



PROJECT REPORT
ANALYSING TWITTER TRENDING HASHTAG
SENTIMENT ABOUT RACISM AND BIGOTRY
USING *k-NEAREST NEIGHBOUR* (k-NN)
ALGORITHM

Revival Danisaputra
15.K1.0065

Faculty of Computer Science
Soegijapranata Catholic University
2020

APPROVAL AND RATIFICATION PAGE

ANALYSING TWITTER TRENDING HASHTAG SENTIMENT ABOUT RACISM
AND BIGOTRY USING k-NEAREST NEIGHBOR (k-NN) ALGORITHM

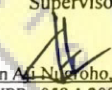
by

REVIVAL DANISAPUTRA – 15.K1.0065

This project report has been approved and ratified
by the Faculty of Computer Science on January, 7, 2020


With approval,

Supervisor,

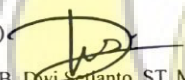

Robertus Setiawan Aji Nugroho, ST., MCompIT., Ph.D
NPP : 058.1.2004.264

Examiners,

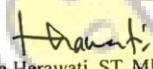
1.)



Hironimus Leong, S.Kom., M.Kom
NPP : 058.1.2007.273


2.)


YB. Dwi Setianto, ST., M.Cs
NPP : 058.7.2017.021

3.)


Rosita Herawati, ST., MIT
NPP : 058.1.2004.263


Dean of Faculty of Computer Science,


Robertus Setiawan Aji Nugroho, ST., MCompIT., Ph.D
NPP : 058.1.2004.264

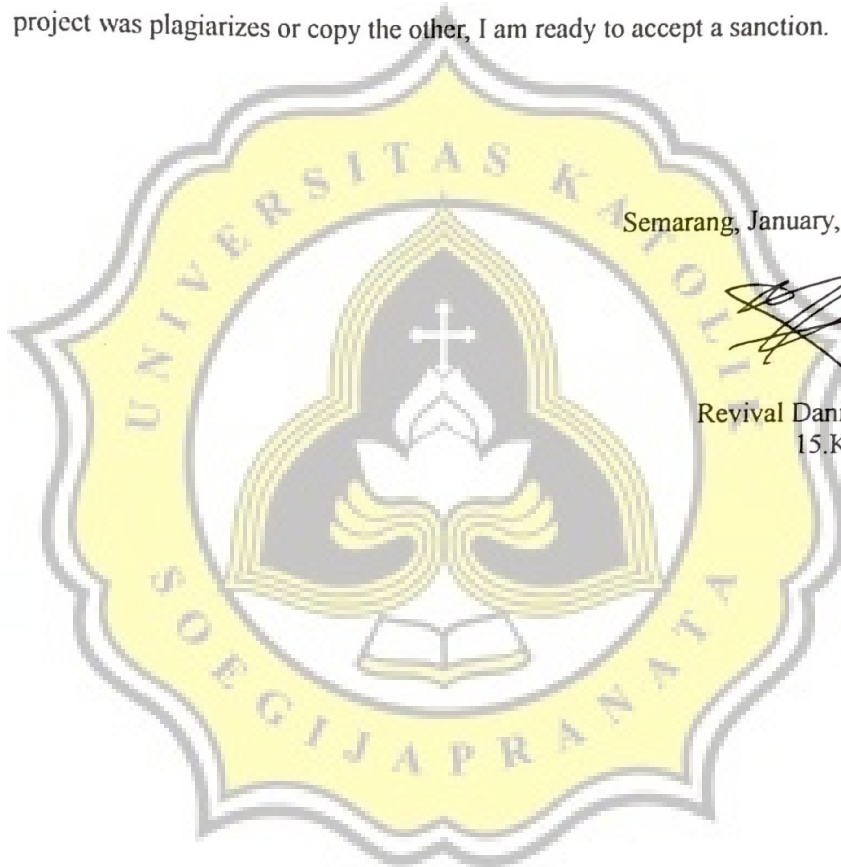
STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Revival Danisaputra

ID : 15.K1.0065

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, 7, 2020

Revival Danisaputra
15.K1.0065

ABSTRACT

Racism, bigotry, and fascism are now a big deal around the world especially on social media. Begin when religion blasphemy from fascists and race humiliation were a perfect subject to put down people personally on social media. Overall, this project can analysed and classified whether it is the negative tweet or not by using k-NN implementation.

