

PROJECT REPORT
COMPARISON BETWEEN KNN ALGORITHM
AND RANDOM FOREST ALGORITHM TO
PREDICT STUDENT GRADUATION



ROMUALDUS MANTARI PADANG ALLO
16.K1.0018

Informatics Engineering Study Program
Faculty of Computer Science
Soegijapranata Catholic University
2021

PROJECT REPORT
COMPARISON BETWEEN KNN ALGORITHM
AND RANDOM FOREST ALGORITHM TO
PREDICT STUDENT GRADUATION

Submitted in Compliance Order
One of the Requirements to Get
Degree S.Kom



ROMUALDUS MANTARI PADANG ALLO
16.K1.0018

Informatics Engineering Study Program
Faculty of Computer Science
Soegijapranata Catholic University
2021

HALAMAN PENGESAHAN



Judul Tugas Akhir: : COMPARISON BETWEEN KNN ALGORITHM AND RANDOM FOREST ALGORITHM TO PREDICT STUDENT GRADUATION

Diajukan oleh : Romualdus Mantari P A

NIM : 16.K1.0018

Tanggal disetujui : 28 Januari 2021

Telah setuju oleh

Pembimbing : Rosita Herawati S.T., M.I.T.

Penguji 1 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Penguji 2 : Rosita Herawati S.T., M.I.T.

Penguji 3 : Hironimus Leong S.Kom., M.Kom.

Penguji 4 : Y.b. Dwi Setianto

Penguji 5 : Yonathan Purbo Santosa S.Kom., M.Sc

Ketua Program Studi : Rosita Herawati S.T., M.I.T.

Dekan : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Halaman ini merupakan halaman yang sah dan dapat diverifikasi melalui alamat di bawah ini.

sintak.unika.ac.id/skripsi/verifikasi/?id=16.K1.0018

STATEMENT OF ORIGINALITY

I, the undersigned:

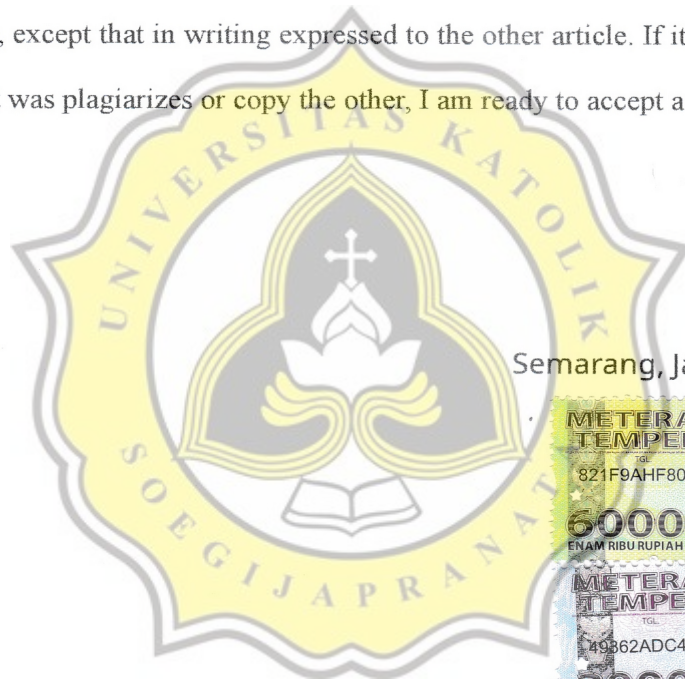
Name : ROMUALDUS MANTARI PADANG ALLO

Nim : 16.K1.0018

Progdi : Informatics Engineering

Faculty : Computer Science

Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.



