

---

**RANCANG BANGUN SISTEM INFORMASI PORTAL BERITA  
MENGUNAKAN METODE CLUSTERING TEXT MINING (STUDI  
KASUS : KANTOR KEMENTERIAN AGAMA KOTA SURABAYA)**

Ahmad Aunul Bari Hayiz<sup>1)</sup>, Ahmad  
Habib<sup>2)</sup>, Agus Darwanto<sup>3)</sup>  
Teknik Informatika, Universitas 17  
Agustus 1945 Surabaya<sup>1</sup>

\*Email : [hvengez@gmail.com](mailto:hvengez@gmail.com)<sup>1)</sup>,  
[habib@untag-sby.ac.id](mailto:habib@untag-sby.ac.id)<sup>2)</sup>,  
[agusdarwanto@untag-sby.ac.id](mailto:agusdarwanto@untag-sby.ac.id)<sup>3)</sup>

**ABSTRAK**

Tujuan dirancangnya portal berita diharapkan dapat dengan mudah menyampaikan informasi yang secara cepat & akurat kepada semua lapisan masyarakat guna memenuhi integritas informasi kelas sosial saat ini. Portal ini dirancang dengan menggunakan PHP dan MySQL, kemudian proses clustering digunakan untuk mencari kata-kata atau kalimat apa yang paling sering muncul atau dibahas pada topik agama. Setelah proses cluster dilakukan, hasilnya berupa kata-kata yang sering muncul pada topik yang sedang dilakukan, selanjutnya dapat dilakukan input berita berdasarkan hasil cluster tersebut. Portal berita dan proses cluster ini dirancang sedemikian rupa untuk membangkitkan minat pengunjung untuk mendapatkan informasi yang lebih cepat dan menarik. Adanya proses analisis sentimen yang nantinya diharapkan dapat berfungsi sebagai evaluator dokumen dan konten berita yang bersifat positif, negatif atau netral. Pada tahapan cluster dan proses sentimen memiliki beberapa tahapan yang sama dalam proses preprocessing. Dimana preprocessing terdiri dari Tokenizing, Case folding, Stemming dan filtering.

Kata-kata kunci: Sistem Informasi, Portal Berita, Website, Cluster.

**ABSTRACT**

*The purpose of designing a news portal is expected to be able to easily convey information quickly and accurately to all levels of society in order to meet the information integrity of today's social class. This portal is designed using PHP and MySQL, then the clustering process is used to find what words or sentences most often appear or are discussed on religious topics. After the cluster process is carried out, the results are in the form of words that often appear on the topic being carried out, then news can be input based on the results of the cluster. This news portal and cluster process are designed in such a way as to arouse the interest of visitors to get information more quickly and interestingly. There is a sentiment analysis process which is later expected to function as an evaluator of documents and news content that is positive, negative or neutral. At the cluster stage and the sentiment process have some of the same stages in the preprocessing process. Where preprocessing consists of tokenizing, case folding, stemming and filtering.*

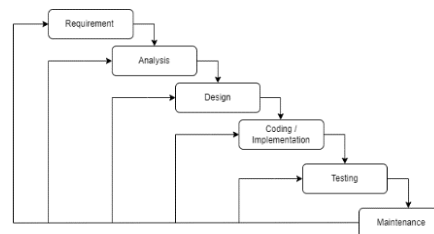
*Keywords: Information Systems, News Portals, Websites, Clusters.*

**Pendahuluan**

Berdasarkan latar belakang yang telah diuraikan secara abstrak, terdapat suatu permasalahan yang bisa dirumuskan antara lain, Bagaimana merancang dan mengimplementasikan sistem informasi portal berita Kementerian Agama. Masyarakat kesulitan menemukan situs portal berita yang memuat berita terkait Agama dan Kementerian Agama. Masyarakat dan pengguna belum dapat memberikan informasi dalam menunjang penambahan pengetahuan untuk kepentingan bersama. Penelitian ini bertujuan untuk membangun portal berita yang nantinya dapat membangun keinginan masyarakat untuk membaca berita. Dalam proses clustering akan berfungsi sebagai dasar untuk setiap berita yang harus dimasukkan dalam portal berita agar disukai oleh pembaca, kemudian dilakukan pencarian kata atau kalimat yang sering muncul pada judul berita di website berita lain. Selanjutnya dilakukan proses analisis sentimen untuk meninjau kalimat atau kata yang akan digunakan nantinya memiliki nilai positif, negatif atau netral.

