**Sentiment Analysis Pemilihan Capres 2024 Menggunakan Algoritma Support Vector Machine (SVM)**

**SKRIPSI**

Sebagai Salah Satu Syarat untuk Meraih

Gelar Sarjana Komputer

Oleh :

Michael Alfonso

32190039

****

**Fakultas Teknologi dan Desain**

**Program Studi Teknik Informatika**

**Universitas Bunda Mulia**

**Tangerang**

**2022**

# **ABSTRAK**

# **PRAKATA**

# **DAFTAR ISI**

# **DAFTAR TABEL**

# **DAFTAR GAMBAR**

**BAB 1**

# **PENDAHULUAN**

## **Latar Belakang**

Pemilu terjadi setiap 5 tahun sekali dengan jumlah pemilih yang semakin bertambah setiap periodenya, tercatat penambahan pemilih tetap sejak tahun 2004 sampai 2019 sebanyak 28,90% (kutip dari BPS dan data KPU). Media sosial seperti Facebook, Twitter, LinkedIn, YouTube sudah menjadi bagian dari kehidupan manusia, dibuktikan dengan cara berinteraksi kebanyakan dapat dilakukan dengan sosia media, sampai menjadi identitas diri ke masyarakat umum (kutip IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET).

Pada sosial media terutama *Twitter*, banyak penyebaran informasi, hiburan, bahkan opini politik, hal ini tidak luput dari terjadinya pemilu yang merupakan kegiatan politik yang akan dilaksanakan pada tahun 2024, disebut sebagai pesta demokrasi, hal tersebut pastinya ramai diperbincangkan di sosial media. Salah satu cara menyampaikan opini dan lewat media sosial adalah dengan menggunakan tanda pagar tertentu pada *tweet* yang dipost di *Twitter*. *Twitter* sendiri digunakan karena kemudahan dalam membalas *tweet* dengan *like*, komentar bahkan *reply*, sehingga topik yang sedang ramai diperbincangkan dapat dengan mudah tersebar (Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization).

Oleh karena itu, maka diperlukannya analisa lebih lanjut untuk melihat seberapa besar kesetujuan dan ketidaksetujuan pengguna *Twitter* terhadap pemilihan calon presiden 2024, lalu data akan diperoleh dari *tweet* pengguna *Twitter* yang berkaitan dengan topik ini yaitu pemilihan umum 2024. *Sentiment Analysis* atau dapat disebut *Opinion Mining* merupakan studi yang mempelajari opini, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas (IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET) (Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization).

*Sentiment Analysis* ini digunakan untuk mengetahui *tweet* yang diposting mengandung kata positif, negatif atau netral (Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis).

Pada penelitian ini, akan menggunakan metode Support Vector Machine (SVM) yaitu salah satu metode klasifikasi pada *machine learning* (*supervised learning*) yang memprediksi kelas berdasarkan pola dari hasil training. Klasifikasi dilakukan dengan membuat garis pemisah (*hyperlane*) antara kelas positif dan negatif (IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET). Sehingga hasil dari penelitian ini adalah sentiment pengguna *Twitter* terhadap pemilihan calon presiden pada tahun 2024.

## **Rumusan Masalah**

Berdasarkan latar belakang yang telah disebutkan, maka penulis mendapati rumusan masalah:

1. Bagaimana penerapan *Sentiment Analysis* menggunakan algoritma *Support Vector Machine* (SVM) diimplementasikan pada topik pemilu 2024?
2. Bagaimana akurasi *Sentiment Analysis* untuk tweet pemilu 2024 dengan menggunakan algoritma *K-10 Fold Cross Validation*?
3. Apakah ada keterkaitan antara hasil analisa dengan survey website berita?

