

Sistem Pengelompokan Dokumen Otomatis Dengan Metode K-Means Studi Kasus UKM Rade ITB STIKOM Bali

I Kadek Arya Putra Raditya¹⁾, Ni Luh Gede Pivin Suwirmayanti²⁾, Lilis Yuningsih³⁾

Sistem Komputer

Institut Teknologi dan Bisnis STIKOM Bali

Denpasar, Indonesia

e-mail: 210010026@stikom-bali.ac.id, pivin@stikom-bali.ac.id, lilis@stikom-bali.ac.id

Abstrak

Penelitian ini dirancang untuk menciptakan sebuah sistem manajemen dokumen dengan menggunakan dasar dari web yang ditujukan untuk Unit Kegiatan Mahasiswa (UKM) Robotics and Embedded Systems (RADE) di ITB STIKOM Bali. Sistem tersebut memanfaatkan algoritma K-Means Clustering guna mengkategorikan berbagai jenis dokumen, termasuk surat masuk, surat keluar, Rancangan Anggaran Belanja (RAB), proposal kegiatan, dan Laporan Pertanggungjawaban (LPJ), berdasarkan kesamaan nama dokumen. Tujuan pengembangan sistem ini adalah untuk meningkatkan efektivitas pengelolaan dokumen, meminimalkan kesalahan dalam pengarsipan, serta menghindari hilangnya data yang sering terjadi dalam sistem manual. Metode pengembangan sistem mengadopsi model Waterfall, yaitu mencakup tahap menganalisis, merancang, implementasi, serta pengujian. Proses pengujiannya ini dilakukan menggunakan metode blackbox testing guna memverifikasi bahwa semua fungsi sistem beroperasi seperti apa yang ditujukannya. Penelitian memberikan hasil mengindikasikan yaitu pembangunan sistem menggunakan metode K-Means ini mampu mengelompokkan dokumen secara otomatis, sehingga proses pencariannya serta pengelolaannya dokumen menjadi mudah, dan meningkatnya memudahkan operasional yang efisien di UKM RADE.

Kata kunci: Pengelolaan Dokumen, K-Means Clustering, Sistem Berbasis Web, Laravel

1. Pendahuluan

Di Denpasar terdapat institusi swasta perguruan tinggi yang bernama Institut Teknologi dan Bisnis (ITB) STIKOM Bali, bergerak di bidang Teknologi Informasi dan Bisnis [1]. Kampus ini mendukung pengembangan minat dan bakat mahasiswa melalui 31 organisasi mahasiswa (ORMAWA), seperti BEM PM, DPM PM, HIMAPRODI, dan UKM. Salah satu UKM aktif, UKM Robotics and Embedded Systems (RADE), fokus pada pelatihan, penelitian, lomba, dan pengelolaan dokumen. Namun, pengelolaan dokumen RADE masih manual menggunakan Microsoft Word dan Excel, yang berisiko merusak atau kehilangan data. [2].

Suatu teknik *clustering* guna mengelompokkan data terhadap beberapa *cluster* berdasarkan kesamaan atribut yang dimiliki oleh setiap data dinamakan metode K-Means.[3]. Metode K-Means membagi data ke dalam kelompok didasarkan pada persamaan karakteristiknya, sehingga data dalam satu *cluster* lebih mirip dibandingkan dengan data di *cluster* lain. Dalam pengelolaan dokumen UKM RADE ITB STIKOM Bali, metode ini dapat digunakan untuk mengelompokkan dokumen seperti surat masuk, surat keluar, RAB, dan LPJ berdasarkan kesamaan nama dokumen. Pengelompokan ini mempermudah pencarian, meningkatkan efisiensi, menghindari kesalahan pengarsipan, dan meminimalkan risiko kehilangan atau kerusakan dokumen penting.

