

IMPLEMENTASI ALGORITMA NAIVE BAYES PADA DATA ULASAN MAHASISWA TENTANG SARANA DAN PRASARANA KAMPUS

Adrianus Neonub¹, Yasinta Oktaviana Legu Rema², Debora Chrisinta^{3*}
Teknologi Informasi, Universitas Timor, NTT

*Penulis korespondensi: deborachrisinta@unimor.ac.id

ABSTRAK

Universitas merupakan tempat bagi para siswa untuk dapat melanjutkan pendidikan tinggi. Oleh karena itu, salah satu aspek yang mendukung kualitas pendidikan salah satunya adalah sarana dan prasarana yang ada. Tujuan penelitian ini adalah untuk melakukan analisis sentimen terhadap persepsi mahasiswa terkait sarana dan prasarana di Universitas Timor menggunakan algoritma naïve bayes. Data yang digunakan diperoleh dari 250 jawaban kuisioner mahasiswa yang di sebarakan secara acak. Hasil yang diberikan menunjukkan bahwa pada sentimen kelas positif sarana dan prasarana umumnya dianggap berfungsi dengan baik, lengkap, dan membantu proses belajar mengajar. Namun, pada sentimen kelas negatif ada beberapa masalah yang perlu diperhatikan yaitu tempat parkir, lcd dan dataran yang berlumpur pada saat hujan. Kinerja algoritma naïve bayes dalam melakukan prediksi kelas sentiment menunjukkan akurasi sebesar 88.89%.

Kata kunci: sarana dan prasarana, ulasan, naïve bayes, akurasi

1 PENDAHULUAN

Dalam era digital saat ini, ulasan atau *feedback* dari pengguna mengenai suatu layanan atau produk telah menjadi salah satu sumber informasi yang sangat berharga bagi pengambil keputusan (Herman dkk, 2018). Di lingkungan akademik, ulasan mahasiswa mengenai sarana dan prasarana kampus memiliki peran penting dalam upaya peningkatan kualitas layanan pendidikan (Purwandani dkk, 2016). Informasi ini dapat digunakan oleh pihak universitas untuk memahami kebutuhan dan ekspektasi mahasiswa, serta untuk merencanakan perbaikan yang diperlukan.

Namun, tantangan utama dalam menganalisis ulasan mahasiswa adalah volume data yang besar dan beragam, yang sering kali memerlukan teknik analisis teks yang efektif dan efisien. Salah satu metode yang dapat digunakan untuk menangani masalah ini adalah algoritma Naive Bayes, yang merupakan salah satu teknik klasifikasi yang sederhana namun kuat dalam pemrosesan teks. Algoritma Naive Bayes didasarkan pada Teorema Bayes dan asumsi independensi antar fitur. Meskipun asumsi independensi ini jarang benar dalam konteks dunia nyata, algoritma ini telah terbukti berhasil dalam berbagai aplikasi klasifikasi teks, termasuk analisis sentimen, spam filtering, dan deteksi topik.

Penelitian terdahulu telah menunjukkan bahwa algoritma Naive Bayes efektif dalam klasifikasi data teks. Misalnya, studi oleh Simarmata dan Chrisinta (2023) menunjukkan bahwa Naive Bayes dapat digunakan untuk analisis sentimen dengan tingkat akurasi sebesar 65%. Penelitian lain oleh Iqbal dkk (2023) memperoleh akurasi 66% dalam melakukan klasifikasi sentimen pada data ulasan aplikasi Threads dan Twitter. Penerapan lain klasifikasi aplikasi Digital Korlantas POLRI menggunakan Naive Bayes pada Google Play Store (Putra dkk, 2024), yang semuanya menunjukkan kemampuan algoritma ini dalam menangani berbagai jenis teks.

