OPTIMIZATION OF NAÏVE BAYES USING LEVENSHTEIN DISTANCE FOR TYPOGRAPHICAL ERROR CORRECTION IN SENTIMENT ANALYSIS

THESIS

By:

Aziz Yogo Utomo

18.61.0149

BACHELOR OF INFORMATICS

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS AMIKOM YOGYAKARTA

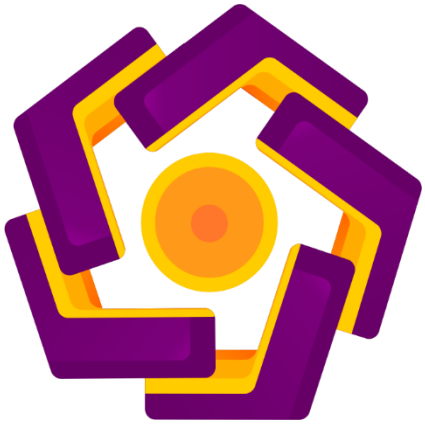
YOGYAKARTA

2022

OPTIMIZATION OF NAÏVE BAYES USING LEVENSHTEIN DISTANCE FOR TYPOGRAPHICAL ERROR CORRECTION IN SENTIMENT ANALYSIS

THESIS

to fulfil the requirements for a Bachelor's degree
in the Informatics study program



By:

Aziz Yogo Utomo

18.61.0149

BACHELOR OF INFORMATICS

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

APPROVAL

THESIS

OPTIMIZATION OF NAÏVE BAYES USING LEVENSHTEIN DISTANCE FOR TYPOGRAPHICAL ERROR CORRECTION IN SENTIMENT ANALYSIS

prepared and arranged by:

Aziz Yogo Utomo

18.61.0149

has been approved by undergraduate thesis supervisor

on 27 January 2022

Supervisor,

Anna Baita, M.Kom.
NIK. 190302290

VALIDATION

THESIS

OPTIMIZATION OF NAÏVE BAYES USING LEVENSHTEIN DISTANCE FOR TYPOGRAPHICAL ERROR CORRECTION IN SENTIMENT ANALYSIS

prepared and arranged by:

Aziz Yogo Utomo

18.61.0149

has been maintained by examiners

on 23 February 2022

The Examiners

Examiner	Signature
Yuli Astuti, M.Kom. **NIK. 190302146**	_____
Supriatin, M.Kom. **NIK. 190302239**	_____
Anna Baita, M.Kom. **NIK. 190302290**	_____

This thesis has been accepted as one of
the requirements for obtaining a Bachelor of Computer degree
on 26 February 2022

DEAN OF FACULTY OF COMPUTER SCIENCE

Hanif Al Fatta,S.Kom., M.Kom.
NIK. 190302096

DECLARATION

I, the undersigned below, state that this thesis is my work (ORIGINAL). The contents of this thesis have never been applied by any other person to receive an academic degree at a particular educational institution. As far as I know, there are no works or thoughts written and/or published by anyone, except those in writing listed in this manuscript and mentioned in the reference list.

Anything that applies to the manuscripts and works that have been made is my responsibility.

