Interface for the Classification of Students' Depression Levels	93

LIST OF TABLES

Table 2. 1 Summary of Previous Research	7
Table 2. 2 PHQ-9 Questions	18
Table 2. 3 GAD-7 Questions.....	20
Table 2. 4 ERQ Questions.....	22
Table 2. 5 Hyperparameter XGBoost.....	29
Table 2. 6 Hyperparameter CatBoost.....	30
Table 3. 1 Dataset Attributes	45
Table 3. 2 Initial dataset	47
Table 3. 3 Transformation dataset	48
Table 3. 4 Depression Severity Categories	48
Table 3. 5 Data Labeling	49
Table 3. 6 Initial data.....	52
Table 3. 7 one-hot encoding.....	53
Table 3. 8 Gradient and Hessian Results.....	53
Table 3. 9 Gradient and Hessian of feature X1	54
Table 3. 10 Example of Left-Hand Leaf Data.....	57
Table 3. 11 Example of Right-Hand Leaf Data.....	57
Table 3. 12 Example of depression data	58
Table 3. 13 Examples of depression data	59
Table 3. 14 OTS Results	60
Table 3. 15 Data Gain	61
Table 3. 16 Gain Data Based on Age	61
Table 3. 17 Example of a Confusion Matrix	64
Table 3. 18 Hyperparameters and Value Ranges.....	67
Table 4. 1 Test Scenario Results	90