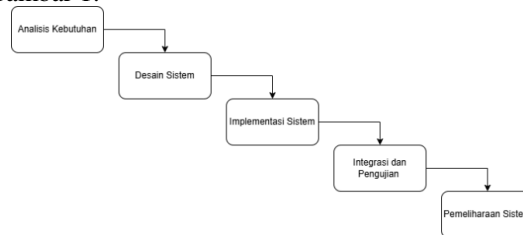
Penelitian sebelumnya, seperti yang dilakukan oleh Ahmad Husaein pada 2020, mengembangkan Sistem Informasi Manajemen Surat (SIMURAT) berbasis web di AKAKOM Stephen Jambi dengan metode *Waterfall*. Sistem ini meningkatkan kecepatan dan akurasi dalam pengelolaan surat masuk dan keluar [4]. Penelitian lain pada 2020, berjudul "Pengembangan Sistem Informasi Manajemen Arsip Dinamis dalam Mendukung Tata Kelola Kearsipan berbasis Web menggunakan Metode SDLC", menghasilkan sistem berbasis web untuk Universitas Negeri Surabaya. Sistem ini meningkatkan efisiensi pengelolaan arsip manual dengan menggunakan metode SDLC [5]. Penelitian ini menunjukkan sistem tersebut memenuhi aspek *usability* yang memiliki nilai rata-ratanya melebihi 3 dalam skala 5 untuk tingkat pembelajarannya (*learnability*), efisiensinya (*efficiency*), memorabilitasnya (*memorability*), kesalahan sistemnya (*errors*), dan kepuasannya (*satisfaction*) [5].

Penelitian terkait pada tahun 2021 juga membahas pengembangan sistem arsip surat berbasis web dengan *Framework* Laravel. Sistem tersebut meningkatkan efisiensi pengelolaan arsip, mempercepat pencarian data, dan menjaga keamanan arsip. *Framework* Laravel dipilih karena fleksibilitasnya dalam pengembangan, keamanan data yang lebih baik, dan kompatibilitas tinggi. Sistem ini dibangun dengan metode *Waterfall*, mencakup analisis kebutuhannya, desainnya, implementasinya, dan pengujiannya. Hasilnya menunjukkan peningkatan efisiensi dalam pengelolaan arsip surat pada instansi tersebut [6]. Penelitian lainnya pada tahun 2022 menggunakan metode *K-Means Clustering* untuk menganalisis dokumen tugas akhir di Universitas Nasional. Hasilnya menunjukkan data dapat dikelompokkan dengan didasarkan terhadap temanya, objeknya, serta metode penelitiannya itu, menghasilkan pola distribusi data yang lebih terorganisir [7].

Sistem pengelolaan dokumen otomatis untuk UKM RADE ITB STIKOM Bali akan diuji dengan metode *blackbox testing*. Pengujian ini melalui pendekatan *equivalence partitions* guna dilakukannya verifikasi fungsi utama seperti pengelolaan surat masuk dan keluar, serta *cause-effect relationship testing* untuk memastikan logika sistem berjalan dengan benar [8]. Berdasarkan penelitian sebelumnya, *blackbox testing* efektif untuk menguji keandalan fungsi dan kinerja sistem, termasuk mengukur respon terhadap beban pengguna seperti *loading time* dan *delay*, sehingga cocok untuk memastikan sistem ini berfungsi optimal.

2. Metode Penelitian

Dalam mengembangkan sistem, dilakukan melalui metode *Waterfall*, yang diinisiasikan melalui Winston Royce pada 1970 [9], adalah model untuk mengembangkan *software* secara sistematis serta urut. Setiap tahap, dimulai melalui perencanaan hingga pemeliharaan, wajib dilakukan penyelesaiannya sebelum tahap selanjutnya dilakukan, dan tidak memungkinkan untuk kembali ke tahap sebelumnya [10]. Model *Waterfall* digambarkan pada Gambar 1.



Gambar 1. Model *Waterfall*

2.1 Analisis Sistem

Pada tahap analisis sistem, fokusnya adalah mengumpulkan dan memahami kebutuhan pengguna melalui wawancara, observasi, dan studi literatur. Tujuannya untuk mendapatkan gambaran jelas mengenai keinginan pengguna dan cara sistem berfungsi [11]. Hasilnya adalah dokumen spesifikasi kebutuhan yang menjadi panduan untuk tahap selanjutnya.

2.2 Desain Sistem

Setelah analisis kebutuhan, tahap desain sistem dimulai untuk menerjemahkan kebutuhan menjadi desain rinci. Desain mencakup arsitektur, struktur *database*, diagram alir data, dan antarmuka pengguna. Desain ini menjadi panduan untuk tahap implementasi, memastikan setiap komponen terdefinisi jelas dan siap dikembangkan. Hasilnya berupa diagram dan spesifikasi teknis rinci.

