

ANALISIS SENTIMEN TERHADAP PEMILIHAN GUBERNUR PROVINSI SULAWESI SELATAN TAHUN 2018 MENGGUNAKAN DATA TWITTER DAN METODE SENTIMENT LEXICON

ANALYSIS OF SENTIMENT AGAINST THE ELECTION OF THE GOVERNOR OF THE PROVINCE OF SOUTH SULAWESI, THE YEAR 2018 USING THE TWITTER DATA AND METHODS SENTIMENT LEXICON

FarahFatimah¹, Riza Agustiansyah, ST., M. Kom², Ir. Ahmad Musnansyah, M.Sc³

^{1,2,3}Prodi Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom

¹Farahfatimah03@gmail.com, ²riza.agustiansyah@gmail.com, ³ahmadanc@gmail.com

Abstrak

Pemilihan umum tahun 2018 dilakukan serentak di seluruh Indonesia termasuk wilayah provinsi Sulawesi Selatan dalam pemilihan calon gubernur dan wakil gubernur 2018. Banyak masyarakat yang menyampaikan opini atau pendapat mereka tentang sosok calon gubernur dan wakil gubernur Sulawesi Selatan sehingga banyak opini positif dan negatif. *Twitter* masih menjadi media sosial favorit untuk melakukan kampanye karena efisien dan efektif. Penelitian ini akan melakukan analisis sentimen masyarakat terhadap pemilihan umum calon gubernur dan wakil gubernur provinsi Sulawesi Selatan 2018 dengan mengambil data *Twitter*. Ada beberapa tahapan dalam menganalisis sentimen diantaranya: pengumpulan data, *preprocessing* data, menentukan *class sentiment lexicon* dan klasifikasi opini menggunakan SVM *Support Vector Machine* (SVM). Data *Twitter* dikumpulkan selama kurang lebih 4 bulan dari tanggal 25 Maret 2018 sampai 30 Juni 2018 yang berjumlah 165,779 *tweet*. Secara umum dapat disimpulkan pasangan Nurdin Abdullah – Andi Sulaiman memiliki sentimen positif paling tinggi yaitu 51% jika dibandingkan dengan empat pasangan lainnya pasangan Nurdin Halid – Aziz Mudzakkar dengan sentimen positif 21%, pasangan Ichsan Yasin – Andi Mudzakkar dengan sentimen positif sebesar 22%, pasangan Agus Arifin – Tanribali dengan sentimen positif sebesar 8%.

Kata kunci *Twitter*, Analisis Sentimen, Metode *Sentiment Lexicon-Base*.

Abstract

The length Election year 2018 conducted simultaneously throughout Indonesia including South Sulawesi province in the election of Governor and Vice Governor candidate for 2018. Many communities that deliver opinions or their opinion about the figure of the candidate Governor and Vice Governor of South Sulawesi, so many positive and negative opinions. *Twitter* is still a favorite to do social media campaign because of the efficient and effective. This research will conduct an analysis of the public sentiment against the general election candidate for Governor and Vice Governor of South Sulawesi province 2018 with *Twitter* data. There are several stages in analyzing sentiment include: data collection, data preprocessing, specify the class sentiment lexicon and classification opinions using SVM *Support Vector Machine* (SVM). *Twitter* data collected for a little over 4 months from March 25 until 30 June 2018 totaling 165.779 *tweet*. In general, it can be concluded the couple Nurdin Abdullah.

Keywords: *Twitter*, Sentiment Analysis, Methods of *Sentiment Lexicon-Base*.

1. Pendahuluan

Perkembangan informasi yang pesat dan luas menyebabkan munculnya berbagai macam media sosial yang memberikan kemudahan bagi pengguna untuk melakukan komunikasi. Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menyebutkan pengguna internet di Indonesia pada tahun 2017 mencapai 143,26 juta jiwa angka tersebut terus meningkat dibandingkan tahun sebelumnya yang mencapai 132,7 juta jiwa. Indonesia tercatat memiliki 130 juta pengguna aktif media sosial. Adapun, setidaknya 49% dari total populasi penduduk memiliki dan aktif menggunakan media sosial. Dengan 120 juta pengguna di antaranya menggunakan ponsel untuk terhubung dengan akun media sosialnya[1]. Indonesia masuk dalam peringkat kelima dengan 29

juta warga Indonesia yang memiliki akun *Twitter*. Rata-rata orang Indonesia mem-posting sebanyak 32 *tweet* per harinya.

