



UNDERGRADUATE THESIS

APPLICATION OF EXPLAINABLE BOOSTING MACHINE TO BEST MATCHING 25 SEARCH RESULTS FOR BOOK QUALITY RECOMMENDATION WITH FEATURE-LEVEL EXPLANATIONS

BAGUS SATRIO WICAKSONO

NPM 22081010252

THESIS ADVISORS

Dr. Intan Yuniar Purbasari, S.Kom, MSc

Henni Endah Wahanani, ST, 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

APPLICATION OF EXPLAINABLE BOOSTING MACHINE TO BEST MATCHING 25 SEARCH RESULTS FOR BOOK QUALITY RECOMMENDATION WITH FEATURE-LEVEL EXPLANATIONS

By:
BAGUS SATRIO WICAKSONO
NPM. 22081010252

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 February 24, 2026:

Approved,

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



(Advisor I)

Henni Endah Wahanani, ST, M.Kom.
NIP. 19780922 202121 2 005



(Advisor II)

Dr. Faisal Muttagin, S.Kom, M.T
NIP. 19851231 202121 1 009



(Head Assessor)

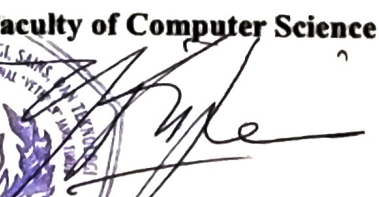
Muhammad Muharrom Al Haromainy,
S.Kom, M.Kom
NIP. 19950601 202203 1 006



(Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science


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

APPROVAL SHEET

APPLICATION OF EXPLAINABLE BOOSTING MACHINE TO BEST MATCHING 25 SEARCH RESULTS FOR BOOK QUALITY RECOMMENDATION WITH FEATURE-LEVEL EXPLANATIONS

By:

BAGUS SATRIO WICAKSONO

NPM. 22081010252

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 : Bagus Satrio Wicaksono
NPM : 22081010252
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, July 20, 2026
Declarant,



BAGUS SATRIO
WICAKSONO
NPM. 22081010252

ABSTRACT

Student Name / NPM : Bagus Satrio Wicaksono / 22081010252
Thesis Title : APPLICATION OF EXPLAINABLE BOOSTING
MACHINE TO BEST MATCHING 25 SEARCH
RESULTS FOR BOOK QUALITY
RECOMMENDATION WITH FEATURE-
LEVEL EXPLANATIONS
Advisor : 1. Dr. Intan Yuniar Purbasari, S.Kom, MSc
2. Henni Endah Wahanani, ST., M.Kom.

ABSTRACT

The rapid growth of digital book collections creates significant challenges for readers identifying high-quality books during exploratory search. Although Best Matching 25 (BM25) effectively retrieves textually relevant books, it ignores quality indicators like popularity and ratings, while existing re-ranking models remain uninterpretable black boxes. This study proposes a two-stage ranking architecture deploying an Explainable Boosting Machine (EBM) as a glass-box re-ranker to optimize BM25 search results while providing intrinsic explanations. The EBM was trained on 13.4 million user-book interactions from the UCSD Goodreads dataset. To evaluate ranking quality and feature impacts, EBM variants—evaluating main effects (int0) and five feature interactions (int5)—were tested. Prediction evaluations showed the int0 model achieved an RMSE of 0.9782 and an MAE of 0.7632. Ranking evaluations across 50 test queries demonstrated that the int5 model achieved an NDCG@10 of 0.9373, outperforming the BM25 baseline (0.9284) on exploratory queries. Specifically, the framework successfully enhanced ranking quality by elevating candidate books that exhibit strong popularity signals. Human judgment evaluation revealed a marginal 0.05 decrease in Precision@1 (0.80 for EBM vs. 0.85 for BM25), presenting a valid trade-off where EBM filters out text-heavy box sets in favor of community-validated literature. Global explanations identified popularity as the most influential individual feature, while local explanations directly corresponded to global shape functions. An ablation study further confirmed explanation faithfulness. This framework enhances BM25 retrieval quality through verified quality re-ranking while maintaining transparent, interconnected explanations to support exploratory search.

Keywords: book search, Goodreads, Best Matching 25, Explainable Boosting Machine, interpretable machine learning.

ACKNOWLEDGEMENTS

Praise be to Allah SWT for all His graces, guidance, and gifts to the author so that the thesis proposal with the title "**Application of Explainable Boosting Machine to Best Matching 25 Search Results for Book Quality Recommendation with Feature-Level Explanations**" can be completed properly.