### Metode

Metode penelitian yang akan digunakan pada penelitian terdiri dari Pengumpulan data dengan cara melakukan pengamatan langsung terkait objek penelitian pada objek yang sedang berlangsung khususnya objek yang diteliti. Melakukan proses tanya jawab kepada pihak-pihak terkait yang memberikan informasi tentang objek penelitian. Metode Pengembangan Sistem: Metode SDLC dengan model Waterfall. Ada 6 tahapan dalam metode waterfall, yaitu Requirement, Analysis, Design, Coding, Testing, Maintenance. Gambar di bawah ini adalah deskripsi dari metode waterfall.



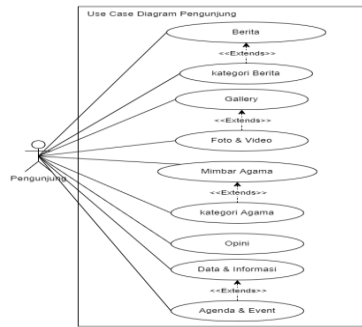
Gambar 1. Waterfall Method

#### A. Menjalankan Analisis Sistem

Analisis sistem adalah tahapan memahami suatu proses untuk mengetahui apa saja proses yang terlibat seperti bagaimana cara kerja suatu proses yang terlibat pada sistem dan hubungan antara suatu proses dan proses lainnya.

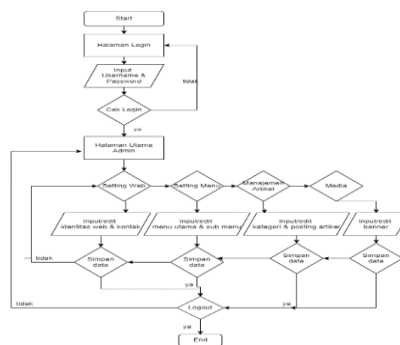
#### B. Analisis Sistem yang Diusulkan

Gambar di bawah ini adalah sistem yang diusulkan sebelumnya.



Gambar 2. Analisis sistem Yang diusulkan

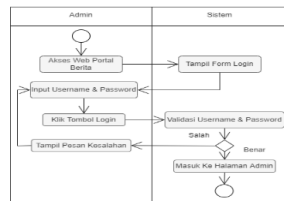
C. Flowchart System Admin



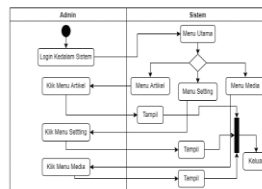
Gambar 3. Flow Chart System Admin

D. Diagram Activity Login Admin

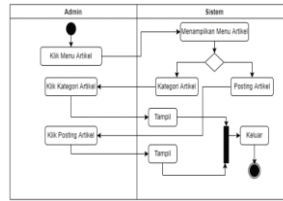
Hal pertama yaitu Admin melakukan login, sistem melakukan validasi username dan password benar atau salah, jika login berhasil akan masuk ke halaman admin yang nantinya dapat dikelola oleh admin.