Semarang, January, 29, 2021



ROMUALDUS MANTARI PADANG ALLO
16.K1.0018

**SCIENTIFIC PUBLICATION STATEMENT PAGE
FOR ACADEMIC INTEREST**

I, the undersigned:

Name : ROMUALDUS MANTARI PADANG ALLO

Nim : 16.K1.0018

Progdi : Informatics Engineering

Faculty : Computer Science

Agree to give Soegijapranata Catholic University Semarang the right to non-exclusive royalty free for a scientific work entitled "Comparison Between KNN Algorithm and Random Forest Algorithm To Predict Student Graduation" along with existing devices. With this Non-exclusive Royalty Free Right, Soegijapranata Catholic University has the right to save, transfer / format media, manage in the form of a database, maintain, and publish this final project as long as I still include my name as the author / creator and as the copyright owner.

This statement I made in truth

Semarang, January, 29, 2021



ROMUALDUS MANTARI PADANG ALLO

16.K1.0018

PREFACE

The purpose of writing this thesis is to fulfill some of the requirements for obtaining a bachelor's degree in education for undergraduate students in the Informatics Engineering study program, Soegijapranata Catholic University, Semarang. The author realizes that this thesis paper is still far from perfect, therefore the author expects constructive criticism and suggestions from all parties for the perfection of this thesis.

The end of this thesis paper cannot be separated from the help of various parties, so that the author on this occasion with all humility and respectfully expresses his deepest gratitude to all parties who have been involved and provide moral and material assistance directly or indirectly to the author In the preparation of this thesis paper to completion, especially to those I respect:

1. Mrs. Rosita Herawati S.T., M.I.T. as the supervisor who has given enthusiastic support, and never gets tired of being asked for guidance from the beginning of the title formation to the end of the thesis report.
2. Especially the family of the Faculty of Computer Science, who from the beginning of the study to the end, always provide knowledge that is very, very useful for writers.

3. Especially for the author's parents, who always support in prayer, and sacrifice from a moral and material perspective so that the author can complete this thesis.
4. For all 2016 Informatics Engineering friends who have accompanied and supported them during the lecture period until the thesis writing is complete.

Finally, the authors would like to thank all those who have helped and the authors hope that this thesis paper can be useful and useful for all of us and become input for the world of education.



Semarang, January, 29, 2021

ROMUALDUS MANTARI PADANG ALLO

16.K1.0018

ABSTRACT

This study aims to compare the results of the accuracy of the KNN Algorithm and the Random Forest Algorithm based on the classification of student data from the student achievement index and student graduation status. The algorithms used in data classification are KNN (K-Nearest Neighbor) and Random Forest. The data used in this study are dummy data. For data processing, the KNN algorithm uses a split data method. Meanwhile, for data processing the Random Forest Algorithm uses car k folds cross validation split. To get the best k value in the KNN Algorithm using the brute force search method. For testing using a lot of training data and testing data used. In the Knn algorithm, the aim is to find the closest neighbors with the tested data. While the Random Forest algorithm is based on data samples using a random variable in the formed decision tree. The amount of data can affect the resulting accuracy value. The speed of classification of the two algorithms is influenced by the amount of data used. Each algorithm also calculates accuracy, precision, recall and f1 score.

Keyword: data classification, comparison classification algorithm, Random Forest, KNN algorithm



TABLE OF CONTENTS

| | |
|--|-----------|
| Cover..... | |
| APPROVAL AND RATIFICATION PAGE..... | |
| STATEMENT OF ORIGINALITY..... | |
| ABSTRACT..... | |
| TABLE OF CONTENTS..... | |
| ILLUSTRATION INDEX..... | |
| INDEX OF TABLES..... | |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 Background..... | 1 |
| 1.2 Problem Formulation..... | 2 |
| 1.3 Scope..... | 2 |
| 1.4 Objective..... | 3 |
| CHAPTER 2 LITERATURE STUDY..... | 4 |
| CHAPTER 3 RESEARCH METHODOLOGY..... | 9 |
| CHAPTER 4 ANALYSIS AND DESIGN..... | 17 |
| 4.1 Analysis..... | 17 |
| 4.2 Desain..... | 35 |
| CHAPTER 5 IMPLEMENTATION AND TESTING..... | 39 |
| 5.1 Implementation..... | 39 |
| 5.2 Testing..... | 42 |
| CHAPTER 6 CONCLUSION..... | 48 |
| REFERENCES..... | |
| APPENDIX..... | A |