Penelitian ini bertujuan untuk mengimplementasikan algoritma Naive Bayes dalam menganalisis ulasan mahasiswa mengenai sarana dan prasarana kampus. Dengan menggunakan teknik ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai persepsi mahasiswa terhadap fasilitas yang disediakan oleh universitas, serta mengidentifikasi area-area yang memerlukan perbaikan. Dalam penelitian ini, data ulasan mahasiswa akan dikumpulkan, diproses, dan dianalisis menggunakan algoritma Naive Bayes. Hasil analisis ini akan memberikan wawasan yang berguna bagi pihak universitas dalam upaya peningkatan kualitas layanan dan fasilitas kampus. Implementasi algoritma Naive Bayes diharapkan dapat menjadi solusi yang efektif dalam memproses dan menganalisis data ulasan secara cepat dan akurat. Dengan demikian, penelitian ini tidak hanya berkontribusi dalam pengembangan metode analisis ulasan mahasiswa, tetapi juga memberikan manfaat praktis bagi peningkatan kualitas sarana dan prasarana di lingkungan akademik.

2 METODE

Penelitian ini dilakukan untuk mengimplementasikan dan mengevaluasi algoritma Naive Bayes dalam analisis ulasan mahasiswa tentang sarana dan prasarana kampus. Metode penelitian yang digunakan terdiri dari beberapa tahapan, yaitu pengumpulan data, pra-pemrosesan data, implementasi algoritma Naive Bayes, dan evaluasi kinerja model (Chrisinta dan Simarmata, 2024).

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah ulasan mahasiswa mengenai sarana dan prasarana kampus. Ulasan tersebut dikumpulkan dari berbagai sumber seperti kuesioner online (*google form*). Data ulasan yang diperoleh kemudian disimpan dalam format teks untuk proses selanjutnya.

2.2 Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan data yang digunakan dalam analisis berada dalam format yang sesuai dan bersih dari noise. Tahapan pra-pemrosesan meliputi:

- Pembersihan teks yaitu menghapus karakter khusus, angka, dan tanda baca yang tidak relevan.
- Tokenisasi yaitu memisahkan teks menjadi unit-unit kata atau token.
- Normalisasi yaitu mengubah semua teks menjadi huruf kecil untuk konsistensi.
- Stopword removal yaitu menghapus kata-kata umum (stopwords) yang tidak memiliki makna penting.
- Stemming dan lemmatization yaitu mengubah kata-kata ke bentuk dasar untuk mengurangi keragaman kata.

2.3 Implementasi Algoritma Naive Bayes

Setelah data siap, langkah berikutnya adalah mengimplementasikan algoritma Naive Bayes untuk klasifikasi ulasan mahasiswa. Tahapan ini meliputi:

- Pembagian data yaitu data dibagi menjadi dua set, yaitu data latih (training set) dan data uji (test set) dengan proporsi tertentu, misalnya 80% untuk pelatihan dan 20% untuk pengujian.
- Pelatihan model yaitu model naive bayes dilatih menggunakan data latih. pada tahap ini, algoritma mempelajari distribusi probabilitas dari fitur-fitur yang ada dalam data.
- Pengujian model yaitu model yang telah dilatih kemudian diuji menggunakan data uji untuk mengevaluasi kinerjanya.

2.4 Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan beberapa metrik evaluasi untuk mengukur seberapa baik model dapat mengklasifikasikan ulasan mahasiswa. Metrik evaluasi yang digunakan meliputi:

- Akurasi yaitu proporsi ulasan yang diklasifikasikan dengan benar oleh model.
- Presisi yaitu proporsi ulasan yang diklasifikasikan sebagai positif yang benar-benar positif.
- Recall yaitu proporsi ulasan positif yang berhasil diidentifikasi dengan benar oleh model.
- F1-score yaitu harmonik rata-rata dari presisi dan recall, memberikan keseimbangan antara keduanya.

2.5 Analisis Hasil

Hasil dari implementasi dan evaluasi algoritma Naive Bayes dianalisis untuk memberikan wawasan mengenai persepsi mahasiswa terhadap sarana dan prasarana kampus. Hasil analisis ini digunakan untuk mengidentifikasi area-area yang perlu perbaikan dan untuk memberikan rekomendasi kepada pihak universitas.

