

**IMPLEMENTASI PENGEMBANGAN APLIKASI BERBASIS
WEB MENGGUNAKAN R SHINY SEBAGAI APLIKASI
TOOL ANALISIS DATA DENGAN METODE ALGORITMA
PCA**



Disusun Oleh:

N a m a : Muhd. Humam Rhamadhani

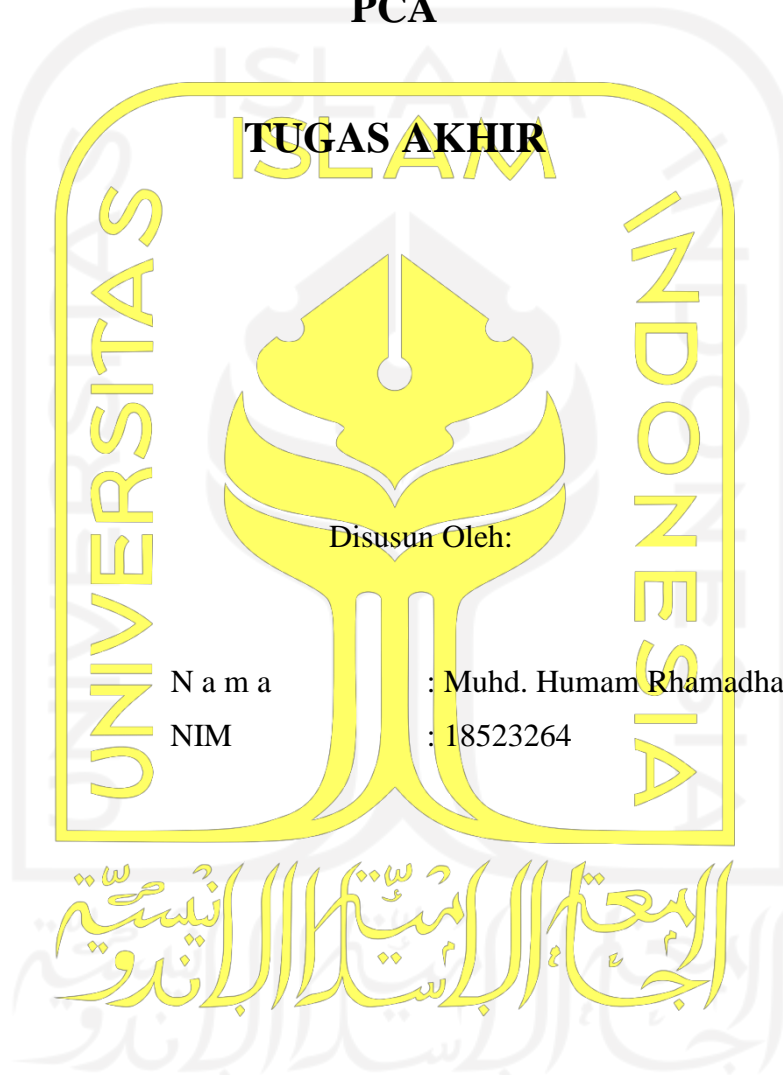
NIM : 18523264

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI UNIVERSITAS ISLAM
INDONESIA**

2022

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**IMPLEMENTASI PENGEMBANGAN APLIKASI BERBASIS
WEB MENGGUNAKAN R SHINY SEBAGAI APLIKASI
TOOL ANALISIS DATA DENGAN METODE ALGORITMA
PCA**



Yogyakarta, 18 Juli 2022

Pembimbing,

(Lizda Iswari, S.T., M.Sc.)

HALAMAN PENGESAHAN DOSEN PENGUJI

**IMPLEMENTASI PENGEMBANGAN APLIKASI BERBASIS
WEB MENGGUNAKAN R SHINY SEBAGAI APLIKASI
TOOL ANALISIS DATA DENGAN METODE ALGORITMA
PCA**

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 1 Agustus 2022

Tim Penguji

Lizda Iswari, S.T., M.Sc.

Anggota 1

Nur Wijyaning Rahayu, S.Kom., M.Cs.

Anggota 2

Chandra Kusuma Dewa, S.Kom., M.Cs.,
Ph.D.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dr. Raden Teduh Dirgahayu, S.T., M.Sc.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Muhd. Humam Rhamadhani

NIM : 18523264

Tugas akhir dengan judul:

**IMPLEMENTASI PENGEMBANGAN APLIKASI BERBASIS
WEB MENGGUNAKAN R SHINY SEBAGAI APLIKASI
TOOL ANALISIS DATA DENGAN METODE ALGORITMA
PCA**

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 18 Juli 2022



(Muhd. Humam Rhamadhani)

HALAMAN PERSEMBAHAN

Dengan menyebut asma Allah. Bismillahirrahmanirrahim, saya persembahkan karya ini untuk,

Kedua orang tua

Terutama kepada Ibu saya yang telah mendukung penuh hasil studi saya dengan doa dan kasih sayangnya, sehingga saya dapat menyelesaikan laporan akhir ini dengan baik walaupun sering terhalang kendala dan cobaan yang sering ditemui dalam proses pengerjaan.

Kakak kandung

Seorang kakak kandung saya, Mbak Tika yang juga mendukung dalam hal materil dan formil dalam mengatasi permasalahan yang saya hadapi.

Dosen Pembimbing

Ibu Lizda Iswari juga menjadi andil yang sangat penting dalam membimbing saya dalam mengerjakan karya berupa laporan akhir penelitian ini.

Dosen Pembimbing Akademik

Pak Yudi Prayudi yang selalu mendukung kegiatan studi saya baik yang berhubungan dengan penelitian maupun tidak.

Akun YouTube “Statkomat”

Yang telah membantu saya terhadap panduan pengembangan aplikasi menggunakan R Shiny.

Sahabat Seperjuangan

Saudara Muhammad Daffa Muafa, yang selalu menghabiskan waktu bersama demi menuntaskan program dan laporan akhir secara bersama-sama.

Dosen Program Studi Informatika FTI UII

Seluruh dosen yang secara tidak langsung terlibat dalam proses pengerjaan laporan ini karena masih terdapat ilmu yang saya terapkan dalam laporan ini.

Mahasiswa Informatika 2018

Dalam peran sebagai teman dan sahabat saya yang juga secara tidak langsung membantu dalam pengerjaan proses laporan akhir untuk penelitian ini.

HALAMAN MOTO

“Fear he, who fears nothing”

(Terrorblade)

“Ilmu tanpa akal ibarat seperti memiliki sepatu tanpa kaki. Dan akal tanpa ilmu ibarat seperti memiliki kaki tanpa sepatu.”

(Ali bin Abi Thalib)



KATA PENGANTAR

Allhamdulillahirobbil'alamin, Puji dan syukur dipanjatkan sebesar-besarnya kepada Allah SWT. yang telah memberi kemudahan dan kelancaran atas segalanya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul “Implementasi Pengembangan Aplikasi Berbasis Web Menggunakan R Shiny Sebagai Analisis Data dengan Metode Algoritma PCA”. Penulis menyadari bahwa penelitian ini belum bisa dikatakan sempurna, oleh karena itu diharapkan dari laporan akhir penelitian ini, topik yang relevan terhadap penelitian ini nantinya dapat dikembangkan lebih lanjut.

Dalam penyusunan laporan akhir ini, tentunya banyak kendala dan hambatan yang dihadapi oleh penulis, namun dari sisi ini penulis menganggapnya sebagai sebuah tantangan sehingga hal itu dapat diselesaikan secara baik dan terorganisir. Oleh karena itu, penulis ingin menyampaikan terimakasih yang sebesar-besarnya kepada :

1. Ibu Lizda Iswari selaku dosen pembimbing saya yang pertama kali mengarahkan untuk topik untuk laporan akhir penelitian ini. Beliau juga memberikan saran, ilmu, dan juga bimbingan yang telah berjasa dalam pembuatan laporan akhir ini.
2. Ibu dan ayah saya tercinta, Umi Kalsum & Suharto yang telah mendukung dan mendoakan saya dalam menyelesaikan laporan akhir ini demi menuntaskan studi saya dengan semangatnya dan dedikasinya yang luar biasa.
3. Teman seperjuangan saya, Muhammad Daffa Muafa yang bekerja sama dengan saya dan membantu dalam proses pengerjaan penelitian ini disebabkan dalam bimbingan yang sama dan juga topik yang relevan.
4. Teman-teman angkatan Informatika 2018, dimana nama-namanya tidak bisa disebutkan satu-satu yang turut memberikan dukungan terhadap penelitian ini.

Sehingga dalam penulisan laporan akhir penelitian ini, penulis berharap topik/tema yang diangkat dapat bermanfaat bagi siapapun termasuk dari diri saya sendiri. Terimakasih.

Yogyakarta, 18 Juli 2022



(Muhd. Humam Rhamadhani)

SARI

Machine learning merupakan salah satu dari cabang disiplin ilmu dari ilmu sains data. PCA merupakan salah satu algoritma *machine learning* yang memiliki teknik multivariat di dalam perhitungannya pada sebaran suatu data. PCA memiliki karakteristik tersendiri untuk proses komputasinya yaitu menggunakan asas reduksi data atau biasa disebut *dimensionality reduction*.

Terbatasnya aplikasi analisis data menggunakan *machine learning* terkhusus penerapan menggunakan algoritma PCA saat ini menjadi alasan penelitian ini dilakukan. Nantinya hasil dari penelitian akan ditujukan menjadi implementasi dari *machine learning* PCA itu sendiri ke dalam sebuah aplikasi berbasis web menggunakan R Shiny. Sehingga pada penelitian ini akan dilakukan proses penerapan algoritma PCA ke dalam sebuah aplikasi yang berbasis web dengan basis R Shiny. Alasan penerapan aplikasi ini menggunakan R Shiny karena adanya *library* yang menyediakan penerapan algoritma PCA di dalamnya sehingga pengembangan tidak dilakukan secara kompleks ditambah Shiny merupakan salah satu *framework* penyedia pengembangan aplikasi web berbasis bahasa R.

Sehingga nantinya dari hasil penelitian ini akan didapatkan bagaimana menerapkan algoritma PCA ini yang akan diimplementasikan ke dalam aplikasi berbasis web sebagai aplikasi *tool* analisis data. Diharapkan hasil dari penelitian ini merupakan aplikasi yang dapat berguna bagi para pengguna yang ingin melakukan maupun mencoba untuk menganalisis sebuah data yang dimilikinya menggunakan *machine learning* terutama algoritma PCA yang ingin mengetahui sebaran dari suatu data dengan algoritma tersebut menjadi informasi yang berguna.

Kata kunci: *machine learning*, sains data, *dimensionality reduction*, PCA, R Shiny, *tool*

GLOSARIUM

<i>Data Science</i>	Disiplin ilmu yang mempelajari data secara ilmiah
<i>Machine Learning</i>	Salah satu cabang ilmu pembelajaran dari <i>data science</i>
R Shiny	<i>Framework</i> berbasis web yang menggunakan operasi bahasa R
PCA	Salah satu algoritma dari <i>machine learning</i>
<i>Dataset</i>	Kumpulan dari sebuah banyak data
<i>Plot</i>	Bentuk visualisasi pada data



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI.....	viii
GLOSARIUM.....	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	2
1.1 Latar Belakang	2
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.4.1 Manfaat bagi pengembang	5
1.4.2 Manfaat bagi pengguna aplikasi	5
1.4.3 Manfaat dalam aplikasi	5
1.4.4 Manfaat bagi penulis	5
1.5 Batasan Masalah	5
1.6 Metodologi Penelitian	6
1.7 Sistematika Penulisan	7
BAB II LANDASAN TEORI	8
2.1 Analisis Data	8
2.2 Sains Data	9
2.3 <i>Machine Learning</i>	9
2.4 Algoritma PCA	11
2.4.1 Tujuan PCA	11
2.4.2 <i>PCA interpreting</i>	12
2.4.3 <i>PCA plotting</i>	14
2.5 R Shiny.....	15
2.6 Pengujian Aplikasi	16
2.6.1 <i>Black Box Testing</i>	16
2.6.2 <i>Software Testing</i>	16
2.7 Tinjauan Pustaka	17
BAB III METODOLOGI PENELITIAN	19
3.1 Subjek & Objek Penelitian.....	19
3.2 Sumber Data.....	19
3.3 Metode Pengumpulan Data.....	19
3.3.1 Studi Literatur.....	19
3.3.2 Peninjauan <i>Dataset</i>	20
3.4 Indikator Keberhasilan	20
3.5 Metode Penelitian	20
3.5.1 Analisis Kebutuhan	21
3.5.2 Rancangan Sistem	22

3.5.3	Implementasi Sistem	22
3.5.4	Hasil Aplikasi	22
3.5.5	Pengujian Hasil.....	23
	BAB IV HASIL DAN PEMBAHASAN	24
4.1	Analisis Kebutuhan	24
4.1.1	Gambaran Umum	24
4.1.2	Kebutuhan Fungsional.....	25
4.1.3	Kebutuhan Non-fungsional	29
4.1.4	Prioritas Kebutuhan.....	30
4.2	Rancangan Sistem	31
4.2.1	<i>Activity Diagram</i>	32
4.2.2	<i>Class Diagram</i>	33
4.2.3	Rancangan Antarmuka (<i>Mockup</i>).....	33
4.3	Implementasi Hasil	38
4.3.1	Inisiasi Aplikasi.....	38
4.3.2	Library R	40
4.3.3	Implementasi Kode Program.....	40
4.4	Hasil Aplikasi.....	42
4.4.1	Halaman utama (<i>Dataset</i>).....	43
4.4.2	<i>Pre-Processing</i>	43
4.4.3	PCA	44
4.4.4	<i>Data Report</i>	47
4.5	Skenario Pengujian Hasil Aplikasi	47
4.5.1	Tahap Demonstrasi.....	47
4.5.2	Rangkuman Hasil	49
4.6	Hasil Pengujian	50
4.6.1	<i>Black Box Testing</i>	51
4.6.2	<i>Software Testing</i>	53
4.7	Pembahasan.....	56
4.7.1	Data PCA.....	56
4.7.2	Komputasi PCA.....	57
4.7.3	Visualisasi PCA.....	57
4.7.4	<i>Data Report PCA</i>	57
	BAB V KESIMPULAN DAN SARAN.....	59
5.1	Kesimpulan	59
5.2	Saran.....	59
	DAFTAR PUSTAKA	60
	LAMPIRAN	63

DAFTAR TABEL

Tabel 2.1 Perbandingan algoritma <i>supervised</i> dan <i>unsupervised</i>	10
Tabel 2.2 Tabel analisis tinjauan pustaka	17
Tabel 4.1 Kebutuhan non-fungsional	30
Tabel 4.2 Tabel prioritas kebutuhan aplikasi	30
Tabel 4.3 Hasil pengujian <i>black box</i>	51
Tabel 4.4 Hasil pengujian <i>software</i>	56



DAFTAR GAMBAR

Gambar 1.1 Contoh <i>Principal Component</i> (Komponen Utama)	3
Gambar 2.1 Proses Analisis Data Penelitian Kualitatif	9
Gambar 2.2 Skema Metode <i>Machine Learning</i>	10
Gambar 2.3 Tujuan dari PCA	12
Gambar 2.4 Tampilan RStudio	15
Gambar 2.5 Tampilan R Shiny dalam RStudio	15
Gambar 2.6 Representasi dari <i>black box testing</i>	16
Gambar 3.1 Alur metode penelitian	21
Gambar 4.1 <i>Use case diagram</i> aplikasi	25
Gambar 4.2 Aplikasi draw.io	31
Gambar 4.3 <i>Activity diagram</i> aplikasi	32
Gambar 4.4 <i>Class diagram</i> aplikasi	33
Gambar 4.5 Rancangan antarmuka halaman “Dataset”	34
Gambar 4.6 Rancangan antarmuka halaman “Pre-Processing”	34
Gambar 4.7 Rancangan antarmuka halaman “Data PCA”	35
Gambar 4.8 Rancangan antarmuka “Komputasi PCA”	36
Gambar 4.9 Rancangan antarmuka halaman “Visualisasi PCA”	37
Gambar 4.10 Rancangan antarmuka halaman “Data Report”	38
Gambar 4.11 Membuat R Shiny	39
Gambar 4.12 Membuat R Shiny (2)	39
Gambar 4.13 <i>Publish</i> aplikasi ke Shinyapps	39
Gambar 4.14 <i>Library</i> pada aplikasi	40
Gambar 4.15 Kode program untuk <i>dashboard</i> pada UI	41
Gambar 4.16 Kode program untuk <i>dataset</i> pada server	42
Gambar 4.17 Hasil tampilan <i>dataset</i>	43
Gambar 4.18 Hasil tampilan <i>pre-processing</i>	43
Gambar 4.19 Hasil tampilan data PCA	44
Gambar 4.20 Hasil tampilan komputasi PCA	45
Gambar 4.21 Hasil tampilan visualisasi PCA	46
Gambar 4.22 Hasil tampilan aplikasi <i>data report</i>	47
Gambar 4.23 Data <i>pre-processing</i> Premier League table 19-20	48
Gambar 4.24 Data <i>pre-processing</i> COVID-19 di Indonesia	48

Gambar 4.25 Data <i>pre-processing</i> Iris.....	49
Gambar 4.26 Biplot <i>dataset</i> Premier League <i>table</i> 19-20	49
Gambar 4.27 Biplot <i>dataset</i> COVID-19 di Indonesia	50
Gambar 4.28 Biplot <i>dataset</i> Iris.....	50
Gambar 4.29 Penggunaan aplikasi WEKA.....	54
Gambar 4.30 Hasil komputasi PCA pada WEKA	55
Gambar 4.31 Hasil analisis PCA pada aplikasi PCA <i>online</i>	55
Gambar 4.32 Contoh hasil <i>Data Report</i> (1).....	58
Gambar 4.33 Contoh hasil <i>Data Report</i> (2).....	58



BAB I

PENDAHULUAN

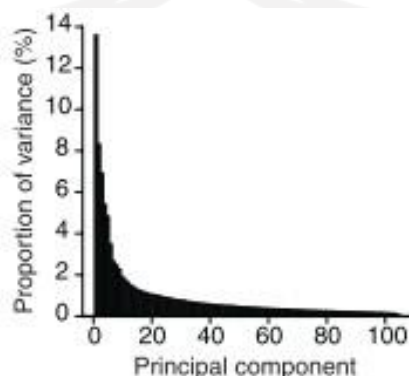
1.1 Latar Belakang

Manusia pada era teknologi industri pada saat ini aktivitasnya sangat bergantung dan membutuhkan data dalam proses kerja kehidupan manusia di dalamnya. Banyak subjek yang membutuhkan data untuk mendapatkan informasi baik dalam skala individu maupun organisasi. Data saat ini menjadi salah satu faktor utama dalam pengambilan keputusan secara cepat (Adhisyanda Aditya et al., 2020). Menurut Nuzulla Agustina, data adalah keterangan mengenai sesuatu hal yang sudah sering terjadi dan berupa himpunan fakta, angka, grafik, tabel, gambar, lambang, kata, huruf-huruf yang menyatakan sesuatu pemikiran, objek, serta kondisi dan situasi. Dengan adanya data, manusia diberikan kemudahan karena data menjadi sebuah fungsi nilai yang merepresentasikan deskripsi dari suatu objek dan kejadian (Irmansyah, 2003). Data-data ini akan berperan penting dan memiliki sebuah nilai yang bersifat faktual dikarenakan ini sudah dikumpulkan sebelumnya dengan berbagai metode yang sudah dilakukan oleh para pencari atau pengumpul data sesuai dengan kejadian, objek, dan peristiwa yang didapatkannya. Ketika sebuah data telah terkumpul dari berbagai macam objek ataupun kejadian yang telah diriset maka akan membentuk sebuah dataset. Dataset inilah yang akan menjadi awal pemrosesan atau cikal bakal dari suatu kesimpulan data yang telah diambil. Munculnya data ke dalam dunia teknologi dan informasi tentunya akan membuat pengolahan/pengumpulan data maupun dataset menjadi berkembang sesuai dengan beriringnya waktu. Salah satu yang menjadi faktor berkembangnya suatu data adalah munculnya sains data, dimana kategori data disini berfungsi kegiatan yang bersifat ilmiah.

