



PROJECT REPORT
CAFFEINE CONTENT OF DRINK CLASSIFICATION USING
KNN AND SVM

GEOVANNY FIRDAUS ATMAJA
16.K1.0035

Faculty of Computer Science
Soegijapranata Catholic University
202

APPROVAL AND RATIFICATION PAGE



HALAMAN PENGESAHAN

Judul Tugas Akhir: : Caffeine Content of drink classification using knn and svm
Diajukan oleh : GEOVANNY FIRDAUS A
NIM : 16.K1.0035
Tanggal disetujui : 30 Januari 2023
Telah setuju oleh
Pembimbing : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 1 : Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 2 : Hironimus Leong S.Kom., M.Kom.
Penguji 3 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 4 : Rosita Herawati S.T., M.I.T.
Penguji 5 : Y.b. Dwi Setianto S.T., M.Cs.
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=16.K1.0035

DECLARATION OF AUTHORSHIP (Heading Plain)

I, the undersigned:

Name : Geovanny Firdaus Atmaja

ID : 16.K1.0035

declare that this work, titled "**CAFFEINE CONTENT OF DRINK CLASSIFICATION USING KNN AND SVM**", 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, 25 Januari 2023



Geovanny Firdaus Atmaja

16.K1.0035

HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini:

Nama : Geovanny Firdaus Atmaja

Program Studi : Teknik Informatika

Fakultas : Ilmu Komputer

Jenis Karya : Skripsi

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneksklusif atas karya ilmiah yang berjudul "**CAFFEINE CONTENT OF DRINK CLASSIFICATION USING KNN AND SVM**". 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, 25 Januari 2023

Yang menyatakan



Geovanny Firdaus Atmaja

16.K1.0035

ACKNOWLEDGMENT

All praise and thanks, the author goes to God Almighty for His mercy and grace, so that the writer can complete the final assignment entitled "Caffeine Content of Beverages Using Classification Using the KNN and SVM Methods" smoothly and on time. This final project is one of the requirements that must be fulfilled as an academic requirement in obtaining a computer bachelor's degree in the Informatics Study Program at Soegijapranata Catholic University.

In the preparation of this final project, the author received assistance from various parties. Therefore, on this occasion the author would like to express his deepest gratitude to :

1. God Almighty, who has provided assistance, strength and assistance during the process of making this final project
2. Mother and sister who always provide prayers and support to the author during the work of this Final Project.
3. R. Setiawan Aji Nugroho S.T. McomperT. Ph.D as a supervisor who has provided guidance and input to the author, so that this final project can be completed properly
4. And other parties that the author cannot mention one by one who have provided support and assistance to the author during the preparation of this final project.

ABSTRACT

Caffeine is popular worldwide and has many health benefits due to the presence and number of compounds. The compound is a xanthine alkaloid in the form of crystals and tastes bitter which works as a psychoactive stimulant and mild diuretic. Caffeine was discovered by a German chemist, Friedrich Ferdinand Runge in 1819. Caffeine is found naturally in foods such as coffee beans, tea leaves, colas, guarana, and maté. In plants, it acts as a natural pesticide that paralyzes and kills certain insects that eat these plants. The first process is how to process some drink data that will be classified using the KNN and SVM algorithms to get accurate results for the comparison of the two algorithms by programming processing using the Orange data mining application and how the application works to produce which accuracy has higher accuracy. The results of the implementation of the two algorithms with Orange have a conclusion. Orange has deficiencies when determining detailed results on ca, f1-score, precision, and recall. Then there are many features whose functions are unknown and only show instantly. Therefore, there are many features that must be studied further according to the needs we want. Then when processing the application it only states that the data is true and false. So it can't be as detailed as using processing with coding. So we cannot determine where the TN, FN, TP, and FP are located. But the strengths of Orange are that it's easy to use and does it instantly. The processing results reveal that the two methods provide virtually identical outputs. It's only that the K-NN method has a better degree of precision than SVM since the two algorithms have the same qualities, particularly the "Supervised Algorithm" property that both algorithms require training data.