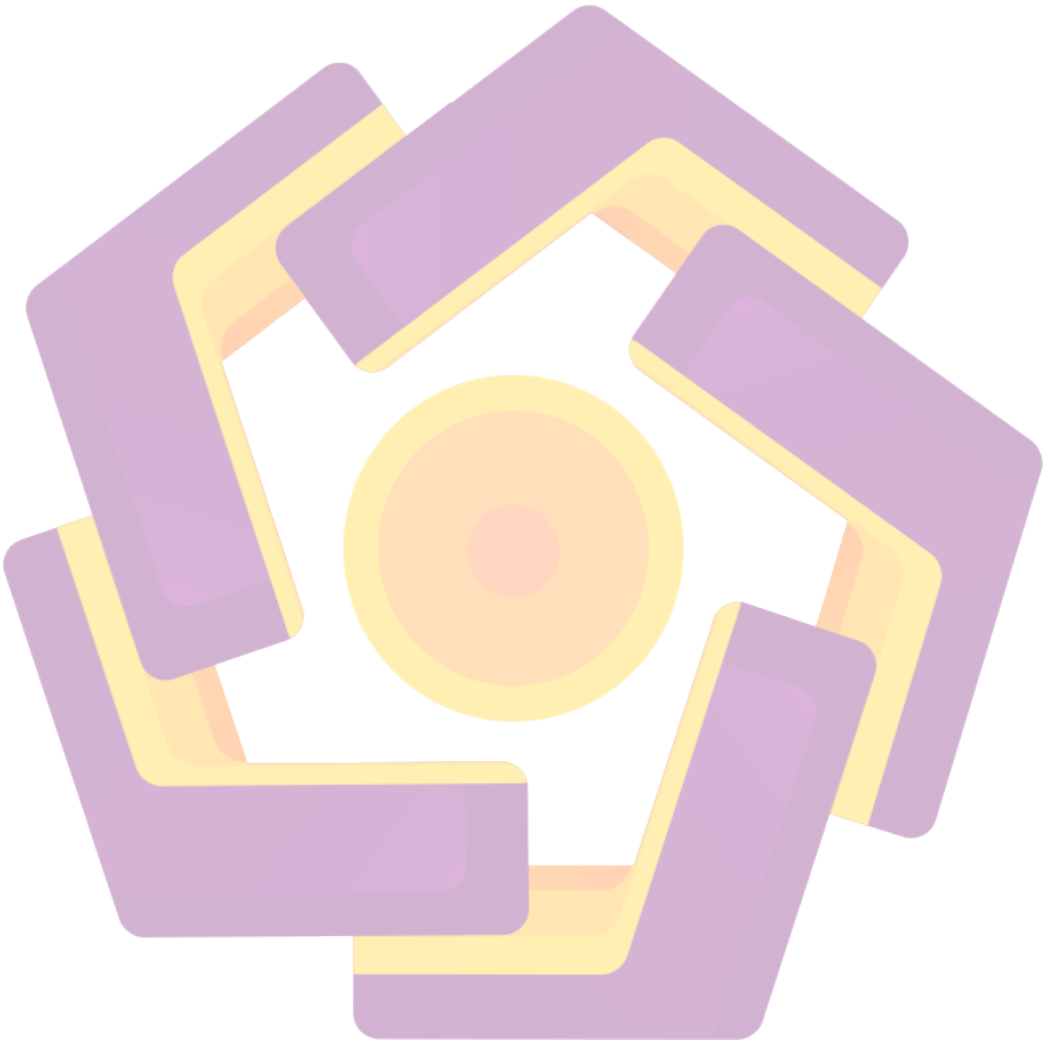
Yogyakarta, 24 February 2022

AZIZ YOGO UTOMO

18.61.0149

MOTTO

“Break your limits.”



DEDICATION

Praise be to Allah SWT with His blessing; this thesis can be written and completed correctly. With this, the researcher will dedicate this thesis to:

1. Allah S.W.T who always gives love and affection and guidance for His servants.
2. My beloved parents.
3. Thesis Supervisor and The Examiner.
4. All parties who have helped the researcher in completing this thesis.

ACKNOWLEDGEMENTS

Praise and gratitude I pray to Allah SWT, Lord of the worlds, who controls all things on earth and in the sky, for all the abundance of His graces, favors, and gifts. Only to Him do I surrender and ask for help, and only to Him is the place for me to complain during this research. Not left behind, I also want to say thank you to:

1. Big thanks to my parents, who have provided all kinds of support that might I cannot return it. Hopefully, what has been given can be helpful for me in my future life.

2. Mrs. Anna Baita, M. Kom., a supervisor for this research, was swift in replying to messages and giving feedback. Thank you for sparing your time to guide me in this research, and my apologies if I bothered you for the last few months. Hopefully, you will be given more from Allah.

3. Nabila, *partner in everything*.

4. All parties who have helped the researcher in completing this thesis.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Now we live in an era with a tremendous amount of unstructured data such as text data. Naïve Bayes is an algorithm that is well-performed for dealing with text data. In processing text data, there are several problems, such as the vast amount and the data dimension. Text data derived from human fingers allows typographical errors in writing; this typographical error becomes another problem because it will make the data dimension bigger and change the semantics of the word itself. In addition, typographical errors liable the calculation of the Naïve Bayes Likelihood to be 0 and affect the performance of the model. Levenshtein Distance is one method to correct typographical errors. This research indicates that the Levenshtein Distance has succeeded in increasing the performance of the Naïve Bayes model with optimal results using distance 2 of 5,9%.

***Keywords**: Text Mining, Classification, Naive Bayes Classifier, Levenshtein Distance, Typographical Error Correction*