Sains data adalah istilah dari gabungan dua kata yaitu “sains” dan “data” dimana dari dua kata tersebut menjelaskan suatu kegiatan ilmiah di sekitar dan berhubungan dengan data, mulai dari pengumpulan, pengolahan, hingga informasi yang dapat berguna nantinya sebagai pengambil keputusan dan dapat berguna bagi pihak yang membutuhkannya dengan data tersebut terkhusus bagi *data scientist* atau *data analyst* (Nasution, 2019). Sains data merupakan disiplin suatu ilmu khusus yang mempelajari suatu data yang dimana data tersebut merupakan data-data yang bersifat kuantitatif atau yang biasa disebut data numerik secara terstruktur maupun tidak terstruktur (Adhisyanda Aditya et al., 2020).

Mereka kemudian dapat menggunakan data-data hasil dari ilmu sains data tersebut menjadi ke dalam sebuah bentuk *machine learning*. Definisi dari *machine learning* sendiri merupakan suatu bidang studi yang bisa dikatakan baru dan bersumber dari algoritma kuantum yang dinamakan *supervised* (terpandu; terstruktur) dan *unsupervised* (mandiri; tidak terstruktur) (Mutiarra & Refianti, 2018). Algoritma *supervised* adalah suatu pencarian algoritma yang bersumber dari contoh yang sudah tersedia atau diambil secara eksternal untuk menghasilkan suatu hipotesa yang mana hipotesis tersebut nantinya dapat membuat prediksi untuk kejadian masa depan yang biasanya akan dianalisis ke dalam bentuk klasifikasi (Kotsiantis, 2007). Sedangkan algoritma *unsupervised* adalah algoritma yang mengurangi jumlah fitur-fitur yang meningkatkan pemahaman dan memperbaiki beberapa masalah yang perlu diselesaikan (Dy & Brodley, 2004).

Salah satu metode algoritma machine learning yang digunakan dalam penelitian ini adalah *Principal Component Analysis* (PCA) dimana ini merupakan salah satu bagian dari algoritma *unsupervised learning*. PCA merupakan algoritma dengan teknik multivariat yang menganalisis sebuah tabel data yang mana nantinya data tersebut akan menghasilkan beberapa sebaran variabel yang bersifat independen dengan sebaran data lainnya yang saling berkorelasi (Abdi & Williams, 2010). Secara garis besar cara kerja algoritma ini adalah mempertahankan sebagian besar variasi dari dalam kumpulan data. Hal inilah yang menjadi kunci utama dari profil algoritma PCA yaitu membuat pengurangan sebagai identifikasi arah yang disebut sebagai komponen utama (M Ringnér, 2008) seperti yang terlihat pada Gambar 1.1.



Gambar 1.1 Contoh *Principal Component* (Komponen Utama)

Sumber : (M Ringnér, 2008)

PCA kembali banyak digunakan dalam banyak penelitian ilmiah yang beragam dari berbagai bidang sesuai dengan tujuan dari penelitian yang sebanding dengan tujuan algoritma PCA yaitu sebagai reduksi data. Hasil dari algoritma ini antara lain menghasilkan

kebanyakan data tereduksi dari kebanyakan penelitian sains yang dilakukan (Wold et al., 1987).

Dengan adanya PCA ini, penelitian ini akan mencoba membangun sebuah sistem aplikasi berbasis web dengan sebuah *tool* yang digunakan yaitu R Shiny. R Shiny merupakan aplikasi web *development* dengan basis bahasa pemrograman R dimana bahasa pemrograman yang sesuai untuk menganalisis dan memproses sebuah dataset dengan algoritma *machine learning* yang akan digunakan.

Penelitian ini akan mengarah dalam pengembangan aplikasi sebagai alat atau *tool* analisis data yang dimiliki pengguna nantinya. Terlebih aplikasi ini akan mengkombinasikan analisis data dari pengguna aplikasi dengan menggunakan salah satu teori dari *machine learning* yaitu PCA dimana dari aplikasi *tool* yang dikembangkan dengan R Shiny yang disediakan oleh bahasa pemrograman R. Lalu dari hasil analisis tersebut, pengguna nantinya akan mendapatkan *insight* hasil dan juga proses dari penerapan algoritma PCA di dalam aplikasi *tool* tersebut. Nantinya sistem dari aplikasi tersebut dapat memvisualisasikan dan memperhitungkan dari hasil model menggunakan metode algoritma PCA dari dataset yang diinput oleh pengguna aplikasi. Alasan penggunaan R Shiny sebagai subjek dari pengembangan aplikasi yang dibuat karena R Shiny merupakan sebuah *library* yang disediakan oleh R Studio dimana teknologi ini sangat interaktif terhadap pengembangan web yang akan mengimplementasikan *machine learning* di dalamnya (Sievert, 2020).

Dari seluruh rincian yang sudah dijelaskan, penelitian ini akan berfokus untuk membangun sebuah aplikasi web untuk salah satu *tool* analisis data dengan algoritma PCA. Harapannya aplikasi dapat digunakan pengguna yang berasal dari kalangan apapun mulai dari yang masih belum mengetahui apa itu sains data hingga profesional seperti *data scientist*, *data analyst*, dan sejenis pekerjaan yang mengolah data. Aplikasi ini layak digunakan para pengguna ke dalam bidang seperti bisnis, sosial, manufaktur, dan lainnya dalam upaya mengetahui penyebaran *instance* dari suatu data yang diinput seperti apa dan dapat menghasilkan suatu nilai ataupun kesimpulan yang didapatkan dengan menggunakan perhitungan metode algoritma PCA ini.

1.2 Rumusan Masalah

Rumusan masalah berdasarkan latar belakang yang sudah dipaparkan pada penelitian ini adalah bagaimana mengembangkan aplikasi analisis data dengan metode algoritma PCA menggunakan R Shiny ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengembangkan aplikasi analisis data dan juga siap digunakan untuk menganalisis data menggunakan algoritma PCA.

1.4 Manfaat Penelitian

1.4.1 Manfaat bagi pengembang

- a. Mengetahui bagaimana mengolah sebuah data dalam basis *machine learning* menggunakan algoritma PCA sebagai bentuk analisis data.
- b. Mendapatkan pengetahuan baru membuat aplikasi berbasis web menggunakan R Shiny.

1.4.2 Manfaat bagi pengguna aplikasi

- a. Mendapatkan *insight* dari penggunaan aplikasi dengan R Shiny
- b. Mengetahui perhitungan menggunakan analisis algoritma PCA.
- c. Memperoleh bentuk grafik dari hasil analisis data algoritma PCA.

1.4.3 Manfaat dalam aplikasi

- a. Dapat digunakan untuk mereduksi data yang berdimensi besar.
- b. Dapat digunakan untuk mengetahui dengan lebih etil hasil pengolahan data dalam bentuk informasi secara statistik maupun secara visual.
- c. Dapat digunakan untuk menganalisis profil data.

1.4.4 Manfaat bagi penulis

Mengetahui sebagai sumber dari referensi maupun untuk pengembangan selanjutnya dengan kaitan dari penelitian ini menggunakan metode algoritma PCA.

1.5 Batasan Masalah

Batasan masalah yang akan menjadi fokus pada aplikasi analisis data menggunakan R Shiny dengan PCA akan dipaparkan sebagai berikut :

- a. Aplikasi ini dapat menghasilkan sebuah data dari hasil analisis menggunakan algoritma PCA.
- b. Bentuk hasil analisis yang digunakan pada aplikasi ini merupakan perhitungan algoritma PCA dan bentuk grafik *plot*.
- c. Aplikasi ini dirancang sebagai aplikasi yang berbasis web menggunakan R Shiny.

- d. Hasil akhir dari aplikasi ini adalah menggambarkan fungsionalitas aplikasi sebagai aplikasi analisis data.

1.6 Metodologi Penelitian

Pada penelitian ini, metodologi yang akan dilakukan untuk mencapai tujuan dari penelitian ini yang akan dilakukan sebagai berikut :

- a. Analisis Kebutuhan

Analisis kebutuhan menjadi tahap awal pada metode penelitian ini. Pada tahap ini akan dilakukan analisa terhadap informasi-informasi yang akan dibutuhkan dalam mengembangkan aplikasi dimana nantinya aplikasi dapat berjalan sebagaimana mestinya sesuai dengan kebutuhan dari penelitian ini. Kebutuhan untuk penelitian ini akan didasari dengan materi dari landasan teori.

- b. Rancangan Sistem

Selanjutnya setelah melakukan analisis kebutuhan, maka hasil informasi dari analisis tersebut akan menjadi rancangan dari pengembangan sistem yang ingin dibangun. Rancangan akan didasari dengan membuat *mockup* untuk sederhana untuk tampilan aplikasi dan juga diagram-diagram yang akan digunakan seperti *activity* dan *class* diagram. Pasca rancangan desain dan sistem, maka akan dimulai perencanaan pengembangan *framework* Shiny yang dimulai dari “UI” sebagai *frontend* dan “Server” sebagai *backend*.

- c. Implementasi Sistem

Selanjutnya hasil dari rancangan sistem aplikasi yang sudah dilakukan akan diimplementasikan. Disini sistem akan mulai diimplementasikan dari rancangan aplikasi dengan kode-kode yang akan digunakan untuk pengembangan aplikasi menggunakan algoritma PCA seperti *dataset*, perhitungan algoritma, hingga sampai ke visualisasi plot.

- d. Hasil Aplikasi

Tahap ini merupakan tahap setelah implementasi terhadap kode-kode program yang telah dibuat. Disini menjadi *core point* untuk tampilan aplikasi sebagai bentuk hasil dari aplikasi dengan fungsi yang telah dibuat pada tahap implementasi.

- e. Pengujian Hasil

Langkah akhir pada metode ini merupakan pengujian dari aplikasi. Pengujian dilakukan sebagai tolak ukur dari tujuan penelitian ini apakah aplikasi dapat berjalan sesuai atau tidak.

1.7 Sistematika Penulisan

Penelitian ini ditulis dengan terstruktur agar pembaca dapat memahami secara mudah apa saja yang ada di dalam tiap bab untuk pembahasannya. Sistematika dari penulisan penelitian ini terbagi menjadi lima bab, yaitu :

BAB I PENDAHULUAN

Bab ini berisi tentang pembahasan masalah umum yang diangkat dari subjek penelitian menjadi sebuah objek penelitian. Pada bab ini ditulis beberapa subbab meliputi latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tinjauan pustaka dan teori dasar yang diambil dari berbagai sumber yang relevan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang langkah-langkah yang dilakukan dalam penelitian mulai dari tahap awal inisiasi aplikasi hingga ke tahap finalisasi pada aplikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dari seluruh proses awal hingga akhir dari metodologi penelitian beserta dengan pembahasannya.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi mengenai kesimpulan dari seluruh hasil tahap penelitian yang telah ditulis dan juga saran untuk referensi pada penelitian selanjutnya yang didapatkan dari hasil penelitian ini.

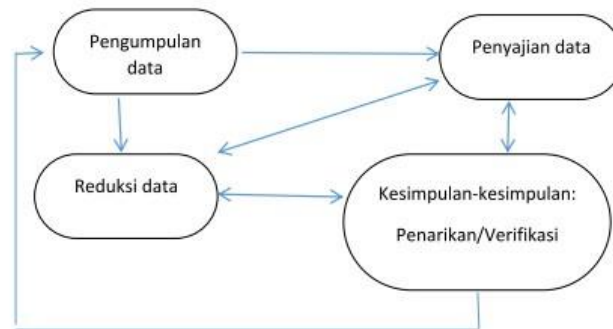
BAB II LANDASAN TEORI

2.1 Analisis Data

Data adalah sebuah satuan terkecil dari sebuah fakta dasar yang tidak bisa dipisahkan dari dunia realitas kehidupan manusia (Nasution, 2002). Dari sebuah konsep data ini akan terus berkembang yang selanjutnya akan menjadi sebuah informasi. Sehingga dari bentuk “data” hingga menjadi “informasi” ini perlu dilakukannya sebuah teknik yang dinamakan analisa atau analisis.

Analisis pada sebuah data menurut pendapat dari Noeng Muhadjir adalah suatu proses dalam mencari dan menata sistematis catatan hasil observasi, wawancara, dan segala teknik lainnya dalam upaya meningkatkan pemahaman dari suatu peneliti tentang kasus atau masalah yang diteliti dan disajikan sebagai suatu penemuan yang berguna bagi orang lain (Rijali, 2018). Kegiatan ini tidak langsung instan dilakukan oleh peneliti melainkan juga memiliki proses di dalamnya. Namun di dalam proses analisis data, terdapat dua jenis sebuah perbedaan dimana perbedaan disebut dibedakan dari jenis kategori data, kuantitatif dan kualitatif.

Proses analisis data penelitian kuantitatif diawali dengan perumusan masalah, perancangan hipotesa, instrumen pengumpulan data, lalu pengumpulan data, hingga dilakukan analisis data yang kemudian menjadi rangkuman dari akhir proses analisis penelitian data ini. Pada proses ini harus dilakukan secara linier, lalu untuk proses analisis data penelitian kualitatif berfokus pada konseptualisasi, kategorisasi, dan deskripsi yang langsung dikembangkan atas dasar sebuah kejadian atau peristiwa yang sedang berlangsung sehingga tidak memungkinkan pengumpulan dan analisis data penelitian kualitatif ini dipisahkan satu sama lain seperti yang ditunjukkan pada Gambar 2.1 (Rijali, 2018).



Gambar 2.1 Proses Analisis Data Penelitian Kualitatif

Sumber : (Rijali, 2018)

2.2 Sains Data

Suatu ringkasan pengetahuan dari data yang memiliki nilai/volume tinggi dimana dalam proses ilmunya menggunakan ilmu komputasi, statistik, dan juga domain spesialis dari para ahli merupakan definisi dari sains data (Nasution, 2019). Secara konseptual, sains data merupakan salah satu disiplin ilmu yang mempelajari dan memproses data. Perkembangan sains data sampai saat ini selalu berkembang menjadi cabang dari ilmu sains data sendiri seperti *internet of things* (IoT), *big data*, *deep learning*, dan juga *machine learning*.

Subjek dalam sains data sangat beragam yang dimana meliputi seluruh aktivitas dari proses data dimulai dari awal pengumpulan hingga menjadi suatu informasi yang didapatkan sesuai dari kesimpulan analisa yang telah dilakukan (Adhisyanda Aditya et al., 2020). Sains data memiliki tahap atau *staging* dalam proses implementasinya. Terdapat 5 tahap *staging* yang diimplementasikan oleh sains data, yaitu :

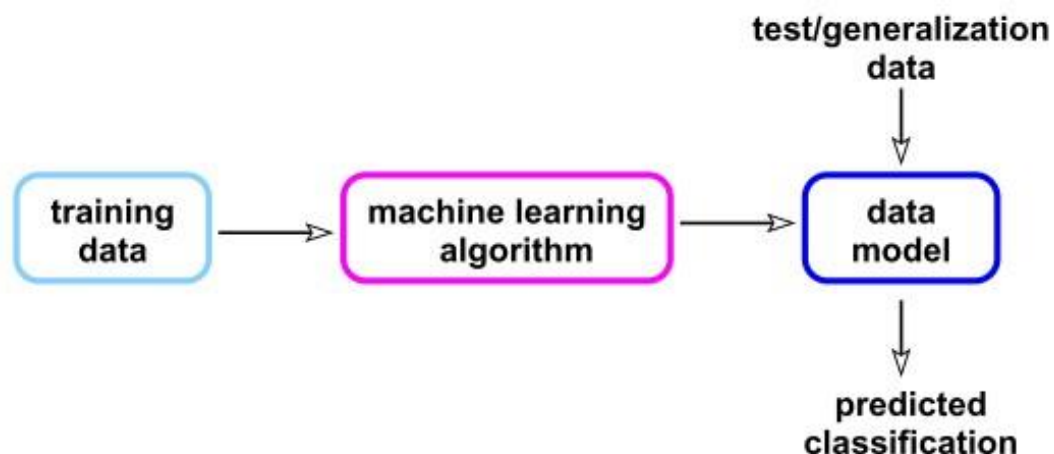
- a. *Capture* : awal pengumpulan data
- b. *Maintain* : tahap arsitektur data (pengolahan data)
- c. *Process* : *modelling* dan *summarizing* (pemrosesan data)
- d. *Analyze* : analisa data
- e. *Communicate* : pengambilan keputusan/kesimpulan

2.3 Machine Learning

Machine learning atau biasa disebut dengan pembelajaran mesin merupakan salah satu disiplin ilmu yang mempunyai fokus pada dua pertanyaan yang saling terkait. Pertama, Bagaimana seseorang dapat membangun sebuah sistem dengan basis komputer yang secara otomatis selalu meningkat karena pengalaman yang dilalui. Lalu dasar statistik, teori,

informasi, dan segala aspek yang berhubungan dalam mengatur semua *learning system* termasuk komputer, manusia, dan perkumpulan atau organisasi (Jordan & Mitchell, 2015).

Sesuai dengan pembahasan di atas, *machine learning* mempelajari algoritma yang meningkat seiring dengan pengalaman yang akan terus digunakannya (Mitchell, 1997). Skema dari metode *machine learning* untuk memahami analisa dari pengalamannya sendiri secara sederhana dimulai dari data yang didapatkan akan di-*training* (dilatih). Kemudian hasil dari *training data* yang sudah dilakukan sekali maupun berkali-kali akan menghasilkan sebuah algoritma *machine learning*. *Machine learning* ini akan menjadi sebuah metode baru untuk memasukkan sebuah data *testing/generalization* yang dimana nanti akan menghasilkan sebuah *data model*. Dari *data model* inilah akan menghasilkan sebuah *value* atau kesimpulan yang didapatkan seperti contoh pada Gambar 2.2 yang menghasilkan nilai klasifikasi atau prediksi dari hasil algoritma *machine learning*.