2.3 Implementasi Sistem Dengan Penerapan Metode K-Means

Implementasi sistem melibatkan penerjemahan desain ke kode program dengan *K-Means* untuk mengelompokkan dokumen. Dokumen diproses melalui *preprocessing* seperti tokenisasi dan *TF-IDF*, kemudian dikelompokkan berdasarkan jarak *Euclidean*. Hasilnya diintegrasikan ke sistem *Laravel* dan disimpan dalam *database*. Pengujian akurasi dilakukan menggunakan *Silhouette Score*.

2.4 Pengujian Sistem

Setelah pengembangan, tahap pengujian dilakukan untuk memastikan semua fungsi bekerja sesuai spesifikasi. Dalam metode *Waterfall* [12], pengujian menggunakan *blackbox testing* [13], yang fokus pada pengujian fungsi tanpa melihat struktur kode internal. Penguji memeriksa *input* dan *output* untuk memastikan sistem berperilaku seperti yang diharapkan, serta menemukan dan memperbaiki *bug* sebelum sistem diluncurkan. Tujuannya adalah memastikan sistem memenuhi kebutuhan pengguna yang telah diidentifikasi pada tahap analisis.

2.5 Pemeliharaan Sistem

Pada tahap pemeliharaan, sistem diperiksa secara berkala untuk menemukan dan memperbaiki kesalahan program. Tahap ini juga mencakup penyesuaian atau modifikasi sistem sesuai dengan kebutuhan yang muncul setelah sistem mulai digunakan.

2.6 K-Means

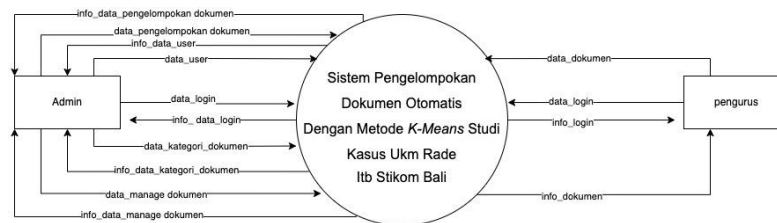
K-Means adalah algoritma pengelompokan data yang membagi *dataset* ke dalam sejumlah *cluster* (kelompok) berdasarkan kemiripan karakteristik. Algoritma ini bekerja dengan menentukan *centroid* (pusat *cluster*) dan mengelompokkan data ke *centroid* terdekat. Proses ini diulang hingga *centroid* tidak lagi berubah atau *cluster* menjadi stabil. K-Means berguna untuk analisis data, segmentasi pasar, atau pengenalan pola, dengan tujuan meminimalkan variasi dalam setiap *cluster* [3].

3. Hasil dan Pembahasan

3.1 Perancangan Sistem

1. Data Flow Diagram

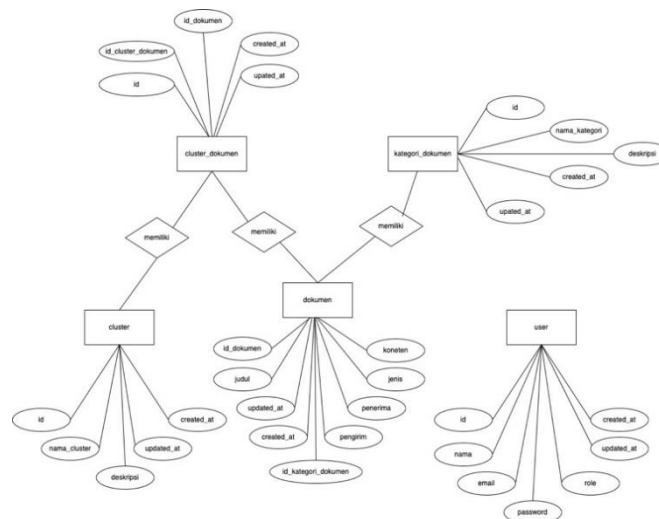
Data Flow Diagram (DFD) [14] adalah diagram yang menunjukkan pergerakan informasi antar lokasi dalam suatu proses. Berikut DFD dari sistem ini yang ditampilkan pada Gambar 2.



Gambar 2. *Data Flow Diagram*

2. Entity Relationship Diagram

Entity Relationship Diagram (ERD) [15] adalah alat untuk menggambarkan data atau objek dunia nyata (entitas) dan hubungan antar entitas menggunakan notasi, simbol, dan bagan. ERD digunakan untuk menyusun struktur data dan relasi antar data [16]. ERD yang digunakan dalam konteks ini terdiri dari 5 entitas. Rincian lebih lanjut mengenai ERD dapat ditemukan pada Gambar 3.