Twitter adalah sebuah situs *micro blogging* yang memungkinkan pengguna untuk berbagi informasi dalam maksimum 140 bentuk karakter tulisan. *Twitter* itu sendiri merupakan media sosial yang sangat akrab bagi orang Indonesia. Keunikan *Twitter* adalah ia merupakan jejaring media sosial yang dapat mendeteksi topik secara cepat dibandingkan dengan jejaring media sosial lainnya. Diketahui bahwa *Twitter* populer dipakai untuk menulis opini atau pendapat dari penggunanya, bahkan dewasa ini tokoh-tokoh politik lazim menggunakan *Twitter* untuk mengungkapkan pandangan politik atau tanggapan terhadap suatu hal yang sedang marak. Keunikan *Twitter* ini menjadi latar belakang dilakukannya analisis sentimen pengguna *Twitter* terhadap calon gubernur Sulawesi Selatan dalam pilkada 2018, selain itu juga *twitter* mungkin memiliki semua permasalahan politik yang ada dan data *twitter* bersifat transparan[2]. Di samping itu *Twitter* juga menyediakan API (Application Programming Interface) yang memudahkan kita mengakses data *tweet* yang ada.

Analisa sentimen adalah cabang ilmu pembelajaran di domain *text mining* yang mempelajari analisis terhadap suatu opini, sentimen, emosi, sikap, evaluasi yang dituangkan ke dalam bentuk tekstural[3]. Banyak yang melakukan penelitian tentang analisis sentimen karena dapat memberikan penilaian atau evaluasi terhadap suatu produk, pelayanan dan keadaan yang sedang terjadi saat ini dari sebuah opini yang berbentuk text, sms, *tweet* atau dokumen. Data *Twitter* yang diambil berupa *tweet* dan dilakukan *text mining*. Melakukan *text mining* pada *tweet* yang ditulis, maka kita dapat lihat apakah *tweet* tersebut bernilai positif atau negatif. Penelitian ini menggunakan metode *sentimen lexicon* untuk menentukan atau mengungkapkan sentimen politik dari masing-masing kandidat. Data *tweet* yang telah dianalisis hasil akhirnya berupa klasifikasi *tweet* positif dan negatif serta persentase yang akan dibandingkan dengan hasil pemilihan gubernur Sulawesi Selatan.

Bahasa pemrograman *python* digunakan pada penelitian ini untuk menganalisis *tweet* para calon pasangan gubernur. Adanya penelitian tersebut diharapkan dapat diketahui hasil sentimen masyarakat untuk pasangan calon Gubernur Sulawesi Selatan pada tahun 2018 karena pada tahun 2014 ibukota Sulawesi Selatan Makassar menjadi kota ke tujuh pengguna *twitter* terbanyak. Algoritma *Support Vector Machine* juga digunakan untuk menguji tingkat akurasi metode *Sentimen Lexicon*. Penelitian ini diharapkan dapat mengetahui sentimen masyarakat untuk pasangan calon Gubernur.

2. Dasar Teori

2.1 Data

Data adalah sekumpulan data yang memiliki fakta apa adanya. Dalam komputasi, data adalah informasi yang telah diterjemahkan dalam bentuk yang bisa diproses oleh komputer [4]

2.2 Twitter

Twitter adalah situs *micro blogging* yang memungkinkan pengguna untuk menulis teks singkat yang disebut *tweet* yang memiliki batasan 140 karakter dan dapat dilihat oleh pengguna lainnya. Pada tahun 2017 Twitter resmi melakukan *update* batasan karakter *tweet* yang awalnya 140 menjadi 280 karakter yang ditampilkan di halaman penggunanya[5].

2.2 Python

Python menjadi bahasa pemrograman yang amat populer untuk melakukan analisis data, alasan utamanya adalah ketersediaan pustakanya untuk memudahkan tugas-tugas analisis data [6] Beberapa library python yang digunakan pada penelitian yaitu, *tweepy* yang berfungsi untuk mengambil data twitter, *pandas* yang berfungsi untuk memuat teks Twitter ke dalam sebuah table virtual seperti *spreadsheet* sebelum data diolah menggunakan *Numpy*. *matplotlib* sebagai alat visualisasi *default* dan *Numpy* menyediakan struktur data, algoritma, dan berbagai piranti *library* yang dibutuhkan untuk sebagian besar aplikasi ilmiah yang melibatkan data numerik dengan Python.