## **Tujuan dan Manfaat**

### **Tujuan Penelitian**

Berdasarkan masalah yang dinyatakan diatas, maka tujuan dari penelitian ini adalah:

1. Mengembangkan aplikasi web sederhana sebagai bentuk implementasi *Sentiment Analysis* menggunakan algoritma *Support Vector Machine* (SVM) dengan topik pemilu 2024
2. Menguji akurasi *Sentiment Analysis* dengan algoritma *Support Vector Machine* (SVM) untuk *tweet* pemilu 2024

Melakukan perbandingan hasil prediksi dengan website berita terpercaya seperti detik.com

### **Manfaat Penelitian**

1. Manfaat bagi penulis:
   1. Hasil penelitian dapat memberikan manfaat bagi penulis tentang penggunaan *Text Mining* dalam memprediksi hasil pemilu 2024 berdasarkan data *Twitter*
   2. Mengimplementasi ilmu dan teori yang penulis dapatkan selama perkuliahan dalam bidang *Machine Learning*
2. Manfaat bagi akademis:
   1. Penulis berharap dengan riset yang dilakukan dapat menjadi referensi bacaan dengan topik *Sentiment Analysis* dan *Text Mining*
3. Manfaat bagi masyarakat:
   1. Sebagai prediksi pemilihan calon presiden 2024 berdasarkan data Twitter
   2. Memberikan referensi dan bacaan bagi masyarakat yang berkaitan dengan tema politik

## **Ruang Lingkup**

Pada penelitian ini, ruang lingkup yang dibatasi oleh penulis adalah sebagai berikut:

1. Perancangan aplikasi ini dijalankan berbasis website
2. Penerapan *Sentiment Analysis* menggunakan algoritma *Support Vector Machine* (SVM)
3. Tema pemilu yang diambil hanya seputar pemilihan calon presiden, bukan calon lainnya.
4. Pengujian algoritma akan dilakukan dengan metode *10-fold* *cross validation*
5. Bahasa pemrogaman yang digunakan dalam perancangan aplikasi ini adalah bahasa pemrograman python.
6. Data diambil dari *Twitter* dengan batas 1 Januari 2023 sampai dengan 31 Maret 2023

## **Metodologi Penelitian**

Metode yang akan dipakai pada penelitian ini adalah algoritma *Support Vector Machine* (SVM) yang akan diterapkan pada data *tweet* dari media sosial *Twitter* dengan tema pemilihan capres 2024, lalu data tersebut akan dilakukan beberapa proses diantaranya:

1. *Data Labeling* melakukan pelabelan setiap data menjadi 3 jenis, positif, netral, dan negatif.
2. *Preprocessing* yaitu pembersihan data dari noise sehingga siap digunakan pada tahap *Transformation*
3. *Transformation* menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) sehingga data diubah dari kalimat atau kata menjadi sebuah nilai
4. *Klasifikasi* dilakukan pemrosesan data menggunakan algoritma *Support Vector Machine* (SVM).
5. *Evaluasi* akan dilakukan untuk mengetahui nilai akurasi, presisi, *recall*, dan F1-*score* menggunakan *10*-*Fold Cross Validation* dengan membagi dataset kedalam data *training* dan data *testing* secara bertahap.

## **Sistematika Penulisan**

Sistematika penulisan yang akan digunakan adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah yang diangkat, rumusan masalah yang didapatkan, batasan masalah yang dibuat, manfaat dan tujuan penelitian bagi berbagai pihak serta sistematika penulisan yang akan digunakan di penelitian ini.

BAB 2: LANDASAN TEORI

Bab ini menjelaskan tentang teori yang berkaitan tentang S*entiment Analysis* dan algoritma *Support Vector Machine* yang diambil dari berbagai sumber sebagai landasan dalam pembuatan penelitian.

BAB 3: ANALISIS DAN PERANCANGAN

Pada bagian ini berisi tahapan rancangan aplikasi *Sentiment Analysis* dengan algoritma *Support Vector Machine* berbasis website sederhana.

BAB 4: IMPLEMENTASI

Bab ini berisikan hasil pengembangan yang diimplementasi dalam website dan uji akurasi yang diukur dengan metode 10-*Fold Cross Validation*

BAB 5: KESIMPULAN DAN SARAN

Bab ini berisikan pemaparan kesimpulan dari hasil uji coba dan pendapat penulis tentang riset yang telah dilakukan.

