

# Impelementasi Metode *Latent Semantic Analysis* pada Sistem Informasi Ujian *Online* Berbasis Web

Syahfrialief<sup>a,1,\*</sup>, Irawati<sup>a,2</sup> dan Muh. Aliyazid Mude<sup>a,3</sup>

<sup>a</sup> Universitas Muslim Indonesia, Jl. Urip Sumoharjo, Makassar 90231, Indonesia  
<sup>1</sup> Syahfrialief14@gmail.com; <sup>2</sup> irawan2804@mail.com; <sup>3</sup> aliyazid.mude@umi.ac.id;  
\*corresponding author

INFORMASI ARTIKEL	ABSTRAK
Diterima : 14 – 02 – 2023 Direvisi : 26 – 05 – 2023 Diterbitkan : 31 – 05 – 2023	Esai adalah pertanyaan yang menuntut peserta didik untuk menjawab dalam bentuk menguraikan, menjelaskan, mendiskusikan, membandingkan, memberikan alasan, dan bentuk lain yang sejenis sesuai dengan tuntutan pertanyaan dengan menggunakan kata-kata dan bahasa sendiri. Permasalahan kerap terjadi pada penilaian soal esai. Guru kadang mengalami ketidak konsistenan saat menilai soal-soal esai. Hal ini bisa terjadi ketika jumlah soal yang akan dinilai sangat banyak dan waktu untuk menyelesaikannya sangatlah sedikit. Tujuan penelitian ini untuk mempermudah guru dalam membrtikan nilai pada jawaban murid yang bervariasi. Penelitian ini menggunakan metode <i>Latent Semantic Analysis</i> (LSA) untuk membandingkan hubungan antar dokumen teks, menghitung bobot kata menggunakan algoritma <i>Term Frequency - Inverse Document Frequency</i> (TF-IDF), diukur kesamaan data jawaban antara guru dengan murid menggunakan algoritma <i>euclidean distance</i> . Adapun data yang digunakan terdiri dari 5 soal dengan jumlah penjawab 15 murid. Referensi pembuatan soal dari buku IPS kelas 8 kurikulum 2013 revisi 2017, Hasil dari penilaian otomatis jawaban murid dengan TF-IDF, LSA dan <i>euclidian distance</i> didapatkan nilai korelasi sebesar 84%.
<i>Kata Kunci:</i> Ujian Online <i>Latent Semantic Analysis</i> <i>Euclidean Distance</i>	
	This is an open access article under the <a href="#">CC-BY-SA</a> license
	

## I. Pendahuluan

Pada setiap tengah semester dan akhir semester sekolah akan melaksanakan Penilaian Tengah Semester (PTS), Penilaian Akhir Semester (PAS) dan Penilaian Akhir Tahun (PAT). Ujian dilaksanakan untuk mengukur seberapa jauh pengetahuan murid yang sudah dikuasai. Hasil ujian PTS, PAS dan PAT digunakan sebagai penilaian rapor di tiap semesternya.

Pada Mts Attaufiq Padaelo jenis soal yang digunakan pada ujian akhir semester adalah soal Pilihan Ganda (PG) dan soal esai, dalam [1] soal pilihan ganda merupakan suatu soal yang jawabannya dapat dipilih dari beberapa kemungkinan jawaban yang ada. Setiap soal PG berisi pokok soal dan pilihan jawaban. Pilihan Jawaban terdiri dari kunci jawaban dan pengecoh. Pada penilaian ujian pilihan ganda cukuplah mudah dikarenakan setiap bobot soal sama yaitu jika benar bernilai satu dan jika salah bernilai nol, dalam [2] soal esai adalah soal yang mengandung permasalahan dan menuntut penguraian atas jawaban yang diberikan. Penguraian yang diberikan harus sesuai dengan apa yang ada pada bahan ajar dan diajarkan. Permasalahan kerap terjadi pada penilaian soal esai. Guru kadang mengalami ketidak konsistenan saat menilai soal-soal esai. Hal ini terlihat dari nilai yang diberikan guru. Kadang murid tidak puas dengan nilai yang diberikan guru setelah mereka bandingkan dengan nilai pada jawaban murid lainnya. Kondisi ini bisa terjadi ketika jumlah soal yang akan dinilai sangat banyak dan waktu untuk menyelesaikannya sangatlah sedikit.