Gambar 2.2 Skema Metode *Machine Learning*

Sumber : (Mitchell, 1997)

Algoritma *machine learning* saat ini terbagi menjadi dua yaitu *supervised* dan *unsupervised*. *Supervised learning* merupakan sebuah metode pembelajaran yang menyimpulkan fungsi dari serangkaian contoh dari *training* sedangkan *unsupervised* merupakan metode pembelajaran yang menemukan struktur tersembunyi dalam data yang tidak berlabel (Lloyd et al., 2013). Berikut dari Tabel 2.1 akan disajikan perbandingan dari algoritma *supervised* dan *unsupervised* (Donalek, 2011).

Tabel 2.1 Perbandingan algoritma *supervised* dan *unsupervised*

Aspek	Supervised	Unsupervised
Teknik	<i>Classification</i>	<i>Clustering</i>

(kebanyakan)		
<i>Label Data</i>	Harus digeneralisasi terlebih dahulu	Dapat dilabelkan jika label hanya tersedia untuk jumlah objek yang mewakili kelas diinginkan
<i>Training Data</i>	Input dan hasil sesuai dengan yang diinginkan	Tidak diberikan hasil yang benar
Hasil proses	Probabilitas (persentase)	Akurat (mutlak)

Sekarang, sudah banyak fungsi dari *machine learning* yang telah digunakan dalam berbagai dunia. Beberapa bidang yang konsentrasi yang menggunakan metode ini antara lain pengembangan *artificial intelligence* (AI), *natural language processing* (NLP) baik teks atau suara, *computer security*, *software engineering*, dan yang lainnya.

2.4 Algoritma PCA

PCA merupakan salah satu cabang algoritma *machine learning* yang masuk ke dalam kelas *unsupervised* (Ghojogh & Crowley, 2019). Algoritma ini merupakan salah satu algoritma yang cukup menarik karena algoritma *unsupervised learning* kebanyakan mengambil teori klusterisasi atau *clustering* sedangkan pada algoritma PCA ini mengambil teknik dengan asas “*dimensionality reduction*” atau yang biasa dikenal dengan teknik pengurangan dimensi pada suatu data. Data yang diproses ini tidak akan mengurangi karakteristik yang cukup signifikan dan juga mempertahankan informasi yang masih sesuai di dalam datanya.

Sehingga ini merupakan sebuah pilihan yang tepat bahwa algoritma PCA termasuk ke dalam algoritma *unsupervised* sesuai dengan dasar dari teknik yang dimilikinya (Boutsidis et al., 2008).

2.4.1 Tujuan PCA

Untuk mencapai tujuan dari proses kerja algoritma ini, PCA menghitung variabel baru yang disebut komponen utama yang diperoleh sebagai kombinasi linier dari variabel asli (Abdi & Williams, 2010). Banyak tujuan PCA yang bersangkutan dengan menemukan hubungan antara objek komponen pada data. Sehingga pengguna algoritma mungkin nantinya tertarik misalnya dalam menemukan pada kelas pada objek komponen yang serupa. Komponen pada kelas mungkin sudah diketahui sebelumnya, tetapi juga dapat ditemukan dengan eksplorasi menggunakan algoritma ini data yang tersedia (Wold et al., 1987).

GOALS OF PCA



Gambar 2.3 Tujuan dari PCA

Sumber : (Wold et al., 1987)

2.4.2 PCA interpreting

Interpreting pada PCA ini akan mewakili dari hasil proses komputasi pada algoritma PCA itu bagaimana dan seperti apa. Berikut akan dijelaskan tiap-tiap penafsirannya untuk poin.

a. *Eigenvalue*

Untuk yang pertama, dari hasil *interpreting* pada suatu hasil pemrosesan data menggunakan algoritma salah satunya adalah nilai. Nilai yang dihasilkan pada algoritma PCA disebut *eigenvalue*. *Eigenvalue* atau nilai eigen merupakan suatu nilai hasil analisis faktor yang berupa persentase varians dengan penjelasan oleh masing-masing *principal component* (komponen utama). Nilai ini menggambarkan jumlah varians yang dipertahankan oleh masing-masing komponen utama (*principal component*) yang telah diukur.

Eigenvalue adalah nilai-nilai faktor variabel yang dipertahankan dalam bentuk dimensi yang dijelaskan oleh masing-masing *principal component* dari *dataset*. Nilai *eigenvalue* akan menghasilkan suatu varians kumulatif dari *principal component* nantinya didapatkan dari persamaan (4.1).

$$P_k = \left(\sum_{i=1}^p v_i^2 \right) \quad (2.1)$$

PC_k menggambarkan nilai kumulatif dari *principal component* sedangkan PC_i menggambarkan nilai tiap *instance* dari *principal component*. Lalu pada nilai p menggambarkan dimensi fitur pada *dataset*.

b. *Coordinates Component*

Setelah mendapatkan *eigenvalue*, maka akan dihasilkan beberapa komponen. Komponen pertama yang akan menjadi hasil dasar komponen utama/*principal component* pada algoritma PCA yaitu *coordinates component*. *Coordinates* disini

akan mewakili nilai koordinat pada tiap variabel pada komponen PCA. Lalu, nilai ini akan menjadi komputasi awal selanjutnya untuk komponen-komponen selanjutnya pada algoritma ini (Kassambara, 2018).

Coordinates merupakan nilai acuan yang menjadi nilai acuan sebaran sesuai dengan titik asal dari *principal component*.

c. *Squared Cosine (Cos²) Component*

Squared cosine component menunjukkan pentingnya komponen untuk pengamatan yang diberikan. Pada bagian ini menunjukkan kontribusi suatu komponen terhadap kuadrat jarak pengamatan ke titik asal (*coordinates*) (Abdi & Williams, 2010).

Square cosine merepresentasikan jarak kuadrat observasi variabel dari kurva kosinus ke dalam *coordinates* pada PC. Dari sinilah asal muasal adanya *squared* pada fungsinya karena menggunakan nilai kuadrat seperti yang ditunjukkan pada persamaan (4.2).

$$v_{i1}^2 \cos^2 \theta_{i1} = v_{i1}^2 \frac{(\sum_{j=1}^p x_{ij} - \bar{x}_j)^2}{\sum_{j=1}^p (x_{ij} - \bar{x}_j)^2} \quad (2.2)$$

d. *Contribution Component*

Sedangkan pada *contribution component* merupakan sebuah komponen yang diperoleh dari perbandingan nilai komponen *squared cosine* dari suatu pengamatan data dengan *eigenvalue* yang berkaitan dengan komponen tersebut (Abdi & Williams, 2010). Dari hasil perbandingan ini menjadi sebuah nilai pada komponen *contribution*.

Contribution adalah persentase nilai yang membandingkan dari variabel *squared cosine* dengan nilai *eigenvalue*. Nilai *eigenvalue* dalam perhitungan ini bisa juga mewakili nilai seluruh dari *squared cosine* seperti yang ditunjukkan pada persamaan (4.3).

$$\text{Contribution (\%)} = \frac{v_{i1}^2 \cos^2 \theta_{i1}}{\sum_{i=1}^n v_{i1}^2 \cos^2 \theta_{i1}} \times 100 \quad (4.3)$$

e. *Summary PCA*

Summary pada PCA ini merupakan sebuah ringkasan nilai yang didapatkan setelah perhitungan PCA ke dalam bentuk dimensinya. Salah satu hasil perhitungan yang di dapat di bagian ini adalah *standard deviation*, *proportion of variance*, dan *cumulative proportion*.

Summary pada bagian ini sama seperti pada data mendeskripsikan rangkuman nilai statistik berupa standar deviasi, proporsi tiap varians, dan proporsi kumulatif untuk masing-masing *principal component*.

f. *Model PCA*

Sedangkan pada *model PCA* disini akan menjadi salah satu *output* dari sebuah hasil analisis suatu data dari hasil perhitungan algoritma PCA. *Model PCA* akan menjelaskan nilai yang bersifat sebuah *data frame* yang merupakan nilai pada *instance* pada data yang telah distandarisasi ke dalam dimensi komponen utama atau *principal component*.

Model PCA merupakan sebaran nilai data yang telah distandarisasi dari hasil *modeling dataset* sebelumnya pada tiap dimensi PC.

g. *Predict PCA*

Predict PCA menjadi hasil *output* selanjutnya setelah *model PCA* untuk komputasi/perhitungan pada dataset yang digunakan dalam hasil data analisis dari PCA. Pada *output* ini merupakan perhitungan yang didapatkan dari nilai *model PCA* yang menjelaskan untuk sebaran nilai pada data baru (karena bersifat prediksi) yang diambil dari data tiap dimensi *principal component* juga.

Setelah mendapatkan data standarisasi hasil *modeling* maka bisa dihasilkan *output* data sebelumnya yang berupa *predict PCA* dimana nilai ini merupakan sebaran data baru hasil prediksi yang telah dianalisis dari hasil tiap dimensi PC.

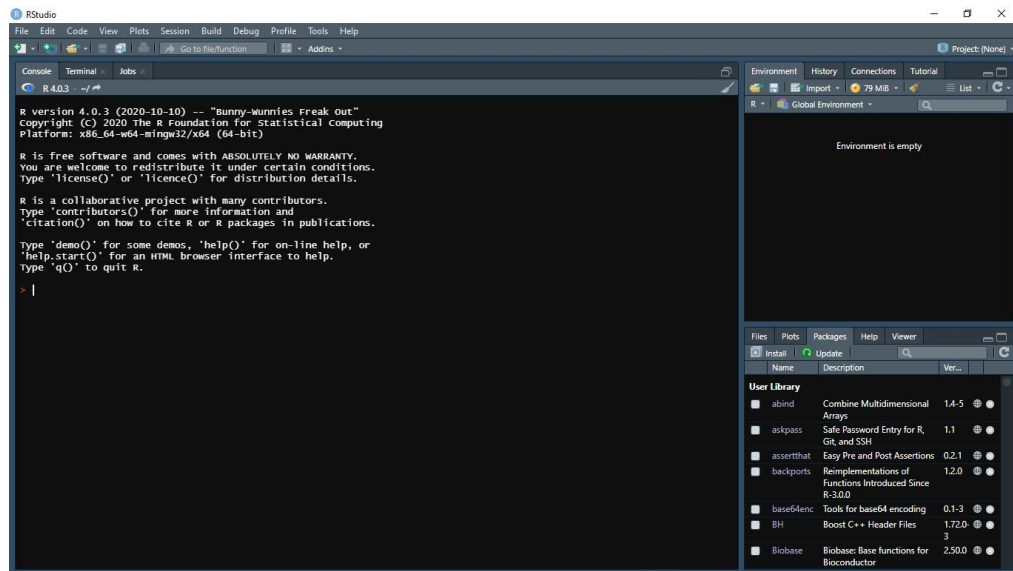
2.4.3 PCA plotting

PCA *plotting* merupakan gambaran *plot* yang akan menjadi bentuk visualisasi dari perhitungan algoritma PCA. *Plot* disini akan mewakili setiap nilai-nilai yang telah dihasilkan pada interpretasi algoritma PCA. Terdapat 5 *plot* disini yang akan digunakan sesuai dengan hasil dari interpretasi setelah perhitungan PCA dilakukan, yaitu :

1. *Scree Plot*
2. *Cos² Plot*
3. *Individuals Plot*
4. *Contrib Plot*
5. *Variables Plot*
6. *Biplot*

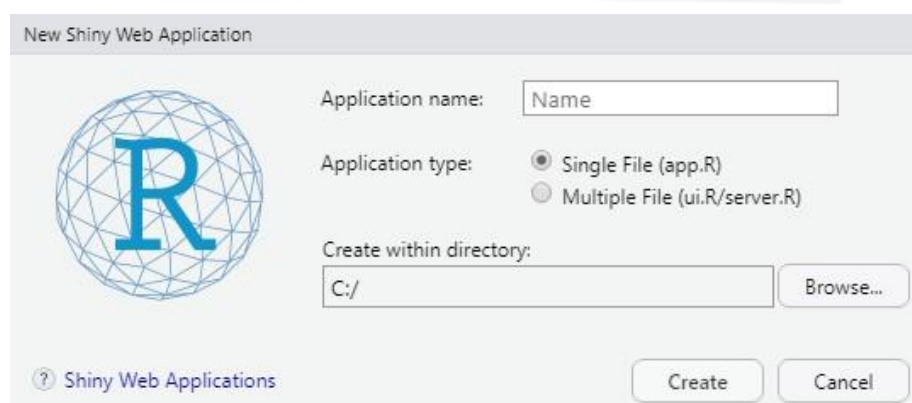
2.5 R Shiny

Pengembangan aplikasi ini untuk penelitian ini menggunakan sebuah *tool* yang disebut R Shiny. Shiny merupakan suatu framework aplikasi pengembangan web yang mana kebutuhan *environment* pada aplikasi tersebut menggunakan bahasa pemrograman R yang dikembangkan oleh RStudio (Potter et al., 2016).



Gambar 2.4 Tampilan RStudio

Shiny merupakan aplikasi yang mudah dan reliabel penggunaannya sebagai pengembangan web karena langsung terintegrasi dengan bahasa R. Dengan aplikasi ini, sistem yang dibangun diharapkan akan menjadi lebih dinamis dan mudah digunakan ke depannya bagi para *user*. Penggunaan R Shiny pada R hanya perlu memanggil suatu *library* yang disebut “**Shiny**”. Pada Gambar 2.4 akan ditunjukkan tampilan pada R Shiny untuk awal inisiasi pembuatannya.



Gambar 2.5 Tampilan R Shiny dalam RStudio

Dengan Shiny, aplikasi ini dapat membangun aplikasi data analisis yang interaktif, dinamis, ramah pengguna, menarik secara visual, dan, dengan fungsionalitas yang mirip

dengan applet Java/JavaScript satu-satunya persyaratan adalah beberapa keakraban dalam R (Potter et al., 2016).

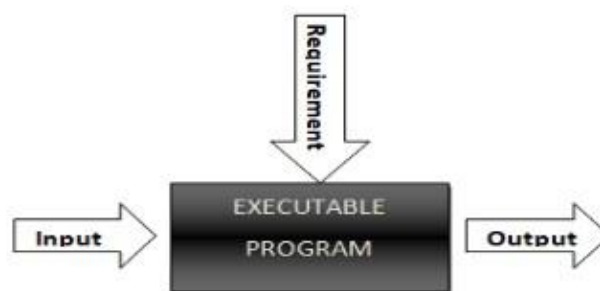
2.6 Pengujian Aplikasi

Pengujian aplikasi merupakan sebagai pembuktian metode untuk tindakan/peristiwa yang diminta oleh aplikasi itu sendiri oleh pengguna (Lucca & Penta, 2003). Ada beberapa pengujian yang akan dilakukan pada aplikasi ini. Hal ini dilakukan dalam upaya untuk menunjukkan kualitas dari aplikasi ini apakah berjalan sesuai yang diharapkan atau tidak. Pengujian yang akan dilakukan sebagai berikut.

2.6.1 *Black Box Testing*

Black box testing merupakan salah satu bentuk metode pada *software/system testing* yang paling sering digunakan. *Testing* ini didasarkan pada kebutuhan dan spesifikasi dari aplikasi. Ini merupakan teknik pengujian yang dimana cara kerja internal dari item yang diuji tidak diketahui oleh penguji (Verma et al., 2017).

Pada *testing* ini tidak perlu dilakukan pengecekan secara internal. Maksud dari internal adalah pengecekan langsung ke dalam bentuk *code* aplikasinya. Seperti yang ditunjukkan pada Gambar 2.6, *testing* ini hanya untuk melakukan pengujian apakah input dari output sudah sesuai dengan hasil yang diharapkan dan kebutuhan pada aplikasi.



Gambar 2.6 Representasi dari *black box testing*

Sumber : (Verma et al., 2017)

2.6.2 *Software Testing*

Pengujian ini merupakan pengujian yang melibatkan aktivitas apa pun yang ditujukan untuk mengevaluasi atribut atau kemampuan program atau sistem dan menentukan bahwa itu memenuhi hasil yang diperlukan (Pan, 1999). Fokus pada pengujian ini tidak hanya pada proses *debugging/running* program aplikasi tetapi juga terdapat validasi dan verifikasi terhadap aplikasi.

2.7 Tinjauan Pustaka

Tinjauan pustaka ini akan mengangkat penelitian yang relevan (*similar*) dengan penelitian ini. Fokus pada tinjauan pustaka yang diangkat sesuai dengan penelitian ini adalah algoritma PCA sebagai konteks utamanya terlebih jika terdapat topik yang mengangkat juga implementasi PCA ke dalam aplikasi.

Penyajian kesimpulan tinjauan pustaka untuk penelitian ini akan dirincikan menggunakan tabel analisis seperti yang terlihat pada Tabel 2.2 untuk memudahkan pembaca untuk melihat hasil laporan penelitian ini.

Tabel 2.2 Tabel analisis tinjauan pustaka

Literatur	Judul	Studi Kasus	Sitasi
1	Turning Big data into tiny data: Constant-size coresets for k-means, PCA and projective clustering	Mengubah data besar menjadi reduksi ke dalam bentuk data kecil	(Feldman et al., 2013)
2	Power quality disturbance monitoring and classification based on improved PCA and convolution neural network for wind-grid distribution systems	Pemantauan dan klasifikasi gangguan kualitas daya	(Shen et al., 2019)
3	Penerapan PCA dalam algoritma K-Means untuk Menentukan Centroid pada <i>Data Clustering</i>	Menentukan Centroid pada <i>Data Clustering</i>	(Tadi, 2018)
4	PCA untuk Dimensi Reduksi <i>Data Clustering</i> Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia	Pemetaan Persentase Sertifikasi Guru di Indonesia	(Rahayu, 2017)
5	Convex sparse PCA for unsupervised feature learning	<i>Convex sparse PCA</i>	(Chang et al., 2016)

Tinjauan yang didapatkan dari sesuai judul dari beberapa kajian pustaka diatas adalah :

- a. PCA merupakan salah satu cabang *machine learning* yang berada dalam kategori *unsupervised learning*.
- b. Pada penelitian-penelitian sebelumnya menggunakan algoritma PCA sebagai analisis/*mapping* dengan memiliki studi kasus penelitian terhadap dataset yang diteliti.
- c. Dua penelitian yang diambil dari untuk sumber pustaka ini dikombinasikan menggunakan *unsupervised machine learning* menggunakan algoritma K-Means.
- d. Fokus pada penggunaan algoritma PCA didasari menggunakan teknik reduksi data sehingga kompleksitas data yang berjumlah besar dapat disimpulkan menjadi bentuk data yang lebih kecil.
- e. Aplikasi analisis data yang menggunakan algoritma PCA sejatinya sudah tersedia sebelum penelitian ini. Namun aplikasi itu sendiri sangat terbatas sebagai subjek penelitian yang ingin dikembangkan, sehingga alasan inilah tinjauan pustaka yang diambil menjadi lebih sulit. Sehingga aplikasi yang sudah ada menjadi subjek pada *software testing* aplikasi ini sebagai bentuk validasi dari perbandingan hasil aplikasi ini dengan aplikasi yang sudah ada daripada diarahkan sebagai tinjauan pustaka.