2.3 Text Mining

Text Mining adalah suatu teknologi yang digunakan untuk menganalisis data teks yang tidak terstruktur. Fase utama dalam *text mining* yaitu *Preprocessing* dari kalimat yang tidak terstruktur menjadi kalimat yang lebih terstruktur. *Text mining* adalah suatu transformasi dari data teks menjadi data numerik, dengan kata lain *text mining* mengubah *unstructured* data menjadi *structured* data[7].

2.4. Analisis Sentimen

Sentimen Analisis bisa juga diartikan sebagai cara untuk mendapatkan sentimen dari sebuah teks [8].

Analisis Sentimen juga dikenal sebagai *Opinion Mining* adalah bidang dalam *Natural Language Processing* (NLP) yang membangun sistem yang mencoba mengidentifikasi dan mengekstrak pendapat dalam teks. Saat ini, analisis sentimen adalah topik yang sangat menarik dan berkembang karena memiliki banyak aplikasi praktis.

2.5 Sentimen Lexicon

Leksikon dapat diartikan sebagai kosa kata, yang di mana di dalamnya terdapat atau terkandung kata-kata dan ekspresi dalam mengutarakannya[9]. Sedangkan sentimen lexicon sebuah klasifikasi yang menggunakan frasa dan idiom. Sesuai dengan gambar II-II di atas, ada dua pendekatan umum bagaimana melakukannya: 1) Machine Learning Approach (MLA)-pendekatan berbasis pengelompokan dan 2) Lexicon Based Approach (LBA)-pendekatan berbasis kamus/kata. Pada pendekatan MLA teks diklasifikasi dan disimpulkan opininya menggunakan metode standar pada machine learning. Sedangkan LBS, memberi label sentimen pada setiap kosa katanya

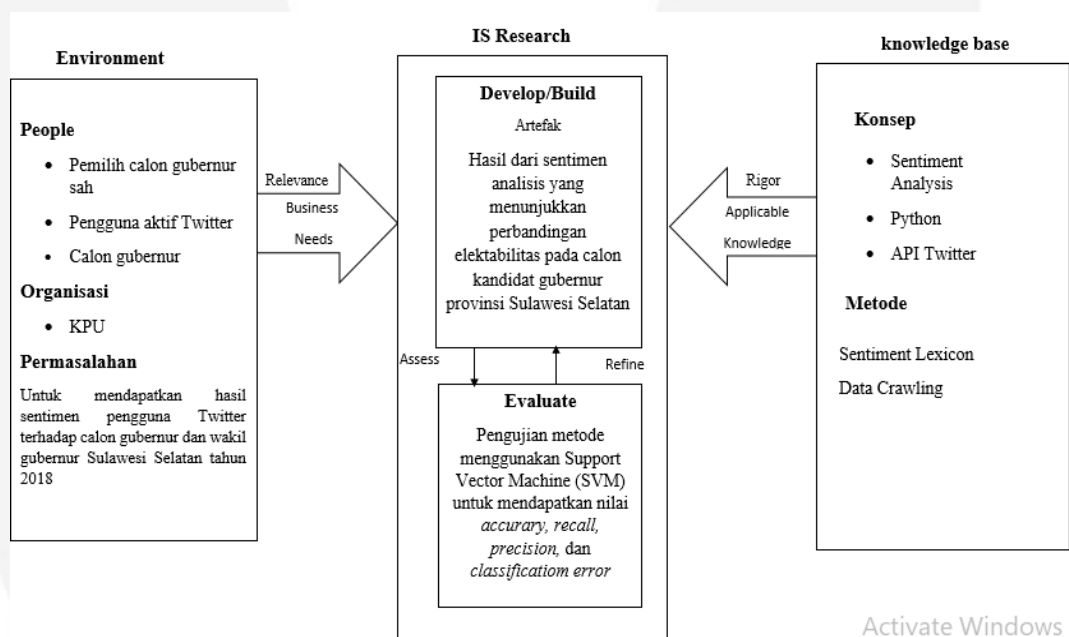
2.6 Support Vector Machine (SVM)

Support Vector Machine merupakan model pembelajaran metode algoritme *machine learning* yang menganalisis data untuk klasifikasi dan analisis regresi. SVM memiliki kelebihan yaitu mampu mengidentifikasi *hyperplane* (garis) terpisah yang memaksimalkan margin antara dua kelas yang berbeda[10].

