



IMPLEMENTASI ALGORITMA LEVENSTHEIN DISTANCE GUNA PENGELOMPOKAN DOKUMEN ARSIP ADMINISTRASI PADA SMK PANGUDI LUHUR SEPUTIH MATARAM

Agung Prabowo¹, Saniati², Dedi Darwis³

^{1,2,3}Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia

Corresponding Author : agung.prb1108@gmail.com

Received: (10 January 2022) Accepted: (15 February 2022) Published: (30 March 2022)

Abstract

The administrative document grouping application was built using the Microsoft Visual Studio application with the C # programming language. There are 3 types of documents that will be grouped, namely SPP, Semester, and Re-registration. At this time the administrative staff in classifying documents was still manual. This causes document loss when the principal asks for some student payment data. This application is designed to facilitate administrative staff in grouping administrative archive documents. The grouping process is carried out using the Levensthein Distance Algorithm by comparing the strings of the Test Text with the existing strings in the document. The results of this grouping will be stored in a predefined folder. This is useful for making it easier for administrative staff when searching for documents. In this study, the authors used a testing aspect based on the accuracy of the results of document grouping. From the results of the research, the accuracy obtained is 100%, by counting the number of correct documents divided by the number of documents to be grouped then multiplied by 100%. So document grouping using the Levensthein Distance Algorithm can be categorized as good.

Keywords: Levensthein Distance, Document Grouping, C#, Desktop, Visual Studio

Abstrak

Aplikasi pengelompokan dokumen administrasi dibangun menggunakan aplikasi Microsoft Visual Studio dengan bahasa pemrograman C#. Terdapat 3 jenis dokumen yang akan dikelompokkan yaitu SPP, Semester, dan Daftar Ulang. Pada saat ini staff administrasi dalam mengelompokkan dokumen masih secara manual. Hal ini menyebabkan sering terjadi kehilangan dokumen saat kepala sekolah meminta beberapa data pembayaran siswa. Aplikasi ini dirancang guna memudahkan staff administrasi dalam mengompokan dokumen arsip administrasi. Proses pengelompokan dilakukan dengan Algoritma Levensthein Distance yaitu dengan membandingkan string dari Teks Uji dengan String yang ada didokumen. Hasil pengelompokan ini akan disimpan pada folder yang telah ditentukan sebelumnya. Hal ini berguna mempermudah staff administrasi saat melakukan pencarian dokumen. Dalam penelitian ini penulis menggunakan aspek pengujian berdasarkan akurasi hasil dari pengelompokan dokumen. Dari hasil penelitian akurasi yang didapatkan yaitu 100%, dengan menghitung jumlah dokumen benar dibagi dengan jumlah dokumen yang akan dikelompokkan kemudian dikali 100%. Jadi pengelompokan dokumen menggunakan Algoritma Levensthein Distance dapat dikategorikan baik.

Kata Kunci: Levensthein Distance, Pengelompokan Dokumen, C#, Desktop, Visual Studio

To cite this article:

Prabowo et.al. (2022). Implementasi algoritma levensthein distance guna pengelompokan dokumen arsip administrasi pada smk pangudi luhur. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, Vol 3. No. 1, 57-65.

PENDAHULUAN

Saat ini perkembangan teknologi sangat cepat sekali, banyak inovasi baru yang lahir terutama dibidang teknologi dan informasi yang bertujuan untuk mempermudah manusia dalam melakukan pekerjaannya. Dalam dunia pendidikan, teknologi dan informasi yang digunakan sangatlah beragam, misalnya ada yang menciptakan sebuah aplikasi untuk mempermudah pembayaran dan pengelompokan dokumen administrasi sekolah dan masih banyak lagi. SMK Pangudi Luhur Seputih Mataram merupakan salah satu sekolah swasta yang ada di Lampung Tengah. SMK Pangudi Luhur Seputih Mataram memiliki siswa yang akan selalu bertambah setiap tahunnya. Berdasarkan hasil wawancara dengan pihak SMK Pangudi Luhur Seputih Mataram, banyaknya siswa membuat data yang diperlukan juga semakin bertambah, hal ini dapat menjadi salah satu kendala dalam proses administrasi keuangan nantinya, seperti arsip keuangan yang tercampur satu sama lain. Hal ini menyebabkan staff tata usaha sering kali mengalami kehilangan data saat akan memberikan dokumen kepada kepala sekolah.