Gambar 3. *Entity Relationship Diagram*

3.2 Penerapan K-Means

Data dikelompokkan berdasarkan jarak *euclidean* ke *centroid*, lalu *centroid* diperbarui hingga konvergensi tercapai. Hasil *clustering* dikembalikan setelah proses selesai. Potongan kode dapat dilihat pada Gambar 4.

```

1 private function kMeans($data, $centroids, $k)
2 {
3     $clusters = [];
4     $converged = false;
5
6     while (!$converged) {
7         // Assign data ke cluster terdekat
8         $clusters = array_fill(0, $k, []);
9         foreach ($data as $doc) {
10            $distances = array_map(fn($centroid) => $this->euclideanDistance($doc, $centroid, $centroids);
11            $closestCentroid = array_search(min($distances), $distances);
12            $clusters[$closestCentroid][] = $doc;
13        }
14
15        // Update centroid
16        $newCentroids = [];
17        foreach ($clusters as $cluster) {
18            $newCentroids[] = $this->calculateCentroid($cluster);
19        }
20
21        // Periksa konvergensi
22        $converged = $centroids == $newCentroids;
23        $centroids = $newCentroids;
24    }
25
26    return $clusters;
27 }

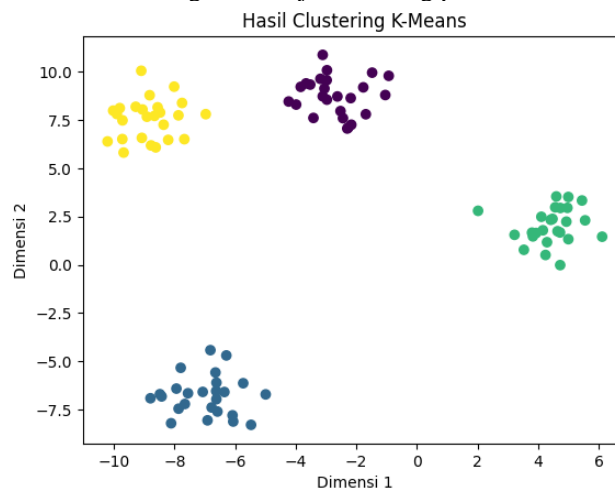
```

Gambar 4. Potongan kode K-Means

Hasil *clustering* berhasil mengelompokkan dokumen seperti surat masuk, surat keluar, RAB, dan LPJ. Pengelompokan ini memudahkan *admin* mengelola dan mencari dokumen. Evaluasi menggunakan *Silhouette Score* menunjukkan dokumen dalam satu *cluster* memiliki kemiripan tinggi, sementara antar *cluster* memiliki perbedaan signifikan.

3.3 Hasil Clustering

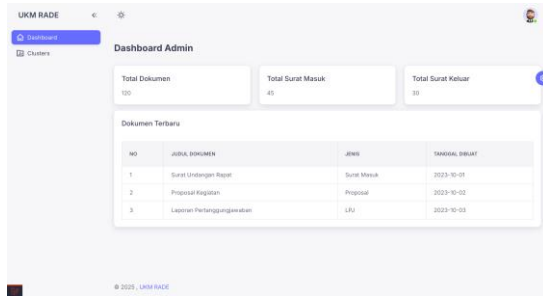
Grafik *clustering* K-Means menunjukkan distribusi dokumen ke dalam empat *cluster*, yaitu surat masuk, surat keluar, RAB, dan LPJ. *Scatter plot* pada dimensi 1 dan 2 mengilustrasikan pengelompokan dokumen berdasarkan karakteristiknya, dengan beberapa area tumpang tindih yang menunjukkan kesamaan fitur antar *cluster*. Hasil ini memberikan gambaran jelas tentang pola distribusi dokumen dalam *dataset*.