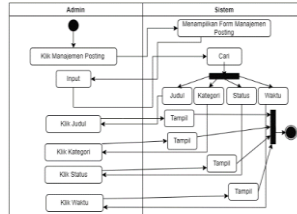
Gambar 4. Diagram Activity Login Admin



Gambar 5. Diagram Activity Main Menu



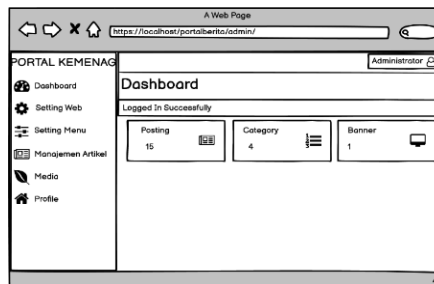
Gambar 6. Diagram Activity Menu Articles



Gambar 7. Diagram Activity Posting Articles

### E. Design Interface

Design Interface merupakan hal utama dalam pembuatan perangkat lunak. Dalam proses perancangan ini penulis membagi kebutuhan ke dalam perangkat lunak. Proses ini menghasilkan arsitektur perangkat lunak sehingga dapat diterjemahkan ke dalam kode program dan antarmuka.

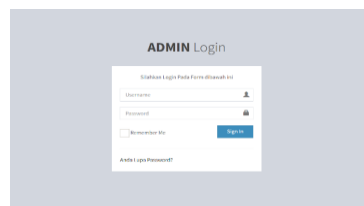


Gambar 8. Halaman Utama Admin

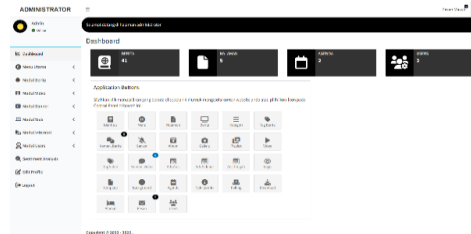
## Hasil dan Pembahasan

Perancangan dan Pembuatan Website di Kantor Kementerian Agama Kota Surabaya, adapun proses perancangan yaitu: kebutuhan fungsional, kebutuhan nonfungsional, perancangan model atau sistem, perancangan input dan output dan perancangan database. Website pada portal berita dirancang yang dapat digunakan oleh dua pengguna yaitu editor dan admin.

### A. Interface Halaman Admin

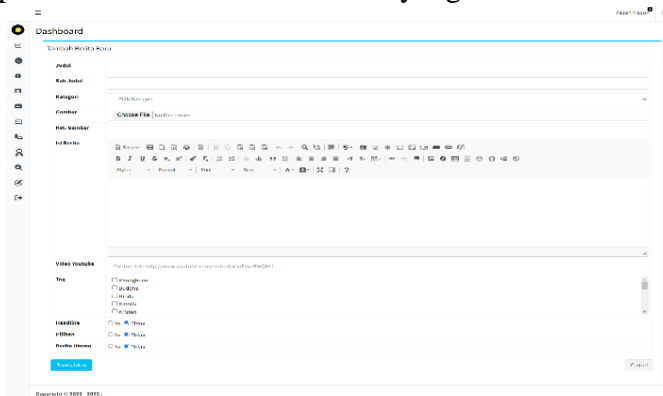


Gambar 9. Login Page Admin



Gambar 10. Interface Dashboard Admin

Terdapat banyak menu pada halaman admin seperti menu utama, modul berita, modul video, modul interaksi, modul pengguna dan lain-lain. Sedangkan untuk tampilan pada menu bisa dikatakan memiliki fungsi yang sama yaitu memiliki fitur edit, delete dan add. adapun tampilan halaman berita tambahan yang telah dibuat.



Gambar 11. Menu Add News

Tampilan halaman utama merupakan tampilan awal dari website portal berita dimana pada halaman ini pengunjung disediakan beberapa pilihan menu yang kemudian akan menuju ke halaman yang diinginkan menggunakan pilihan yang tersedia.