The project use python as its programming language. The first thing came with scrapping tweets by #hashtag trending from twitter; data will be saved in csv form and its title has # symbol on the front. The next step is pre-processing which is tokenize, stopwords, and stemming. Afterwards, the document is calculated with Tf-Idf and compared with data training which also had been calculated with Tf-Idf. The Final step is all the result input into k-NN algorithm to find out the nearest neighbour of negative words which is the lowest score.

Final conclusion of this project is the result which has less point has the nearest score on negative text. The document which declared negative is the document which has more than 40% negative texts inside. The accuracy of this calculation is 82.3%.

Keyword: Tf-Idf, k-NN, racism, bigotry, hashtag.

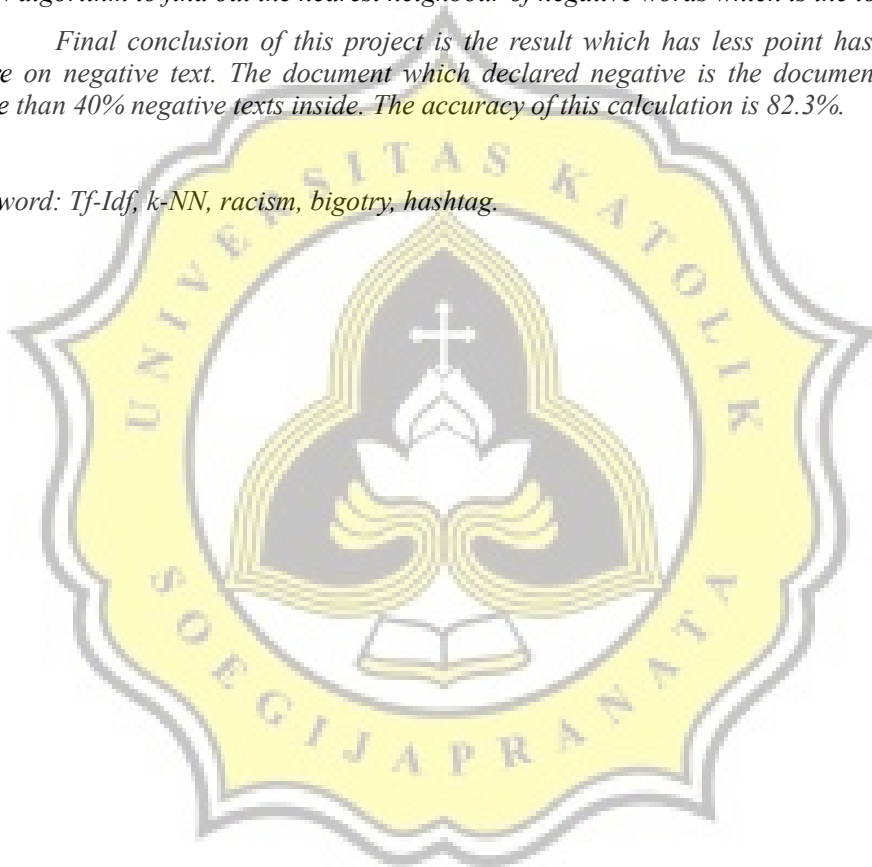
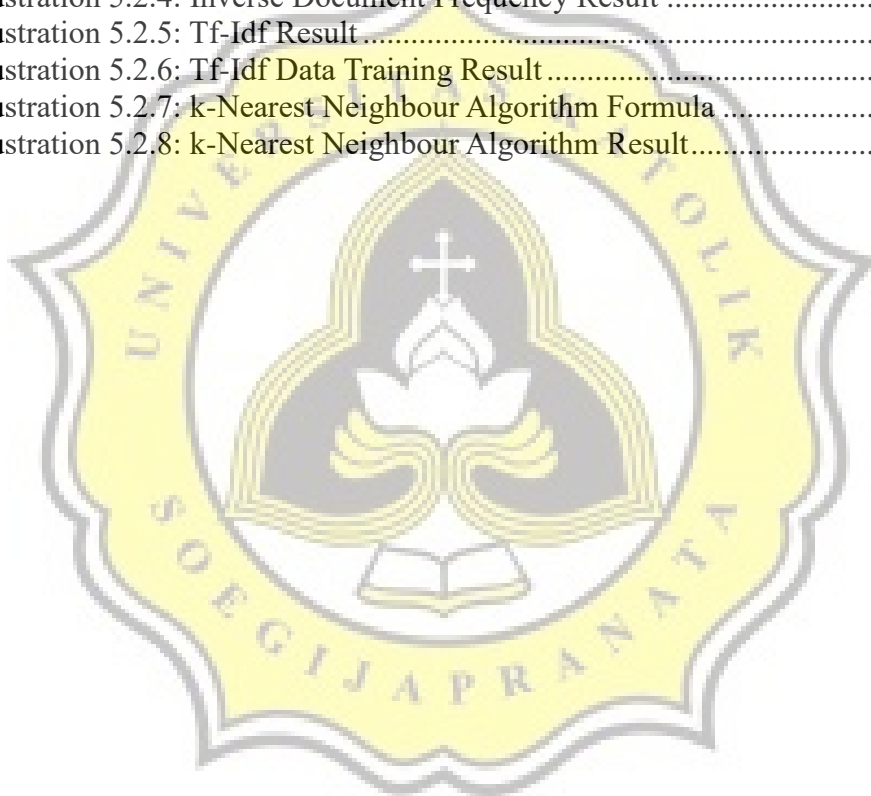