**BAB 2**

# **LANDASAN TEORI**

## **Landasan Teori**

1. **Twitter**

Twitter merupakan salah satu sosial media yang digunakan untuk berinteraksi secara daring dengan pengguna lain dan atau bot. Pada twitter sendiri, terdapat istilah *tweet* yaitu teks status pengguna yang digunakan untuk memberikan informasi (Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Berbasis Metode Lexicon Dan Support Vector Machine).

1. **Text Mining**

Text mining menggunakan proses preprocessing dokumen, proses ini dapat menghasilkan sentimen kelas positif, negatif, dan netral (PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA). Text mining menurut Nurhuda dan Sihwi, proses menemukan informasi dari sekumpulan dokumen teks menggunakan metode analisis tertentu (Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Berbasis Metode Lexicon Dan Support Vector Machine). Umumnya, tahapan pada text mining adalah sebagai berikut:

1. *Tokenizing* merupakan tahapan pemotongan string input per kata penyusunnya ([Microsoft PowerPoint - Text Mining (pens.ac.id)](http://tessy.lecturer.pens.ac.id/kuliah/dm/6Text%20Mining.pdf)). Contohnya “Saya Ingin Makan” akan menjadi “Saya”, “Ingin”, “Makan”.
2. *Filtering* merupakan tahapan mengambil kata-kata penting dari hasil *Tokenizing*, salah satu algoritma yang digunakan adalah algoritma *stop list* atau *word list*.
3. *Stemming* merupakan tahapan mencari kata dasar dari tiap kata hasil *filtering*. Contohnya “Mengambil” menjadi “Ambil”
4. *Tagging* merupakan tahapan mencari bentuk awal dari tiap kata lampai atau kata hasil *stemming*.
5. *Analyzing* merupakan tahap penentuan keterhubungan kata-kata antar dokumen.
6. **Sentiment Analysis**

*Sentiment analysis* juga bisa disebut *Opinion Mining* merupakan sebuah riset yang meneliti opini, sentiment dan emosi secara tekstual lalu diklasifikasi menjadi kelompok sentiment negatif, netral dan positif (PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE).

1. **Pembobotan TF-IDF**

Bobot TF-IDF merupakan penilaian bobot setiap kata yang ada pada sebuah dokumen (PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE), TF-IDF merupakan salmetode pembobotan fitur yang kuat yang mengukur frekuensi dan distribusi kata untuk menentukan bobotnya. TF-IDF terdiri dari TF (*Term Frequency*) yaitu frekuensi kemunculan kata dalam dokumen dan IDF (*Inverse Document Frequency*) yaitu kebalikan dari frekuensi dokumen, nilai TF-IDF dapat diperoleh dengan menggunakan persamaan berikut:

1. Nilai *Term Frequency* (TF) diperoleh dengan nilai frekuensi kemunculan fitur *t* pada dokumen *d*
2. Nilai Inverse Document Frequency (IDF) diperoleh dari logaritma banyaknya dokumen *n* dibagi dokumen *df* yang mengandung fitur *t*
3. Nilai Term Frequency Inverse Document Frequency (TF-IDF) didapatkan dengan mengalikan nilai *TF* dan *IDF*

(PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DENGAN TF-IDF N-GRAM UNTUK TEXT CLASSIFICATION).