Gambar 12. Landing Page News Portal

B. Clustering Judul Berita Menggunakan Algoritma K means

Proses KMeans dimulai dengan menentukan himpunan cluster. Tujuannya adalah untuk menentukan sebuah jumlah cluster yang digunakan dan jumlah semua grup yang dicari. Jumlah cluster yang digunakan memiliki jumlah yang sama dengan jumlah judul berita yang digunakan. Proses selanjutnya menentukan centroid, dimana proses penentuan centroid ditentukan dengan cara acak dari titik-titik data. Terdapat beberapa proses dalam proses cluster yaitu Pengumpulan Data dan Analisis Persyaratan, Praproses Dokumen (Filtering, Tokenizing, Case Folding, Stemming). Proses cluster dapat dirujuk pada gambar di bawah ini.

	title	author	pagelink	article
0	Izinkan Pernikahan Beda Agama, PN Surabaya Dig...	Ikhwani Hastanto	https://www.vice.com/id/article/gpq58/pengedi...	Penggunaan beleid bikin bingung masyarakat pas...
1	Putusan PN Surabaya Buktikan Kalian Bisa Nikah...	VICE Staff	https://www.vice.com/id/article/5d3zjq/pn-sura...	"Ini ketidaktauhuan aparat hukum, Mereka penyal...
2	Guru Madrasah Tewas Digorek Gara-Gara Mimpi Ba...	Rimal Farukh	https://www.vice.com/id/article/5dga85/guru-ma...	Laporan Al Jazeera menentusikan, 1990, 62 Paki...
3	Begini Rasanya Jadi Anak Muda yang Harus Antre...	Prima Sulisty	https://www.vice.com/id/article/akezkk/antrean...	Melihat runggu 40-90 tahun menunaikan Rukun...
4	Sulit Nikahi Kekasihnya yang Beda Agama, Lelak...	Mahisa Cempaka	https://www.vice.com/id/article/z3nq95/uu-perk...	"Di [UU Perkawinan 1974] diatur pernikahan dil...

Gambar 13. Pengumpulan Data

Stemming adalah proses dimana menghilangkan perubahan kata ke bentuk dasarnya, misalnya kata "pernikahan", "dinikah", "menikah" akan diubah menjadi kata "nikah". Pada proses case folding kali ini dilakukan dengan tujuan untuk mengubah text menjadi lowercase. Tokenizing adalah proses pemisahan kata menjadi potongan potongan yang disebut sebagai token.

```

from Sastrawi.StopwordRemover.StopwordRemoverFactory import StopwordRemoverFactory
from nltk.tokenize import word_tokenize
factory = StopwordRemoverFactory()
stopword = factory.create_stop_word_remover()
kalimat = "izinkan pernikahan beda agama, pn surabaya digugat orang yang tak suka putusan itu"
stop = stopword.remove(kalimat)
tokens = nltk.tokenize.word_tokenize(stop)
print(tokens)

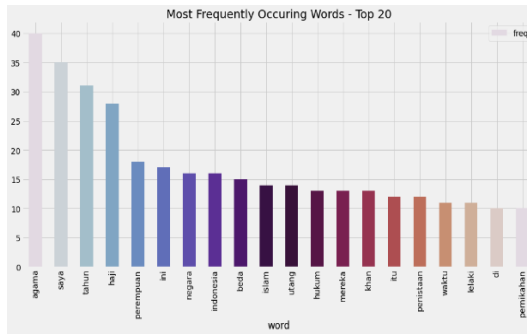
```

Gambar 14. Proses Filtering

Proses ini adalah proses dimana sistem akan mengambil kata-kata penting dari hasil token sebelumnya, pada hasil token menghasilkan kalimat "Izinkan', 'Pernikahan', 'Beda', 'Agama,', 'PN', 'Surabaya', 'Digugat', 'Orang', 'yang', 'Tak', 'Suka', 'Putusan', 'Itu']" maka pada proses filtering dihasilkan "[Izinkan', 'Pernikahan', 'Beda', 'Agama', 'PN', 'Surabaya', 'Digugat', 'Orang', 'Tak', 'Suka', 'Putusan', 'Itu']".