In addition, during the preparation of the thesis proposal, the author also received a lot of assistance from various parties. For this the author would like to thank the following:

1. Prof. Dr. Ir. Novirina Hendrasarie, M.T., Dean of the Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" East Java.
2. Dr. Intan Yuniar Purbasari, S.Kom., MSc., as the Coordinator of Informatics Study Program, Faculty of Computer Science, National Development University of East Java. As author's first advisor, my gratitude sent for her invaluable guidance, encouragement, and continuous support throughout the completion of this thesis.
3. The lecturers of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" East Java, for their knowledge and guidance throughout my academic journey.
4. Mrs. Henni Endah Wahanani, ST, M.Kom., as author's second advisor. Words cannot fully express my gratitude for her guidance, and kindness throughout this research. Her advice, guidance, and help lead me to my thesis defense approval. May God always bless her with good health and happiness.
5. Author's dearest Father, who has always played an important role in the author's journey. Thanks to his prayer and upbringing, I have been given a lot of freedom for choosing what's good for me. His comforting words, advice, and encouragement have always given me the strength to overcome every challenge throughout this education journey. Thank you very much for everything.

6. Author dearest Mother who leaves too soon. I always believe that her love and prayers continue to accompany me. I hope she is smiling proudly as she witnesses this achievement from heaven. Thank you Mi.
7. Love of my life, Ammara Nayla Fauzi, who always back me up in every step and struggles. Her love, patience, and support played a big role in this thesis completion. Everytime i frustated or overwhelmed when working on this thesis, her calming voice and words did cool me down. We spent a lot of time stressing and laughing together, thank you for being my love. I love you forever, let's chase our dreams together.
8. Author's college friends, "Empang"; Akbar , M, Aziz, Rajawali, Bayu, and others, which im grateful for spending this incredible 4 years with you guys. My house, Wiguna, it will miss the atmosphere when we do our homework, projects, and watching movies. So long my friends, thanks.
9. Author's highschool friends, "Wes Life"; Oddie, Rey, Brian, Fitra, Boyil, Faros. Thanks for the fun, support and prayers, we've been a good friends for a long long time. I hope all of you achieve everything you are striving for.
10. The greatest K-pop idol of all time, TWICE. I would like to thank them for bringing joy, comfort, and motivation throughout the past five years. Your songs; Up No More, Be as One, What is Love, meant a lot to me. Now that you all are hiatus, i just hope to watch your concert one more time, Thanks.
11. Other parties, friends, and families. Thank you for the support.

The author realizes that in the preparation of the following thesis there are many shortcomings. Finally, with all the limitations that the author has, hopefully the following report can be useful for all parties in general and the author in particular.

Surabaya, July 16th 2026

BAGUS SATRIO WICAKSONO

TABLE OF CONTENTS

APPROVAL SHEET	ii
APPROVAL SHEET	iii
STATEMENT OF ORIGINALITY	iv
ABSTRACT	v
ACKNOWLEDGEMENTS.....	vi
TABLE OF CONTENTS.....	viii
LIST OF FIGURES	x
LIST OF TABLES	xii
LIST OF CODES	xiv
CHAPTER I INTRODUCTION.....	1
1.1 Background	1
1.2 Research Questions	4
1.3 Research Objectives	4
1.4 Research Contributions	4
1.5 Scope and Limitations.....	5
CHAPTER II LITERATURE REVIEW	6
2.1 Previous Studies	6
2.2 Theoretical Framework	9
2.2.1 Explainable Boosting Machine	9
2.2.2 Information Retrieval	11
2.2.3 Digital Library.....	13
2.2.4 Okapi Best Matching 25.....	13
2.2.5 Explainable AI.....	15
2.2.6 Evaluation.....	17
CHAPTER III METHODOLOGY.....	20
3.1 System Design and Architecture	20
3.2 Stage 1: Data Collection and Preparation	22
3.2.1 Data Reduction.....	25
3.2.2 Data Cleaning.....	27
3.2.3 Corpus Preparation for BM25	27

3.2.4	Feature Engineering EBM.....	28
3.3	Indexing and Model Training.....	31
3.3.1	Stage 2: BM25 Indexing	32
3.3.2	Stage 3: EBM Training	34
3.4	Two-Stage Ranking.....	38
3.4.1	Tahap 4: Retrieval BM25.....	39
3.4.2	Stage 5: EBM Re-ranking	41
3.5	Stage 6: Local and Global Explanation.....	43
3.5.1	Local Explanation	44
3.5.2	Global Explanation.....	44
3.6	Stage 7: System Testing and Evaluation	45
3.6.1	Hyperparameter.....	45
3.6.2	Ranking Evaluation Scenarios	46
3.6.3	EBM Prediction Accuracy Evaluation	48
3.6.4	Ranking Quality Evaluation.....	48
3.6.5	Local and Global Explanation Analysis.....	49
CHAPTER IV RESULT AND DISCUSSION.....		51
4.1	System Implementation Results	51
3.1.1	Stage 1: Data Collection and Preparation	51
3.1.2	Stages 2–3: BM25 Indexing and EBM Training Results.....	55
3.1.3	Stages 4–5: Two-Stage Ranking	58
3.1.4	Stage 6: EBM Feature Explanations	67
4.2	Discussion	75
4.2.1	UCSD Dataset Distributio.....	75
4.2.2	Testing Scenarios and Evaluation	76
4.2.3	Feature Explanations Provided by EBM.....	80
CHAPTER V CONCLUSION		84
5.1	Conclusion.....	84
5.2	Future Work	85
REFERENCES.....		86