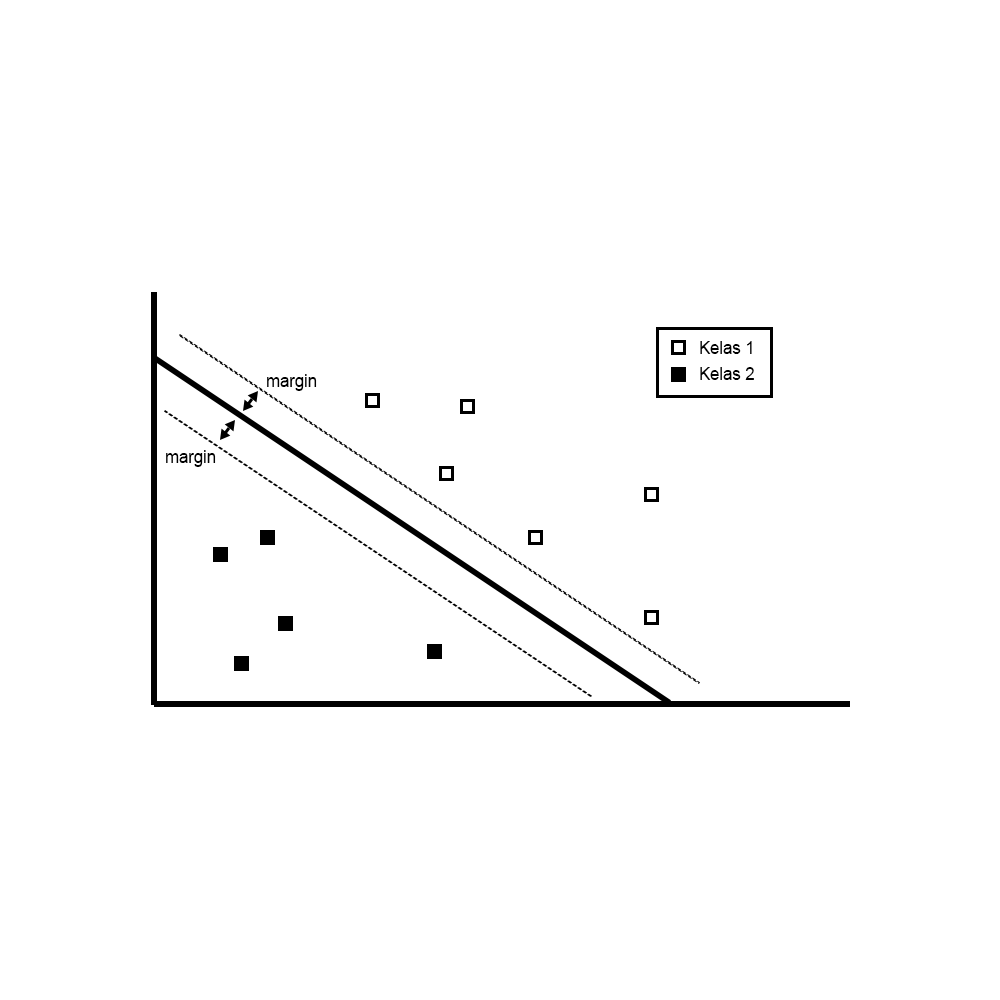
1. **Support Vector Machine (SVM)**

Support Vector Mahine diciptakan oleh Vladimir Vapnik yang merupakan salah satu metode klasifikasi dengan metode machine learning (Supervised Learning) yang akan melakukan prediksi kelas berdasarkan hasil *training* (PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE). Algoritma ini akan menghasilkan *hyperplane* yang memisahkan sebuah plane menjadi 2 bagian pada masing-masing kelas yang dapat dilihat pada gambar 1. Berikut adalah rumus perhitungan *hyperplane* pada algoritma SVM:

Dengan keterangan:

W = parameter *hyperlane* yang dicari (garis tegak lurus antara garis *hyperlane* dengan titik support vector)

x = titik data masukan SVM

b = parameter hyperlane yang dicari (nilai bias)

Beberapa kernel yang dapat diterapkan pada SVM:

1. *Linear* Kernel merupakan fungsi kernel yang memisahkan 2 kelas dengan persamaan sebagai berikut:
2. *Polynomial* Kernel merupakan fungsi kernel yang tidak terpisah secara *linear*, persamaan sebagai berikut:
3. *Radial Basis Function* (RBF) Kernel merupakan fungsi untuk menganalisis data yang tidak terpisah secara linear dan memerlukan 2 jenis parameter (Gamma dan Cost).

(Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Berbasis Metode Lexicon Dan Support Vector Machine)

1. **Python**

Python merupakan bahasa pemrograman digunakan oleh banyak platform oleh developer seperti website, mobile dan desktop (PEMBELAJARAN PEMROGRAMAN PYTHON DENGAN PENDEKATAN LOGIKA ALGORITMA). Namun sejak berkembangnya *Machine Learning* python pun turut digunakan terutama dengan ketersediaan banyaknya library seperti scikit-learn, keras, tensorflow, openCV, dan lain-lain.