The screenshot shows a document preprocessing interface with two columns. The left column, titled 'scraped\_headlines', contains a list of news titles. The right column, titled 'headlines\_cleaned', contains the corresponding processed tokens for each title. The titles include: 'Izinkan Pernikahan Beda Agama, PN Surabaya Digugat Orang yang Tak Suka Putusan Itu', 'Putusan PN Surabaya Buktikan Kalian Bisa Nikah Beda Agama (Asal Direvisi Keluarga)', 'Guru Madrasah Tewas Digorek Gara-Gara Mimpi Bocah Soal Peristiwa Agama', 'Begini Rasanya Jadi Anak Muda yang Harus Antre Tahun untuk Nikah Haji', 'Sulit Nikahi Kekasihnya yang Beda Agama, Lelaki Papua Gugat UU Perkawinan ke PK', 'Dituduh Nistakan Agama Jelang Chat WA, Seorang Perempuan Divonis hukuman mati', 'Apas! Keherasan seksual pada murid Indonesia mayoritas terjadi di Sekolah Agama', 'Pemilik Restoran yang Viral Karena Soal Makanan Padang Non-Halal Diungguli Polisi', 'Kelasok ekstremis hindu di India gencar memarang cipta beda agama', 'Pencelakaan Nasik Terbalak 145 Berjulang Ancaman 212 Gertak Kedes Singapura', 'Alok Tak Akan Diperjara untuk kasus Peristiwa Agama', 'Pencus Orang-Orang Maluku Punya Keberanian Menjalani Profesi Penagih utang', 'Begini Rasanya Disandera Gerombolan Penculik Berperilaku Terpuji', 'Bagaimana Penyandang Tunanetra Merasakan Daya Tarik Seksual Pasanganya?', 'Trik-trik jitu menjawab pertanyaan soal gaji saat wawancara kerja', 'Agama itu kanibal dan Tuhan bisa berganti-ganti dalam serial Netflix asli India 'Sacred'.

Gambar 15. Proses Preprocessing Document



Gambar 16. Jumlah Term Yang Sering Muncul



Gambar 17. Word Cloud

Semakin banyak kata yang muncul dalam suatu kalimat ataupun dokumen, maka semakin tinggi nilai atau bobot kata tersebut.

No	Title	Cluster
1.	izinkan pernikahan beda agama pn surabaya digugat orang yang tak suka putusan itu	0
2.	putusan pn surabaya buktikan kalian bisa nikah beda agama asal direstui keluarga	0
3.	guru madrasah tewas digorok garagara mimpi bocah soal penistaan agama	0
4.	begini rasanya jadi anak muda yang harus antri tahun untuk naik haji	1
5.	dituduh nistakan agama lewat chat wa seorang perempuan divonis hukuman mati	0
6.	begini rasanya disandera gerombolan penculik berperilaku terpuji	1

Gambar 18. Cluster Result 1

Top terms per cluster	
Cluster 0	Cluster 1
Agama	saya
Penistaan	tahun
Hukum	haji
Perempuan	agama
Mati	ini
Tuduhan	negara

Gambar 19. Cluster Result 2

Hasil cluster judul berita dibagi menjadi 2 kelompok. Setelah proses cluster dilakukan, dapat menjadi pertimbangan bagi pengelola konten berita, bahwa judul berita yang disukai adalah topik agama yang kategorinya adalah berita tentang Haji.

C. Sentiment Analysis

Tujuannya sentiment analisis adalah untuk mengetahui bagaimana pendapat pengguna twitter di Indonesia tentang topik agama di Indonesia, apakah bisa dikatakan lebih “contra” daripada “pro” atau sebaliknya. Oleh karena itu, analisis dan klasifikasi (negatif, netral atau positif) akan dilakukan pada setiap tweet pengguna tentang

"agama" di Indonesia. Selain itu, hasil dari proses analisis sentimen tersebut nantinya akan menghasilkan sebuah kata yang paling sering digunakan atau diperbincangkan tentang topik keagamaan di kalangan masyarakat Indonesia.

