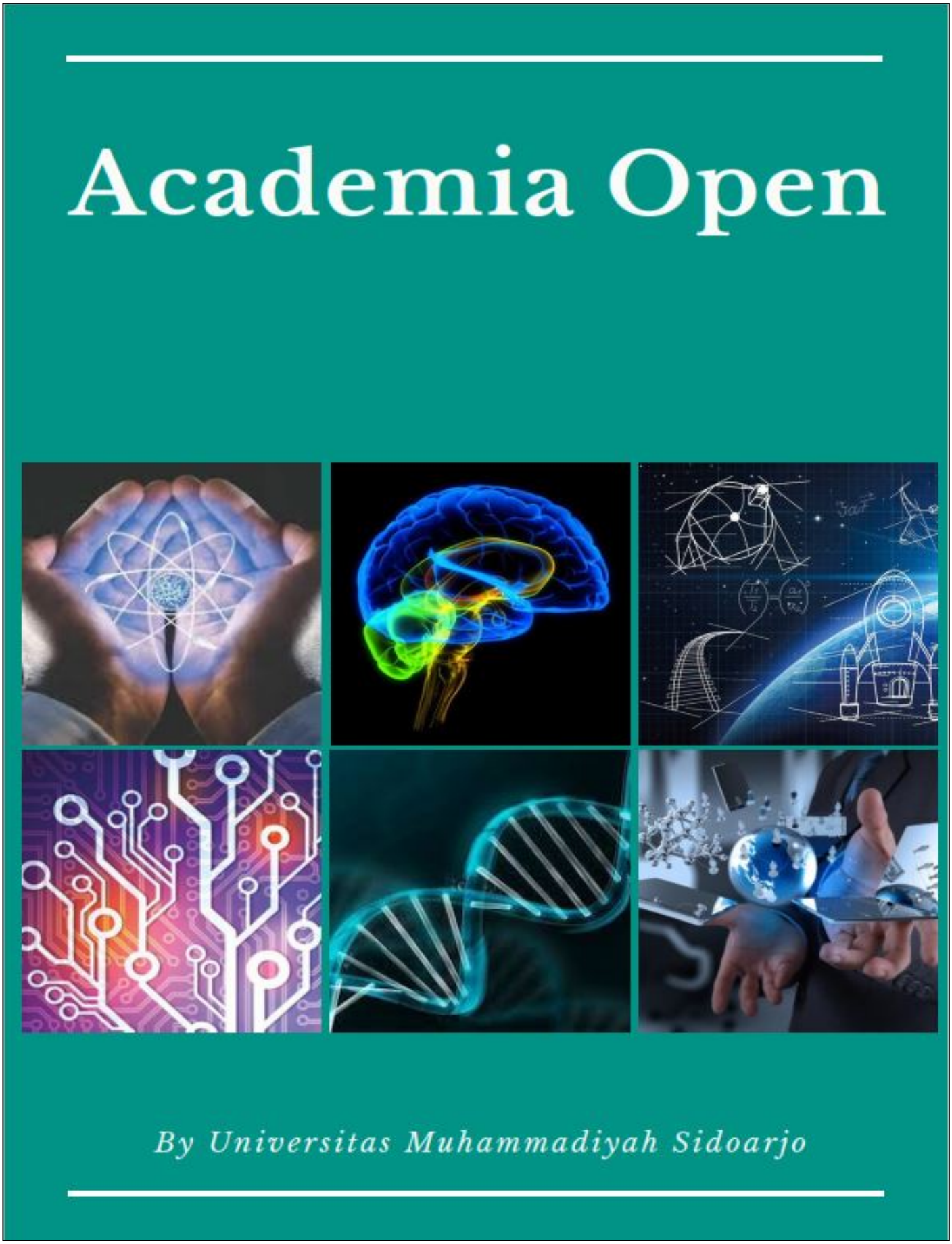
Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

Table Of Content

Journal Cover 2
Author[s] Statement 3
Editorial Team 4
Article information 5
Check this article update (crossmark) 5
Check this article impact 5
Cite this article 5
Title page 6
Article Title 6
Author information 6
Abstract 6
Article content 7

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)



ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

Copyright © Author(s). This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

EDITORIAL TEAM

Editor in Chief

Mochammad Tanzil Multazam, Universitas Muhammadiyah Sidoarjo, Indonesia

Managing Editor

Bobur Sobirov, Samarkand Institute of Economics and Service, Uzbekistan

Editors

Fika Megawati, Universitas Muhammadiyah Sidoarjo, Indonesia

Mahardika Darmawan Kusuma Wardana, Universitas Muhammadiyah Sidoarjo, Indonesia

Wiwit Wahyu Wijayanti, Universitas Muhammadiyah Sidoarjo, Indonesia

Farkhod Abdurakhmonov, Silk Road International Tourism University, Uzbekistan

Dr. Hindarto, Universitas Muhammadiyah Sidoarjo, Indonesia

Evi Rinata, Universitas Muhammadiyah Sidoarjo, Indonesia

M Faisal Amir, Universitas Muhammadiyah Sidoarjo, Indonesia

Dr. Hana Catur Wahyuni, Universitas Muhammadiyah Sidoarjo, Indonesia

Complete list of editorial team (link)

Complete list of indexing services for this journal (link)

How to submit to this journal (link)

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

Article information

Check this article update (crossmark)

Check this article impact (*)

Save this article to Mendeley

(*) Time for indexing process is various, depends on indexing database platform

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

Classification of Words of Wisdom in Indonesian on Twitter Using Naïve Bayes and Multinomial Naive Bayes

Klasifikasi Kalimat Mutiara Berbahasa Indonesia Pada Twitter Dengan Menggunakan Naïve Bayes dan Multinomial Naive Bayes

Andry Rachmadany, rachmadany@umsida.ac.id, (1)
Universitas Muhammadiyah Sidoarjo, Indonesia

Yuliana Melita Pranoto, ymp@stts.edu, (0)
Sekolah Tinggi Teknik Surabaya, Indonesia

Gunawan Gunawan, gunawan@stts.edu, (0)
Sekolah Tinggi Teknik Surabaya, Indonesia

(1) Corresponding author

Abstract

Quote is a sentence made with the hope that someone becomes a strong personality, an individual who always improves himself to advance and achieve success. Social media is a place for people to express their hearts to the world which is sometimes a heart expression in the form of quotes. The purpose of this study is to classify Indonesian quotes on Twitter using Naïve Bayes and Multinomial Naïve Bayes. This experiment uses text classification from Twitter data written by Twitter users whether the quotes are then classified again into 6 categories (Love, Life, Motivation, Education, Religion, Others). The language used is Indonesian. The methods used are Naive Bayes and Multinomial Naïve Bayes. Results of this experiment is a classified Indonesian quote collection web application. This classification makes it easy for users to search for quotes based on class or keyword. For example, when a user wants to search for 'motivational' quotes, this classification can be very useful.

Published date: 2020-10-15 00:00:00

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

Pendahuluan

Twitter yang merupakan salah satu blog mikro yang terkenal adalah layanan jejaring sosial yang memungkinkan pengguna untuk berbagi informasi dalam bentuk teks singkat 280 karakter. Pengguna Twitter sangat banyak sekali dan dari berbagai kalangan, setiap pengguna dapat berlangganan untuk menerima pesan dari pengguna lain dengan menjadi 'pengikut'. Twitter telah berperan penting dalam memberikan informasi kepada pengguna dan juga telah mengubah cara orang menyampaikan informasi salah satunya yaitu penyampaian kutipan atau kalimat mutiara, yaitu kalimat yang dibuat dengan harapan seseorang bisa menjadi pribadi-pribadi yang kuat, pribadi-pribadi yang selalu memperbaiki diri untuk maju dan meraih kesuksesan. Kalimat mutiara di tulis dan di sampaikan agar dapat memberi manfaat bagi para pembacanya, sehingga akan selalu menjadi orang orang yang selalu maju menuju kesuksesan tanpa harus kehilangan jadi diri.

Namun, kalimat mutiara yang diposting pada Twitter memerlukan upaya untuk menemukannya. Meskipun mencari dengan kata kunci pada Twitter, hasil tweet yang di tampilkan tidak semua merupakan kalimat mutiara.

1. Untuk mendapatkan kalimat mutiara pengguna harus membaca tweet tersebut satu persatu untuk memastikan apakah tweet tersebut merupakan kalimat mutiara atau bukan.
2. Pengguna kemudian juga perlu menyeleksi apakah tweet tersebut merupakan kalimat mutiara yang sesuai dengan yang dicari misalnya ingin mencari kalimat mutiara tentang cinta, kehidupan, motivasi, agama, pendidikan dan lainnya.