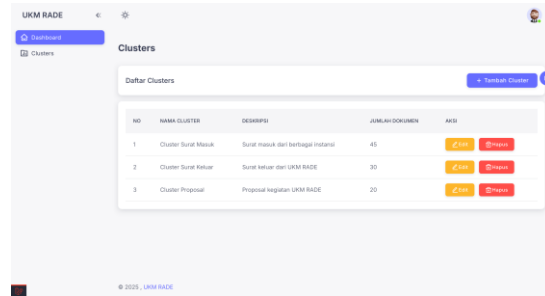
Gambar 5. Hasil Clustering

3.4 Desain Antarmuka

Halaman Dashboard Admin memberikan gambaran cepat tentang aktivitas pengelolaan dokumen di UKM RADE ITB STIKOM Bali, menampilkan statistik seperti total dokumen, surat masuk, dan surat keluar, serta daftar dokumen terbaru. Sementara itu, halaman *clusters* memungkinkan *admin* mengelompokkan dokumen berdasarkan kesamaan nama dokumen, seperti surat masuk, surat keluar, dan proposal, dengan fitur tambah, edit, dan hapus *cluster* untuk pengelolaan yang lebih terstruktur dan efisien.



Gambar 6 Antarmuka Dashboard



Gambar 7 Antarmuka Cluster

3.5 Pengujian Sistem

Proses pengujian sistem dilakukan untuk menilai kualitas dan memastikan bahwa sistem dapat beroperasi sesuai dengan kebutuhan dan harapan pengguna. Dalam penelitian ini, metode *black box testing* digunakan untuk mengevaluasi perangkat lunak dengan fokus pada hasil *output* yang dihasilkan berdasarkan data *input* yang diberikan. Hasil pengujian sistem dapat dilihat pada Tabel 1.

Tabel 1. Pengujian Sistem

Skenario Uji	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
Menu Dashboard Admin	Admin dapat melihat statistik total dokumen, surat masuk, dan surat keluar.	Sistem menampilkan statistik dengan akurat.	Sesuai
Menu Clusters	Admin dapat menambahkan, mengedit, dan menghapus cluster dokumen.	Admin berhasil menambahkan, mengedit, dan menghapus cluster.	Sesuai
Tampilan Dokumen Terbaru	Admin dapat melihat daftar dokumen terbaru beserta detailnya.	Sistem menampilkan daftar dokumen terbaru dengan informasi lengkap.	Sesuai
Pengelompokan Dokumen	Dokumen dapat dikelompokkan berdasarkan cluster yang telah ditentukan.	Dokumen berhasil dikelompokkan sesuai dengan cluster yang dipilih.	Sesuai

4. Kesimpulan

Penelitian ini mengembangkan sistem pengelolaan dokumen berbasis web untuk UKM *Robotics and Embedded Systems* (RADE) di ITB STIKOM Bali, menggunakan metode *K-Means Clustering* untuk mengelompokkan dokumen berdasarkan kemiripan nama dokumen. Sistem ini bertujuan meningkatkan efisiensi, meminimalkan risiko kesalahan pengarsipan, dan menghindari kehilangan data yang terjadi pada pengelolaan manual sebelumnya. Dengan menggunakan model *Waterfall* untuk pengembangan dan *blackbox testing* untuk pengujian, sistem ini diharapkan dapat mempermudah pencarian dan pengelolaan dokumen, serta meningkatkan keandalan pengarsipan di UKM RADE.

Daftar Pustaka

- [1] I. Ayu, C. P. Dewi, N. L. Gede, P. Suwirmayanti, M. Arya, and B. Saputra, "Sistem Informasi Pengelolaan Himpunan Mahasiswa Program Studi Teknologi Informasi Berbasis Website," *Seminar Hasil Penelitian Informatika dan Komputer (SPINTER) | Institut Teknologi dan Bisnis STIKOM Bali*, vol. 1, no. 1, pp. 558–563, Nov. 2023, Accessed: Jan. 18, 2025. [Online]. Available: <https://spinter.stikom-bali.ac.id/index.php/spinter/article/view/118>
- [2] "GAMBARAN GAYA KEPEMIMPINAN KETUA ORGANISASI MAHASISWA... - Google Scholar." Accessed: Jan. 18, 2025. [Online]. Available: https://scholar.google.com/scholar?hl=id&as_sdt=0%2C5&q=GAMBARAN+GAYA+KEPEMIMPINAN+KETUA+ORGANISASI+MAHASISWA+UNIVERSITAS+JENDERAL+ACHMAD+YANI+YOGYAKARTA&btnG=