TABLE OF CONTENTS

APPROVAL AND RATIFICATION PAGE...Error! Bookmark not defined.	
STATEMENT OF ORIGINALITYError! Bookmark not defined.	
ABSTRACT	iv
TABLE OF CONTENTS	v
ILLUSTRATION INDEX.....	vi
INDEX OF TABLES	vii
CHAPTER 1 INTRODUCTION	1
1.1. Background	1
1.2. Problem Formulation	1
1.3. Scope.....	1
1.4. Objective	1
CHAPTER 2 LITERATURE STUDY	2
CHAPTER 3 RESEARCH METHODOLOGY	9
CHAPTER 4 ANALYSIS AND DESIGN.....	11
4.1. Analysis	11
4.2. Design	13
CHAPTER 5 IMPLEMENTATION AND TESTING	14
5.1. Implementation	14
5.2. Testing	15
CHAPTER 6 CONCLUSION	19
References	A
Appendix	C

ILLUSTRATION INDEX

Illustration 2.1.1: Clustering Class Example.....	2
Illustration 2.1.2: Comparison of k-Nearest Neighbour class decision lines (grey lines) with linear classification models (purple lines)	3
Illustration 4.1.1: Idf scores	8
Illustration 4.1.2: Euclidian Distance Formula	9
Illustration 4.2: Procedure of calculating k-Nearest Neighbour	10
Illustration 5.2.1: Scraped Document	12
Illustration 5.2.2: Document After Pre-Processing Stage	13
Illustration 5.2.3: Term Frequency Result.....	13
Illustration 5.2.4: Inverse Document Frequency Result	13
Illustration 5.2.5: Tf-Idf Result	13
Illustration 5.2.6: Tf-Idf Data Training Result.....	14
Illustration 5.2.7: k-Nearest Neighbour Algorithm Formula	14
Illustration 5.2.8: k-Nearest Neighbour Algorithm Result.....	15



INDEX OF TABLES

Table 4.1: k-Nearest Neighbour Analytical Table12