Di era teknologi saat ini telah banyak diterapkannya ujian berbasis komputer, hanya saja hal tersebut berlaku pada ujian pilihan ganda. Oleh karena itu dibutuhkan sistem informasi penilaian untuk mempermudah para guru dalam memeriksa jawaban murid yang bervariasi, sehingga meningkatkan objektivitas dan efisiensi dalam penilaian jawaban murid [3], [4]. Dengan ini penulis akan membuat sistem informasi penilaian otomatis menggunakan metode *Latent Semantic Analysis* (LSA) untuk membandingkan hubungan antar dokumen teks, menghitung bobot kata menggunakan algoritma *Term Frequency - Inverse Document Frequency* (TF-IDF), diukur kesamaan data jawaban antara guru dengan murid menggunakan algoritma *euclidean distance*.

## II. Metode

### A. Text Preprocessing

*Text preprocessing* adalah suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur lagi dengan melalui serangkaian tahapan [5]. Adapun tahapan-tahapan umum pada *text preprocessing*, yaitu :

1. *Cleaning* melakukan proses mengurangi *noise* dengan menghapus setiap tanda baca.
2. *Case Folding* melakukan proses mengubah semua huruf menjadi huruf kecil (*lowercase*).
3. *Tokenizing* melakukan proses memecah semua kalimat menjadi kata atau biasa disebut token.
4. *Filtering* mengambil kata-kata penting dari hasil taken tadi lalu membuang kata-kata yang tidak penting.
5. *Stemming* mencari kata dasar dari setiap kata yang telah melalui tahap *filtering* dan mengubahnya menjadi kata dasar

### B. Term Frequency–Inverse Document Frequency

*Term Frequency - Inverse Document Frequency* (TF – IDF) adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen [6].

$$\text{idf}_j = \ln \frac{1+N}{1+df_{ij}} + 1 \quad (1)$$

$$W_{ij} = \text{tf}_{ij} \times \text{idf}_j \quad (2)$$

### C. Latent Semantic Analysis

*Latent Semantic Analysis* (LSA) merupakan sebuah metode yang memanfaatkan model statistik matematis untuk menganalisa struktur semantik suatu teks. LSA bisa digunakan untuk menilai esai dengan mengkonversikan esai menjadi matriks-matriks yang diberi nilai pada masing-masing term untuk dicari kesamaan dengan term referensi. Secara umum [7], LSA dalam perhitungannya menggunakan *Singular Value Decomposition* (SVD). SVD merepresentasikan *semantic space* ke dalam bentuk matriks yang memiliki ordo lebih kecil dibandingkan ordo matriks aslinya, namun perhitungan matriks tetap menghasilkan matriks yang bernilai hampir sama.

*Singular Value Decomposition* (SVD) adalah salah satu teknik reduksi dimensi yang bermanfaat untuk memperkecil nilai kompleksitas dalam pemrosesan *term-document matrix*. SVD merupakan teorema aljabar linear yang menyebutkan bahwa persegi panjang dari *term-document matrix* dapat dipecah/didekomposisikan menjadi tiga matriks yang dirumuskan dengan [8]:

$$A_{mm} = U_{mm} \times S_{mn} \times V_{nn}^T \quad (3)$$

### D. Euclidean Distance

*Euclidean distance* adalah perhitungan jarak dari dua buah titik dalam *euclidean space* untuk mempelajari hubungan antara sudut dan jarak [9]. Metode *euclidean distance* dapat digunakan untuk mengetahui kemiripan antar dokumen, *euclidean distance* menghitung jarak paling pendek antara dua titik apabila digunakan dalam dua dimensi. Yang dirumuskan dengan:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (4)$$

### E. Teknik Penilaian Jawaban Murid

Berikut rumus yang digunakan untuk perhitungan penilaian jawaban murid:

$$N = \frac{a}{b} \times c \quad (5)$$

## F. Analisis Korelasi

Analisis korelasi adalah suatu bentuk analisis data dalam penelitian yang bertujuan untuk mengetahui kekuatan atau bentuk arah hubungan di antara dua variable dan besarnya pengaruh yang disebabkan oleh variable yang satu terhadap variable lainnya. Koefisien korelasi adalah bilangan yang menyatakan kekuatan hubungan antara dua variable, juga dapat menentukan arah hubungan dari kedua variabel [10]. Berikut merupakan kalsifikasi koefisean korelasi:

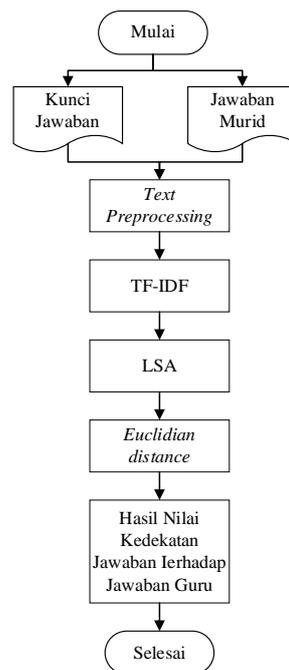
Tabel 1. Klasifikasi Koefisean Korelasi

Nilai Korelasi (r)	Tingkat Hubungan
0.00 – 0.19	Sangat Lemah
0.20 – 0.39	Lemah
0.40 – 0.59	Cukup
0.60 – 0.79	Kuat
0.80 - 100	Sangat Kuat

Adapun jenis koefisien korelasi yang akan digunakan pada penelitian ini adalah koefisien korelasi pearson yang dirumuskan dengan:

$$r = \frac{n \sum xy - (\sum x \sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2) (n \sum y^2 - (\sum y)^2)}} \quad (6)$$

## G. Flowchart Penilaian Jawaban Murid



Gambar 1. Flowchart Penilaian Jawaban Murid

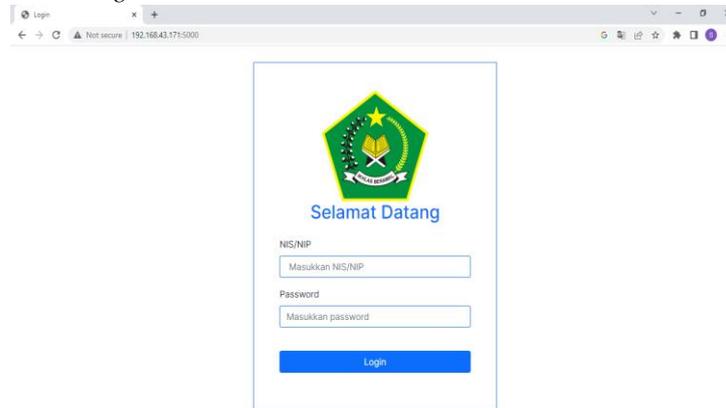
Pada Gambar 4 di atas merupakan implementasi metode TF-IDF, LSA, dan *euclidian distance* pada penilaian jawaban murid, terlebih dahulu dokumen kunci jawaban dan jawaban murid akan melalui tahap *text preprocessing* yaitu menghapus setiap tanda baca, mengubah semua huruf menjadi huruf kecil, memecah semua kalimat menjadi kata, mengambil kata-kata penting lalu membuang kata-kata yang tidak penting dan mencari kata dasar dari setiap kata dan mengubahnya menjadi kata dasar agar data text menjadi lebih terstruktur. Setelah itu menghitung bobot setiap kata yang ada pada dokumen menggunakan algoritma TF-IDF. Selanjutnya menganalisa hubungan antarkata dalam dokumen kunci jawaban dan jawaban murid menggunakan metode LSA kemudian menghitung jarak antara vektor kunci jawaban dan jawaban murid menggunakan metode *euclidean distance*.

### III. Hasil dan Pembahasan

#### A. Hasil Penelitian

Pada penelitian ini menghasilkan sebuah sistem informasi ujian online berbasis web yang dapat melakukan penilaian otomatis terhadap jawaban murid menggunakan metode TF-IDF, LSA dan *euclidian distance*. Adapun data yang digunakan terdiri dari 5 soal dengan jumlah penjawab 15 murid. Referensi pembuatan soal dari buku IPS kelas 8 kurikulum 2013 revisi 2017, berdasarkan perhitungan dengan TF-IDF, LSA dan *euclidian distance* didapatkan nilai korelasi sebesar 84%. Berikut merupakan tampilan dari sistem:

##### 1. Tampilan Halaman Login

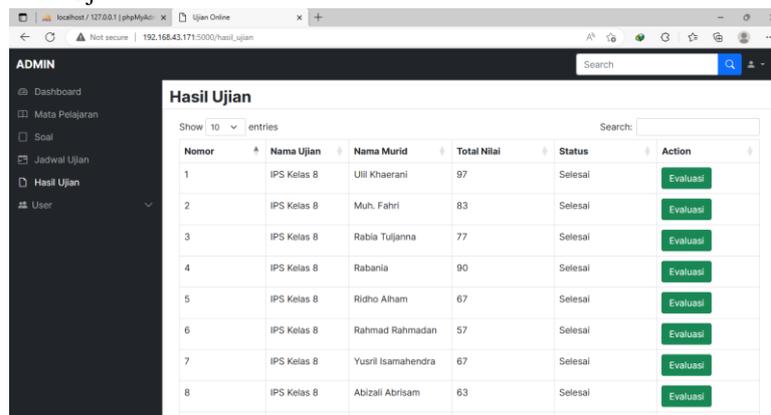


The screenshot shows a web browser window with the URL 'localhost/127.0.0.1/phpMyAdmin/ujian'. The page features a green pentagonal logo with a book and a star. Below the logo, the text 'Selamat Datang' is displayed. There are two input fields: 'NIS/NIP' with the placeholder 'Masukkan NIS/NIP' and 'Password' with the placeholder 'Masukkan password'. A blue 'Login' button is positioned at the bottom.

Gambar 2. Tampilan Halaman Login

Pada Gambar 2 diatas menampilkan halaman *login*. Pada halaman ini *users* memasukkan NIS/NIP dan password untuk bisa masuk ke halaman utama.

##### 2. Tampilan Hasil Ujian



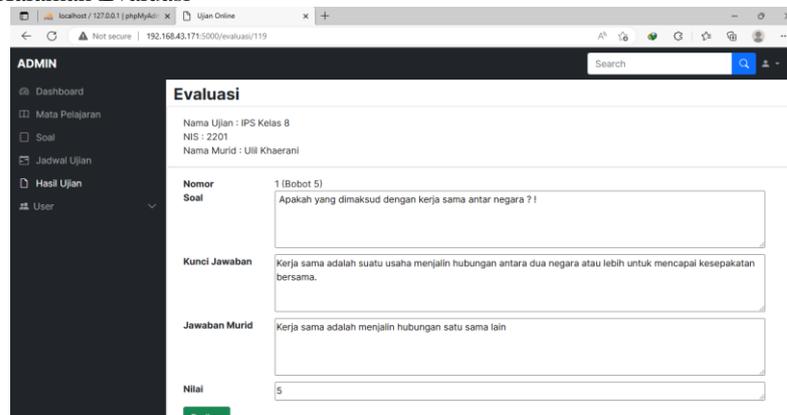
The screenshot shows the 'ADMIN' dashboard with a sidebar menu. The main content area is titled 'Hasil Ujian' and displays a table of exam results. The table has columns for 'Nomor', 'Nama Ujian', 'Nama Murid', 'Total Nilai', 'Status', and 'Action'. There are 8 rows of data, each with a green 'Evaluasi' button in the 'Action' column.

Nomor	Nama Ujian	Nama Murid	Total Nilai	Status	Action
1	IPS Kelas 8	Ulli Khaerani	97	Selesai	Evaluasi
2	IPS Kelas 8	Muh. Fahri	83	Selesai	Evaluasi
3	IPS Kelas 8	Rabia Tujanna	77	Selesai	Evaluasi
4	IPS Kelas 8	Rabania	90	Selesai	Evaluasi
5	IPS Kelas 8	Ridho Ailham	67	Selesai	Evaluasi
6	IPS Kelas 8	Rahmad Ramadhan	57	Selesai	Evaluasi
7	IPS Kelas 8	Yusril Isamahendra	67	Selesai	Evaluasi
8	IPS Kelas 8	Abizall Abrisam	63	Selesai	Evaluasi

Gambar 2. Tampilan Halaman Hasil Ujian

Pada Gambar 2 diatas menampilkan halaman hasil ujian. Pada halaman ini menampilkan data hasil ujian seperti nama ujian, nama murid, total nilai dan status . Admin dan guru dapat melihat data murid yang sedang atau yang telah mengerjakan ujian.