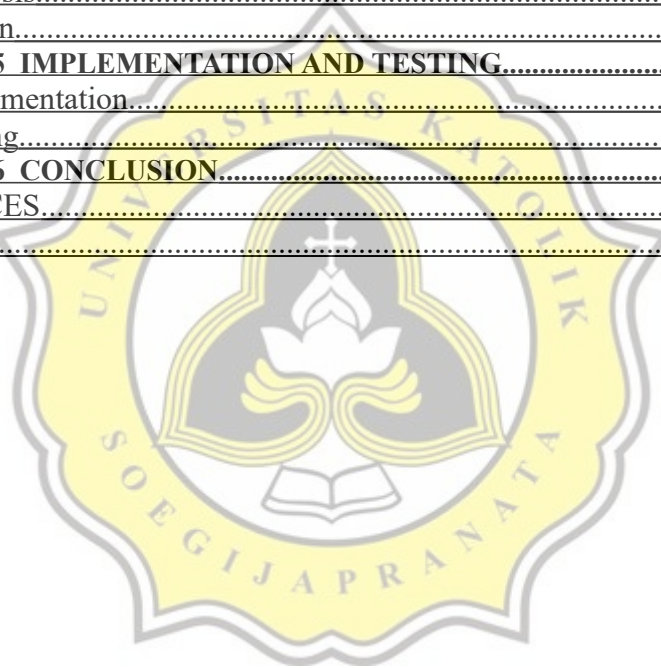


ILLUSTRATION INDEX

| | |
|---|----|
| Illustration 1: Image Euclidean Distance..... | 18 |
| Illustration 2: Figure K value results with data split as 50%..... | 19 |
| Illustration 3: Figure K value results with data split as 60%..... | 20 |
| Illustration 4: Figure K value results with data split as 70%..... | 21 |
| Illustration 5: Figure K value results with data split as 80%..... | 21 |
| Illustration 6: Figure K value results with data split as 90%..... | 22 |
| Illustration 7: Results of Accuracy, Precision, Recall and F1 Score with New Dataset with data split as 50%..... | 24 |
| Illustration 8: Results of Accuracy, Precision, Recall and F1 Score with New Dataset with data split as 60%..... | 25 |
| Illustration 9: Results of Accuracy, Precision, Recall and F1 Score with New Dataset with data split as 70%..... | 26 |
| Illustration 10: Results of Accuracy, Precision, Recall and F1 Score with New Dataset with data split as 80%..... | 27 |
| Illustration 11: Results of Accuracy, Precision, Recall and F1 Score with New Dataset with data split as 90%..... | 28 |
| Illustration 12: Random Forest Classifier..... | 29 |
| Illustration 13: Decision Tree..... | 31 |
| Illustration 14: Flowchart Knn Algorithm..... | 35 |
| Illustration 15: Flowchart Random Forest Algorithm..... | 37 |
| Illustration 16: Comparison Precision Between Knn Algorithms and Random Forest Algorithms with 372 data..... | 42 |
| Illustration 17: Comparison Recall Between Knn Algorithms and Random Forest Algorithms with 372 data..... | 43 |
| Illustration 18: Comparison F1 Score Between Knn Algorithms and Random Forest Algorithms with 372 data..... | 44 |
| Illustration 19: Comparison Precision with Add New Dataset..... | 45 |
| Illustration 20: Comparison Recall with Add New Dataset..... | 46 |
| Illustration 21: Comparison F1 Score with Add New Dataset..... | 47 |

INDEX OF TABLES

| | |
|--|----|
| Table 1: The best accuracy, precision, recall and f1 score results from the KNN algorithm..... | 23 |
| Table 2: Gini Index..... | 30 |
| Table 3: Result Best Split..... | 30 |
| Table 4: Result Accuracy Random Forest from tree..... | 33 |
| Table 5: Result Accuracy, Precision, Recall, F1 Score..... | 33 |
| Table 6: Result Accuracy, Precision, Recall, F1 Score with Add New Dataset..... | 34 |

