PROJECT REPORT
K-MEANS ALGORITHM IMPLEMENTATION
FOR NEWS CLUSTERING

DICKY LARSON
13.02.0001

Faculty of Computer Science
Soegijapranata Catholic University
2017
K-MEANS ALGORITHM IMPLEMENTATION FOR NEWS CLUSTERING

by

DICKY LARSON – 13.02.0001

This project report has been approved and ratified

by the Faculty of Computer Science on July, 12, 2017

With approval,

[Signatures of examiners and supervisor]

[Signatures of dean]

Supervisor,

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

Examiners,

1.) Suyanto EA, Jr., M.Sc

2.) Rosita Herawati, ST., MIT
NPP: 058.1.2004.263

3.) Shinta Estri Wabyuningrum, S.Si., M.Cs
NPP: 058.1.2007.272

Dean of Faculty of Computer Science,

Esti Wibowo Nugroho, ST., MT
NPP: 058.1.2002.254
STATEMENT OF ORIGINALITY

I, the undersigned:

Name : DICKY LARSON
ID : 13.02.0001

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, July, 12, 2017

[Signature]

DICKY LARSON
13.02.0001
ABSTRACT

With rapid development of internet, there are a lot of online news which are made. With all the amount of news circulate, it will require a lot of time and effort to find news with the wanted topic. To make it easy for user to find news with the topic user want, the program is created to automate the news clustering so that user can easily find the related news. The program will implement the k-means algorithm.

K-means algorithm is one of the algorithm for clustering which has long existed. It has been proven in many implementations of program. K-means use euclidean distance to count data point distance between data and centroid. The euclidean distance between all the centroid compared to find the nearest distance. All the euclidean with the nearest distance with the same centroid will be searched for average. The average will be the new centroid data. The process is repeated until the new centroid data not changed or similar with the previous centroid data.

The result is the program that can clustering online news article based on similar topic with user news article. User upload their own news article as the centroid. And the online news taken from kompas.com.

Keyword: news clustering, k-means, data mining
PREFACE

This project report discuss about K-means algorithm implementation for news clustering. K-means is algorithm that can cluster data. Cluster of data based on the nearest distance between centroid and data. To specify the data cluster, iterative process is done until new centroid value not changed. K-means is implemented to this news clustering program along with 2-dimensional array. In this report will be discussed about background that contain the news clustering problem, how it happen, how to solve it with K-means with 2-dimensional array and the solution result. Then discuss about scope of this project that contain specific question about the focus of the project.

After that will be discussed about studies that have been done similar to this project topic and tell the difference with this project. Next is discussing the steps undertaken in this project. Then discuss an explanation of the news clustering problem from input, process, to output and a design of solution to the news clustering problem. Then displays K-means algorithm and 2-dimensional array data structures in the form of applied applications and program testing. Few main code and program usage with some data example will be shown. Finally, the conclusion of this project that also answer the scope and suggestion for further research will be discussed.
# TABLE OF CONTENTS

Cover................................................................................................................................................................i
APPROVAL AND RATIFICATION PAGE................................................................................................. ii
STATEMENT OF ORIGINALITY............................................................................................................ iii
ABSTRACT............................................................................................................................................... iv
PREFACE................................................................................................................................................... v
TABLE OF CONTENTS........................................................................................................................... vi
ILLUSTRATION INDEX............................................................................................................................ vii
INDEX OF TABLES.................................................................................................................................. viii
CHAPTER 1 INTRODUCTION.................................................................................................................. 1
  1.1 Background.................................................................................................................................... 1
  1.2 Scope............................................................................................................................................ 1
  1.3 Objective........................................................................................................................................2
CHAPTER 2 LITERATURE STUDY.......................................................................................................... 3
CHAPTER 3 RESEARCH METHODOLOGY.......................................................................................... 5
CHAPTER 4 ANALYSIS AND DESIGN................................................................................................ 7
  4.1 Analysis.......................................................................................................................................... 7
  4.2 Design...........................................................................................................................................10
CHAPTER 5 IMPLEMENTATION AND TESTING............................................................................... 12
  5.1 Implementation............................................................................................................................12
  5.2 Testing..........................................................................................................................................16
CHAPTER 6 CONCLUSION.................................................................................................................. 21
REFERENCES........................................................................................................................................... 
APPENDIX............................................................................................................................................... A
ILLUSTRATION INDEX

Illustration 4.1: Data Flow Diagram.................................................................7
Illustration 4.2: Flowchart...........................................................................10
Illustration 5.1: Form Interface.................................................................16
Illustration 5.2: Filled Form.................................................................17
Illustration 5.3: Clustering Result Displayed on index.php...............19
INDEX OF TABLES

Table 5.1: Euclidean Distance................................................................................17
Table 5.2: Clustering Result...................................................................................19