BAB III

METODOLOGI PENELITIAN

3.1 Subjek & Objek Penelitian

Menurut Darningsih subjek penelitian merupakan suatu hal yang menjadi dasar diteliti baik dari manusia, benda, hingga suatu kumpulan atau organisasi (Darningsih et al., 2016). Subjek penelitian pada penelitian ini akan merujuk pada batasan masalah penelitian dimana untuk subjeknya sendiri berfokus kepada algoritma PCA pada R Shiny. Sedangkan pada objek pada penelitian yang berperan sebagai permasalahan suatu hal yang ingin diteliti (Darningsih et al., 2016). Dalam penelitian ini, objek penelitian akan ditunjukkan ke implementasi algoritma PCA ke dalam aplikasi.

3.2 Sumber Data

Sumber data menjadi dasar metode utama dan pertama untuk penelitian ini. Sumber data yang digunakan pada penelitian yaitu data sekunder. Data sekunder akan didapatkan yang bersumber dari jurnal dan buku yang mendukung untuk topik dari penulisan di penelitian ini dan beberapa *dataset* sebagai subjek *input* pada aplikasi yang telah ditinjau dengan indikasi dari *dataset* tersebut bersifat publik.

3.3 Metode Pengumpulan Data

Pengumpulan data merupakan bagian meliputi seluruh bagian yang perlu untuk menjadikan kelengkapan dari kegiatan analisis data (Rijali, 2008). Pengumpulan data disini akan berdasarkan dari analisis metode yang digunakan pada penelitian ini yaitu analisis data kualitatif. Sehingga pengumpulan data disini akan berkaitan dengan teknik dari subjek penelitian yang diangkat pada penelitian ini. Pada penelitian ini metode pengumpulan data akan berdasarkan metode “kuantitatif” dengan rincian sebagai berikut.

3.3.1 Studi Literatur

Studi literatur atau *literature review* disini akan berperan sebagai landasan subjek penelitian yang memberikan hasil rangkuman dan evaluasi dari kumpulan tulisan tentang topik tertentu (Ridgway et al., 2007). Studi literatur yang digunakan dalam penelitian ini adalah literatur yang berhubungan dengan subjek materi penelitian yaitu algoritma PCA dan implementasi aplikasi web menggunakan R Shiny.

Batasan dalam studi literatur terhadap topik yang diangkat adalah lingkup dari materi yang membahas terkait algoritma PCA yang digunakan dengan tujuan tertentu. Sehingga nantinya ini menjadi dasar pertimbangan terhadap studi literatur ini. Literatur akan dituliskan dalam bahasa Indonesia maupun bahasa Inggris.

3.3.2 Peninjauan *Dataset*

Peninjauan dataset merupakan salah satu langkah pada metode pengumpulan data dalam penelitian ini. Penelitian ini membutuhkan *dataset* karena akan berperan sebagai suatu input pada aplikasi nantinya. Pada tahap pengumpulan ini, peneliti akan berfokus untuk mencari data-data yang dapat diproses menggunakan algoritma PCA. Hal ini dilakukan karena pada algoritma PCA *dataset* yang digunakan tidak bisa dari semua jenis *dataset*, namun data-data yang sudah memiliki karakteristiknya sendiri dengan algoritmanya tersebut.

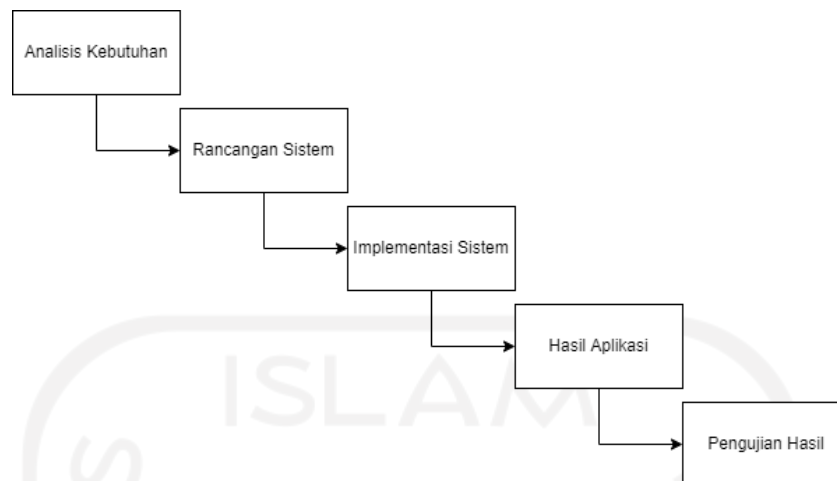
3.4 Indikator Keberhasilan

Penelitian ini membutuhkan indikator keberhasilan sebagai peran dari objek dan tujuan penelitian ini apakah sesuai dengan harapan atau tidak. Indikator keberhasilan untuk penelitian ini yaitu :

1. Aplikasi ini dapat meng-*input* suatu *dataset*.
2. Aplikasi mampu menjabarkan dataset yang telah diinput.
3. Aplikasi bisa memberikan hasil perhitungan dari proses algoritma PCA dari sebuah *dataset*.
4. Aplikasi dapat menyajikan hasil perhitungan menjadi sebuah visualisasi data dalam bentuk *plot*.

3.5 Metode Penelitian

Penelitian ini tentunya akan diharapkan dapat sesuai dengan tujuan yang diinginkan. Oleh sebab itu sebelum tujuan itu tercapai maka terdapat proses-proses yang dilakukan sebelumnya. Proses-proses tersebut adalah sebuah metode penelitian. Metode penelitian ini menerapkan metode *waterfall* di dalamnya dimana langkah-langkah dari metode penelitian akan dilakukan secara sekuensial (berurutan). Sesuai pada Gambar 3.1 yang menjadi gambaran alur dari metode penelitian yang digunakan.



Gambar 3.1 Alur metode penelitian

Penelitian ini diawali dengan menganalisa kebutuhan yang dibutuhkan dalam penelitian terkhusus untuk pengembangan aplikasinya. Setelah itu penelitian diawali dengan perancangan sistem untuk merancang desain dari *interface* aplikasi dan juga diagram-diagram seperti *use case diagram*, *activity diagram*, dan *class diagram*. Dari rancangan yang telah dibuat, kemudian masuk ke langkah implementasi sistem sebagai bentuk hasil aplikasi menggunakan R Shiny dan juga pengimplementasian algoritma PCA. Setelah implementasi maka akan berhasil menjadi sebuah produk aplikasi sesuai kebutuhan. Langkah terakhir dari penelitian ini akan dilakukan pengujian aplikasi sebagai bentuk validasi dan verifikasi terhadap aplikasi apakah hasil perhitungan algoritma PCA sesuai atau tidak.

3.5.1 Analisis Kebutuhan

Analisis kebutuhan menjadi awal dari metode pada penelitian ini. Disini kebutuhan dianalisis menggunakan analisis *system requirements specification* yang memuat *functional requirements* dan *non-functional requirements*.

System requirements specification atau biasa disingkat dengan SRS merupakan salah satu bentuk analisis kebutuhan yang dibutuhkan pada suatu sistem. Menurut Jaroslaw Kuchta, SRS merupakan dasar untuk proyek perangkat lunak sistem aplikasi dalam metodologi perangkat lunak dan aplikasi (Kuchta, 2016). Aplikasi yang dibangun untuk penelitian akan berhubungan juga dengan perangkat lunak *software* seperti R Studio dan lain sebagainya. SRS akan dibagi menjadi dua kebutuhan yaitu kebutuhan sistem fungsional dan non- fungsional serta kebutuhan prioritas pada sistem.

a. *Functional Specification Requirements*

Pada kebutuhan ini akan memuat sebagian besar fungsi utama pada aplikasi dan berfokus terhadap algoritma PCA yang digunakan untuk aplikasi ini. Landasan fitur

fungsionalitas untuk aplikasi ini berlandaskan pada bagian landasan teori sebelumnya seperti *upload* dataset, komputasi PCA (interpretasi), dan juga bentuk visualisasi data menggunakan algoritma PCA.

b. *Non-functional Specification Requirements*

Untuk kebutuhan non-fungsional pada aplikasi ini akan membahas masalah fungsi yang berada di luar fungsi utama pada sistem namun tetap dibutuhkan. Fungsi yang dibutuhkan antara lain UI untuk aplikasi, kebutuhan *software*, kebutuhan *hardware*, dan juga Git untuk hal dalam *versioning* aplikasi.

c. *Prioritization Requirements*

Kebutuhan prioritas menjadi sub bagian terakhir yang digunakan pada *specification requirements system* untuk aplikasi ini. *Prioritization requirements* adalah penomoran (rangking) suatu hal berdasarkan persyaratan dalam urutan tertentu (Yaseen et al., 2019). Hal ini mengacu sebagai penanda dari fitur yang dinilai paling penting untuk di-*develop* diawal dan diakhir pengembangan aplikasi

3.5.2 Rancangan Sistem

Setelah hasil analisis kebutuhan telah terkumpul, maka disini aplikasi sistem mulai dirancang. Perancangan dimulai dengan pemahaman terhadap algoritma PCA dan juga aplikasi R Studio. Setelah itu dilakukan merancang aplikasi *mockup interface* untuk aplikasi dan juga membuat diagram dimana dalam aplikasi ini hanya membutuhkan *activity diagram* dan *class diagram*. Penggunaan *database diagram* seperti ERD dan sejenisnya tidak digunakan karena aplikasi ini tidak membutuhkan sebuah *database* dan juga aplikasi ini dikembangkan sebagai aplikasi web statis.

3.5.3 Implementasi Sistem

Implementasi merupakan tahap penerapan setelah rancangan selesai dibuat. Pada tahap ini berfokus untuk melakukan *development* aplikasi menggunakan R Shiny. *Development* atau pengembangan pada aplikasi ini dilakukan dengan cara memasukkan kode-kode program yang sesuai dengan kebutuhan algoritma PCA dan juga pada R Shiny.

3.5.4 Hasil Aplikasi

Hasil aplikasi menjadi bentuk *output* dari implementasi pada aplikasi penelitian. Tentunya untuk hasil pada aplikasi ini, diharapkan sesuai dengan tujuan pada penelitian ini . Lalu juga pada tahap ini menjadi tolak ukur untuk fungsionalitas pada aplikasi sebelum dilakukan ke tahap pengujian.

3.5.5 Pengujian Hasil

Langkah ini menjadi langkah terakhir pada metode penelitian yang digunakan. Pengujian hasil menjadi keberhasilan aplikasi ini dapat diterima oleh target *user* yang telah ditentukan. Metode pada pengujian hasil akan dilakukan dengan tiga metode, yaitu :

1. *Black Box Testing* : menguji keseluruhan sistem fungsionalitas , non-fungsionalitas, dan prioritas pada aplikasi.
2. *Software Testing* : menguji dengan studi komparasi dengan aplikasi lain sebagai kualifikasi perbandingan hasil dengan menggunakan *software* dari aplikasi WEKA dan juga aplikasi PCA *online* berbasis *web* dengan situs yang sudah tersedia (lucianoabriata.altervista.org/jsinscience/pca/pca5.html).



BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Kebutuhan

Analisis kebutuhan menjadi langkah awal dalam membangun aplikasi dalam penelitian ini. Kebutuhan yang akan digunakan dalam penelitian ini berdasarkan metode dari analisis kebutuhan yang sesuai yaitu menggunakan *system requirements specification* (SRS). SRS pada penelitian ini akan dirincikan ke dalam gambaran umum, kebutuhan fungsional, kebutuhan non-fungsional, dan juga kebutuhan prioritas.

4.1.1 Gambaran Umum

Aplikasi dalam penelitian ini memiliki gambaran umum untuk proses *input* dan hasil *output* untuk aplikasi ini nantinya seperti apa. Berikut perinciannya :

1. Tentang Aplikasi

Aplikasi ini merupakan aplikasi analisis data dimana analisis tersebut berbasis aplikasi web. Aplikasi ini tidak hanya sebatas menganalisa sebuah data begitu saja namun terdapat perhitungan *machine learning* di dalamnya yang menggunakan algoritma *machine learning* PCA.

2. Fungsi Aplikasi

Aplikasi ini secara garis besar memiliki fungsi sebagai analisis terhadap suatu data yang dimiliki oleh *user*. Fungsi analisis ini tidak hanya sekedar menganalisa tetapi terdapat fungsi *output* untuk tutorial penggunaan aplikasi ini dalam bentuk video, fungsi informasi keterangan dataset yang dirangkum dengan nilai statistik, fungsi perhitungan algoritma PCA, fungsi visualisasi *plot* menggunakan algoritma PCA, dan yang terakhir fungsi *reporting* sebagai bentuk hasil rangkuman untuk kesimpulan analisis data pada aplikasi.

3. Pengguna Aplikasi

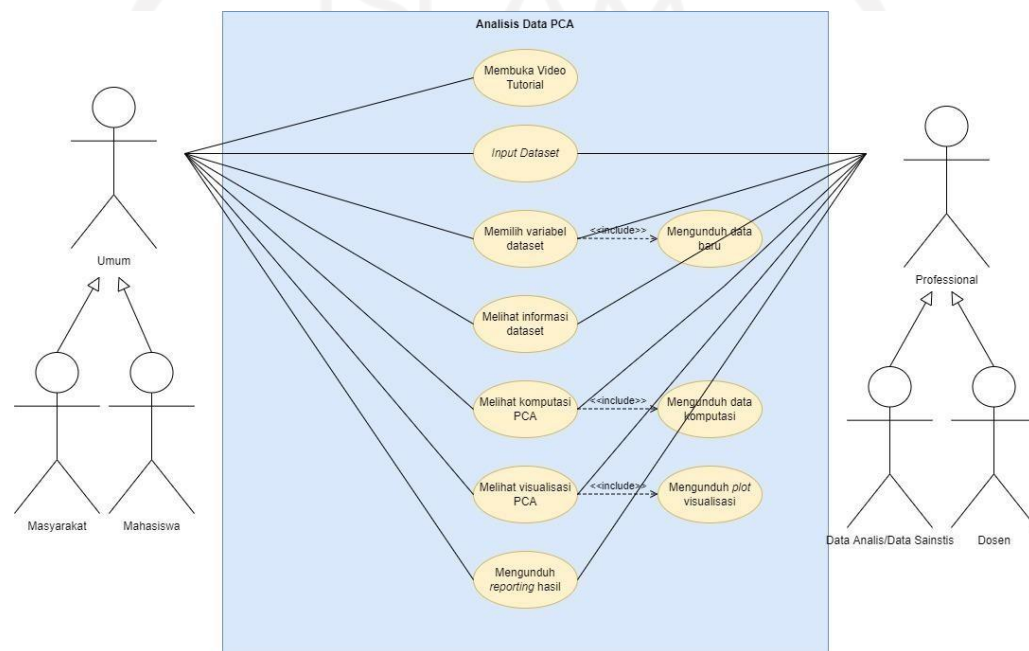
Target pada pengguna untuk aplikasi ini adalah masyarakat umum, mahasiswa, dan profesional untuk mengetahui bagaimana hasil dari analisis pada suatu data menggunakan *machine learning* terkhusus dengan algoritma PCA.

4. Kendali Aplikasi

Seluruh hal di atas yang telah disebutkan, kendali pada aplikasi ini dapat digunakan pada web *browser* aplikasi manapun baik pada *mobile device* atau PC.

4.1.2 Kebutuhan Fungsional

Pada Gambar 4.1 dijelaskan *use case* aplikasi sebagai fitur dan *actor* sebagai pengguna. Terdapat dua pengguna pada aplikasi ini dimana pada masing-masing pengguna memiliki *inheritance actor* yaitu umum dengan masyarakat dan mahasiswa lalu professional dengan data analis/saintis dan dosen. Dalam *use case diagram* tersebut hampir keseluruhan dari dua aktor memiliki *use case* yang sama hanya saja untuk aktor “umum” memiliki *use case* “membuka video tutorial” sebagai video pengantar sebelum menggunakan aplikasi.



Gambar 4.1 *Use case diagram* aplikasi

Kebutuhan fungsional akan dirincikan sesuai dengan *use case* pada Gambar 4.1. Rincian pada kebutuhan ini akan dijelaskan dengan deskripsi, langkah, dan kebutuhan dari sistem aplikasi ini. Berikut rincian kebutuhannya :

1. Membuka video tutorial

a. Deskripsi

Fungsi ini sebagai fitur dari bentuk awal pengantar sebelum menggunakan aplikasi. Sehingga aplikasi dapat terarah dengan jelas dari awal penggunaannya bagi pengguna aplikasi.

b. Langkah

- 1) Pengguna membuka tab “Tutorial”.
- 2) Pengguna membaca gambaran umum aplikasi dan menekan tombol video.
- 3) Pengguna diarahkan ke *link* YouTube sebagai referensi untuk video tutorial.

4) Selesai.

c. Kebutuhan

- 1) Sistem dapat membuka halaman dari tutorial untuk menampilkan gambaran umum aplikasi dan juga video tutorial.
- 2) Sistem dapat mengarahkan ke *link* YouTube sebagai bentuk referensi dari video tersebut.

2. *Input* dataset

a. Deskripsi

Fungsi ini merupakan fitur utama aplikasi dimana dataset merupakan subjek pada aplikasi untuk dianalisis. Setelah dataset ter-*input* maka dataset tersebut akan tampil di dalam aplikasi.

b. Langkah

- 1) Pengguna membuka tab “Dataset”.
- 2) Pengguna memilih dataset yang ingin dianalisis dengan format yang telah ditentukan.
- 3) Dataset yang telah dipilih akan ditampilkan.
- 4) Selesai.

c. Kebutuhan

- 1) Sistem mampu meng-*input* dataset.
- 2) Sistem dapat menampilkan dataset yang telah di-*input*.
- 3) Sistem mampu membaca data dengan separator yang berbeda.
- 4) Sistem mampu membaca data dengan format yang berbeda.

3. Memilih variabel dataset

a. Deskripsi

Fungsi ini adalah sebagai bentuk *pre-processing* sebelum masuk ke perhitungan algoritma PCA. Pada tahap ini merupakan tahap selektif dimana dataset yang telah di-*input* mungkin tidak semua kolom pada dataset ingin dianalisa oleh pengguna nantinya. Selain itu dataset yang telah proses ini dapat diunduh menjadi data baru.

b. Langkah

- 1) Pengguna telah memasukkan dataset sebelumnya.
- 2) Pengguna memilih tab “Pre-Processing”.