3 HASIL DAN PEMBAHASAN

Implementasi algoritma Naive Bayes di RStudio melibatkan beberapa langkah, termasuk pemrosesan data, pelatihan model, pengujian model, dan evaluasi kinerja model. Berikut merupakan langkah penerapan Naive Bayes pada data yang digunakan pada penelitian ini yaitu sentiment mahasiswa terhadap Dosen, Sarana dan Prasarana, Kepegawaian dan Sistem Akademik:

1. Instalasi dan mengaktifkan paket yang digunakan
Skrip dimulai dengan memuat beberapa paket yang diperlukan, termasuk `readxl` untuk membaca file Excel, `ggplot2` untuk membuat visualisasi data, `tm` untuk memproses teks, dan `SnowballC` untuk stemming.

```
library(readxl)
library(ggplot2)
library(tm)
library(SnowballC)
```

2. Membaca Data dari Excel
Data dibaca dari file Excel "Text.xlsx" pada lembar "Filter2" menggunakan fungsi `read_excel()` dari paket `readxl`.

```
data <- read_excel("Naive Bayes/Text.xlsx", sheet = "Filter2")
```

3. Membuat Variabel Faktor
Variabel "Semester" diubah menjadi faktor menggunakan fungsi `factor()`.

```
data$Semester <- factor(data$Semester)
```

4. Menampilkan Data Awal
Data pertama dari setiap variabel ditampilkan menggunakan fungsi `head()`.

```
head(data)
```

5. Ringkasan Deskriptif
Deskripsi ringkas dari variabel "Prodi" dan "Semester" ditampilkan menggunakan fungsi `summary()`

```
summary(data$Prodi)
summary(data$Semester)
```

6. Membuat Diagram Batang
Diagram batang menampilkan jumlah mahasiswa per semester dan prodi. Ini dilakukan dengan menggunakan `ggplot()` dari paket `ggplot2`.

```
ggplot(data, aes(x = Semester, fill = Prodi)) +
  geom_bar(position = "dodge", color = "black") +
  geom_text(stat='count', aes(label=..count..), vjust=-0.5,
           position=position_dodge(width=0.9)) +
  labs(title = "Jumlah Mahasiswa per Semester dan Prodi",
       x = "Semester",
       y = "Jumlah Mahasiswa",
```

```
fill = "Prodi")
```

7. Analisis Sentimen

Analisis sentimen dilakukan untuk setiap aspek (dosen, sarana dan prasarana, kepegawaian, dan sistem akademik) dengan menggunakan daftar kata positif dan negatif yang dibaca dari file teks, serta daftar stopwords. Fungsi `detect_sentiment()` mengembalikan sentimen (positif atau negatif) dari setiap teks.

```
detect_sentiment <- function(text) {  
  # Preprocessing Teks  
  text_corpus <- Corpus(VectorSource(text))  
  text_corpus <- tm_map(text_corpus, content_transformer(tolower))  
  text_corpus <- tm_map(text_corpus, removePunctuation)  
  text_corpus <- tm_map(text_corpus, removeNumbers)  
  text_corpus <- tm_map(text_corpus, removeWords, stopwords)  
  text_corpus <- tm_map(text_corpus, stemDocument)  
  # Analisis Sentimen  
  sentiment <- sapply(text_corpus, function(x) {  
    positive_count <- sum(unlist(strsplit(x, " ")) %in% positive_words)  
    negative_count <- sum(unlist(strsplit(x, " ")) %in% negative_words)  
    if (positive_count > negative_count) {  
      return("Positif")  
    } else {  
      return("Negatif")  
    }  
  })  
  return(sentiment)}
```

8. Naive Bayes

Pada setiap aspek, dilakukan analisis sentimen menggunakan metode Naive Bayes. Data teks dipreproses terlebih dahulu, kemudian dilakukan pemodelan menggunakan fungsi `naiveBayes()` dari paket `e1071`.

```
#Naive Bayes Sarana dan Prasarana2  
# Preprocessing Teks dengan menggunakan stopwords dari file teks  
corpus2 <- Corpus(VectorSource(data$`Sarana dan Prasarana`))  
corpus2 <- tm_map(corpus2, content_transformer(tolower))  
corpus2 <- tm_map(corpus2, removePunctuation)  
corpus2 <- tm_map(corpus2, removeNumbers)  
corpus2 <- tm_map(corpus2, removeWords, stopwords)  
corpus2 <- tm_map(corpus2, stemDocument)  
# Create a Document-Term Matrix (DTM)  
dtm2 <- DocumentTermMatrix(corpus2)  
# Mengonversi DTM ke data frame  
dtm_df2 <- as.data.frame(as.matrix(dtm2))  
dtm_df2$sentimen_sarana_prasarana <- data$Sentiment_Sarana_Prasarana  
  
# Membagi data menjadi set pelatihan dan pengujian  
set.seed(123)  
indeks_pelatihan2 <- createDataPartition(dtm_df2$sentimen_sarana_prasarana, p = 0.8, list  
= FALSE)  
data_pelatihan2 <- dtm_df2[indeks_pelatihan2, ]  
data_pengujian2 <- dtm_df2[-indeks_pelatihan2, ]  
# Model Naive Bayes  
model_nb2 <- naiveBayes(sentimen_sarana_prasarana ~ ., data = data_pelatihan2)  
#prediksi  
prediksi_nb2 <- predict(model_nb2, data_pengujian2)
```