3. Metodologi Penelitian

3.1. Model Konseptual

Gambar 1 Model Konseptual



Dalam penelitian dilakukan atas dasar oleh adanya keinginan untuk menganalisis opini masyarakat terhadap calon gubernur dan wakil gubernur Provinsi Sulawesi Selatan, dari tweet yang di post oleh pengguna twitter secara luas. Pada gambar 1[11], dengan memanfaatkan teknologi yang ada untuk mendapatkan data tweet dari *twitter* dengan menggunakan *API Twitter* serta library *python* yang bernama *tweepy*, Maka kita dapat membuat sebuah analisis sentimen yang di mana nantinya akan dapat menunjukan hasil perbandingan persentase dukungan terhadap setiap kandidat calon gubernur dan wakil gubernur provinsi Sulawesi Selatan, Dalam penelitian ini menggunakan konsep analisis sentiment dengan metode *sentiment lexicon* serta data yang di ambil dengan cara *crawling* di *Twitter*, *tweet* yang di ambil berdasarkan kata kunci setiap kandidat nama calon gubernur dan wakil gubernur, tagline dan slogan masing-masing kandidat. Hasil akhir yang di dapat dari hasil analisis pada penelitian akan dibandingkan dengan hasil Komisi Pemilihan Umum (KPU) dan akan dilakukan evaluasi metode dengan menggunakan *Support Vector Machine (SVM)*.

4. Pembahasan Dan Analisis

4.1 DataSet

Dataset yang digunakan dalam penelitian ini diambil dari tanggal 25 Maret - 30 Juni 2018 dengan menggunakan *Tweepy*. Untuk mendapatkan *tweet* yang dibutuhkan kata kunci atau keyword yang berkaitan dengan kandidat gubernur dan wakil gubernur sebagai contohnya “NH-Aziz” atau “Prof Andalan” Tabel 1 menunjukkan rincian informasi yang didapatkan dari dataset:

Table 1 Jumlah Tweet yang dikumpulkan

Bulan	Jumlah Dataset
Maret	11.439
April	12.082
Mei	17.171
Juni	125.087
Total	165,779

Dataset yang digunakan berasal dari *tweet* pengguna Twitter yang berasal dari Indonesia. Dalam satu *tweet* yang ditulis oleh pengguna untuk pasangan calon akan mewakili satu kalimat komentar.

4.2 Preparing Corpus

Tahap ini dilakukan untuk mengambil seluruh data text yang didapat dari *tweet* hasil *crawling* yang kemudian dimasukkan kedalam sebuah dokumen yang berisi kumpulan data text.

4.3 Preprocessing

Tahap *preprocessing* dilakukan untuk pemilihan domain dataset, yaitu dengan menghilangkan *tweet* yang tidak dibutuhkan yang tidak memiliki kalimat opini (hanya terdapat URL) dalam *tweet* tersebut. Dalam tahap ini akan juga menghilangkan stopwordnya, seperti “yg”, “jg,”, “atau”, “karena”, “yaitu” dan lain sebagainya. Proses *preprocessing*, parameter data yang akan digunakan hanya id *Twitter* dan *tweet*.

4.4 Splitting Corpus

Tahapan *splitting corpus* ini bertujuan untuk memisahkan data *tweet* yang terkumpul berdasarkan dari masing-masing kandidat pasangan gubernur dan wakil gubernur. *Tweet* dari dataset hasil pre-processing akan dipisahkan menjadi empat corpus yaitu corpus kandidat pasangan Nurdin Halid - Aziz Mudzakkar, corpus kandidat pasangan Nurdin Abdullah – Andi Sulaiman, corpus kandidat Ichsyan Yasin – Andi Mudzakkar dan corpus kandidat Agus Arifin – Tanribali.