- 3) Pengguna memilih kolom yang ingin digunakan untuk tahap selanjutnya yaitu komputasi PCA.
 - 4) Pengguna dapat melihat tampilan dataset yang telah dipilih kolomnya.
 - 5) Selesai.
- c. Kebutuhan
- 1) Sistem mampu memilih kolom dari dataset yang diinginkan.
 - 2) Sistem dapat menampilkan hasil dari proses pemilihan kolom dataset yang telah dipilih.
 - 3) Sistem dapat mengunduh hasil dari proses pemilihan kolom dataset menjadi data baru.
 - 4) Sistem dapat melanjutkan proses ke penerapan algoritma PCA nantinya.
4. Melihat informasi dataset
- a. Deskripsi
- Fungsi ini menggambarkan informasi dataset yang telah di-*input* dataset yang dipilih oleh *user*. Mulai dari tampilan hingga bentuk grafik batang pada dataset akan disajikan pada fungsi sistem ini.
- b. Langkah
- 1) Pengguna telah melakukan proses *input* pada dataset yang dipilihnya.
 - 2) Pengguna juga telah memilih kolom variabel dataset yang telah dipilih.
 - 3) Pengguna melihat hasil dataset yang telah diproses.
 - 4) Pengguna melihat informasi dari dataset yang tersedia.
 - 5) Selesai
- c. Kebutuhan
- 1) Sistem dapat menampilkan informasi dari dataset yang telah di-*input*.
 - 2) Sistem mampu memahami detail dari dataset yang di-*input*.
 - 3) Sistem dapat mengunduh hasil informasi dari dataset yang telah di-*input*.
5. Melihat komputasi PCA
- a. Deskripsi
- Fungsi ini merupakan fungsi utama pada algoritma PCA yang berperan sebagai komputasi penerapan algoritmanya. Disini terdapat beberapa hasil komputasi yang akan disajikan.
- b. Langkah
- 1) Pengguna telah melakukan proses *input* pada dataset yang telah dipilih.

- 2) Pengguna juga telah memilih kolom variabel dataset yang telah dipilih.
- 3) Pengguna memilih tab “Komputasi”.
- 4) Pengguna melihat informasi dari komputasi algoritma PCA yang tersedia pada aplikasi.
- 5) Selesai.

c. Kebutuhan

- 1) Sistem mampu menampilkan hasil komputasi PCA dari dataset yang telah di-input.
- 2) Sistem dapat menghasilkan komputasi yang sesuai dengan *dataset* baru hasil dari kolom yang telah dipilih sebelumnya pada tahap memilih variabel dataset.
- 3) Sistem juga dapat mengunduh hasil data komputasi tersebut menjadi sebuah *file* eksternal.

6. Melihat visualisasi PCA

a. Deskripsi

Fungsi ini merupakan hasil interpretasi dari perhitungan/komputasi algoritma PCA pada *dataset* yang telah dipilih. Penyajian untuk visualisasi data disajikan dalam bentuk *plot*.

b. Langkah

- 1) Pengguna telah melakukan proses input pada dataset yang telah dipilih.
- 2) Pengguna juga telah memilih kolom variabel dataset yang telah dipilih.
- 3) Pengguna memilih tab “Visualisasi”.
- 4) Pengguna melihat informasi dari hasil visualisasi komputasi algoritma PCA dalam bentuk *plot* yang tersedia pada aplikasi.
- 5) Selesai.

c. Kebutuhan

- 1) Sistem mampu menampilkan hasil visualisasi PCA dari dataset yang telah di-input.
- 2) Sistem dapat menghasilkan hasil visualisasi yang sesuai dengan dataset baru hasil dari kolom yang telah dipilih sebelumnya pada tahap memilih variabel dataset.
- 3) Sistem juga mampu menghasilkan visualisasi yang sesuai dan sinkron terhadap hasil pada komputasi algoritma PCA sebelumnya.

- 4) Sistem juga dapat mampu mengunduh hasil dari visualisasi berupa format gambar (png) dalam *file* eksternal.

7. Mengunduh *reporting* hasil

a. Deskripsi

Fungsi ini merupakan bentuk *summary* atau kesimpulan dari keseluruhan hasil proses penerapan algoritma PCA dari awal hingga akhir. Nantinya hasil yang diberikan berupa dokumen yang bisa di-*download* berupa format R Markdown.

b. Langkah

- 1) Pengguna telah melakukan proses input pada dataset yang telah dipilih.
- 2) Pengguna juga telah memilih kolom variabel dataset yang telah dipilih.
- 3) Pengguna telah memahami hasil penerapan dari algoritma PCA dari fungsi awal hingga akhir.
- 4) Pengguna memilih tab “Data Report”.
- 5) Pengguna melihat informasi keseluruhan dari algoritma PCA yang telah dilakukan menggunakan *dataset* yang telah dipilih sebelumnya.
- 6) Selesai.

c. Kebutuhan

- 1) Sistem dapat membuat suatu rangkuman keseluruhan dari penerapan *dataset* yang di-*input*, komputasi PCA, dan visualisasi PCA ke dalam bentuk dokumen *report*.
- 2) Sistem mampu menampilkan bentuk dokumen dalam *format* HTML yang telah disediakan pada R Studio berupa R Markdown.
- 3) Sistem dapat mengunduh hasil *report* dari data hasil penerapan algoritma PCA tersebut.

4.1.3 Kebutuhan Non-fungsional

Kebutuhan non-fungsional pada sistem aplikasi ini berupa fungsi yang di luar dari penerapan algoritma PCA. Namun fungsi tersebut menjadi sebuah kebutuhan yang perlu pada *development* untuk aplikasi dalam penelitian ini.

Terdapat beberapa faktor yang akan menjadi kebutuhan non-fungsional yang digunakan pada aplikasi. Berikut penyajiannya yang akan disajikan dalam bentuk tabel seperti yang terlihat pada Tabel 4.1.

Tabel 4.1 Kebutuhan non-fungsional

No.	Kebutuhan	Deskripsi	Tool
1.	Antarmuka	UI pada tampilan aplikasi menggunakan ui.r sebagai peran <i>frontend</i> pada aplikasi	R Shiny (R Studio)
2.	Server	<i>logic code</i> program pada tampilan aplikasi menggunakan server.r sebagai <i>backend</i> aplikasi	R Shiny (R Studio)
3.	Git	sebagai <i>versioning</i> dan <i>controlling</i> kode program aplikasi	GitHub Desktop
4.	Perangkat lunak	Berperan dalam eksekusi program dalam bentuk yang tidak berwujud	Browser, OS
5.	Perangkat keras	Berperan dalam eksekusi program dalam bentuk fisik	PC

4.1.4 Prioritas Kebutuhan

Setelah menentukan kebutuhan fungsional dan non-fungsional, maka akan dilakukan penentuan kebutuhan sesuai prioritas pada aplikasi. Dalam hal ini yang akan dijadikan prioritas pada kebutuhan ini adalah fitur aplikasi. Penentuan prioritas dipilih berdasarkan pada *developer* aplikasi yang berlandaskan pada teori yang sudah ditentukan. Seluruh prioritas kebutuhan untuk fitur pada aplikasi terdapat pada Tabel 4.2.

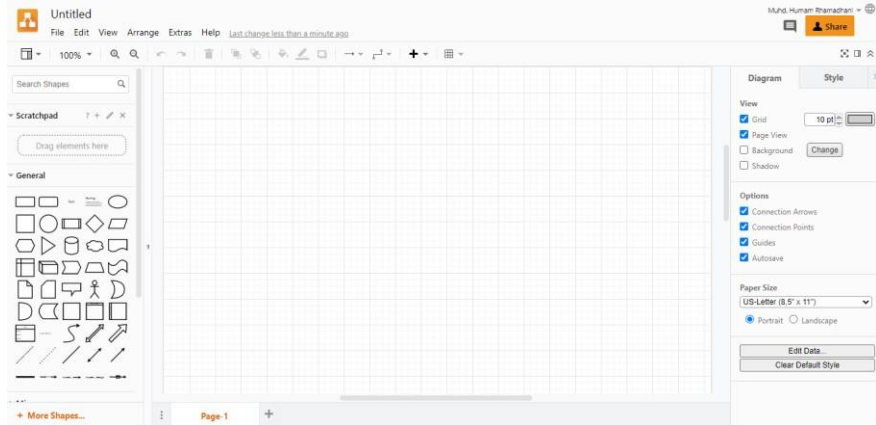
Tabel 4.2 Tabel prioritas kebutuhan aplikasi

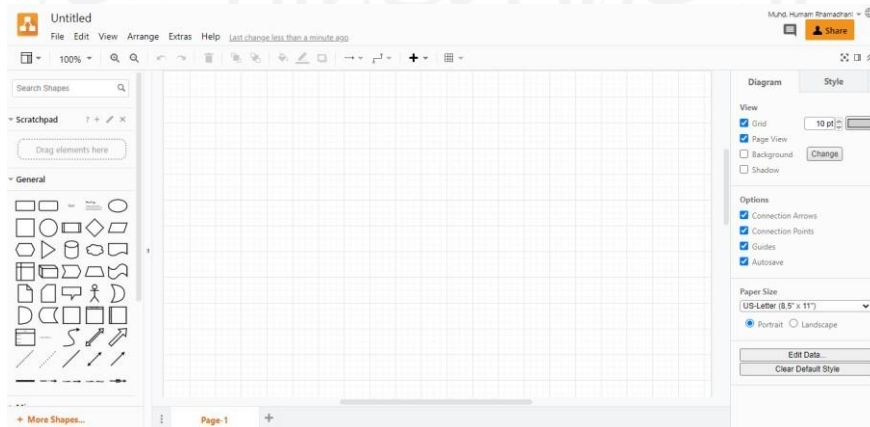
No.	Fitur	Bagian pada Aplikasi	Prioritas
1.	Upload <i>dataset</i>	Dataset	Prioritas 1 (acuan)
2.	Pemilihan variabel	Pre-Processing	Prioritas 2
3.	Tampilan <i>dataset</i>	Dataset, Pre-Processing	Prioritas 4
4.	Profil <i>dataset</i>	Data PCA	Prioritas 18
5.	<i>Summary dataset</i>	Data PCA	Prioritas 19
6.	Histogram <i>dataset</i>	Data PCA	Prioritas 20
7.	<i>Eigenvalue</i>	Komputasi PCA	Prioritas 5
8.	<i>Coordinates PCA</i>	Komputasi PCA	Prioritas 6
9.	<i>Squared Cosine PCA</i>	Komputasi PCA	Prioritas 7
10.	<i>Contribution PCA</i>	Komputasi PCA	Prioritas 8

11.	<i>Summary PCA</i>	Komputasi PCA	Prioritas 9
12.	<i>Model PCA</i>	Komputasi PCA	Prioritas 11
13.	<i>Predict PCA</i>	Komputasi PCA	Prioritas 10
14.	<i>Scree Plot</i>	Visualisasi PCA	Prioritas 12
15.	<i>Squared Cosine Plot</i>	Visualisasi PCA	Prioritas 16
16.	<i>Contribution Plot</i>	Visualisasi PCA	Prioritas 17
17.	<i>Individuals Plot</i>	Visualisasi PCA	Prioritas 13
18.	<i>Variables Plot</i>	Visualisasi PCA	Prioritas 14
19.	<i>Biplot</i>	Visualisasi PCA	Prioritas 15
20.	<i>Data Report</i>	Data Report	Prioritas 3

4.2 Rancangan Sistem

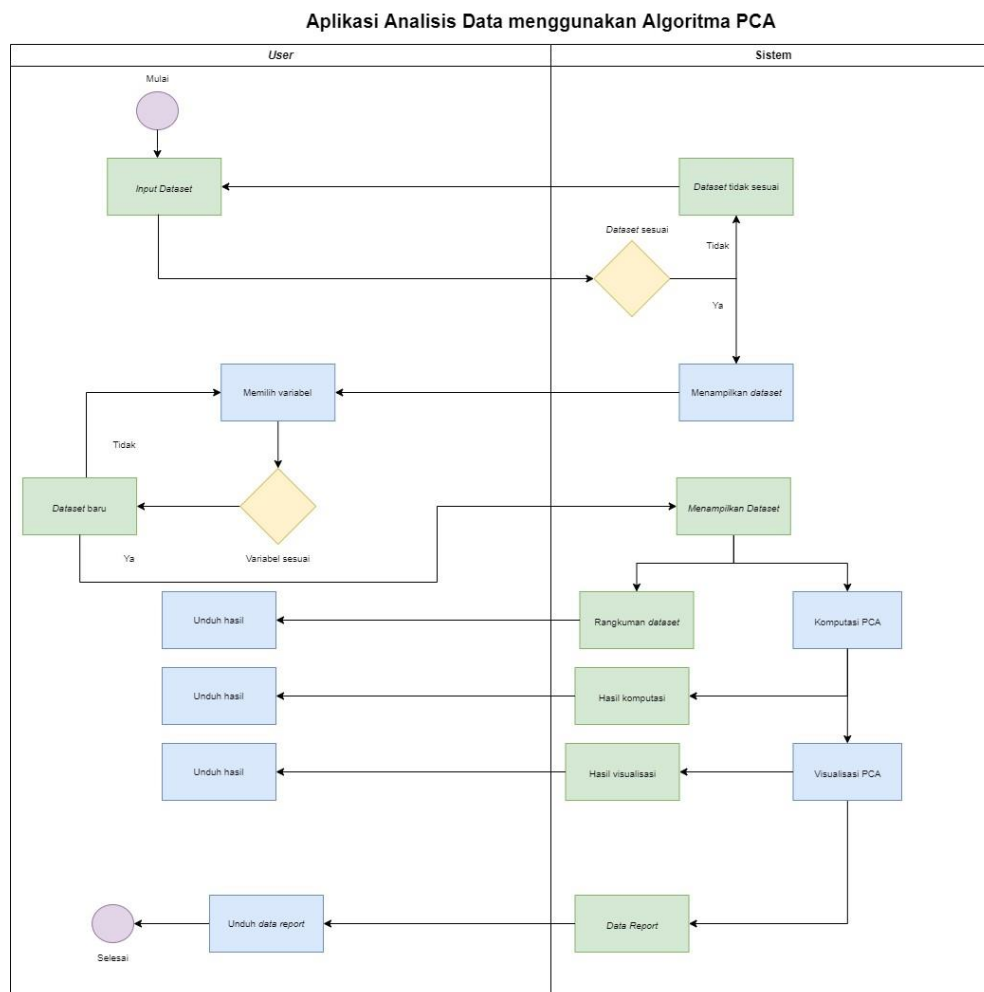
Pada tahap ini, hasil perancangan menjadi dasar sebelum final aplikasi dibuat. Sehingga hal ini menjadikan peneliti atau pengembang aplikasi masih dapat mengeksplorasi hal-hal yang mungkin bisa dijadikan penambahan pada sistem nanti sebelum masuk ke tahap implementasi.

Rancangan sistem aplikasi akan merancang dua diagram dan rancangan antarmuka yang berupa *mockup*. Dua diagram kerangka rancangan sistem akan menggunakan *activity diagram* dan *class diagram*. Untuk *activity diagram* akan menggambarkan bagaimana aktivitas sistem aplikasi bekerja dari awal hingga akhir. Sedangkan pada *class diagram* akan menggambarkan kelas objek dan subjek untuk *field type* yang nantinya akan menjadi rancangan implementasi kode program pada aplikasi. Perancangan dilakukan menggunakan aplikasi berbasis web yaitu **draw.io** yang ditunjukkan seperti pada  dimana dari situs aplikasi tersebut menyediakan *flow diagram maker* dan juga desain untuk *mockup* aplikasi.



Gambar 4.2 Aplikasi draw.io

4.2.1 Activity Diagram



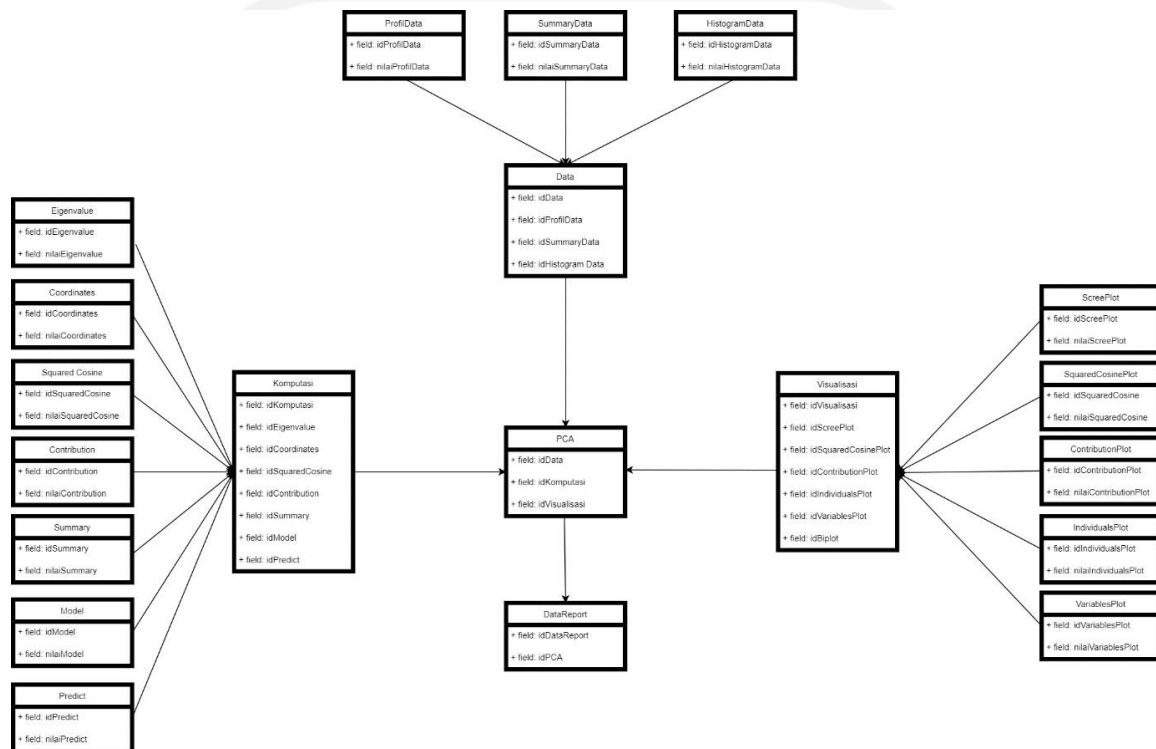
Gambar 4.3 Activity diagram aplikasi

Pada Gambar 4.3 dijelaskan *activity diagram* untuk proses aktivitas aplikasi yang dimana kotak warna biru menggambarkan suatu *input* pada aplikasi sedangkan kotak warna hijau menggambarkan suatu *output* pada aplikasi. *Activity diagram* menggambarkan suatu proses dari awal hingga akhir yang terjadi dalam suatu sistem dalam hal ini sistem merujuk kepada aplikasi. Dalam diagram tersebut terdapat dua *role* yang akan mendukung sistem tersebut berjalan yaitu *role* untuk *user* dan sistem.