9. Evaluasi Naive Bayes

Performa model Naive Bayes dievaluasi menggunakan confusion matrix untuk setiap aspek.

```
# Evaluasi Naive Bayes Sarana dan Prasarana
conf_matrix_nb2 <- confusionMatrix(prediksi_nb2, as.factor(data_pengujian2$
    sentimen_sarana_prasarana))
conf_matrix_nb2
```

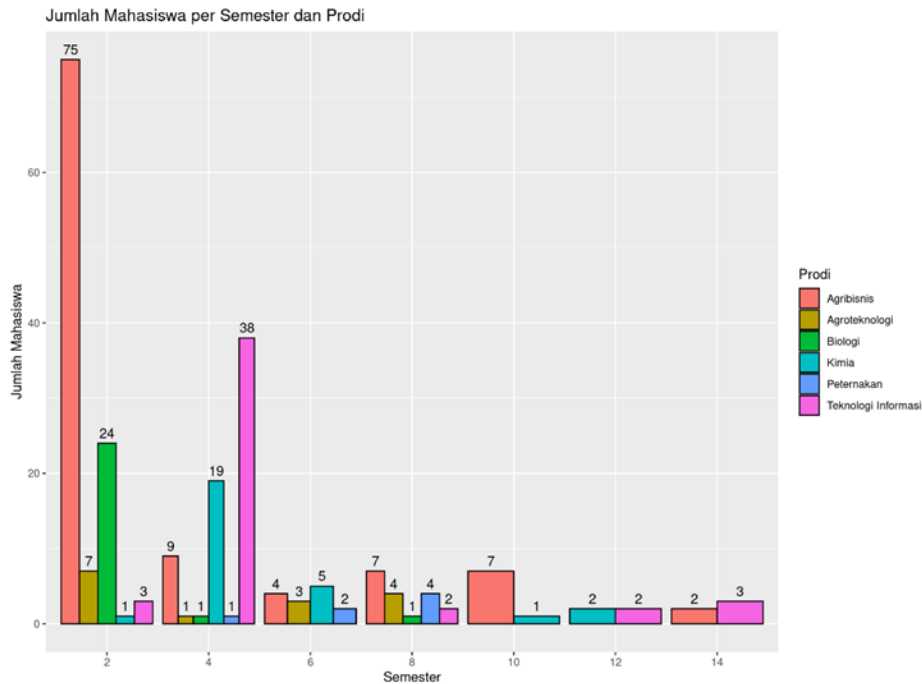
10. Visualisasi

Tahap ini dilakukan visualisasi kata-kata yang paling sering muncul dalam teks dengan membedakan berdasarkan sentimen (positif dan negatif) menggunakan diagram batang. Data kata-kata dan frekuensinya diurutkan, dan hanya 10 kata teratas yang ditampilkan

```
# Plot diagram batang dengan membedakan warna berdasarkan sentimen
ggplot(top_wordsp1, aes(x = word, y = freq)) +
  geom_bar(stat = "identity", fill = "green") +
  geom_text(aes(label = freq), vjust = -0.5, color = "black", size = 3) + # Menambahkan teks
  untuk jumlah kata
labs(x = "Kata", y = "Frekuensi", title = "Dosen (Sentimen Positif)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme_minimal()

ggplot(top_wordsn1, aes(x = word, y = freq)) +
  geom_bar(stat = "identity", fill = "red") +
  geom_text(aes(label = freq), vjust = -0.5, color = "black", size = 3) + # Menambahkan teks
  untuk jumlah kata
labs(x = "Kata", y = "Frekuensi", title = "Dosen (Sentimen Negatif)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme_minimal()
```

