PROJECT REPORT

Online News Classification Using K-Means Algorithm

Sony
11.02.0018
2014/2015

FACULTY OF COMPUTER SCIENCE
SOEGIJAPRANATA CATHOLIC UNIVERSITY
Jl. Pawiyatan Luhur IV/1, Bendan Duwur, SEMARANG 50234
Telp. 024-8441555 (hunting) Web: http://www.unika.ac.id
http://ikomunika.web.id/
PROJECT REPORT

Online News Classification Using K-Means Algorithm

by

11.02.0018 – Sony

This project report has been approved and ratified by the Dean of Faculty of Computer Science and Supervisor on 12 December 2014

With approval,

Examiners,

Suyanto Edward Antonius, Jr., M.Sc

Shinta Estri Wahyuningrum, S.Si, M.Cs
NPP: 058.1.2007.272

Examiners,

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

Rosita Herawati, ST., MIT
NPP: 058.1.2004.263

Dean of Faculty of Computer Science,

Hironimus Leong, S.Kom., M.Kom
NPP: 058.1.2007.273
STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Sony
ID : 11.02.0018

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, 15 December 2014

Sony
11.02.0018
FOREWORD

Praise God Almighty. Thanks to His abundance of blessing so I was able to complete the final project in the Faculty of Computer Science Soegijapranata Catholic University. The final project that I created proposed as a requirement of graduation. My final project titled "Online News Classification Using K-Means Algorithm".

On this occasion I would like to thank:

1. My Parents (Mr. Chadiri and Mrs. Yayuk Zubaedah), for their incessantly love
2. My Project Supervisor (Mr. Hieronimus Leong) were very patient and outstanding in guiding me
3. My Brothers (Thomas and Johan) for their tremendous support
4. My Ikom Friends (especially Steven Sugiarto Wijaya and Pranawa Singgih), for encouragement and motivation
5. Friends of the organizations (Soegijapranata Catholic University Student Senate and Student Executive Board Soegijapranata Catholic University) who have supported me and endlessly giving spirit
6. As well as others that I can not mention one by one
Abstract — Classification is the grouping of data with the greatest degree of similarity. Classification is generally only use the data numbers (math / statistics). The following classification using the data in the form of words of an article.

K-Means algorithm is an algorithm that is commonly used in data mining concepts. K-Means will classify the data according to the words contained therein. The articles that have a high degree of similarity words will occupy the same group.

Final results of this classification process is a recommendation for readers news. Readers can see any news that has a high degree of similarity with the article he was reading.

Keywords: Data Mining, K-Means Clustering.
# Table of Contents

APPROVAL AND RATIFICATION PAGE............................................................. ii
STATEMENT OF ORIGINALITY................................................................. iii
FOREWORD................................................................................................... iv
ABSTRACT...................................................................................................... v
Chapter I Introduction..................................................................................... 1
  1.1 Background............................................................................................ 1
  1.2 Scope..................................................................................................... 1
  1.3 Objective............................................................................................... 2
Chapter II Literature Study.............................................................................. 3
  2.1 Data Mining............................................................................................ 3
  2.2 Clustering Algorithm............................................................................. 5
  2.3 Algoritma K-Means............................................................................... 5
  2.4 Example.................................................................................................. 6
  2.5 Data Structure....................................................................................... 10
    2.5.1 Two Dimensional Array................................................................. 10
Chapter III Planning...................................................................................... 12
  3.1 Research Methodology......................................................................... 12
  3.2 Project Management............................................................................. 13
Chapter IV Analysis and Design................................................................... 14
  4.1 Analysis.................................................................................................. 14
    4.1.1 Use Case Diagram ........................................................................ 14
    4.1.2 Flow Chart.................................................................................... 16
  4.2 Design..................................................................................................... 18
    4.2.1 Class Diagram............................................................................... 18
Chapter V Implementation and Testing......................................................... 20
  5.1 Implementation..................................................................................... 20
  5.2. Testing................................................................................................ 29
Chapter VI Conclusion.................................................................................. 36
  6.1 Conclusion............................................................................................. 36
  6.2 Further Research.................................................................................. 36
TABLE OF TABLE

Table 1. Example Unique Words List.................................................................7
Table 2. Euclidian Distance List.................................................................8
Table 3. Boolean List.................................................................................8
Table 4. First Element............................................................................9
Table 5. n-1 Iteration..........................................................................9
Table 6. n Iteration..............................................................................9
Table 7. Project Management...............................................................13
TABLE OF DRAWING

<table>
<thead>
<tr>
<th>Drawing</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Euclidian Distance</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Two Dimensional Array</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Use Case Diagram</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Flow chart</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>Class Diagram</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Begin Date</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>End Date</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>Curl Command</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>Insert Article</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>Common Words</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>Get The Articles</td>
<td>23</td>
</tr>
<tr>
<td>12</td>
<td>Unique Words List</td>
<td>24</td>
</tr>
<tr>
<td>13</td>
<td>K-Means Variable</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>Euclidian Distance Implementation</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>Boolean Value</td>
<td>26</td>
</tr>
<tr>
<td>16</td>
<td>Build New Centroids</td>
<td>27</td>
</tr>
<tr>
<td>17</td>
<td>Choose The Iteration</td>
<td>27</td>
</tr>
<tr>
<td>18</td>
<td>Choose The Group</td>
<td>28</td>
</tr>
<tr>
<td>19</td>
<td>Choose The Article</td>
<td>28</td>
</tr>
<tr>
<td>20</td>
<td>Home Page</td>
<td>29</td>
</tr>
<tr>
<td>21</td>
<td>Load The Articles</td>
<td>30</td>
</tr>
<tr>
<td>22</td>
<td>Value of Each Article</td>
<td>31</td>
</tr>
<tr>
<td>23</td>
<td>Input Cluster</td>
<td>31</td>
</tr>
<tr>
<td>24</td>
<td>Iteration Done</td>
<td>32</td>
</tr>
<tr>
<td>25</td>
<td>Number of Iteration</td>
<td>32</td>
</tr>
<tr>
<td>26</td>
<td>Last Iteration</td>
<td>33</td>
</tr>
<tr>
<td>27</td>
<td>Group of The Articles</td>
<td>34</td>
</tr>
<tr>
<td>28</td>
<td>Read The Article</td>
<td>35</td>
</tr>
</tbody>
</table>