-
- [3] “View of PENERAPAN TEKNIK CLUSTERING UNTUK PENGELOMPOKAN KONSETRASI MAHASISWA DENGAN METODE K-MEANS.” Accessed: Jan. 18, 2025. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/sintesa/article/view/884/788>
- [4] A. Husaein, M. Informatika, A. Akuntansi, D. Komputer, S. Jambi, and J. Halim, “Perancangan Sistem Informasi Manajemen Surat (SIMURAT) Pada AKAKOM Stephen Jambi,” *ejournal.unama.ac.id* Husaein *Jurnal Ilmiah Media Sisfo*, 2020•*ejournal.unama.ac.id*, vol. 14, no. 2, 2020, doi: 10.33998/mediasisfo.2020.14.2.829.
- [5] R. Suprpto, D. P.-J. of E. Information, and undefined 2020, “Pengembangan Sistem Informasi Manajemen Arsip Dinamis Dalam Mendukung Tata Kelola Kearsipan Berbasis Web Menggunakan Metode SDLC,” *ejournal.unesa.ac.id* R Suprpto, DR Prehanto *Journal of Emerging Information System and Business Intelligence*, 2020•*ejournal.unesa.ac.id*, Accessed: Jan. 18, 2025. [Online]. Available: <https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/36855>
- [6] S. Firdayanti *et al.*, “Model Sistem Informasi Arsip Surat Berbasis Web Menggunakan Framework Laravel,” *ojs.stmik-banjarbaru.ac.id* S Firdayanti, Y Syahidin *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 2021•*ojs.stmik-banjarbaru.ac.id*, Accessed: Jan. 18, 2025. [Online]. Available: <http://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/view/705>
- [7] F. Sept, G. Zentrato, A. Triayudi, and E. Tri, “Analisis Clustering Dokumen Tugas Akhir Mahasiswa Sistem Informasi Universitas Nasional menggunakan Metode K-Means Clustering,” *journal.lembagakita.org* FSG Zentrato, A Triayudi *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 2022•*journal.lembagakita.org*, vol. 6, no. 1, p. 2022, 2022, doi: 10.35870/jti.
- [8] N. Luh *et al.*, “Penerapan Helpdesk System dengan Pengujian Blackbox Testing Implementation Of Helpdesk System With Blackbox Testing,” *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 2, no. 02, pp. 55–64, 2020.
- [9] A. Abdul Wahid Sekolah Tinggi Manajemen Informatika dan Komputer Sumedang, “Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi.” [Online]. Available: <https://www.researchgate.net/publication/346397070>
- [10] E. Listiyan and E. R. Subhiyacto, “Rancang Bangun Sistem Inventory Gudang Menggunakan Metode Waterfall (Studi Kasus Di CV. Aqualux Duspha Abadi Kudus Jawa Tengah).”
- [11] B. Berbasis Website Menggunakan Metode, “SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI PERANCANGAN UI/UX SISTEM INFORMASI INVENTARIS.”
- [12] M. B.-P. J. P. R. dan and undefined 2021, “Penerapan Metode Waterfall Untuk Perancangan Sistem Informasi Inventory Pada Toko Keramik Bintang Terang,” *e-jurnal.lppmunsera.org*, vol. 8, no. 2, 2021, Accessed: Jan. 18, 2025. [Online]. Available: <https://e-jurnal.lppmunsera.org/index.php/PROSISKO/article/view/3852>
- [13] S. S.-I. (International J. of I. System and undefined 2020, “Software testing with the approach of blackbox testing on the academic information system,” *ijistech.org*, Accessed: Jan. 18, 2025. [Online]. Available: <http://ijistech.org/ijistech/index.php/ijistech/article/view/54>
- [14] R. I. preprint arXiv:1011.0278 and undefined 2010, “Formalization of the data flow diagram rules for consistency check,” *arxiv.org*, vol. 1, no. 4, Nov. 2010, doi: 10.5121/ijsea.2010.1406.
- [15] A. Kusuma, T. W.-A. J. Ilmiah, and undefined 2016, “Rancang bangun sistem pendataan nilai akademik siswa berbasis web menggunakan php dan mysql di SMA Islam Hasanuddin Kesamben,” *ejournal.unisbablitar.ac.id*, Accessed: Jan. 18, 2025. [Online]. Available: <http://ejournal.unisbablitar.ac.id/index.php/antivirus/article/view/83>
- [16] D. Edi, S. Betshani, J. Prof, D. Suria, S. N.-J. informatika, and undefined 2009, “Analisis data dengan menggunakan ERD dan model konseptual data warehouse,” *academia.edu*, Accessed: Jan. 18, 2025. [Online]. Available: <https://www.academia.edu/download/15389724/404-1158-1-PB.PDF>
-