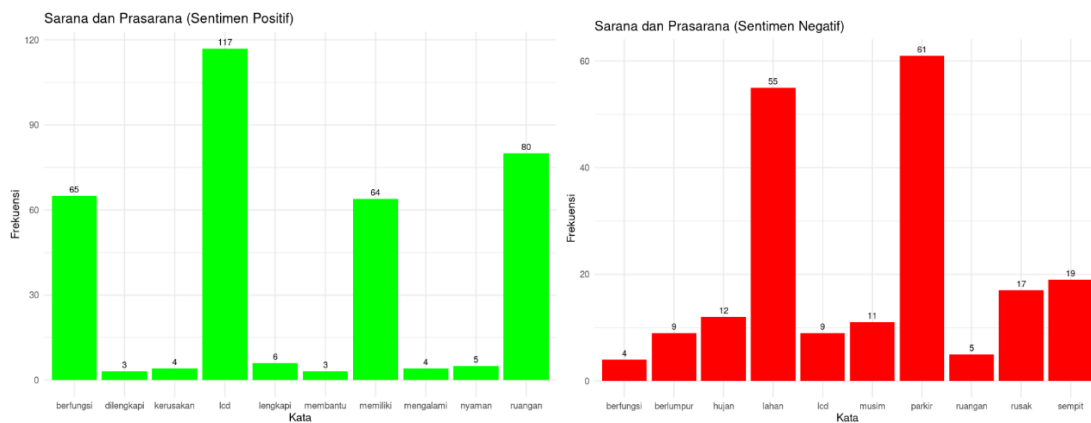
Berikut ini merupakan sebaran jawaban kuisioner yang diperoleh berdasarkan semester dan prodi:



Gambar 1. Distribusi frekuensi data berdasarkan prodi dan semester

Gambar 1 tersebut menunjukkan bahwa distribusi data mahasiswa tidak merata di semua prodi dan semester. Ada sebanyak 6 Prodi yang terdata yaitu Agribisnis, Agroteknologi, Biologi, Kimia, Peternakan dan Teknologi Informasi. Agribisnis dan Teknologi Informasi merupakan prodi yang memiliki jumlah mahasiswa paling banyak dalam mengisi kuisioner, sedangkan Prodi Peternakan memiliki jumlah mahasiswa paling sedikit. Mahasiswa yang mengisi kuisioner berasal dari semester 2, 4, 6, 8, 10, 12, dan 14, dengan mayoritas berasal dari semester 2.

Visualisasi top word berdasarkan sentimen bertujuan untuk mengidentifikasi kata-kata yang paling sering muncul dalam data text dengan sentimen positif dan negatif. Frekuensi kemunculan kata-kata dihitung dan top word divisualisasikan dengan word cloud. Hasilnya menunjukkan bahwa kata-kata yang sering muncul dalam data text dengan sentimen positif dan negatif berbeda. Kata-kata tertentu lebih sering muncul dalam data text dengan sentimen positif, sedangkan kata-kata lainnya lebih sering muncul dalam data text dengan sentimen negatif.



Gambar 2. Distribusi 10 Top Word Berdasarkan Variabel Sarana dan Prasarana

Visualisasi 10 top word dari variabel sarana dan prasarana dengan sentimen positif dan negatif memberikan informasi yang lengkap tentang persepsi terhadap sarana dan prasarana. Sarana dan prasarana umumnya dianggap berfungsi dengan baik, lengkap, dan membantu proses belajar mengajar. Namun, ada beberapa masalah yang perlu diperhatikan, seperti kerusakan, genangan air, lahan yang sempit, dan tempat parkir yang terbatas.