Untuk meringankan masalah ini dan untuk menyampaikan data kalimat mutiara secara efektif, penelitian ini diharapkan dapat menghasilkan aplikasi web yang dapat mengumpulkan dan mengelompokkan kalimat mutiara bahasa Indonesia dari data yang diperoleh melalui Twitter dengan menggunakan metode klasifikasi Naïve Bayes dan Multinomial Naive Bayes.

Metode

Uji coba dilakukan mengambil data dari twitter menggunakan API yang disediakan oleh twitter dengan bahasa pemrograman PHP. Data tersebut akan di ambil secara acak 1700 data tweet, kemudian dibagi menjadi 1500 data digunakan sebagai data label dan 200 data sebagai data testing.

Data label yang berupa pasangan tweet dan kelas dijadikan sebagai sumber pembentukan model analisis. Setiap fitur yang merepresentasikan tweet dihitung probabilitas kemunculannya. Setelah dilakukan pelabelan manual kemudian tweet tersebut diproses untuk tahap awal preprocessing dengan melakukan Tokenisasi, Normalisasi Fitur, Case Folding, Clean Number, Stopword Removal, dan Stemming.

Dari hasil preposesing tersebut akan mendapatkan tweet yang bersih, dan selanjutnya dapat dilakukan proses pembobotan untuk mendapatkan probabilitas. Probabilitas yang didapatkan yaitu probabilitas mengenali kalimat mutiara atau bukan dan probabilitas 6 kategori(Cinta, Kehidupan, Motivasi, Agama, Pendidikan, Lainnya) untuk jenis kalimat mutiara dengan menggunakan metode klasifikasi Naive Bayes dan Multinomial Naive Bayes Figure 1.

Figure 1. *Rumus Probabilitas* Dimana:P = Probabilitasc = KelasN = Jumlahw = Kata pada dokumenv = Jumlah kata unik dalam dokumen kelasd = Dokumen ke-nk = Jumlah kata pada dokumen

Sebelum melakukan klasifikasi maka perlu dilakukan tahap preproses dimana tahap ini adalah pengolahan data mentah yang nantinya akan dipakai baik untuk data label ataupun data uji yang berasal dari blog mikro Twitter yang di ambil menggunakan API Twitter, praproses sangat berguna karena dapat mengurangi ukuran index, menambah akurasi serta mendukung proses klasifikasi. Tahap yang dilakukan meliputi Tokenisasi, Normalisasi Fitur, Case Folding, Clean Number, Stopword Removal, dan Stemming.

Setelah tahap preprocessing langkah selanjutnya pembentukan model analisis menggunakan metode pembelajaran mesin Naïve Bayes dan Multinomial Naïve Bayes untuk pembentukan model. Setiap fitur yang merepresentasikan tweet dihitung probabilitas kemunculannya di kelas yang termasuk tweet kalimat mutiara atau tweet yang termasuk bukan kalimat mutiara, apabila kalimat tersebut termasuk kalimat mutiara kemudian di representasikan probabilitasnya lagi ke setiap dokumen untuk setiap masing-masing kategori kalimat mutiara (Cinta, Kehidupan, Motivasi, Agama, Pendidikan, Lainnya). Ketika ada tweet baru yang ingin diklasifikasikan, maka nilai probabilitas dari setiap fitur untuk masing-masing kelas dikalikan. Nilai probabilitas yang paling besar akan dijadikan kelas tweet baru tersebut Figure 2.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