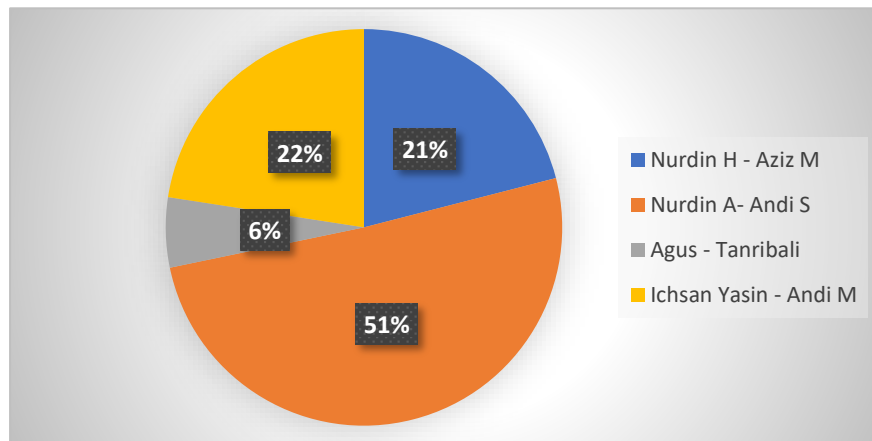
4.5 Klasifikasi Sentimen Kandidat

Pada table 2 menunjukan klasifikasi sentiment dengan menggunakan lexicon based memperlihatkan perolehan *tweet* dan presentase pada setiap kandidat:

Table 2 Klasifikasi Sentimen dengan lexicon

No.	Kandidat	Positif Sentimen		Negatif Sentimen		Jumlah Tweet
		Tweet	Presentase	Tweet	Presentase	
1	Nurdin H-Aziz M	2,610	53%	2,278	48%	4,888 tweets
2	Nurdin A- Andi S	6,306	75%	2,121	25%	8,427 tweets
3	Ichsyan Y- Tanribali	2.804	87%	420	13%	3,224 tweets
4	Agus A-Andi M	709	52%	650	48%	1,359 tweets

4.6 Perbandingan Sentiment Positif Kandidat



Gambar 2 Perbandingan tweet antar empat kandidat

Pada gambar 2 menunjukan jumlah *tweet* positif pasangan Nurdin Halid – Aziz Mudzakkar sebesar 21%, pasangan Nurdin Abdullah – Andi Sulaiman sebesar 51%, pasangan Ichsan Yasin – Andi Mudzakkar sebesar 22 % dan pasangan Agus Arifin - Tanribali sebesar 6 % dari keseluruhan jumlah *tweet* positif sebanyak 17,898 *tweet*.

4.7 Perbandingan dengan Hasil Resmi KPU

Untuk mengevaluasi dari hasil analisis sentimen pada penelitian ini, maka di perlukan perbandingan dengan hasil resmi pemilihan yang di Komisi Pemilihan Umum lakukan KPU. Hasil perbandingan antara hasil analisis sentimen dan hasil resmi KPU dapat dilihat di tabel 3.

Table 3 Perbandingan Hasil Sentimen dengan Hasil KPU

No.	Kandidat	Hasil		Selisih
		Data Crawling	Hasil Resmi KPU	
1.	Nurdin H -Aziz M	21%	27,32%	6,32%
2.	Nurdin A-Andi S	51%	43.87%	7,13%
3.	Ichsan Y-Andi M	22%	18.97%	3.03%
4.	Agus A-Tanribali	6%	9.85%	3.85%
Hasil Selisih				7.69 %

4.8 Evaluasi Hasil Penelitian

Proses evaluasi ada hasil penelitian ini akan menggunakan metode Support Vector Machine (SVM). Pengukuran uji metode ini untuk menentukan nilai *Recall*, *Precession*, *Accuracy* dan *Classification Error* dari data hasil penelitian. Data set yang digunakan ada, yaitu data training dan data testing. Data training yang digunakan adalah dataset yang telah dilakukan proses preprocessing dan analisis sentimen pada bulan 25 Maret – 30 April 2018, yang dimana data training tersebut diberikan sebuah label positif dan negatif berdasarkan role dari hasil analisis sentimen berbasis leksikon. Data testing merupakan dataset yang diambil pada bulan Juni 2018 (1 Juni – 30 Juni 2018) untuk menentukan hasil Recall, Accuracy, Precision dan Classification Error-nya. Pada tabel 4 menunjukan jumlah dataset yang digunakan.

