



**PROJECT REPORT**  
**PREDICTION OF BREAST CANCER USING THE**  
**NAÏVE BAYES ALGORITHM**

**DIKA AVILA WULANDARI**  
**17.K1.0018**

**Faculty of Computer Science**  
**Soegijapranata Catholic University**  
**2022**

## HALAMAN PENGESAHAN



Judul Tugas Akhir: : PREDICTION OF BREAST CANCER USING THE NAÏVE BAYES  
ALGORITHM

Diajukan oleh : Dika Avila Wulandari

NIM : 17.K1.0018

Tanggal disetujui : 30 Juli 2022

Telah setuju oleh

Pembimbing : Hironimus Leong S.Kom., M.Kom.

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

Penguji 2 : Y.b. Dwi Setianto S.T., M.Cs.

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

Penguji 4 : Yonathan Purbo Santosa S.Kom., M.Sc

Penguji 5 : Rosita Herawati S.T., M.I.T.

Penguji 6 : Yulianto Tejo Putranto S.T., M.T.

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

Dekan : Dr. Bernardinus Harnadi S.T., M.T.

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

[sintak.unika.ac.id/skripsi/verifikasi/?id=17.K1.0018](http://sintak.unika.ac.id/skripsi/verifikasi/?id=17.K1.0018)

## HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini:

Nama : DIKA AVILA WULANDARI

Program Studi : Teknik Informatika

Fakultas : Ilmu Komputer

Jenis Karya : Penelitian

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneksklusif atas karya ilmiah yang berjudul “PREDICTION OF BREAST CANCER USING THE NAÏVE BAYES ALGORITHM” beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Katolik Soegijapranata berhak menyimpan, mengalihkan media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir ini selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 30 Juli 2022

Yang menyatakan



DIKA AVILA WULANDARI

## DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : DIKA AVILA WULANDARI

ID : 17.K1.0018

declare that this work, titled "PREDICTION OF BREAST CANCER USING THE NAÏVE BAYES ALGORITHM", and the work presented in it is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at Soegijapranata Catholic University
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given.
5. Except for such quotations, this work is entirely my own work.
6. I have acknowledged all main sources of help.
7. Where the work is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Semarang, July 30<sup>th</sup> 2022



DIKA AVILA WULANDARI

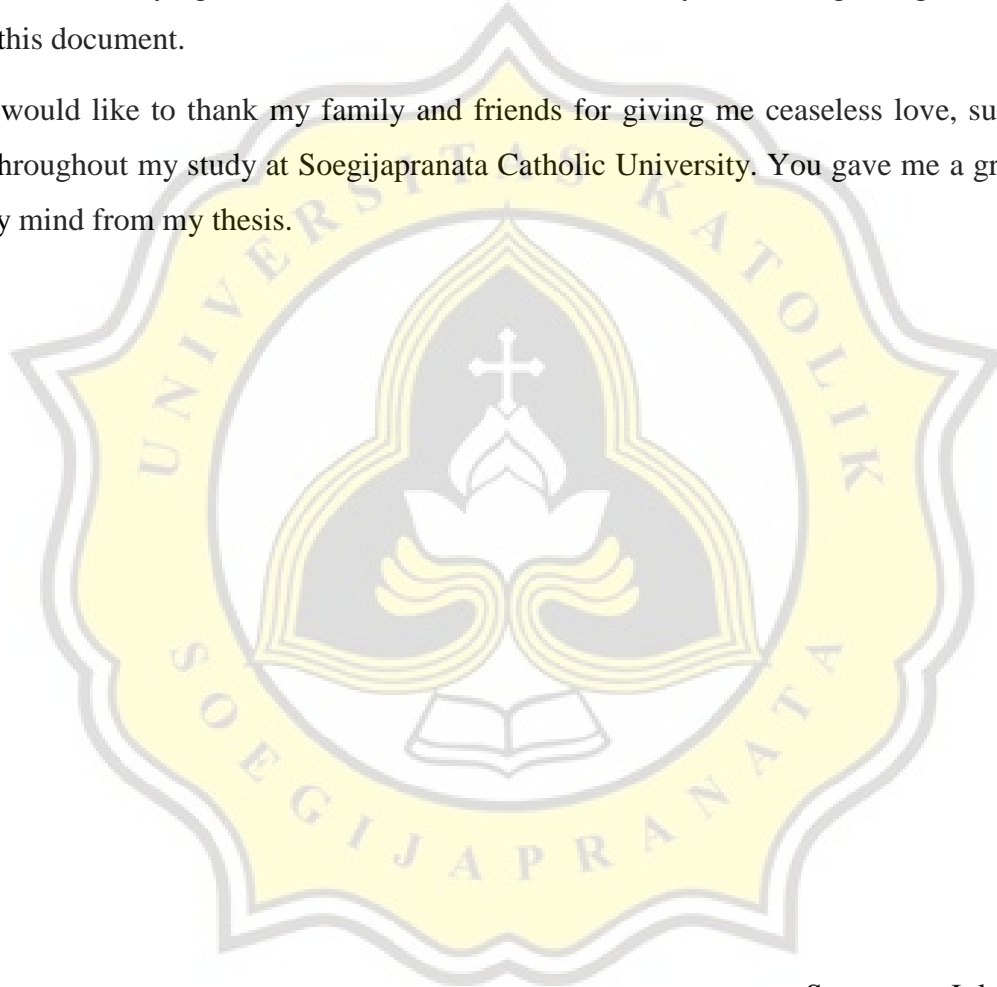
17.K1.0018

## ACKNOWLEDGMENT

I would like to thank God almighty for all His love and grace.