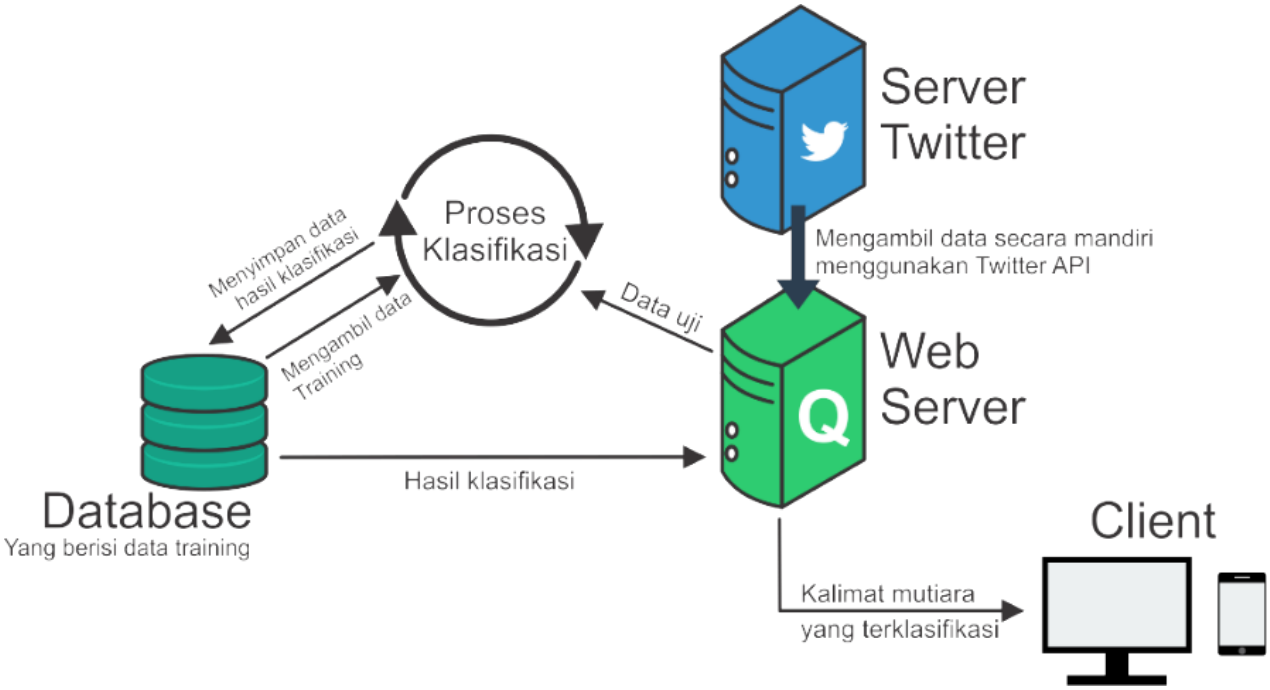


Figure 2. *Arsitektur Sistem Web*

Hasil

Figure 3 menunjukkan proses pembuatan data latih dimana terdapat proses (Tokenisasi, Normalisasi, Case Folding, Clean Number, Stopword Removal, Stemming). Praproses dilakukan pada setiap dokumen tweet yang masuk.