Pada *role user* disini terdapat aktivitas yang krusial dan menjadi awal mulai sistem bisa dijalankan yaitu “*input dataset*” dimana ini akan menjadi subjek pada aplikasi yang akan digunakan dalam perhitungan algoritma PCA. Lalu pada langkah terakhir bisa dilihat pada Gambar 4.3 *user* akan mendapatkan “*data report*” dari aktivitas unduh dimana itu langkah akhir pada sistem yang akan merangkum seluruh informasi *dataset* untuk diterapkan pada algoritma PCA yang terdapat pada aplikasi.

Sedangkan pada *role* sistem akan dimulai penentuan keputusan dari *dataset* yang telah di-*input* oleh pengguna aplikasi sesuai atau tidak karena jika tidak aplikasi akan mengalami *catch error* yang akan dibalikkan ke *user* untuk memasukkan data baru. Jika *dataset* telah sesuai maka sistem akan melanjutkan aktivitas/kegiatannya. Pada akhir *flow*, sistem akan mendapatkan hasil dari *data report* untuk bisa diunduh oleh pengguna.

4.2.2 Class Diagram



Gambar 4.4 Class diagram aplikasi

Kelas-kelas pada diagram ini akan mewakili untuk rancangan subjek/objek pada aplikasi untuk pengimplementasian ke sistem. Masing-masing pada kelas memiliki *field* ID sebagai *identifier* kelas dan juga nilai sebagai *value* untuk tiap masing-masing kelas.

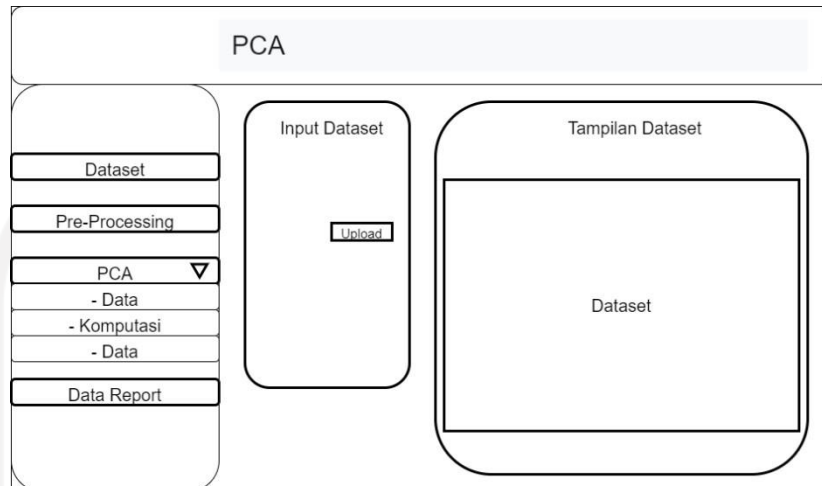
Pada Gambar 4.4 terdapat kelas data memiliki 3 *subclass*, komputasi memiliki 7 *subclass*, dan pada visualisasi memiliki 5 *subclass*. Lalu bisa dilihat kembali bahwa kelas untuk data, komputasi, dan visualisasi akan terhubung ke sistem yang mengarah kepada kelas PCA dimana ini merupakan kelas utama pada sistem aplikasi nantinya untuk menghasilkan *data report*.

4.2.3 Rancangan Antarmuka (Mockup)

Rancangan antarmuka akan dirancang dari dasar *class diagram* yang telah dibuat sebelumnya. Kelas-kelas pada diagram tersebut akan mewakili tiap *mockup session* untuk perancangan antarmuka aplikasi. Rancangan antarmuka juga didasari dengan ketersediaan

pada *layout* yang telah disediakan pada R Shiny. Pengembangan aplikasi ini akan menggunakan tampilan *dashboard* untuk tampilan utama antarmukanya.

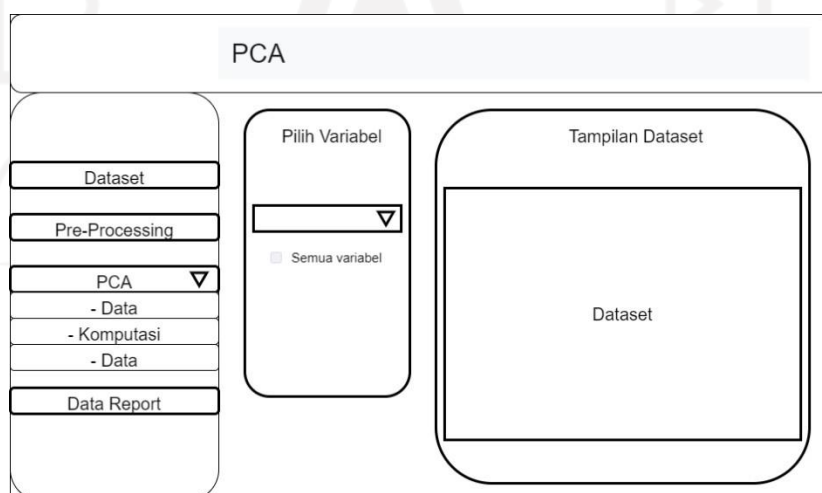
a. *Dataset* (halaman utama)



Gambar 4.5 Rancangan antarmuka halaman “Dataset”

Pada Gambar 4.5 ditampilkan sebuah rancangan antarmuka untuk halaman utama khususnya pada dataset PCA. Tampilan ini menampilkan sebuah *side panel* dan *main panel*. *Side panel* disini berfungsi sebagai *action* untuk *input dataset* oleh pengguna dimana terdapat *button* “upload”. Sedangkan pada *main panel* akan berfungsi sebagai tampilan dataset dari *dataset* yang telah di-*input* dan bentuk dari tampilan *dataset* itu berupa *dataframe*.

b. *Pre-processing* (pemilihan variabel)



Gambar 4.6 Rancangan antarmuka halaman “Pre-Processing”

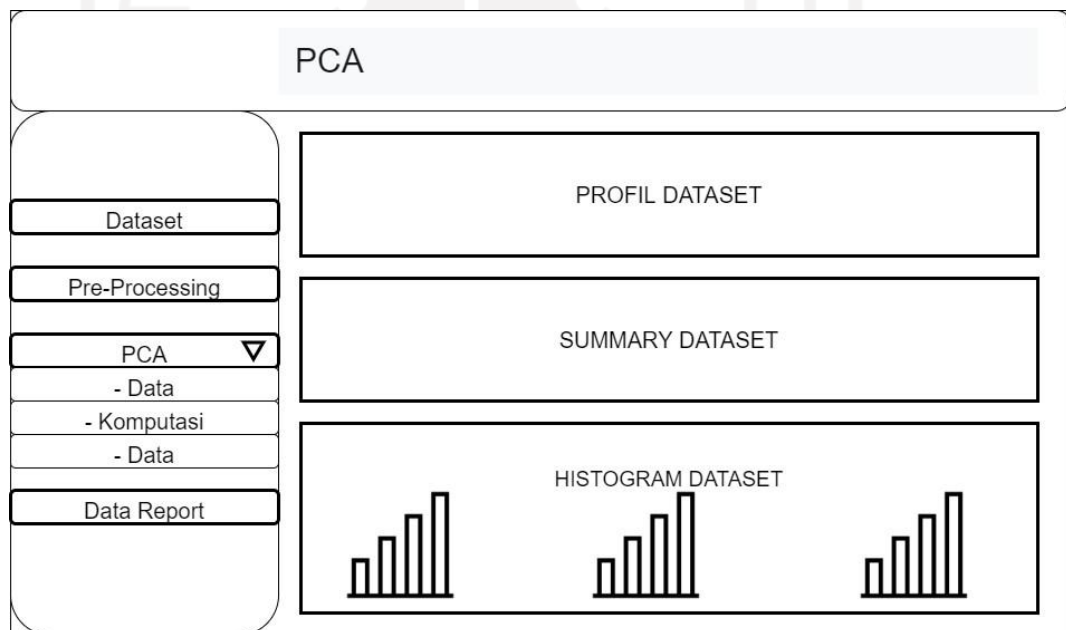
Gambar 4.6 menampilkan rancangan antarmuka untuk halaman *tab* “Pre-Processing”. Bagian ini akan berfokus terhadap pemilihan variabel untuk *dataset* pada proses sebelumnya. Serupa dengan halaman sebelumnya, pada antarmuka

halaman ini hanya memiliki *side panel* dan *main panel*. *Side panel* berfungsi untuk pemilihan variabel untuk kolom yang akan dipilih oleh *user* yang akan digunakan dalam pemrosesan algoritma PCA pada aplikasi. Disini juga tersedia *checkbox button* yang memudahkan *user* jika ingin menggunakan seluruh variabel *dataset*. Lalu pada *main panel* berfungsi sama seperti halaman “Dataset” yang hanya menampilkan *dataset* dalam bentuk *dataframe*.

c. PCA

Pada bagian halaman ini memiliki 3 *submenu* dimana terdapat *tab* untuk menampilkan rangkuman *dataset*, komputasi proses algoritma PCA, dan hasil visualisasi berupa *plotting*. Berikut tampilan dari tiga *tab* bagian tersebut.

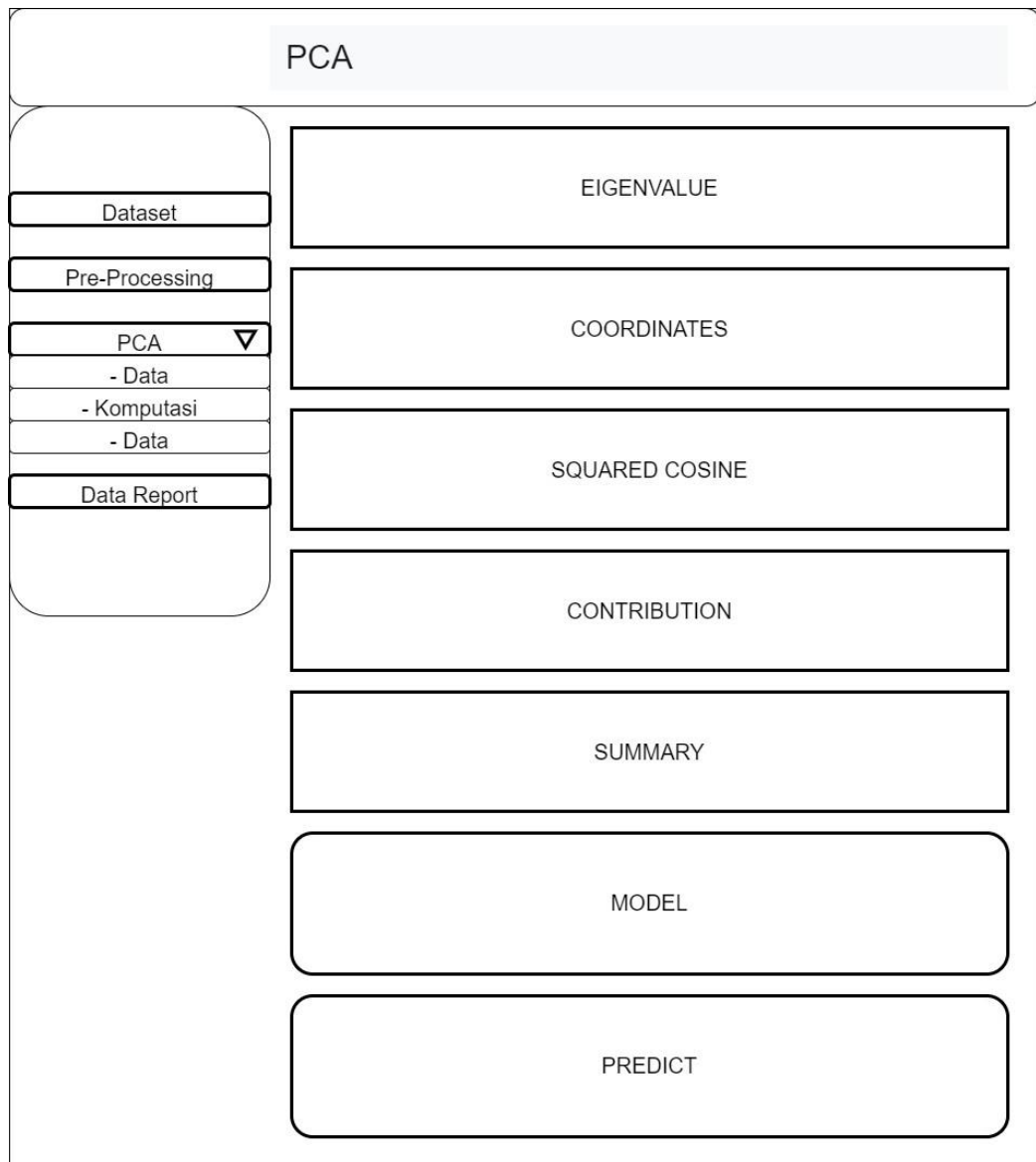
a) Data PCA



Gambar 4.7 Rancangan antarmuka halaman “Data PCA”

Tampilan pada “Data PCA” akan memuat *panel view* yang memiliki tiga *box*. Sesuai pada Gambar 4.7 terdapat *box* untuk profil *dataset*, *summary dataset*, dan histogram *dataset*.

b) Komputasi PCA



Gambar 4.8 Rancangan antarmuka “Komputasi PCA”

Komputasi PCA akan memuat perhitungan proses algoritma PCA. Terdapat lima format *box* dan dua format dalam bentuk *dataframe*. Melihat pada Gambar 4.8, format *box* akan memuat nilai *eigenvalue*, *coordinates*, *squared cosine*, *contribution*, dan *summary* pada PCA. Kemudian untuk *dataframe* memuat nilai *model* dan *predict* PCA.

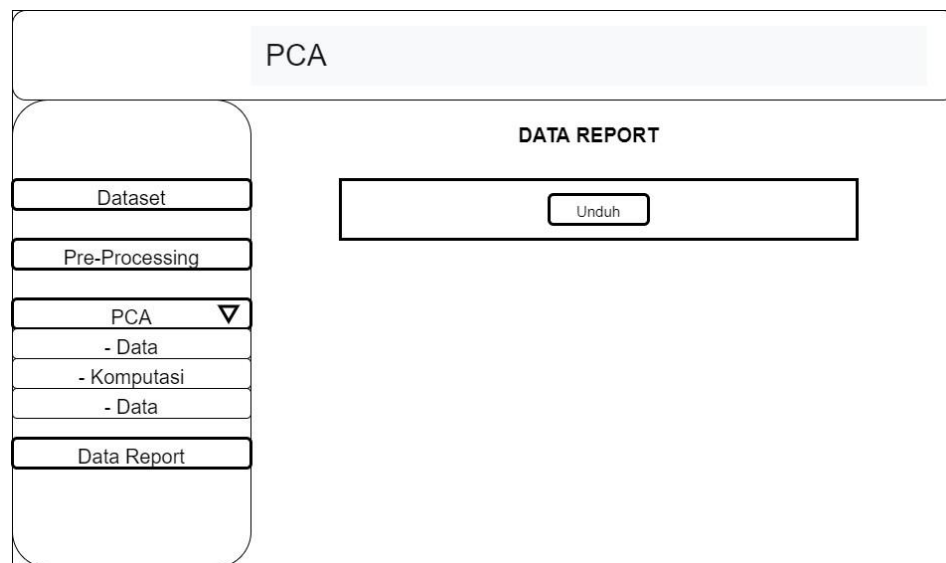
c) Visualisasi PCA



Gambar 4.9 Rancangan antarmuka halaman “Visualisasi PCA”

Terakhir *submenu* untuk PCA adalah visualisasi PCA yang memuat keseluruhan hasil komputasi PCA yang disajikan dalam bentuk *plot*. Format *layout* yang akan ditampilkan pada aplikasi berupa *box* yang di dalamnya terdapat *plot* masing-masing nilai. Terdapat lima *plot* yang akan disajikan yaitu

d. Data report



Gambar 4.10 Rancangan antarmuka halaman “Data Report”

Data report merupakan halaman terakhir pada aplikasi yang akan menampilkan halaman *action* untuk mengunduh *data report* tersebut. *Report* tersebut akan memiliki format HTML yang dimana akan langsung terbuka setelah mengunduhnya. Tampilan halaman ini cukup sederhana seperti yang dilihat pada Gambar 4.10.

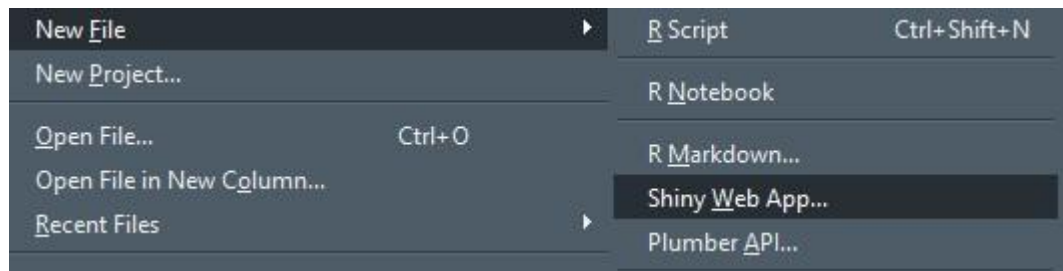
4.3 Implementasi Hasil

Setelah rancangan diagram dan desain kasar antarmuka (*mockup*) telah dilakukan, maka pada tahap implementasi akan memasukkan kode-kode program yang akan menampilkan hasil dari rancangan tersebut. Implementasi akan dilakukan dengan beberapa tahapan. Berikut tahapan-tahapan yang akan dilakukan dalam pengimplementasian sistem aplikasi ini.

4.3.1 Inisiasi Aplikasi

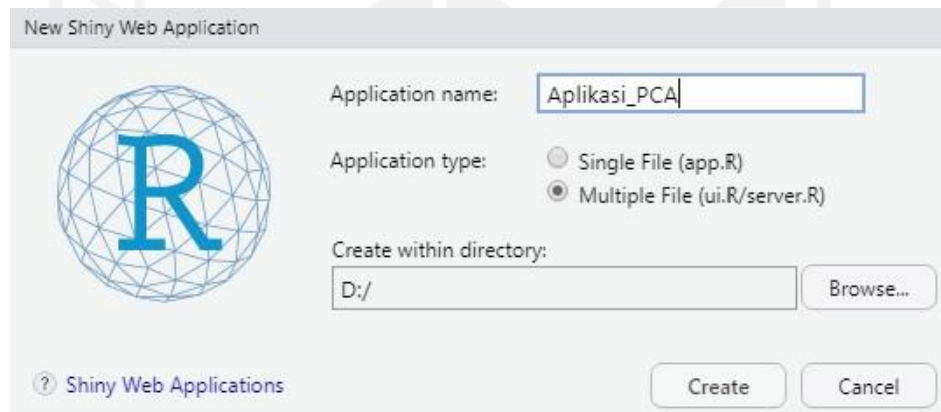
Pada tahap ini akan dilakukan persiapan untuk membuat R Shiny terlebih dahulu menggunakan R Studio. Yang pertama dilakukan adalah membuat *ui.r* sebagai *front-end* aplikasi dan *server.r* sebagai *back-end* aplikasi. Setelah itu dilakukan *publishing* sehingga aplikasi dapat diakses pada web. *Publishing* dilakukan menggunakan *environment* yang sudah disediakan juga oleh R yaitu **Shinyapps**. Tahapan inisiasi akan dilakukan sebagai berikut.