Table 4 Rincian data training dan data testing

Data set	Positif	Negatif	Jumlah
Data Training	8.712	3.053	11.765
Data Testing	8.375	1.320	9.695
Jumlah data set			21.460

Hasil perhitungan *Recall*, *Precession*, *Accuracy* dan *Classification Error* dari pengujian metode SVM pada tabel dibawah:

Table 5 Hasil accuracy, recall, precision, dan classification error dengan SVM

variabel	Hasil Perhitungan
<i>Recall</i>	60.25%
<i>Precession</i>	96.28%
<i>Accuracy</i>	94.58%
<i>Classification Error</i>	5.42%

Pada table 5, variabel *recall* mendapatkan persentase sebesar 60.25% yang dimana menunjukan sebuah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Variabel *precision* mendapatkan persentase sebesar 96.28% yang menunjukkan tingkat ketepatan antara informasi yang ditentukan dengan jawaban yang diberikan SVM. Variabel *accuracy* mendapatkan persentase sebesar 94.58% yang menunjukkan tingkat akurasi antara nilai prediksi dan nilai aktual. *Classification error* mendapatkan persentase sebesar 5.42% yang menunjukkan tingkat kesalahan klasifikasi.

Daftar Pustaka:

- [1] Duwi Setiya Ariyanti, "Separuh Populasi Indonesia Aktif di Medsos," 31-Jan-2018. [Daring]. Tersedia pada: <https://ekonomi.bisnis.com/read/20180131/105/732548/130-juta-penduduk-indonesia-pengguna-media-sosial>. [Diakses: 28-Des-2018].
- [2] J. Tidey, "Social network data: Twitter vs FB vs Google vs everyone else," *Towards Data Science*, 27-Nov-2017. [Daring]. Tersedia pada: <https://towardsdatascience.com/social-network-data-twitter-vs-fb-vs-google-vs-everyone-else-830ea0291c86>. [Diakses: 14-Jan-2019].
- [3] B. Liu, P. S. Yu, dan Xiaowen Ding, "A holistic lexicon-based approach to opinion mining," dalam *Proceedings of the international conference on Web search and web data mining - WSDM '08*, Palo Alto, California, USA, 2008, hlm. 231.
- [4] Dale Janssen, "What is Data? - Definition from Techopedia." [Daring]. Tersedia pada: <https://www.techopedia.com/definition/807/data>. [Diakses: 20-Jan-2019].
- [5] Y. Ratnasari, "Twitter Resmi Menambah Batasan Kicauan Jadi 280 Karakter," *tirto.id*, 08-Nov-2017. [Daring]. Tersedia pada: <https://tirto.id/twitter-resmi-menambah-batasan-kicauan-jadi-280-karakter-czNm>. [Diakses: 19-Jan-2019].
- [6] Wes McKinney, "Python for Data Analysis, 2nd Edition [Book]," Okt-2017. [Daring]. Tersedia pada: <https://www.oreilly.com/library/view/python-for-data/9781491957653/>. [Diakses: 19-Jan-2019].
- [7] Walaamedha, AhmedHassan, dan HodaKorashy, "Sentiment analysis algorithms and applications: A survey - ScienceDirect," Des-2014. [Daring]. Tersedia pada: <https://www.sciencedirect.com/science/article/pii/S2090447914000550>. [Diakses: 20-Jan-2019].
- [8] S. S. Retno Purwani Sari, "LEKSIKON, GAGASAN, DAN MAKNA DALAM KONTEKS BUDAYA: KENDALA PENCARIAN PADANAN (KAJIAN TERJEMAHAN IDIOMATIS)," *SASTRA Ingg.*, vol. Volume 11, Okt 2013.
- [9] B. Liu, P. S. Yu, dan Xiaowen Ding, "A holistic lexicon-based approach to opinion mining," dalam *Proceedings of the international conference on Web search and web data mining - WSDM '08*, Palo Alto, California, USA, 2008, hlm. 231
- [10] H. Honakan, A. Adiwijaya, dan S. A. Faraby, "Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia," *EProceedings Eng.*, vol. 5, no. 1, Apr 2018.
- [11] Alan Hevner, "(PDF) Design Science in Information Systems Research," *ResearchGate*, Mar-2004. [Daring]. Tersedia pada: https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research. [Diakses: 19-Jan-2019].