Pembuatan aplikasi ini bertujuan untuk mengelompokan dokumen berdasarkan jenis pembayaran dan tahun ajaran, agar saat staff tata usaha akan memberikan dokumen kepada kepala sekolah tidak mengalami kesalahan saat mengelompokan dokumen secara manual. Meskipun semua tidak lepas dari peran manusia dalam proses operasionalnya, serta pengelolaan teknologi yang tepat dapat membantu dalam pekerjaannya. Aplikasi Pengelompokan Dokumen Administrasi dibuat menggunakan Algoritma Leventhein Distance. Karena Algoritma ini dapat meminimalisir kesalahan dalam pencarian dan pengelompokan data tertentu.

Dalam penelitian ini ketepatan pengelompokan dokumen dapat diukur menggunakan rumus akurasi. Akurasi dimaksudkan derajat pemenuhan terhadap pengukuran standar, yaitu yang mana menjangkau pengukuran yang tepat sasaran. Akurasi mengukur ketepatan dan kemiripan hasil pada waktu yang sama dengan membandingkannya terhadap nilai yang telah ditentukan. Untuk mengetahui keberhasilan sistem klasifikasi dilakukan pengujian terhadap sistem dengan menghitung akurasinya menggunakan rumus akurasi (Sendiherla, 2018). Pengujian akurasi sistem pada penelitian ini penulis menggunakan rumus Akurasi = (Jumlah Dokumen Benar)/(Dokumen Dikelompokan) \times 100%.

Tujuan dari penelitian ini adalah membangun sistem pengelompokan dokumen dengan mengukur akurasi menggunakan algoritma levensthein distance. Hal ini agar dapat memudahkan staff administrasi mengelompokan dan mencari dokumen.

METODE PENELITIAN

Menurut Putra (2020) SDLC (Systems Development Life Cycle) dalam rekayasa perangkat lunak adalah proses pembuatan dan pengubahan sistem serta model dan metodologi yang digunakan untuk mengembangkan sistem-sistem tersebut. Pembuatan aplikasi ini mengikuti tahapan-tahapan SDLC (System Development Life Cycle) menggunakan waterfall model karena pembuatan aplikasi dikerjakan secara linier. Waterfal Model, salah satu model SDLC paling awal, terdiri dari lima fase berturut-turut dan mereka masing-masing: Analisis bisnis, desain, implementasi, pengujian, dan pemeliharaan (Galandi, 2017).