1. **K-Fold Cross Validation**

K-Fold Cross Validation merupakan salah satu bentuk pengujian yang berfungsi untuk menilai kinerja algoritma dengan cara kerja membagi sampel data secara acak lalu membagi data tersebut sebanyak nilai K dari k-fold. Selanjutnya, salah satu kelompok K tersebut akan dijadikan sebagai data *testing* dan sisa kelompok akan dijadikan sebagai data *training* (Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia).

Untuk perhitungan akurasi akan dihitung dengan rumusan berikut:

## **Penelitian Terdahulu**

Berikut ini merupakan penelitian yang menjadi referensi penulis dalam melakukan penelitian ini:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Penulis | Tahun | Judul | Kesimpulan |
| 1 | Valonia Inge Santoso, Gloria Virginia, Yuan Lukito | 2017 | PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE | Sistem berhasil melakukan *sentiment analysis* dengan metode SVM terhadap evaluasi dosen FTI UKDW Prodi Teknik Informatika tahun ajaran 2014/2015 menggunakan validasi *K-Fold Cross Validation* dengan akurasi tertinggi pada k = 6 sebesar 67.83%. |
| 2 | Fadholi Fat Haranto, Bety Wulan Sari | 2019 | IMPLEMENTASI *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET | Penelitian sentiment pengguna twitter terhadap biznet dan telkom menggunakan algoritma SVM mendapat hasil rata-rata akurasi sebesar 79.6%, precision 76.5%, recall 72.8%, dan f1-score sebesar 74.6% untuk data Telkom. Sedangkan rata-rata akurasi 83.2%, precision 78.8%, recall 71.6%, dan f1-score 75% untuk Biznet. |
| 3 | Sri Diantika, Windu Gata, Hiya Nalatissifa, Mareanus Lase | 2021 | Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi  Kestabilan Jaringan Listrik | Akurasi algoritma *Support Vector Machine* pada data kestabilan jaringan mendapatkan akurasi 98.8% semestara jika menggunakan *Naïve Bayes* sebesar 97.64%. |
| 4 | Fatihah Rahmadayana, Yuliant Sibaroni | 2021 | *Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization* | Penelitian ini menunjukan bahwa penggunakan metode *acronym extension* terhadap topik *Work From Home* mendapatkan hasil F1-score tertinggi sebesar 83.362%. Sedangkan pada penerapan SVM dengan *Randomized Search Optimization* menunjukan hasil sebesar 62.35% *tweet* positif, dan 37.65% *tweet* negatif. |
| 5 | Dedi Darwis, Eka Shintya Pratiwi, A. Ferico Octaviansyah Pasaribu | 2020 | PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA  DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK  INDONESIA | Hasil klasifikasi menggunakan metode SVM dibagi menjadi 3 kelas, yaitu 8% positif, 15% netral, dan 77% negatif. Lalu hasil pengujian nilai akurasi, precession, recall, dan F1-score dapat disimpulkan bahwa sentimen masyarakat terhadap kinerja KPK sangat kurang baik dengan presentase negatif sebesar 77% serta pengujian hasil akurasi sebesar 82%, precission 90%, recall 88%, dan f1-score 89% |

## **Kerangka Berpikir**

Kerangka alur penelitian yang dilakukan dalam membuat sistem *sentiment analysis* terhadap topik pemilu 2024. Dimulai dari mendapatkan dataset *tweet* dari kata kunci yang relevan seperti “Ganjar Pranowo”, “Prabowo Subianto”, dan “Anies Baswedan”, lalu melakukan *labeling* *tweet* secara manual oleh penulis, dilanjutkan dengan melakukan *pre-processing* pada dataset seperti melakukan *Cleansing*, *Tokenizing*, *Filtering*, *Stemming*, *Tagging*, dan *Analyzing*. Selanjutnya baru setiap kata diberi bobot dengan metode *TF-IDF* dan dilanjutkan dengan membagi data kedalam data *training* dan data *testing* menurut metode *10-Fold Cross Validation* untuk dimasukan ke dalam algoritma *Support Vector Machine*. Lalu dari setiap fold akan dilihat untuk nilai K terbaik dari segi akurasi, *precission*, *recall*, dan *f1-score*. Lalu hasil prediksi setiap capres akan dibandingkan secara manual oleh survey yang diterbitkan oleh portal berita untuk melakukan perbandingan hasil prediksi dengan survey yang disebarkan.