Setelah menyelesaikan pemrosesan data pada proses cluster sebelumnya, maka didapatkannya data frame.

id	username	created_at	tweet
0	teddyremen	2022-07-06 15:23:09 SE Asia Standard Time	@onivlectra @hidwanTjan Waduh no comment kalo tentang itu, kubatalan saya Katolik dan selalu menghormati pemuka agama manapun apalagi temen2 Protestant
1	gediswidi_h3	2022-07-06 15:22:59 SE Asia Standard Time	4.ujanya saat dihubungi https://tco/fochVtdqj, Selasa (5/7/2022). Beberapa hari lalu, Kemenag Garut, bersama pemerintah daerah, MAJ, Dewan DSI, dan para pemuka agama, melaksanakan diskusi kembali yang terapan paham radikal itu ke NKRI.
2	fluffygen	2022-07-06 15:22:58 SE Asia Standard Time	@Vivitamin, Perlu ke pengadilan agama gk yk aku?
3	lucy_cali	2022-07-06 15:22:46 SE Asia Standard Time	Tidak ingin melewatkan kesempatan istimewa? jangan khawatir, dimulai dari rumah saja berangkat jadi mudah loh, kami selalu ada untuk umat, InyaaAllah dilaksanakan sesuai dengan syariat agama Islam.
4	inwablr	2022-07-06 15:22:46 SE Asia Standard Time	@RufiSoekoto @AdiPutriNahar @laila_jarim @Him28 Uda bnyk dibatai kok tau kalo jadi kan berguna bwt rakyat, lah ni jela2 menjual agama, penderantn orang, dut donasi ditilap apa bedanya em korupsi? Its salah msh aja dbeba blok

Gambar 20. Data Frame Tweets

Sebelum menghitung polaritas untuk setiap baris, harus dipastikan tidak ada nilai yang hilang pada dataset. Setelah didapatkan nilai polaritas maka nilai tersebut akan didefinisikan sedemikian rupa sehingga jika nilai polaritas = 0 maka akan masuk kelas sentimen netral, < 0 maka akan masuk kelas sentimen negatif dan jika skor > 0 maka akan dikategorikan positif kelas sentimen. Cara menghitung sistem skor sentimen adalah jumlah kata positif dikurangi total kata negatif dalam sebuah kalimat dibagi dengan jumlah kata dalam kalimat.

$$StSc = \frac{\text{number of positive words} - \text{number of negative words}}{\text{total number of words}}$$

Gambar 21. Sentiment Formula

Setelah proses pelabelan dilakukan, ditentukan polaritas dari sentimen data tweet. Berikut hasil dari penentuan polaritas sentimen tweet, tweet negatif sebanyak 777 tweet, tweet positif sebanyak 191 tweet dan tweet netral sebanyak 50 tweet.

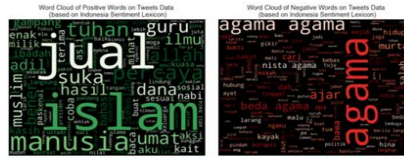
text	class	polarity score	polarity
1. ada beberapa hari ini... (text truncated)	0	0.0	netral
2. (text truncated)	0	0.0	netral
3. (text truncated)	0	0.0	netral
4. (text truncated)	0	0.0	netral
5. (text truncated)	0	0.0	netral
6. (text truncated)	0	0.0	netral
7. (text truncated)	0	0.0	netral
8. (text truncated)	0	0.0	netral
9. (text truncated)	0	0.0	netral
10. (text truncated)	0	0.0	netral