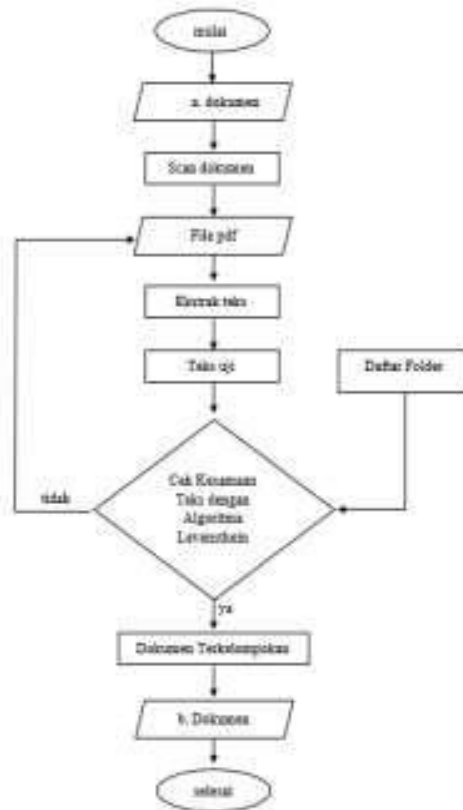
Metode Analisis

(1) Analisis sistem merupakan tahap awal dalam sebuah siklus SDLC, dimana analisis sistem akan melakukan berbagai macam analisis terhadap sebuah sistem yang sudah ada, dan bagaimana nantinya sebuah sistem akan berjalan. Hal ini termasuk di dalamnya adalah sebagai bentuk kelebihan dan kekurangan sistem, fungsi dari sistem, hingga berbagai macam pembaruan yang bisa saja diterapkan pada sebuah sistem. (2) Tahapan berikutnya dari siklus SDLC pada sebuah sistem adalah perancangan terhadap sistem Tahap ini merupakan tahap dimana seluruh hasil analisa diterapkan menjadi sebuah rancangan atau cetak biru dari sebuah sistem. Tahap perancangan sistem ini bisa kita sebut sebagai cetak biru, atau bisa juga kita sebut sebagai prototype, dimana sistem ini sudah siap untuk dikembangkan. (3) Tahapan pengembangan sistem ini merupakan tahapan dimana rancangan atau cetak biru sistem ini mulai dikerjakan dan dibuat atau diimplementasikan menjadi sebuah sistem yang utuh, dan dapat digunakan. Tahap ini merupakan tahapan yang cukup lama, karena dalam prakteknya tahap pengembangan sistem ini bisa saja menemui kendala – kendala baru yang menyebabkan proyek menjadi terhambat, sehingga dibutuhkan analisis tambahan, ataupun perancangan tambahan. (4) Setelah sistem selesai dikembangkan dan juga dibuat, maka sistem tersebut tidak akan langsung digunakan secara umum ataupun secara komersil. Tentu saja harus ada proses pengujian terhadap sistem yang sudah dikembangkan tersebut. Tahap pengujian sistem ini merupakan waktu yang tepat untuk mencoba apakah sistem yang sudah berhasil

dikembangkan memang dapat bekerja dengan optimal dan juga sempurna. Apabila memang dapat bekerja dengan baik dan sempurna, maka sistem siap untuk digunakan. (5) Tahap ini bisa dibilang sebagai tahapan final atau tahapan akhir dari satu buah siklus SDLC. Tahapan ini merupakan tahapan dimana sebuah sistem sudah selesai dibuat, sudah diujicoba, dan dapat bekerja dengan baik dan juga optimal. Ketika tahapan sebelumnya sudah berhasil dilewati, maka ini lah saatnya sistem tersebut mulai diimplementasikan dan digunakan secara real oleh user yang membutuhkan.

Flowchart

Pada bagian ini sebutkan variabel-variabel yang digunakan, susunlah seperti contoh dibawah ini:



Gambar 1 Flowchart Pengelompokan Dokumen

Ket : a dokumen (dokumen sebelum di scan), b dokumen (dokumen setelah di scan). Setelah dokumen discan akan menghasilkan file pdf, dari file pdf ini akan dilakukan ekstraks teks didalam sistem. Untuk ekstraks teks penulis menggunakan pustaka yang sudah disediakan Visual Studio yaitu Asprise OCR. string, teks akan diuji kesamaan dengan daftar folder yang telah dibuat admin. Jika teks tidak sama maka akan diproses kembali ke step sebelumnya, jika teks sama maka dokumen akan dikelompokan.

HASIL DAN PEMBAHASAN

Implementasi Aplikasi

Tahap implementasi aplikasi ini bertujuan agar dapat menguji tingkat keakuratan dari Algoritma Levensthein Distane dalam mengelompokan dokumen. Didalam aplikasi pengelompokan dokumen berikut terdapat beberapa menu, diantaranya adalah menu Admin, Siswa, Input File, dan Laporan. Berikut tampilan menu utama pada aplikasi Pengelompokan Dokumen.