**BAB 3**

# **ANALISIS DAN PERANCANGAN**

# **DAFTAR PUSTAKA**

[1] Badan Pusat Statistik, “Persentase Penduduk yang Memiliki/Menguasai Telepon Seluler Menurut Provinsi dan Klasifikasi Daerah 2019-2021,” *Badan Pusat Statistik*, 2022. https://www.bps.go.id/indicator/2/395/1/persentase-penduduk-yang-memiliki-menguasai-telepon-seluler-menurut-provinsi-dan-klasifikasi-daerah.html (accessed Dec. 11, 2022).

[2] T. Nurseto, “MEMBUAT MEDIA PEMBELAJARAN YANG MENARIK,” *Jurnal Ekonomi & Pendidikan*, vol. 8, no. 1, 2011.

[3] R. R. Ananda, “PENGEMBANGAN MEDIA PEMBELAJARAN SCRUM INTERAKTIF,” UIN ALAUDDIN MAKASAR, Makasar, 2018.

[4] D. G. H. Divayana, P. W. A. Suyasa, and N. Sugihartini, “Pengembangan Media Pembelajaran Berbasis Web Untuk Matakuliah Kurikulum dan Pengajaran di Jurusan Pendidikan Teknik Informatika Universitas Pendidikan Ganesha,” *Jurnal Nasional Pendidikan Teknik Informatika*, vol. 5, no. 3, pp. 1–9, 2016.

[5] B. Holiwarni, “PENGEMBANGAN MEDIA PEMBELAJARAN BERBANTUKAN KOMPUTER (COMPUTER ASSISTED INSTRUCTION/CIA) UNTUK PEMBELAJARAN KIMIA SMA,” *Jurnal Sorot*, vol. 9, no. 1, pp. 1–8, 2014.

[6] R. SUPRIATNA, “IMPLEMENTASI DAN USER ACCEPTANCE TEST (UAT) TERHADAP APLIKASI E-LEARNINGPADA MADRASAH ALIYAH NEGERI (MAN) 3 KOTA BANDA ACEH,” UNIVERSITAS ISLAM NEGERI AR-RANIRY, BANDA ACEH, 2018.

[7] B. P. Nurken, “PERANCANGAN GAME ‘ AJI SAKA LEGENDA TANAH JAWA’ MENGGUNAKAN RPG MAKER BERBASIS ANDROID,” Universitas Bunda Mulia, Tangerang, 2022.

[8] W. A. Prabowo and C. Wiguna, “Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, pp. 149–156, Jan. 2021, doi: 10.30865/mib.v5i1.2604.

[9] G. P. Ariputri, “PENGEMBANGAN APLIKASI ANDROID UNTUK MENDUKUNG PEMBELAJARAN LISTENING BAHASA INGGRIS KELAS XI SMA,” UNIVERSITAS NEGERI SEMARANG, Semarang, 2015.

[10] M. Haris Batubara, A. Hatuaon Sihite, and I. Saputra, “APLIKASI PEMBELAJARAN TEKNIK MESIN OTOMOTIF KENDARAAN RINGAN DENGAN METODE COMPUTER ASSISTED INSTRUCTION (STUDI KASUS : SMK SWASTA KARYA PENDIDIK),” *Majalah Ilmiah INTI*, vol. 12, no. 2, pp. 1–5, 2017.

[11] H. Suhendi and R. Gunawan, “APLIKASI PEMBELAJARAN MULTIMEDIA BERBASIS ANDROID DENGAN MENGGUNAKAN METODE COMPUTER ASSISTED INSTRUCTION (CAI),” *ISU TEKNOLOGI STT MANDALA* , vol. 14, no. 2, pp. 1–11, 2019.

[12] H. D. Hutahaean and P. M. Hasugian, “Aplikasi Pembelajaran Kriptograpiberbasis Mobile menggunakan Computer Assisted Instruction,” *Media Informasi Analisa dan Sistem*, vol. 4, no. 1, pp. 1–4, 2019.