Pengujian algoritma Naïve Bayes pada data sentimen mahasiswa menggunakan confusion matrix merupakan cara untuk mengevaluasi performa model Naïve Bayes dalam mengklasifikasikan sentimen (positif dan negatif) pada data text. Berikut hasil kinerna Naïve Bayes dalam memodelkan data text dengan variable Sarana dan Prasarana:

Tabel 1. Tabel kontingensi kelas sentimen

Prediksi	Aktual	
	Negatif	Positif
Negatif	13 (TN)	3 (FN)
Positif	2 (FP)	27 (TP)

Berdasarkan tabel confusion matrix, berikut adalah perhitungan akurasi, presisi, dan recall untuk variabel Sarana dan Prasarana:

1. Akurasi dihitung dengan membagi jumlah data yang diklasifikasikan dengan benar dengan total data.

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

$$\text{Akurasi} = (27 + 13) / (27 + 2 + 3 + 13)$$

$$\text{Akurasi} = 0.8889 \text{ atau } 88.89\%$$

2. Presisi dihitung dengan membagi jumlah data yang diprediksi positif dan benar-benar positif dengan total data yang diprediksi positif.

$$\text{Presisi} = TP / (TP + FP)$$

$$\text{Presisi} = 27 / (27 + 2)$$

$$\text{Presisi} = 0.9310 \text{ atau } 93.10\%$$

3. Recall dihitung dengan membagi jumlah data yang diprediksi positif dan benar-benar positif dengan total data yang sebenarnya positif.

$$\text{Recall} = TP / (TP + FN)$$

$$\text{Recall} = 27 / (27 + 3)$$

$$\text{Recall} = 0.9 \text{ atau } 90\%$$

4 KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan algoritma Naive Bayes untuk melakukan analisis sentimen terhadap ulasan mahasiswa mengenai sarana dan prasarana di Universitas Timor. Hasil analisis menunjukkan bahwa sebagian besar mahasiswa memberikan ulasan positif, menyoroti bahwa sarana dan prasarana kampus umumnya berfungsi dengan baik, lengkap, dan membantu proses belajar mengajar. Namun, terdapat beberapa ulasan negatif yang mengidentifikasi masalah seperti tempat parkir, kerusakan LCD, dan area berlumpur saat hujan. Kinerja algoritma Naive Bayes dalam melakukan prediksi kelas sentimen menunjukkan hasil yang memuaskan dengan akurasi sebesar 88.89%, presisi sebesar 93.10%, dan recall sebesar 90%. Angka-angka ini mengindikasikan bahwa model Naive Bayes mampu mengklasifikasi sentimen ulasan mahasiswa dengan tingkat keakuratan yang tinggi. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma Naive Bayes merupakan alat yang efektif dan efisien untuk menganalisis sentimen ulasan mahasiswa. Dengan demikian, pihak universitas dapat memanfaatkan hasil analisis ini untuk memahami persepsi mahasiswa dan mengidentifikasi area yang memerlukan perbaikan, sehingga dapat meningkatkan kualitas sarana dan prasarana kampus secara keseluruhan.

DAFTAR PUSTAKA

- Chrisinta, D., & Simarmata, J. E. (2024). Eksplorasi Teknik Web Scraping pada Data Mining: Pendekatan Pencarian Data Berbasis Python. *Faktor Exacta*, 17(1).
- Chrisinta, Debora, and Justin Eduardo Simarmata. "Analisis Sentimen Penilaian Masyarakat Terhadap Pejabat Publik Menggunakan Algoritma Naïve Bayes Classifier." *Komputika: Jurnal Sistem Komputer* 12.1 (2023): 93-101.
- Herman, Lalu Edy, and Handry Sudiarta Athar. "Pengembangan Model Social Media Marketing dan Keputusan Berkunjung: Sebuah Pendekatan Konseptual." *J-IKA: Jurnal Ilmu Komunikasi Fakultas Ilmu Komunikasi Universitas BSI Bandung* 5.2 (2018).
- Iqbal, Muhammad, et al. "Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter." *KLIK: Kajian Ilmiah Informatika dan Komputer* 4.3 (2023): 1799-1807.
- Purwandani, Dela, and Cicih Sutarsih. "Pengaruh Mutu Layanan Sarana dan Prasarana Terhadap Kepuasan Mahasiswa di Fakultas Pendidikan Teknologi dan Kejuruan Universitas Pendidikan Indonesia." *Jurnal Tata Kelola Pendidikan* 1.1 (2016): 80-90.
- Putra, Syarif Hidayatulloh Wazir, and Dimas Febriawan. "Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store." *KLIK: Kajian Ilmiah Informatika dan Komputer* 4.4 (2024): 1962-1971.