Gambar 2 Tampilan Menu

Penjabaran Menu

Didalam aplikasi pengelompokan dokumen ini terdapat 4 menu, diantaranya adalah Admin, Siswa, Input File, dan Laporan. Menu admin berguna untuk menambahkan admin yang akan mengelola/menggunakan aplikasi pengelompokan dokumen ini. Hanya ada beberapa user saja yang menggunakan aplikasi ini diantaranya yaitu Staff Keuangan dan Kepala Staff Keuangan. Berikut adalah tampilan dari menu admin.



Gambar 3 Tampilan Menu Admin

Setelah menu admin, selanjutnya ada menu siswa. Menu siswa ini berguna untuk menginput nama, kelas, jurusan, dan tahun ajaran secara manual. Selebihnya menu siswa ini berguna dalam penggunaan database untuk pengelompokan dokumen. Karena dalam pengelompokan dokumen ini penulis menampung semua data kedalam database. Berikut adalah tampilan dari menu siswa.



Gambar 4 Tampilan Menu Siswa

Selanjutnya yaitu menu yang paling inti yaitu menu input file. Didalam menu input file inilah proses ekstrak dokumen dan pengelompokan dokumen akan dilakukan. Dan dimenu input file ini kita dapat melihat hasil ekstraksi dari dokumen PDF OCR dan hasil dari algoritma saat mengelompokkan dokumen. Jadi user dapat mengetahui dokumen yang telah diinput apakah berhasil dikelompokkan atau tidak. Berikut adalah tampilan dari menu input file.



Gambar 5 Tampilan Menu Input File

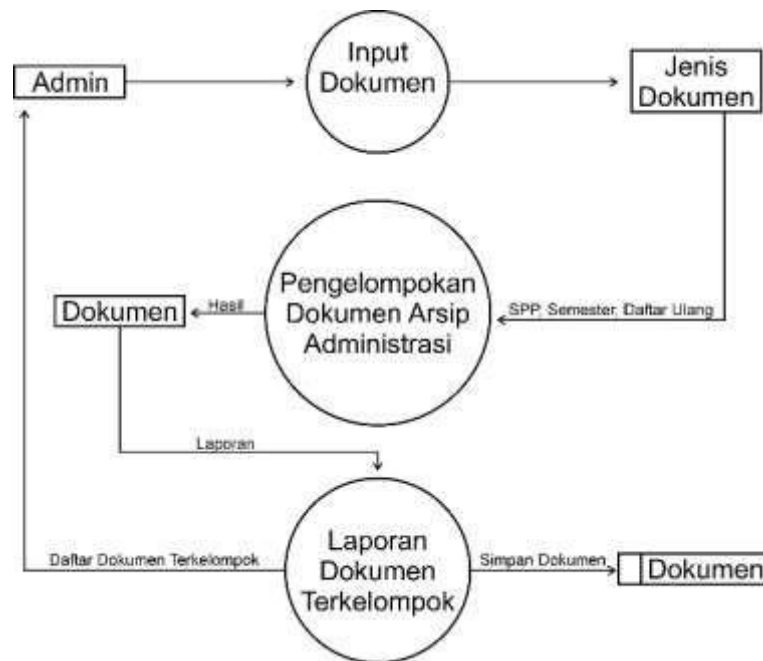
Dan menu yang terakhir yaitu menu laporan. Pada menu laporan ini terdapat 3 submenu, diantaranya adalah submenu SPP, submenu Semester, dan submenu Daftar Ulang. Didalam 3 submenu tersebut terdapat beberapa button yang fungsinya sama. Pada menu laporan ini user dapat melihat hasil dari semua dokumen yang telah terkelompokkan, atau dapat melihat juga dokumen yang telah terkelompokkan berdasarkan Tahun Ajaran, Jurusan, dan Kelas. Dimenu laporan ini juga user dapat mengekspor dokumen yang telah terkelompokkan kedalam folder yang telah ditentukan berdasarkan kriteria dari pengelompokan. Berikut adalah tampilan dari menu laporan.