##### 3. Tampilan Halaman Evaluasi



The screenshot shows the 'ADMIN' dashboard with a sidebar menu. The main content area is titled 'Evaluasi' and displays a form for evaluating a question. The form has fields for 'Nomor', 'Soal', 'Kunci Jawaban', 'Jawaban Murid', and 'Nilai'. There is a green 'Periksa' button at the bottom.

Form details:

- Nomor: 1 (Bobot 5)
- Soal: Apakah yang dimaksud dengan kerja sama antar negara ?!
- Kunci Jawaban: Kerja sama adalah suatu usaha menjalin hubungan antara dua negara atau lebih untuk mencapai kesepakatan bersama.
- Jawaban Murid: Kerja sama adalah menjalin hubungan satu sama lain
- Nilai: 5

Gambar 3. Tampilan Halaman Evaluasi

Pada Gambar 3 menampilkan halaman evaluasi, Pada halaman ini menampilkan data detail ujian murid seperti nama ujian, NIS, nama murid, soal, kunci jawaban, jawaban murid dan nilai. Admin dan guru dapat melakukan penilaian otomatis terhadap jawaban murid.

#### B. Pembahasan

Berikut merupakan tabel korelasi dari jawaban murid soal nomor 1.

Tabel 2. Koefisien Korelasi Pearson

Responden Ke -	Penilaian Guru (X)	Penilaian Sistem (Y)	XY	X <sup>2</sup>	Y <sup>2</sup>
1	4	5	20	16	25
2	3	3	9	9	9
3	5	5	25	25	25
4	3	5	15	9	25
5	4	3	12	16	9
6	1	2	2	1	4
7	4	3	12	16	9
8	3	3	9	9	9
9	3	3	9	9	9
10	5	4	20	25	16
11	5	5	25	25	25
12	5	5	25	25	25
13	4	4	16	16	16
14	4	4	16	16	16
15	3	3	9	9	9
Total	56	57	224	226	231

Dari data pada tabel 6 di atas, maka dapat dilakukan perhitungan dengan menggunakan persamaan (6) sebagai berikut:

$$r = \frac{15 \times 224 - (56 \times 57)}{\sqrt{(15 \times 226 - (56)^2) - (15 \times 231 - (57)^2)}}$$

$$= \frac{168}{\sqrt{(254) - (216)}}$$

$$= 0,717$$

Tabel 3. Korelasi Setiap Soal

No Soal	Korelasi
1	0,717
2	0,754
3	0,991
4	0,810
5	0,956

Berdasarkan tabel 3 di atas, maka nilai rata-rata korelasi pada penelitian ini sebesar 0.84 atau jika dituliskan dalam presentase mencapai 84%.

#### IV. Kesimpulan dan saran

Berdasarkan hasil penelitian yang telah dilakukan selama 6 bulan mulai bulan agustus sampai dengan januari 2023, pada sekolah mts attaufiq dapat dinyatakan: Sistem penilaian esai otomatis dengan metode TF-IDF, LSA dan *euclidian distance* mampu memberikan nilai pada jawaban murid dengan akurasi 84%, kesalahan pengetikan saat menginputkan jawaban sangat berpengaruh terhadap penilaian pada jawaban murid, penilaian jawaban yang dihasilkan sistem dipengaruhi oleh ada tidaknya kata kunci pada jawaban murid, ketidakmampuan sistem dalam membaca urutan kata, Jumlah murid yang mengikuti ujian sangat mempengaruhi penilaian otomatis pada jawaban murid.

Adapun saran yang dapat diberikan untuk penelitian selanjutnya sebagai berikut : pada penelitian selanjutnya disarankan menggunakan metode lain seperti *Generalized Latent Semantic Analysis* untuk membandingkan hubungan antar dokumen teks dan *Language Based Matching* untuk pengukuran kemiripan, menambahkan algoritma yang dapat melakukan pengecekan kesalahan kata seperti rabin-karp atau *levenshtein distance*.