1. Membuka aplikasi RStudio.
2. Membuat *file* baru dengan format Shiny yang dinamakan “Shiny Web App” seperti pada Gambar 4.11.



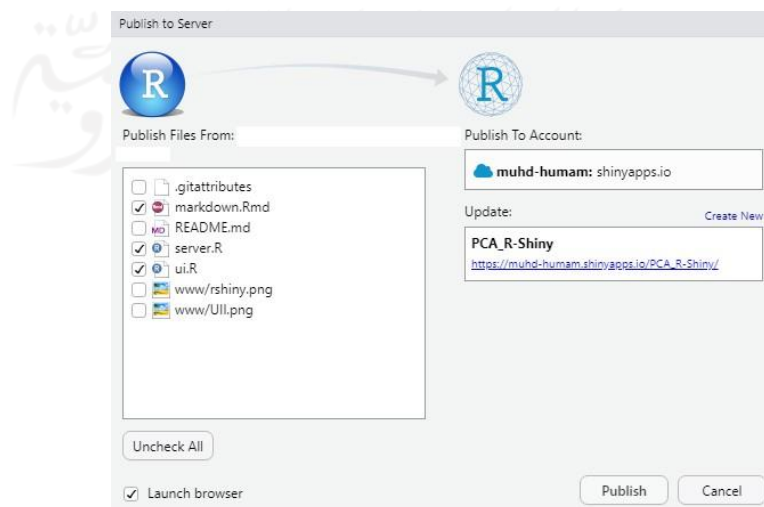
Gambar 4.11 Membuat R Shiny

- Menamakan aplikasi R Shiny tersebut setelah itu memilih tipe aplikasi dimana pada Shiny sudah terdapat *template* untuk *multiple file* dimana sudah dikombinasikan antara ui dan server sebagai peran tampilan dan logika pemrogramannya (*front-end & back-end*) seperti pada terlihat pada Gambar 4.12.



Gambar 4.12 Membuat R Shiny (2)

- Terakhir menentukan direktori *file* tersebut dalam *hardware* (PC).
- Setelah proses di atas selesai, dilanjutkan dengan membuat akun pada Shinyapps.io dan dipublikasikan seperti pada Gambar 4.13.

Gambar 4.13 *Publish* aplikasi ke Shinyapps

4.3.2 Library R

```
library(shiny)
library(shinyAce)
library(shinyWidgets)
library(shinydashboard)
library(shinycssloaders)
library(shinyscreenshot)
library(shinythemes)
library(rsconnect)
library(kableExtra)
library(readr)
library(DT)
library(dplyr)
library(skimr)
library(visdat)
library(factoextra)
library(corrplot)
library(ggplot2)
```

Gambar 4.14 *Library* pada aplikasi

Library disini dimaksudkan merupakan sebuah kumpulan kode-kode program yang memiliki fungsi tertentu dan nantinya dapat dipanggil ke dalam sebuah program untuk menggunakan fungsinya. Dalam hal ini *library* disini akan mengimplementasikan fungsi tersebut ke dalam aplikasi penelitian ini untuk penerapan algoritma PCA. *Library* yang akan digunakan dalam aplikasi penelitian ini terlihat seperti pada Gambar 4.14. Pada implementasi aplikasi ini akan digunakan tiga tema *library*, *library* untuk kebutuhan fungsi Shiny, kebutuhan fungsi *dataset*, dan kebutuhan fungsi algoritma PCA.

4.3.3 Implementasi Kode Program

a. UI

Pada bagian ini akan dijelaskan salah satu implementasi kode program untuk bagian tampilan aplikasi atau *front-end*. Tema *layout* yang akan digunakan pada pengimplementasian ini menggunakan basis tampilan berupa *dashboard*. Gambar 4.15 menunjukkan kode program dari *dashboard* yang akan diimplementasikan pada aplikasi.

```

dashboardPage(
  dashboardHeader(
    title = "PCA", titleWidth = 300
  ), #Header

  dashboardSidebar(width = 300,
    sidebarMenu(
      menuItem("Dataset", tabName = "Dataset"),
      menuItem("Pre-Processing", tabName = "Variabel"),
      menuItem("PCA", dropdownButton = TRUE,
        menuSubItem("Data PCA", tabName = "Data2"),
        menuSubItem("Komputasi PCA", tabName = "Komputasi"),
        menuSubItem("Visualisasi PCA", tabName = "Visualisasi")
      ),
      menuItem("Data Report", tabName = "DataReport"),
    ) #Menu
  ), #Sidebar

  dashboardBody(
    ...
  )
)

```

Gambar 4.15 Kode program untuk *dashboard* pada UI

b. Server

Seerti yang sudah dijelaskan pada bab sebelumnya, server akan berperan sebagai *back-end* untuk logika pemrograman aplikasi yang telah dibuat. Pada aplikasi ini juga server berfungsi sebagai penghubung antara UI dan kode program logika dari algoritma PCA dan yang lainnya. Dalam server, fungsi utama yang menjadi basis keseluruhan fungsi *back-end* pada aplikasi adalah *dataset* yang berupa fungsi *reactive* dimana nantinya *dataset* ini dapat dipanggil ke fungsi yang lain sebagai bentuk korelasi seperti yang terlihat pada Gambar 4.16.

```

server <- function(input, output, session)
{
  data <- reactive(
  {
    upload_data <- input$upload_data

    if(is.null(upload_data))
      return(NULL)

    data = read.csv(upload_data$datapath, sep = input$pemisah_variabel,
                    row.names = 1)
    return(data)
  })
  ...
}

```

Gambar 4.16 Kode program untuk *dataset* pada server

4.4 Hasil Aplikasi

Hasil aplikasi merupakan tahap setelah implementasi dari kode program dilakukan. Pada tahap ini diberikan tampilan dari aplikasi yang sudah berbentuk sempurna. Tampilan dari hasil aplikasi ini akan diperlihatkan berdasarkan halaman yang sesuai dengan rancangan sistem sebelumnya.

Empat halaman akan menjadi fungsi *output* yang penting dalam aplikasi ini karena ini merupakan fungsi dari penerapan algoritma PCA yang saling berkorelasi satu sama lainnya. Empat bagian *output* itu adalah *dataset*, *pre-processing*, PCA, dan *data report*. Hasil implementasi aplikasi ini akan dicoba menggunakan tiga sampel *dataset* yaitu “**Premier League table 19-20**” yang didapatkan dari situs *open-source dataset* yaitu **Kaggle**, “**COVID-19 Indonesia berdasarkan Provinsi**”, dan yang terakhir menggunakan contoh sampel *dataset* “**Iris**” dari *package* R yang telah disediakan (**factoextra**). Rincian dari tampilan hasil tampilan aplikasi akan dirincikan sebagai berikut.

4.4.1 Halaman utama (*Dataset*)

The screenshot shows the main dataset page of the PCA application. The interface includes a sidebar with navigation options like 'Tutorial', 'Dataset', 'Pre-Processing', 'PCA', 'Data Report', and 'Source Code'. The main content area is titled 'Input Data' and features a file upload section for a '.txt/.csv file'. The selected file is 'Premier League 19-20.csv'. Below the upload section, there are links for 'Belum memiliki dataset?', 'Kaggle', 'Data World', and 'Factoextra'. A 'Separator' section allows users to choose between 'Comma', 'Semicolon', and 'Tab'. The 'Tampilan Dataset' section displays a table of 10 football teams with columns for 'PI', 'W', 'D', 'L', 'F', 'A', 'GD', and 'pts'. A search bar and pagination controls are also visible.

	PI	W	D	L	F	A	GD	pts
Liverpool	38	32	3	3	85	33	52	99
Manchester City	38	26	3	9	102	35	67	81
Manchester United	38	18	12	8	66	36	30	66
Chelsea	38	20	6	12	69	54	15	66
Leicester City	38	18	8	12	67	41	26	62
Tottenham Hotspur	38	16	11	11	61	47	14	59
Wolverhampton Wanderers	38	15	14	9	51	40	11	59
Arsenal	38	14	14	10	56	48	8	56
Sheffield United	38	14	12	12	39	39	0	54
Burnley	38	15	9	14	43	50	-7	54

Gambar 4.17 Hasil tampilan *dataset*

4.4.2 *Pre-Processing*

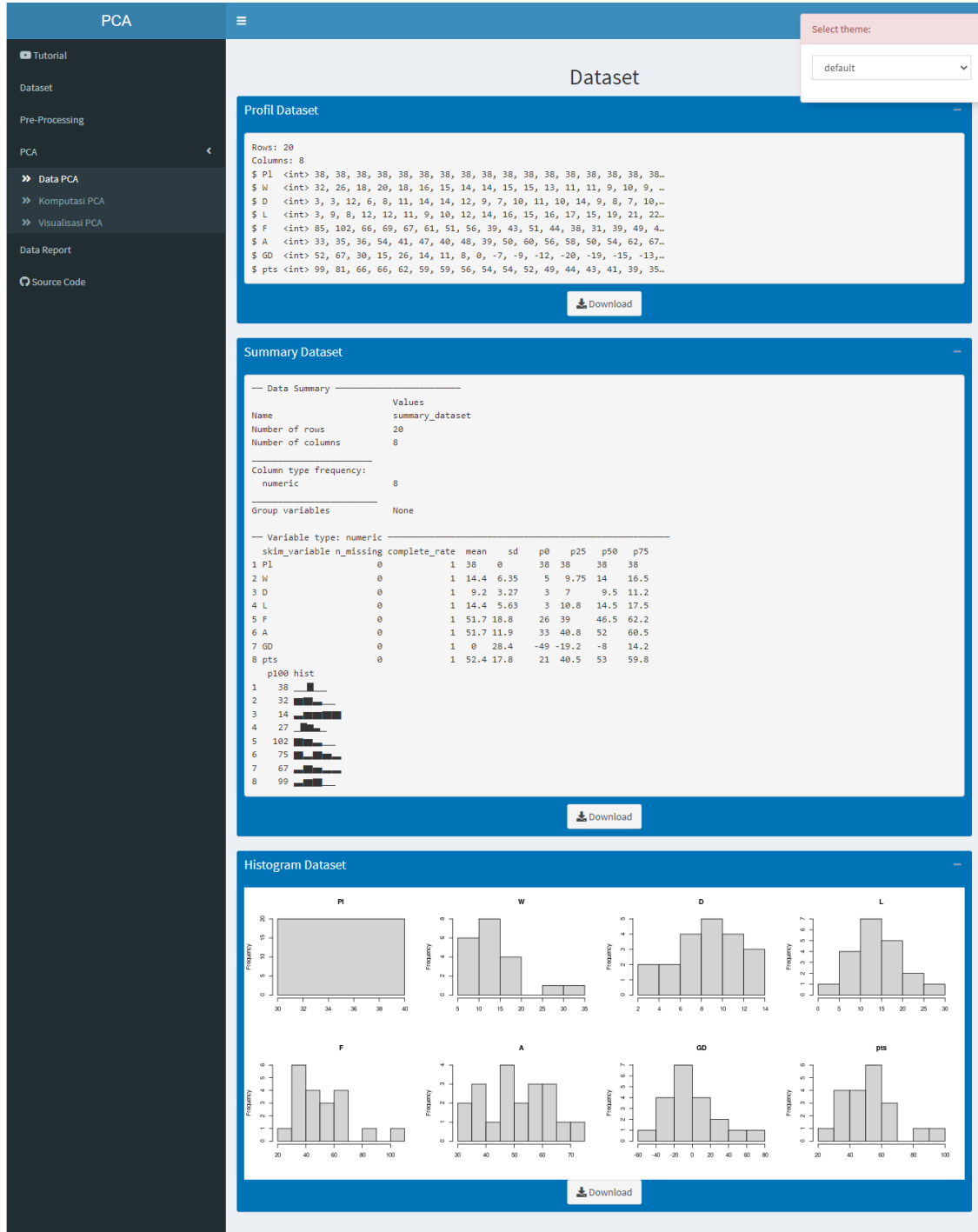
The screenshot shows the pre-processing page of the PCA application. The interface includes a sidebar with navigation options like 'Tutorial', 'Dataset', 'Pre-Processing', 'PCA', 'Data Report', and 'Source Code'. The main content area is titled 'Data Pre-Processing (Pemilihan Variabel)'. It features a 'Variabel' section with a list of variables: 'PI', 'W', 'D', 'L', 'F', 'A', 'GD', and 'pts'. Below this, there is a checkbox for 'Pilih Semua Variabel' which is checked. A 'Download Data' button is also present. The 'Tampilan Dataset' section displays the same table of 10 football teams as in the previous screenshot.

	PI	W	D	L	F	A	GD	pts
Liverpool	38	32	3	3	85	33	52	99
Manchester City	38	26	3	9	102	35	67	81
Manchester United	38	18	12	8	66	36	30	66
Chelsea	38	20	6	12	69	54	15	66
Leicester City	38	18	8	12	67	41	26	62
Tottenham Hotspur	38	16	11	11	61	47	14	59
Wolverhampton Wanderers	38	15	14	9	51	40	11	59
Arsenal	38	14	14	10	56	48	8	56
Sheffield United	38	14	12	12	39	39	0	54
Burnley	38	15	9	14	43	50	-7	54