Nama Siswa	Nilai	Jurusan	Tahun Ajaran	Jumlah	Februari	Maret	April
Muhammad Gayu	81	TKR	2019	170.000	170.000	170.000	170.000
Muhammad Gayu	81	TKR	2019	170.000	170.000	170.000	170.000
Razi Mahesa	81	TKR	2019	170.000	170.000	170.000	170.000
Riky Ay P	81	TKR	2019	170.000	170.000	170.000	170.000
Soneta	81	TKR	2019	170.000	170.000	170.000	170.000
Suhel	81	TKR	2019	170.000	170.000	170.000	170.000
Ta Rahma A	81	TKR	2019	170.000	170.000	170.000	170.000
Zain Han	81	TKR	2019	170.000	170.000	170.000	170.000
Aprianto	81	TKR	2019	170.000	170.000	170.000	170.000
Al Maulana	81	TKR	2019	170.000	170.000	170.000	170.000
Alu Maulana	81	TKR	2019	170.000	170.000	170.000	170.000
Beriz Joppano	81	TKR	2019	170.000	170.000	170.000	170.000
Christa Fiantar	81	TKR	2019	170.000	170.000	170.000	170.000

Gambar 6 Tampilan Menu Laporan

DFD (Data Flow Diagram)

Data Flow Diagram (DFD) memberikan tampilan secara visual tentang aliran data dan informasi dari suatu sistem. Adapun bentuk data flow diagram dari pengelompokan dokumen arsip administrasi dapat dilihat pada gambar dfd berikut.

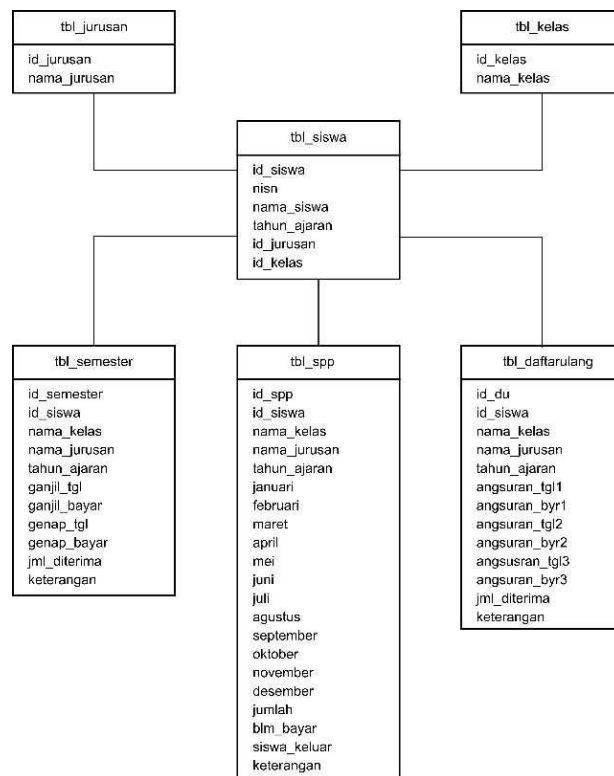


Gambar 7 Data Flow Diagram

ERD (Entity Relationship Diagram)

Entity Relationship Diagram (ERD) adalah suatu diagram yang digunakan untuk merancang suatu basis data, dipergunakan untuk memperlihatkan hubungan atau relasi antar entitas atau objek yang terlihat beserta atributnya.

pada database yang telah penulis rancang mempunyai beberapa tabel yang berelasi yaitu tbl_siswa, tbl_jurusan, tbl_kelas, tbl_spp, tbl_semester, tbl_daftarulang. Berikut adalah rancangan dari ERD.