#### Daftar Pustaka

- [1] Nodia Gustri Somahhida and Imam Makruf, "Multiple Choice Objective Test Arabic Subject: Analysis & Implementation of the Edmodo Application/ Tes Objektif Pilihan Ganda Mata Pelajaran Bahasa Arab : Analisis & Implementasi pada Aplikasi Edmodo," *ATHLA J. Arab. Teaching, Linguist. Lit.*, vol. 3, no. 2, pp. 162–176, Dec. 2022, doi: 10.22515/athla.v3i2.5310.
- [2] S. Raupu, "Analisis Hasil Belajar Integral dengan Menggunakan Soal Pilihan Ganda dan Soal Essay (Studi Perbandingan pada Mahasiswa Semester II Prodi Tadris Matematika IAIN Palopo)," *J. Ilm. Ilmu Pendidik.*, vol. 2, no. 2, pp. 47–58, 2017.
- [3] E. Merindasari, T. Widyaningtyas, and M. Z. Arifin, "Sistem Informasi penilaian akademik siswa kurikulum 2013 berbasis web di SMAN 1 Trenggalek," *Sesindo 2015*, 2015, [Online]. Available: [http://is.its.ac.id/pubs/oajis/index.php/home/detail/1599/SISTEM-INFORMASI-PENILAIAN-AKADEMIK-SISWA-KURIKULUM-2013-BERBASIS-WEB-DI-SMAN-1-TRENGGALEK%0Ahttp://is.its.ac.id/pubs/oajis/index.php/file/download\\_file/1599](http://is.its.ac.id/pubs/oajis/index.php/home/detail/1599/SISTEM-INFORMASI-PENILAIAN-AKADEMIK-SISWA-KURIKULUM-2013-BERBASIS-WEB-DI-SMAN-1-TRENGGALEK%0Ahttp://is.its.ac.id/pubs/oajis/index.php/file/download_file/1599)
- [4] Endah Mastuti, "Pemanfaatan Teknologi Dalam Menyusun Evaluasi Hasil Belajar : Kelebihan Dan Kelemahan ' Tes Online ' Untuk Mengukur Hasil Belajar Mahasiswa," *J. Penelit. Psikol.*, vol. 7, no. 1, pp. 10–19, 2016.
- [5] Y. F. Alfiansah, A. M. Siregar, and A. F. N. Masruriyah, "Penerapan Algoritma K-Nearest Neighbors untuk Analisis Sentimen pada Buletin APTIKOM," *Sci. Student J. Information, Technol. Sci.*, vol. 3, no. 1, 2022.
- [6] D. Septiani and I. Isabela, "Analisis term frequency inverse document frequency (tf-idf) dalam temu kembali informasi pada dokumen teks," *SINTESIA J. Sist. Dan Teknol. Inf. Indones.*, vol. 1, no. 2, 2022.
- [7] B. O. Karo Karo, D. S. Naga, and V. C. Mawardi, "Perancangan Aplikasi Pendeteksi Kemiripan Teks Dengan Menggunakan Metode Latent Semantic Analysis," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 4, no. 1, p. 1, Jul. 2020, doi: 10.24912/computatio.v4i1.7191.
- [8] L. Setyaningsih, E. Setyati, and S. Tjandra, "Klasifikasi Soal Pilihan Ganda Berbahasa Indonesia Berdasarkan Level Kognitif Puspendik Dengan Support Vector Machine," *Insa. Comtech Inf. Sci. Comput. Technol. J.*, vol. 6, no. 2, pp. 30–37, 2020, [Online]. Available: [http://ejournal.unira.ac.id/index.php/insand\\_comtech/article/view/921/710](http://ejournal.unira.ac.id/index.php/insand_comtech/article/view/921/710)
- [9] A. A. Manrang, Y. Salim, and M. A. Asis, "Rancang Bangun Aplikasi Mall Maps Berbasis Mobile Menggunakan Metode Euclidean Distance," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 3, no. 4, pp. 301–310, 2022, doi: 10.33096/busiti.v3i4.1355.
- [10] I. K. A. Mudita, D. K. N. Semara Putra, and I. W. Sujana, "KORELASI ANTARA KEMAMPUAN BERPIKIR KRITIS DENGAN PENGUASAAN KOMPETENSI INTI PENGETAHUAN IPS SISWA," *Mimb. Ilmu*, vol. 24, no. 1, p. 36, Apr. 2019, doi: 10.23887/mi.v24i1.17408.