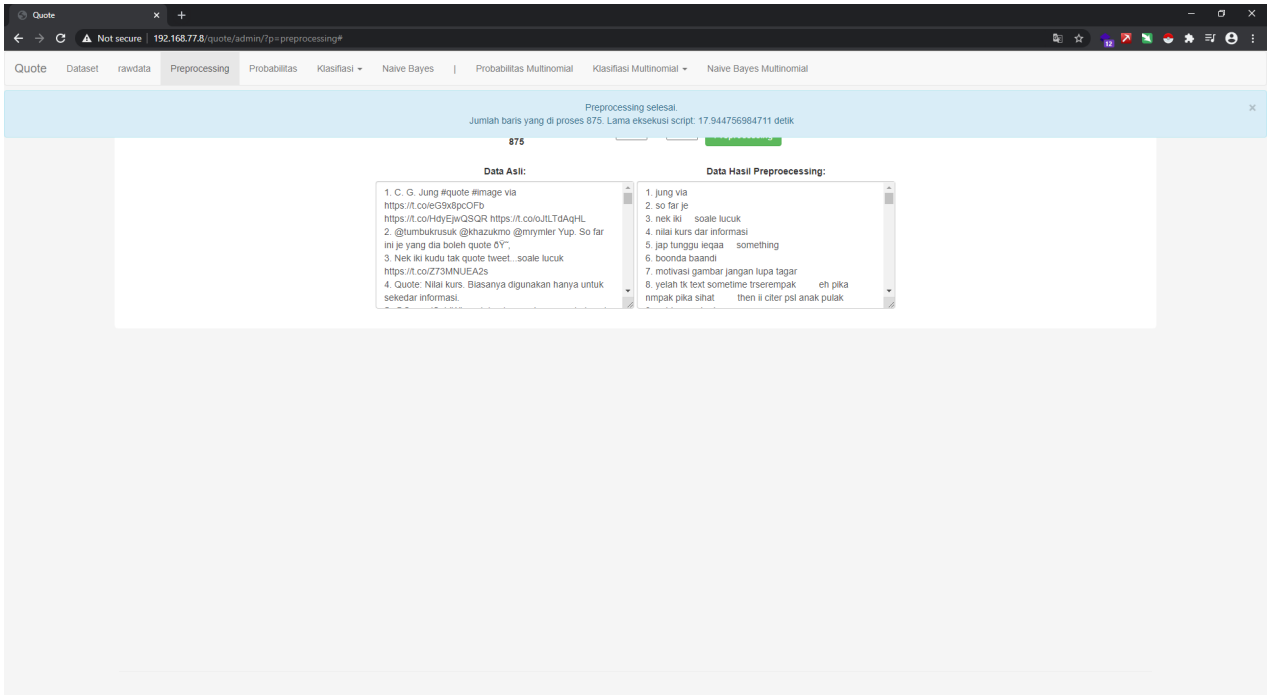


Figure 3. *Praproses Dokumen Tweet*

Figure 4 menunjukkan proses pembuatan nilai probabilitas. Setiap kata dihitung untuk mendapatkan probabilitasnya menggunakan perhitungan Naive Bayes dan Multinomial Naive Bayes.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Academia Open
Vol 3 (2020): December
DOI: 10.21070/acopen.3.2020.787 . Article type: (Engineering)