I have received a myriad of support, advice, and assistance throughout this document writing. I would like to thank my supervisors Mr Hironimus Leong S.Kom., M.Kom. for formulating this topic. I would like to thank all the lecturers who have provided knowledge and helped me while studying here. I would also like to thank all my friend for guiding me with advice to finish this document.

I would like to thank my family and friends for giving me ceaseless love, support, and advices throughout my study at Soegijapranata Catholic University. You gave me a great escape to rest my mind from my thesis.



Semarang, July 30<sup>th</sup> 2022

DIKA AVILA WULANDARI

## ABSTRACT

*Breast cancer is the most common cause of death among women worldwide. Cancer can be classified into two categories, benign and malignant. Cancers are more aggressive than tumors. Therefore, early diagnosis will be very helpful for patients.*

*This report uses the Naïve Bayes data classification method to detect breast cancer and classify whether it is benign or malignant.*

*The conclusion of this study is that the Naïve Bayes algorithm proposed as a classification method for breast cancer prediction is able to get a good score. The average percentage of data that is classified correctly reaches 96.47% and the average percentage of data that is classified incorrectly is only 3.53%. While the level of effectiveness where the average value of precision and recall respectively is 0.956 and 0.9685. For the highest precision and recall values, when the test data uses a 10% split percentage for testing data with values of 0.985 and 0.985.*

*Keyword: Naïve bayes, breast cancer prediction, classification*



## TABLE OF CONTENTS

<b>COVER</b> .....	<b>i</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>ii</b>
<b>HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS</b> .....	<b>iii</b>
<b>DECLARATION OF AUTHORSHIP</b> .....	<b>iv</b>
<b>ACKNOWLEDGMENT</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>TABLE OF CONTENTS</b> .....	<b>vii</b>
<b>LIST OF FIGURE</b> .....	<b>ix</b>
<b>LIST OF TABLE</b> .....	<b>x</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
1.1. Background .....	1
1.2. Problem Formulation .....	3
1.3. Scope.....	3
1.4. Objective .....	3
<b>CHAPTER 2 LITERATURE STUDY</b> .....	<b>4</b>
<b>CHAPTER 3 RESEARCH METHODOLOGY</b> .....	<b>7</b>
3.1. Journal References .....	7
3.2. Data Collection .....	7
3.3. Code .....	8
3.4. Implementation and Analysis.....	8
3.5. Conclusion and Report Writing .....	10
<b>CHAPTER 4 ANALYSIS AND DESIGN</b> .....	<b>11</b>
4.1. Analysis.....	11
4.2. Desain .....	15

<b>CHAPTER 5 IMPLEMENTATION AND RESULTS .....</b>	<b>18</b>
5.1. Implementation .....	18
5.2. Testing.....	20
<b>CHAPTER 6 CONCLUSION.....</b>	<b>24</b>
<b>REFERENCES.....</b>	<b>25</b>
<b>APPENDIX.....</b>	<b>a</b>





## LIST OF FIGURE

<b>Figure 1.1</b> New Cases of Cancer.....	1
<b>Figure 1.2</b> New Cases of Breast Cancer .....	2
<b>Figure 3.1</b> Flowchart of The Implementation of The Naive Bayes Method for Breast Cancer Prediction.....	9
<b>Figure 4.1</b> Example of Original Dataset Snippet .....	11
<b>Figure 4.2</b> After Conversion Dataset.....	12
<b>Figure 4.1</b> Flowchart of Data Cleaning.....	15
<b>Figure 5.1</b> Confusion Matrix 80:20.....	20
<b>Figure 5.2</b> Confusion Matrix 70:30.....	21
<b>Figure 5.3</b> Confusion Matrix 60:40.....	21
<b>Figure 5.4</b> Balanced Data Down.....	22

## LIST OF TABLE

<b>Table 3.1.</b> Dataset Overview .....	7
<b>Table 3.2.</b> Dataset After Cleaning .....	8
<b>Table 4.1.</b> Data Training .....	12
<b>Table 4.2.</b> Mean of Each Attribute .....	15
<b>Table 4.3.</b> Data Testing .....	16
<b>Table 4.4.</b> Confusion Matrix Rules .....	16
<b>Table 5.1.</b> Test Results .....	22
<b>Table 5.2.</b> Test Result for Balanced Data .....	23