Gambar 8 Entity Relationship Diagram

Pengujian Dengan Cannon Scanner

Dengan penggunaan alat scanner jenis Cannon Scanner Lide 300 dapat menghasilkan dokumen dengan format pdf ocr cukup berkualitas HD. Karena dalam proses pengelompokan dokumen hal utama yang paling penting yaitu pengekstrakan dokumen. Dalam pengekstrakan ini semua huruf atau karakter harus dapat dikenali. Jika terdapat huruf huruf yang tidak dapat dikenali itu akan berpengaruh pada proses pengelompokannya. Maka dari itu dokumen PDF OCR yang akan diekstrask harus memiliki kualitas HD.

Dari semua jumlah dokumen yang akan dikelompokan yaitu 60 dokumen dan dokumen yang berhasil dikelompokan yaitu 60 Jadi bisa dilihat dari data tersebut maka dapat dihitung akurasi pengelompokan dokumen dari program ini adalah : $Akurasi = \frac{60}{60} \times 100\% = 100\%$ Setelah melihat berdasarkan presentase hasil uji akurasi dari keseluruhan dokumen yang dapat terkelompokan dapat disimpulkan bahwa pengelompokan dengan dokumen yang discan menggunakan Cannon Scanner Lide 300 dan menggunakan Algoritma Levensthein Distance cukup akurat karena mencapai nilai 100%.

Pengujian Dengan Adobe Scanner

Dengan penggunaan apk Adobe Scanner dan dengan perangkat handphone vivoy12 dapat menghasilkan dokumen dengan format pdf ocr. Pada hasil scan menggunakan Adobe Scanner dokumen yang dihasilkan tidak berkualitas HD. Oleh karena itu saat proses ekstrasi terdapat karakter yang tidak dikenali. Dalam pengekstrakan ini semua huruf atau karakter harus dapat dikenali. Jika terdapat huruf huruf yang tidak dapat dikenali itu akan berpengaruh pada proses pengelompokannya.

Dari semua jumlah dokumen yang akan dikelompokan yaitu 60 dokumen dan dokumen yang berhasil dikelompokan yaitu 60 Jadi bisa dilihat dari data tersebut maka dapat dihitung akurasi pengelompokan dokumen dari program ini adalah : $Akurasi = \frac{60}{60} \times 100\% = 100\%$ Setelah melihat berdasarkan presentase hasil uji akurasi dari keseluruhan dokumen yang dapat terkelompokan dapat disimpulkan bahwa pengelompokan dengan dokumen yang discan menggunakan Apk Adobe Scanner dan menggunakan Algoritma Levensthein Distance cukup akurat karena mencapai nilai 100%.Inti dari pengelompokan dokumen ini adalah berada pada Pengekstrakan file pdf ocr dan Teks Uji. Jika pada pengekstrakan dapat mengenali string yang akan dilakukan perbandingan dengan teks uji, maka dokumen tetap dapat terkelompokan. Yang membedakan file yang discan menggunakan Cannon Scanner Lide 120 dengan Adobe Scanner hanyalah pada hasil kualitas gambarnya. Jika menggunakan Cannon Scanner Lide 120 dokumen yang dihasilkan mempunyai kualitas HD, sehingga semua karakter yang ada didalam file dapat diekstrak dengan sempurna. Sedangkan jika menggunakan Adobe Scanner dokumen yang dihasilkan kualitasnya

kurang bagus, jadi hasil pengestrakan kurang sempurna. Sehingga pada pengelompokan banyak karakter yang tidak rapi.

Analisis Pengujian

Berdasarkan pengujian dengan jumlah keseluruhan 120 dokumen dibagi menjadi 2, 60 dokumen hasil scan Cannon Scanner Lide 120 dan 60 dokumen hasil scan menggunakan perangkat Vivo y12 dengan Apk Adobe Scanner. Berikut tabel hasil dari pengujian.

Tabel 1 Tabel Hasil Analisis Scan Cannon Scanner

No	Jenis Dokumen	Terkelompok	Tidak Terkelompok
1	SPP	20	0
2	Semester	20	0
3	Daftar Ulang	20	0
Total		60	0

Dari tabel diatas dapat dilihat bahwa pengelompokan dengan dokumen yang discan menggunakan Cannon Scanner Lide 120 terbilang akurat, karena dapat mengelompokkan seluruh jumlah dokumen dengan akurasi 100%.