Keyword: Caffeine, Classification, SVM, KNN, Supervised Algorithm.

TABLE OF CONTENTS

COVER	i
CHAPTER 1	INTRODUCTION
	10
1.1. Background	10
1.2. Problem Formulation	11
1.3. Scope	11
1.4. Objective	11
CHAPTER 2 LITERATURE STUDY	12
CHAPTER 3 RESEARCH METHODOLOGY	15
CHAPTER 4 ANALYSIS AND DESIGN	17
4.1. Analysis	17
4.2. K-Nearest Neighbor Analysis (K-NN)	20
4.3. Support Vector Machine Analysis (SVM)	21
4.4. Desain	23
CHAPTER 5 IMPLEMENTATION AND RESULTS	24
5.1. Implementation	24
5.2. K-NN Result Prediction	24
5.3. SVM Result Prediction	31
5.4. Result	37
CHAPTER 6 CONCLUSION	38

LIST OF FIGURE

FIGURE 3. 1 MAPPING ATTRIBUTES (X) INTO CLASS LABELS (Y)	16
FIGURE 4.1. 1 DATASET CAFFEINE.CSV	17
FIGURE 4.1. 2 DATASET SELECTION OF DATA ATTRIBUTES AND COMPOSITION OF DATA DOMAIN	18
FIGURE 4.1. 3 ACCURACY, PRECISION AND RECALL	19
FIGURE 4.1. 4 F1-SCORE	20
FIGURE 4.2 1 PREDICTED K-NN CONFUSION MATRIX	20
FIGURE 4.3 1 PREDICTED SVM CONFUSION MATRIX	22
FIGURE 4.4. 1 FLOWCHART ORANGE DATA MINING	23
FIGURE 5.2 1 TEST AND SCORE 25 % DATA TRAINING AND 75% DATA TESTING	25
FIGURE 5.2 2 CONFUSION MATRIX 25 % DATA TRAINING AND 75% DATA TESTING	26
FIGURE 5.2 3 TEST AND SCORE 50 % DATA TRAINING AND 50% DATA TESTING	27
FIGURE 5.2 4 CONFUSION MATRIX 50 % DATA TRAINING AND 50% DATA TESTING	28
FIGURE 5.2 5 TEST AND SCORE 75 % DATA TRAINING AND 25% DATA TESTING	29
FIGURE 5.2 6 CONFUSION MATRIX 75 % DATA TRAINING AND 25% DATA TESTING	30
FIGURE 5.3 1 TEST AND SCORE 25 % DATA TRAINING AND 75% DATA TESTING	31
FIGURE 5.3 2 CONFUSION MATRIX 25 % DATA TRAINING AND 75% DATA TESTING	32
FIGURE 5.3 3 TEST AND SCORE 50% DATA TRAINING AND 50% DATA TESTING	33
FIGURE 5.3 4 CONFUSION MATRIX 50% DATA TRAINING AND 50% DATA TESTING	34
FIGURE 5.3 5 TEST AND SCORE 75% DATA TRAINING AND 25% DATA TESTING	35
FIGURE 5.3 6 CONFUSION MATRIX 75% DATA TRAINING AND 25% DATA TESTING	36

LIST OF TABLE

TABLE 5.2 1 TEST AND SCORE 25 % DATA TRAINING AND 75% DATA TESTING	25
TABLE 5.2 2 TEST AND SCORE 50 % DATA TRAINING AND 50% DATA TESTING	27
TABLE 5.2 3 TEST AND SCORE 75 % DATA TRAINING AND 25% DATA TESTING	29
TABLE 5.3 1 TEST AND SCORE 25 % DATA TRAINING AND 75% DATA TESTING	31
TABLE 5.3 2 TEST AND SCORE 50 % DATA TRAINING AND 50% DATA TESTING	33
TABLE 5.3 3 TEST AND SCORE 75 % DATA TRAINING AND 25% DATA TESTING	35