Gambar 22. Positive Tweet Result

text	class	polarity score	polarity
1. (text truncated)	-1	-1.0	negatif
2. (text truncated)	-1	-1.0	negatif
3. (text truncated)	-1	-1.0	negatif
4. (text truncated)	-1	-1.0	negatif
5. (text truncated)	-1	-1.0	negatif
6. (text truncated)	-1	-1.0	negatif
7. (text truncated)	-1	-1.0	negatif
8. (text truncated)	-1	-1.0	negatif
9. (text truncated)	-1	-1.0	negatif
10. (text truncated)	-1	-1.0	negatif

Gambar 23. Negative Tweet Result





Gambar 7. Word Cloud Positive and Negative Visualisation

Tujuan dari pembuatan analisis sentimen adalah agar admin dapat dengan mudah menguji konten berita yang nantinya akan diunggah bahwa konten yang akan ditampilkan mengarah pada nilai positif, negatif atau netral. Berikut adalah hasil implementasi tampilan halaman analisis sentimen.



Gambar 8. Interface Sentiment Analysis Page

### Kesimpulan

Berdasarkan penelitian dan sistem portal berita, dapat ditarik kesimpulan, adalah:

Setelah dibuatnya sistem informasi portal berita ini diharapkan masyarakat dapat lebih mudah mengakses berita-berita yang berkaitan dengan agama dan kementerian agama dan diharapkan portal berita ini dapat memberikan informasi dalam menunjang penambahan ilmu pengetahuan untuk kepentingan bersama. Dalam proses uji coba klusterisasi judul berita, beberapa strategi yang dapat dilakukan untuk meningkatkan pengunjung dan masyarakat untuk membaca informasi pada portal berita adapun strateginya yaitu: Tim redaksi dapat melakukan proses klusterisasi dengan data dan topik pada platform media sosial dan situs web berita lainnya untuk mengetahui kata atau kalimat apa yang sedang tren. Tujuannya adalah untuk mengetahui sebuah teks atau kalimat yang paling banyak dibaca atau disukai oleh masyarakat dari waktu ke waktu. Setelah mendapatkan teks yang disukai masyarakat, nantinya tim promosi dan redaksi akan melakukan penyusunan berita dengan menggunakan kalimat yang paling disukai agar pengunjung tertarik untuk membaca berita di portal berita ini. Melihat jumlah kluster judul berita yang telah dilakukan, tim redaksi dapat menentukan judul mana yang paling banyak dibaca dengan topik agama atau kata haji. Selain itu, tim redaksi dapat memanfaatkan fitur news sharing pada platform media sosial yang dinilai memiliki potensi besar untuk membangkitkan minat pembaca untuk membaca di portal berita ini.

### Daftar Pustaka

- Habib, A., Januantoro, A. and Srimarsono, O. C. (2021) 'Development of Management Information System Reservations Using Simple Multi-Attribute Rating Technique (SMART) Based on Android Mobile Applications', *Sisforma*, 8(1), pp. 37–43. doi: 10.24167/sisforma.v8i1.3024.

- Habib, A. and Kindhi, B. Al (2018) 'Rancang Bangun Sistem Informasi Manajemen Keuangan Sekolah', 2(2), pp. 136–146.
- Habib, A., Nova, P. A. and Januantoro, A. (2022) 'Development of Information System Management Customer Satisfaction Using the Simple Additive Weighting Method and Website-Based Google Maps API at ISP XYZ', *Sisforma*, 8(2), pp. 51–59. doi: 10.24167/sisforma.v8i2.3597.
- Habib, A., Satya, R. and Hariadi, B. (2022) 'Development of Management Information System Rental Service Photography Web-Based Using PHP Native Case Study at PT Dwipa Photowork Surabaya', 6(36), pp. 408–416.
- Mailo, F. F. and Lazuardi, L. (2019) 'Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia', *Journal of Information Systems for Public Health*, 4(1).
- Sudrajat, B. (2020) 'Rancang bangun Sistem Informasi Manajemen Asset berbasis WEB', *Jurnal Inovasi Informatika*, 5(2), pp. 100–109. doi: 10.51170/jii.v5i2.92.