Tabel 2 Tabel Hasil Analisis Scan Adobe Scanner

No	Jenis Dokumen	Terkelompok	Tidak Terkelompok
1	SPP	20	0
2	Semester	20	0
3	Daftar Ulang	20	0
Total		60	0

Pada dokumen hasil scan Adobe Scanner juga dapat mengelompokkan keseluruhan dokumen dengan akurasi 100%. Tetapi pada proses ekstrasi, isi yang ada didalam dokumen banyak string atau karakter yang tidak terbaca Hal ini disebabkan karena kualitas citra pdf tidak HD. Sehingga pada pengelompokannya pun banyak juga string atau karakter yang tidak terkelompokkan berdasarkan tabel. Hal ini mengakibatkan isi dari tabel berantakan. Berikut adalah tampilan dari dokumen yang telah dikelompokkan menggunakan hasil scan Adobe Scanner..

PENUTUP

Berdasarkan hasil pembahasan dan penelitian, dapat disimpulkan sebagai berikut :

1. Penelitian ini menghasilkan aplikasi Pengelompokan Dokumen Arsip berbasis desktop. Aplikasi dibangun dengan tool Microsoft Visual Studio menggunakan Bahasa pemrograman C#.
2. Pada Penelitian ini penulis menggunakan 2 jenis hasil scan yang dikelompokkan. Dokumen pertama yang berjumlah 60 discan menggunakan Cannon Scanner Lide 120. Dan dokumen kedua yang berjumlah 60 discan menggunakan perangkat Vivo Y12 dengan Apk Adobe Scanner.
 - a. Hasil pengujian dengan dokumen yang discan menggunakan Cannon Scanner Lide 120 mendapatkan akurasi 100% dan hasil dari pengelompokan rapi.
 - b. Hasil pengujian dengan dokumen yang discan menggunakan perangkat Vivoy12 dengan Apk Adobe Scanner mendapatkan akurasi 100%, tetapi banyak Tahun Ajaran, Jurusan, dan Kelas yang tidak terbaca sehingga hasil dari pengelompokan acak-acakan.

REFERENSI

- Daryanto, H. (2015). Administrasi dan Manajemen Keuangan Sekolah. Jakarta: Rineka Cipta.
- Galandi, F. (2017). Metode Waterfall : Definisi, Tahapan, Kelebihan dan Kekurangan. Diambil kembali dari PengetahuandanTeknologi.com:<http://www.pengetahuandanteknologi.com/2016/09/metode-waterfall-definisitahapan.html>
- Herianto. (2019). Pengelompokan Dokumen, 20.
- Kurniawan, H. (2017). OTORISASI PENGGUNA LABORATORIUM IBI DARMAJAYA BERBASIS CLIENT SERVER. Jurnal Management Sistem Informasi dan Teknologi, 73-74.
- Lesmardin. (2015, Agustus 13). Pengertian Aplikasi. Diambil kembali

- dari <https://lesmardin1988.wordpress.com/2014/08/13/pengertian-aplikasi/>
- [Maulana. (2017). KINERJA ALGORITMA LEVENSTHEIN DISTANCE DALAM MENDETEKSI KEMIRIPAN DOKUMEN TEKS. Jurnal Logika, 133.
- Mundzir, A. (2019). Aplikasi Si Apik pada Masyarakat Desa Santanamekar Tasikmalaya. Volume 2(2018): 972- 980;.
- Ray. (2018). Publikasi Ilmiah. Sistem Pembayaran SPP di SD Muhammadiyah 16 Surakarta, 14.
- S, S. (2020, Juli 26). Pengertian Database dan Perangkat Lunak. Diambil kembali dari gurupendidikan.com:<https://www.gurupendidikan.co.id/pengertian-database/>
- S, S. (2020). Pengertian Dokumen dan Jenis-Jenisnya. 5-7.