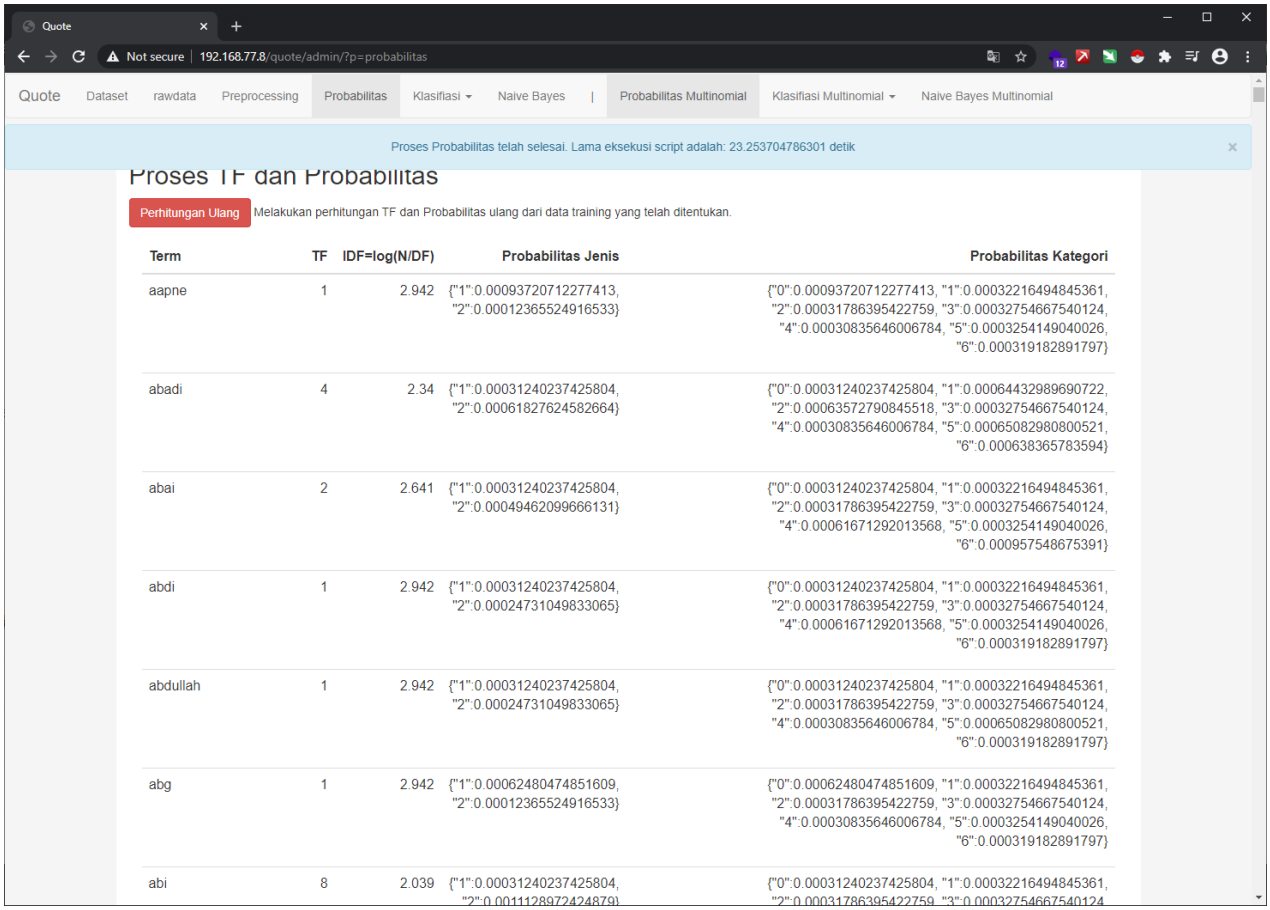
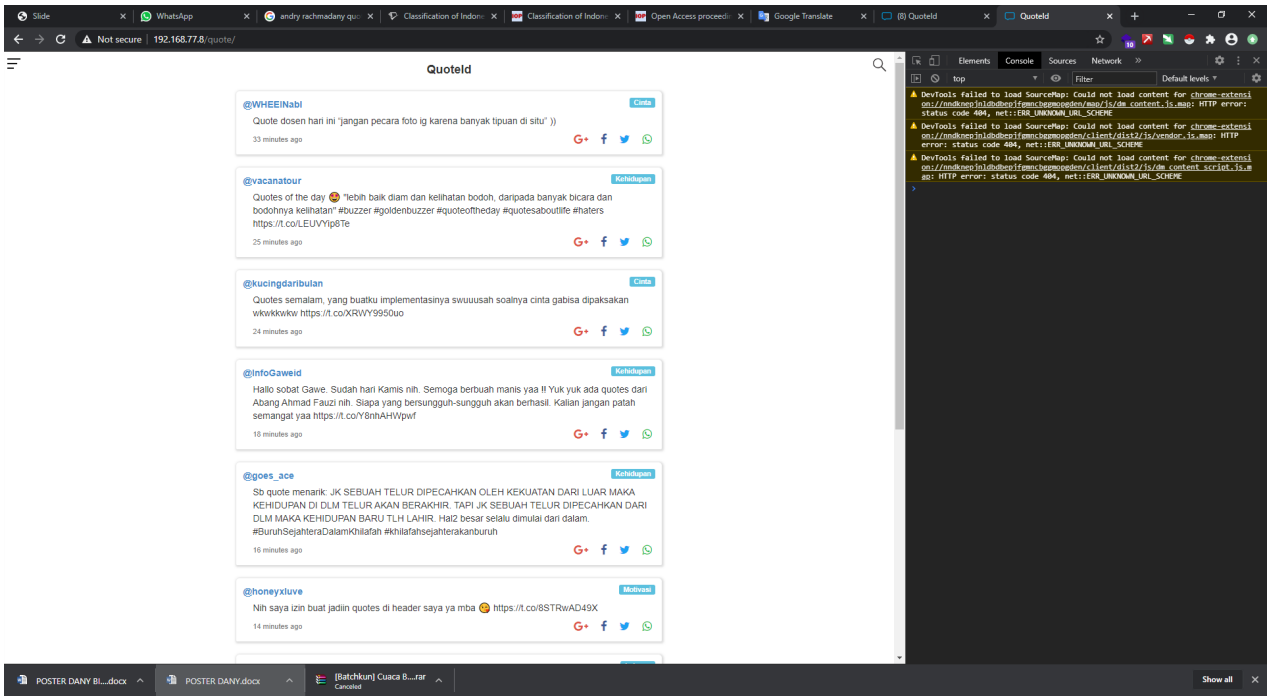


Figure 4. *Praproses Dokumen Tweet 2*

Figure 5 menujukkan hasil dari proses klasifikasi kalimat mutiara pada twitter yang telah di proses menggunakan Naive Bayes dan Multinomial Naive Bayes yang telah di klasifikasi kedalam enam kelas (Cinta, Kehidupan, Motivasi, Religi, Pendidikan, Lainnya).



ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Figure 5. *Aplikasi Website Hasil Klasifikasi*

Figure 5 mengkonfirmasi bahwa penelitian ini efektif untuk mengklasifikasikan kutipan bahasa Indonesia. Hasil percobaan ini juga memberikan aplikasi web kumpulan kutipan bahasa Indonesia yang telah diklasifikasikan. Klasifikasi ini memudahkan pengguna dalam mencari kutipan berdasarkan kelas atau kata kunci. Misalnya, ketika pengguna ingin mencari kutipan 'motivasi', klasifikasi ini akan sangat berguna.

Kesimpulan

Dari hasil penelitian dan berdasarkan atas hipotesa penelitian, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Metode Naïve Bayes dan Multinomial Naïve Bayes yang diusulkan terbukti mampu melakukan klasifikasi dokumen kalimat mutiara pada twitter.
2. Dengan Menggunakan metode klasifikasi teks yaitu metode Naïve Bayes Classifier, dari segi kecocokan jenis kelas yang dihasilkan oleh Naive Bayes Classifier terhadap penentuan jenis kelas pada media sosial twitter tergolong baik. Prosentasi kecocokan jenis kelas terhadap 200 data uji sebesar 81% untuk Naïve Bayes dan 87% untuk Multinomial Naïve Bayes.

Dengan pengklasifikasian data tweet menggunakan metode Naïve Bayes Classifier, dapat meningkatkan efektifitas pengguna dalam menyeleksi kalimat mutiara dan dapat memper cepat pencairan kalimat mutiara yang sebelumnya harus memilih jenis kelas secara manual pada saat pencarian di twitter.

References

1. Akbarisanto Ridho, Danar Wikan, Purwarianti Ayu. Analyzing Bandung Public Mood Using Twitter Data. School of Electrical Engineering and Informatics, Bandung, Indonesia, 2016 https://doi.org/10.1109/icoict.2016.7571910
2. Chris Tseng, Nishant Pateli, Hrishikesh Paranjape, T Y Lin,SooTee Teoh. Classifying Twitter Data with Naive Bayes Classifier. Computer Science Dept., San Jose State University, 2012.
3. Raveena Dayani, Nikita Chhabra, Taruna Kadian and Rishabh Kaushal. Rumor Detection in Twitter: An Analysis in Retrospect. Department of Information Technology Indira Gandhi Delhi Technical University for Women, Delhi, India, 2009. https://doi.org/10.26634/jes.6.2.14757
4. Sari Widya Sihwi, Insan Prasetya Jati, Rini Anggrainingsih. Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier. Universitas Sebelas Maret Surakarta, Indonesia. 2018. https://doi.org/10.1109/isemantic.2018.8549757
5. J. Han, M. Kamber, dan J. Pei, Penambangan Data: Konsep dan Teknik ,Edit Ketiga., Vol. 3. Morgan Kaufmann, 2012.
6. TwitterStreamingAPI.http://apiwiki.twitter.com/Streaming-API-Documentation.AccessedFebruary1,2010.
7. Apache Lucene. http://lucene.apache.org. Accessed February1, 2010.C. D. Manning, P. Raghavan, and H. Schutze.Introdution toInformation Retrieval, pages 108–115. Cambridge UniversityPress, New York, 2008.
8. J.R. Finkel, T. Grenager, and C. Manning.IncorporatingNon-local Information into Information Extraction Systems byGibbs Sampling.Proceedings of the 43nd Annual Meetingof the Association for Computational Linguistics (ACL 2005),pages 363–370, 2005. https://doi.org/10.3115/1219840.1219885
9. V. Qazvinian, E. Rosengren, Dragomir R. Radev, Q. Mei.Rumor hasit: Identifying Misinformation in Microblogs. In Proceedings of the2011 Conference on Empirical Methods in Natural Language Processing(2011).
10. Goncalves Eduardo Correa, NBBR: A Baseline Method for the Evaluation of Bayesian Multi-label Classification Algorithms. Universidade Federal Fluminense (UFF). Niteroi, Rio de Janeiro 24210–240, 2014. https://doi.org/10.1109/iccsa.2014.56

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo
Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).