LIST OF FIGURES

Figure 2.1 Two Stage Ranking Flowchart	11
Figure 3.1 Flowchart System	20
Figure 3.2 Data Preparation Workflow	22
Figure 3.3 Indexing and Training Workflow	31
Figure 3.4 BM25 Indexing Process.....	32
Figure 3.5 EBM Training Process.....	34
Figure 3.6 Search System Pipeline.....	38
Figure 3.7 BM25 Retrieval Flow	39
Figure 3.8 Reranking EBM	41
Figure 3.9 Local Explanation Graphic	44
Figure 3.10 Global Explanation Graphic	45
Figure 3.11 Shape function	45
Figure 4.1 Dataset Loading Results	51
Figure 4.2 Distribution of User Ratings.....	52
Figure 4.3 Distribution of Average Book Ratings	52
Figure 4.4 Results of Numerical Feature Transformation.....	54
Figure 4.5 Data Splitting Results	54
Figure 4.6 Top-10 Genre Identification	55
Figure 4.7 Characteristics of the BM25 Corpus.....	56
Figure 4.8 Prediction Distribution of the EBM int0 Model	57
Figure 4.9 Prediction Distribution of the EBM int5 Model	58
Figure 4.11 BM25 Candidate Books.....	60
Figure 4.12 Precision at 100 from BM25.....	60
Figure 4.13 BM25 Candidate Books.....	61
Figure 4.14 Re-ranked Candidates Using EBM int0.....	62
Figure 4.15 Re-ranked Candidates Using EBM itn5.....	62
Figure 4.16 Alpha 0 Result	63
Figure 4.18 Alpha 1 Result	64
Figure 4.19 BM25 cold-start candidates	64

Figure 4.20 Rerank cold-start candidates	65
Figure 4.21 Open Library Test Queries	65
Figure 4.22 Example of Re-ranked Results	67
Figure 4.23 Local Explanation of a Book	68
Figure 4.24 Global Feature Importance	69
Figure 4.25 Global Feature Interaction Importance	70
Figure 4.26 Shape Function of the Popularity Feature.....	71
Figure 4.27 Shape Function of the Genre Fiction Feature	72
Figure 4.28 Shape Function of the Genre History Feature	72
Figure 4.29 Shape Function of the Number of Pages Feature	73
Figure 4.30 Shape Function Genre non fiction	74

LIST OF TABLES

Table 3.1 Raw Dataset Summary	23
Table 3.2 Book Metadata	23
Table 3.3 Genre Data	24
Table 3.4 User–Book Interaction Records	24
Table 3.5 Primary Dataset for External Evaluation	25
Table 3.6 Data Reduction Results	26
Table 3.7 Example of Text Concatenation	27
Table 3.8 Text Preprocessing and Tokenization Pipeline	27
Table 3.9 Calculation of Proportional Genre Weights	29
Table 3.10 Structural Layout of the Integrated Interaction-Metadata Table.....	30
Table 3.11 Example of an Inverted Index	33
Table 3.12 EBM Feature Matrix (X) and Target Variable (y)	34
Tabel 3.13 Example of User-Item Interaction Input Matrix for EBM	35
Tabel 3.14 Example of Feature Binning	36
Table 3.15 Candidate Documents for Query Q.....	40
Table 3.16 Example of BM25 Search Output	41
Table 3.17 Example of Rank Shifting Post EBM Re-ranking	43
Table 3.18 Structural Layout of the EBM Re-ranking Output.....	43
Table 3.19 Test Query Configuration	46
Table 3.20 Experimental Evaluation Scenarios	47
Table 3.21 Hybrid Scoring Parameter Weights	47
Table 3.22 Graded Relevance Labels.....	49
Table 4.1 Example of Tokenization Results	53
Table 4.2 EBM Prediction Performance	56
Table 4.3 Overall Evaluation Results.....	59
Table 4.4 Scenario B Results	61
Table 4.5 Average NDCG@10 by Query Type	61
Table 4.6 Mean Evaluation Metrics for Scenario D.....	64
Table 4.7 NDCG Comparison on the Open Library Dataset	66
Table 4.8 User Relevance Assessment.....	66

Table 4.9 Precision@1 Based on User Judgments.....	67
Table 4.10 Feature Interaction Contributions	69
Table 4.11 Feature Ablation Results	74

LIST OF CODES

Code 3.1 BM25 retrieval function.....	39
Code 4.1 Merging Genre Information with Book Metadata	53
Code 4.2 Genre Encoding	55