Gambar 4.18 Hasil tampilan *pre-processing*

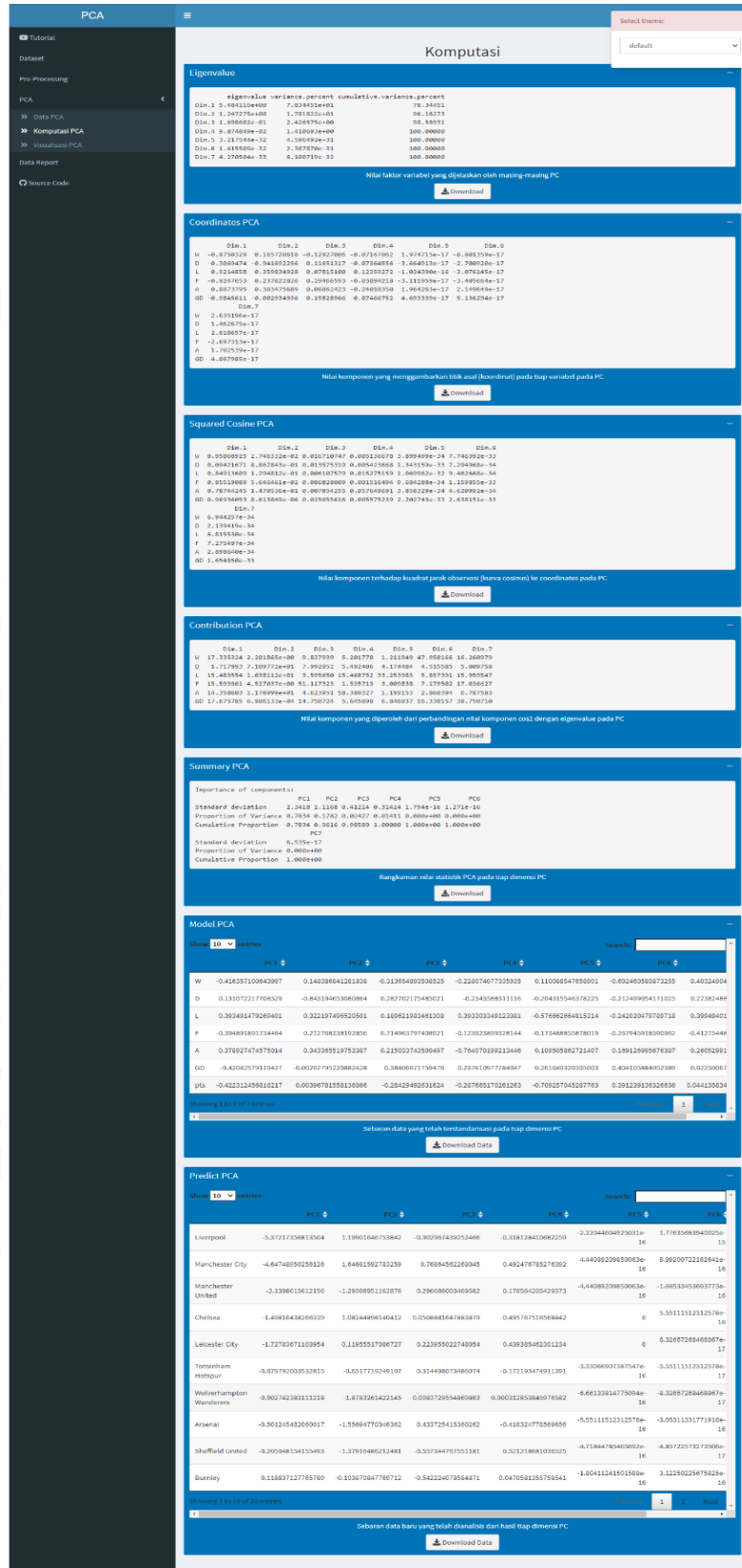
4.4.3 PCA

a. Data



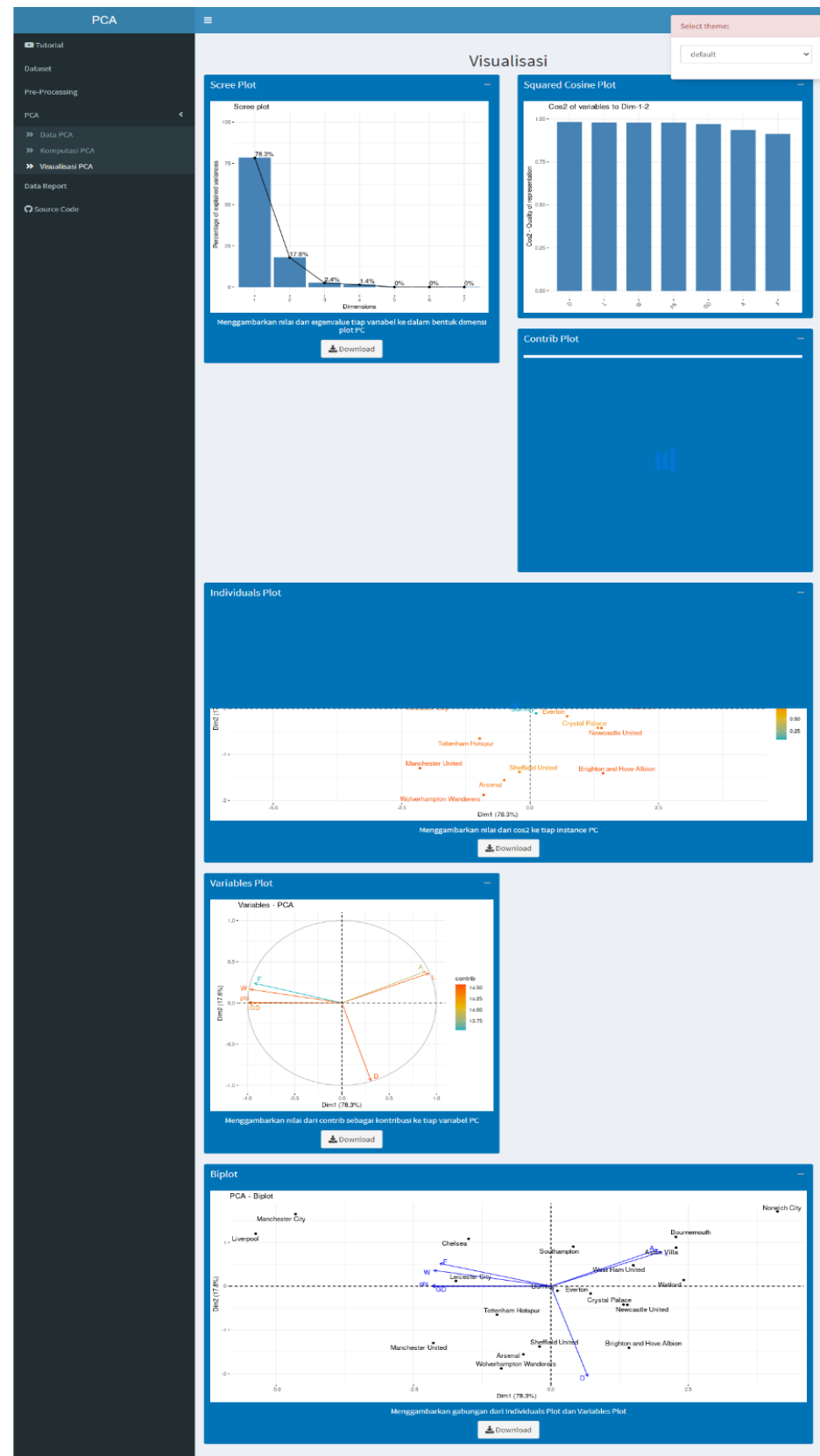
Gambar 4.19 Hasil tampilan data PCA

b. Komputasi



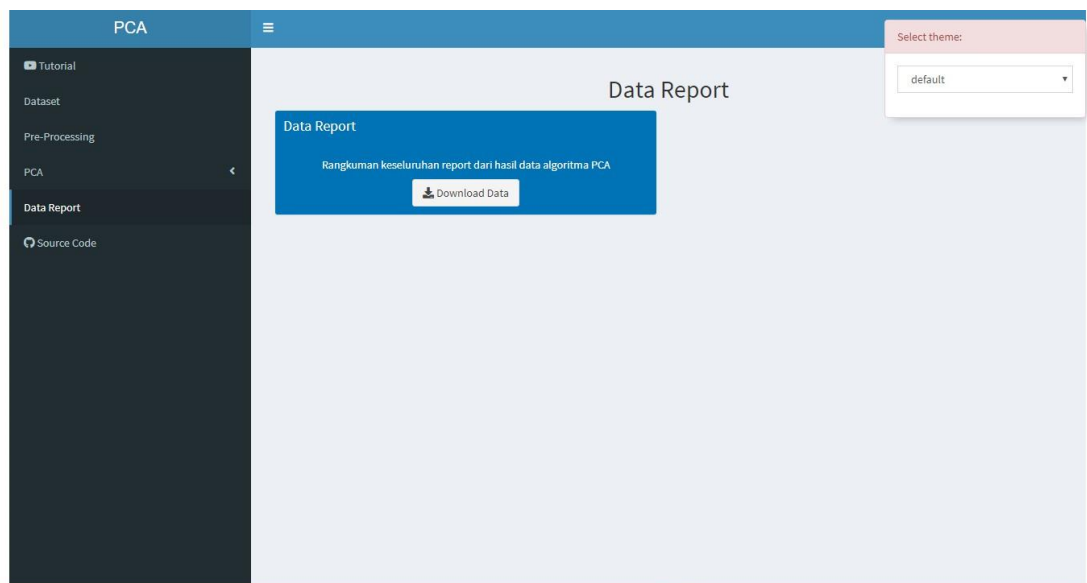
Gambar 4.20 Hasil tampilan komputasi PCA

c. Visualisasi



Gambar 4.21 Hasil tampilan visualisasi PCA

4.4.4 Data Report



Gambar 4.22 Hasil tampilan aplikasi *data report*

4.5 Skenario Pengujian Hasil Aplikasi

4.5.1 Tahap Demonstrasi

Skenario pengujian akan menampilkan hasil perhitungan dan juga visualisasi dari sampel *dataset* yang telah digunakan. Hasil skenario akan memperlihatkan hasil dari analisis *dataset* yang digunakan terlebih juga dapat menampilkan hasil perbedaan analisis yang didapatkan.

Terdapat beberapa syarat *dataset* yang digunakan untuk demonstrasi pada aplikasi ini sehingga *dataset* dapat digunakan sebagai subjek aplikasi analisis data-nya. *Dataset* yang dapat digunakan dapat ditelaah pada bagian Data PCA sehingga disini dapat melihat *profiling* suatu data yang digunakan. Syarat-syarat tersebut adalah :

1. *Dataset* harus terstruktur (tidak berantakan antara objek dan variabel).
2. Memiliki lebih dari 1 kolom (karena menerapkan asas reduksi dimensi pada data).
3. Bersifat numerik (namun pada aplikasi jika terdapat nilai pada kolom sebagai objek yang tidak bersifat numerik maka bisa dieliminasi).
4. Nilainya bervariasi (nilai pada kolom pada objek tidak boleh bernilai sama dari suatu variabel data)

Pada skenario hasil pengujian tidak akan ditunjukkan semua hasil dari aplikasi. Kebutuhan hasil pengujian hanya akan menunjukkan beberapa fitur aplikasi yaitu bagian data *pre-processing* untuk validasi dan verifikasi terhadap *dataset* yang di-*input*. Berikut hasil dari analisis *dataset* yang didapatkan.

a. Premier League *table* 19-20

Variabel data yang digunakan pada *dataset* ini adalah tidak semua variabel karena terdapat 1 objek dari kolom *dataset* yaitu kolom “**PL** (played)” yang tidak memenuhi syarat karena memiliki nilai konstan untuk tiap variabel pada *dataset*. Nilai pada kolom ini menunjukkan nilai yang sama yaitu tetap di angka 38 untuk masing-masing objek seperti yang ditunjukkan Gambar 4.23. Sehingga kolom “**PL**” harus dieliminasi dari pemilihan variabelnya.

Data Pre-Processing (Pemilihan Variabel)

Variabel

Pilih variabel (tekan backspace untuk menghapus) :

PL W D L F A GD pts

Nilai pada seluruh variabel data harus bernilai numerik

Pilih Semua Variabel

Download Data

Tampilan Dataset

Show 10 entries

Search:

	PL	W	D	L	F	A	GD	pts
Liverpool	38	32	3	3	85	33	52	99
Manchester City	38	26	3	9	102	35	67	81
Manchester United	38	18	12	8	66	36	30	66
Chelsea	38	20	6	12	69	54	15	66
Leicester City	38	18	8	12	67	41	26	62
Tottenham Hotspur	38	16	11	11	61	47	14	59
Wolverhampton Wanderers	38	15	14	9	51	40	11	59
Arsenal	38	14	14	10	56	48	8	56
Sheffield United	38	14	12	12	39	39	0	54
Burnley	38	15	9	14	43	50	-7	54

Showing 1 to 10 of 20 entries

Previous 1 2 Next

Gambar 4.23 Data *pre-processing* Premier League *table* 19-20

b. COVID-19 di Indonesia

Variabel data yang digunakan pada *dataset* ini adalah semua variabel karena syarat yang dimiliki oleh objek dari *dataset*, syaratnya terpenuhi semua seperti yang ditunjukkan pada Gambar 4.24.

Data Pre-Processing (Pemilihan Variabel)

Variabel

Pilih variabel (tekan backspace untuk menghapus) :

DIRAWAT TERKONFIRMASI
SEMBUH MENINGGAL

Nilai pada seluruh variabel data harus bernilai numerik

Pilih Semua Variabel

Download Data

Tampilan Dataset

Show 10 entries

Search:

	DIRAWAT	TERKONFIRMASI	SEMBUH	MENINGGAL
DKI Jakarta	662	865688	851418	13608
Jawa Barat	475	708889	693658	14756
Jawa Tengah	1012	486934	455634	30288
Jawa Timur	77	400099	370276	29746
Kalimantan Timur	18	158339	152869	5452
Daerah Istimewa Yogyakarta	94	157008	151645	5269
Banten	51	132848	130102	2695
Riau	331	128936	124488	4117
Bali	55	114397	110282	4060
Sulawesi Selatan	29	110007	107737	2241

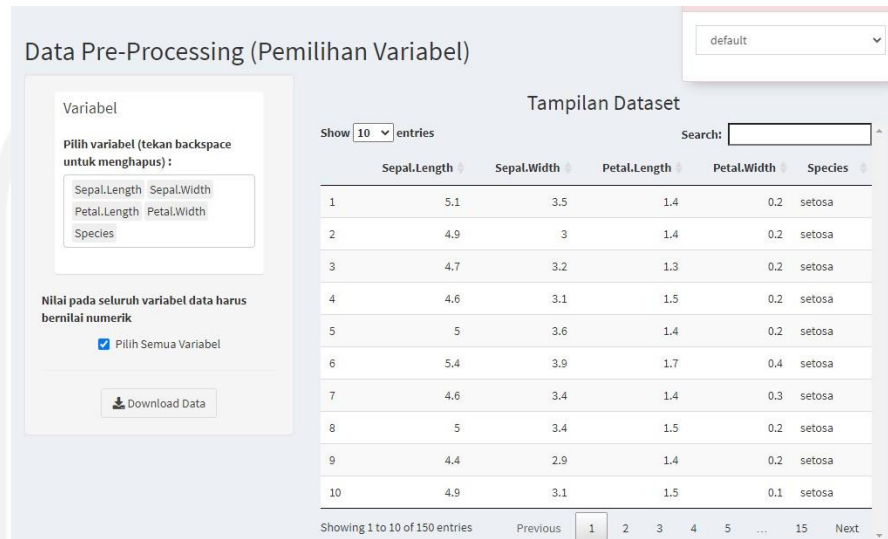
Showing 1 to 10 of 34 entries

Previous 1 2 3 4 Next

Gambar 4.24 Data *pre-processing* COVID-19 di Indonesia

c. Iris

Sedangkan pada *dataset* Iris, objek data yang digunakan pada *dataset* ini adalah sama seperti kasus *dataset* Premier League, namun terdapat perbedaan syarat yang tidak terpenuhi pada *dataset* ini. Pada kolom objek “**Species**” menunjukkan nilai pada kolom yang mempunyai tipe data string pada kolomnya. Maka objek ini wajib dihilangkan untuk memasuki ke arah analisis data selanjutnya.

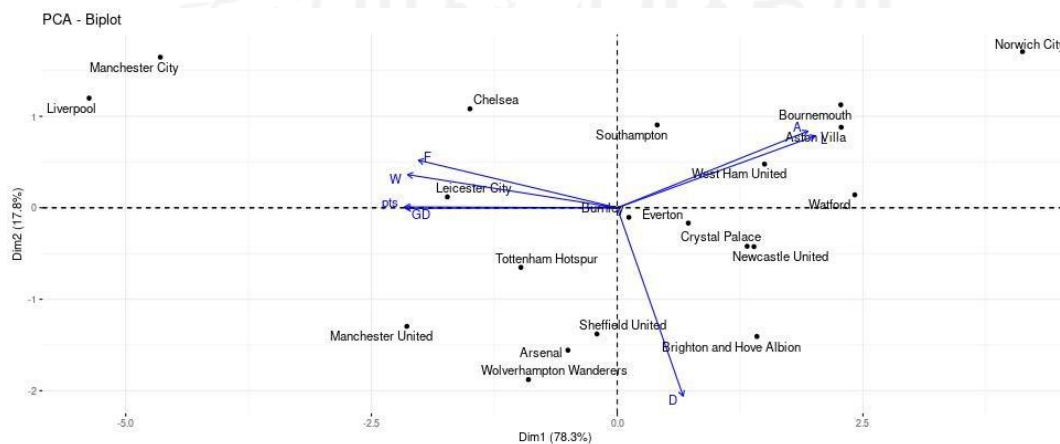


Gambar 4.25 Data *pre-processing* Iris

4.5.2 Rangkuman Hasil

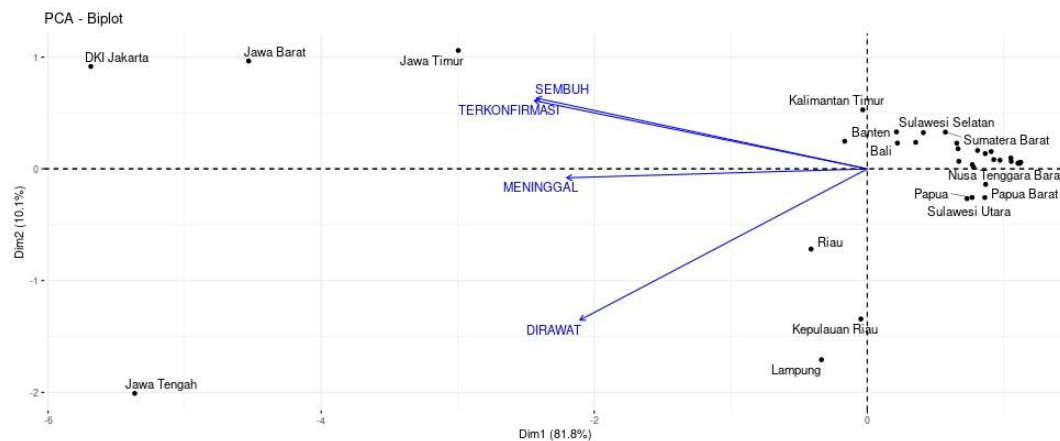
Pada bagian rangkuman akan merujuk ke bagian visualisasi untuk penjelasannya. Hal ini dilakukan karena hasil yang optimal untuk ditunjukkan pada aplikasi adalah bagian visualisasi untuk analisis data pada aplikasi ini. Berikut hasil utama dari visualisasi yaitu *biplot* pada aplikasi.

a. Premier League *table* 19-20



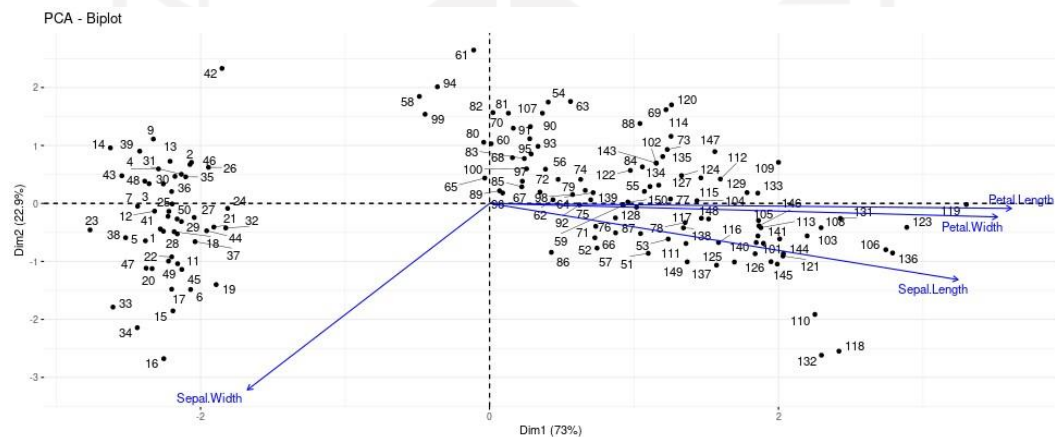
Gambar 4.26 Biplot *dataset* Premier League *table* 19-20

b. COVID-19 di Indonesia



Gambar 4.27 Biplot dataset COVID-19 di Indonesia

c. Iris



Gambar 4.28 Biplot dataset Iris

4.6 Hasil Pengujian

Pengujian hasil aplikasi akan dilakukan dengan dua metode yaitu pengujian dengan pendekatan *black box* dan pengujian dengan pendekatan *software*. *Black box testing* akan menggunakan skenario terhadap hasil yang diharapkan dengan hasil yang terjadi sesungguhnya seperti yang terlihat pada . Sedangkan pada pengujian dengan *software testing* akan membandingkan dengan komparasi aplikasi yang sejenis yaitu, WEKA. WEKA singkatan dari **Waikato Environment for Knowledge Analysis** merupakan aplikasi rangkaian *machine learning software* yang ditulis dengan bahasa pemrograman Java (Sharma et al., 2012). Lalu pada *software testing* juga dilakukan dengan perbandingan aplikasi dengan aplikasi **PCA online** pada *website* (lucianoabriata.altervista.org/jsinscience/pca/pca5.html).

4.6.1 Black Box Testing

Pengujian ini mengarah berdasarkan hasil pengujian yang diharapkan yang akan dibandingkan dengan hasil yang terjadi dengan terhadap pengujian *unit* pada aplikasi seperti yang dijelaskan pada Tabel 4.3.

Tabel 4.3 Hasil pengujian *black box*

No.	Skenario	Pengujian	Hasil yang diharapkan	Hasil yang pengujian	Status
<i>Dataset</i>					
1.	<i>Input Dataset</i>	Mengunggah <i>dataset</i> yang dipilih	<i>Dataset</i> dapat terunggah	Sesuai harapan	Diterima
2.	<i>Separator</i> dari <i>dataset</i>	Memilih <i>separator</i> yang sesuai dengan “radio button”	<i>Dataset</i> dapat ditampilkan	Sesuai harapan	Diterima
<i>Pre-processing</i>					
1.	Variabel kolom yang ingin dipilih	Memilih variabel dari <i>dataset</i> yang telah di- <i>input</i> sebelumnya	<i>Dataset</i> hanya menampilkan variabel yang dipilih	Sesuai harapan	Diterima
2.	Pilihan semua variabel pada <i>dataset</i>	Mengaktifkan <i>checkbox</i> pada tampilan untuk semua variabel dipilih	Semua variabel kolom ditampilkan pada tampilan <i>dataset</i>	Sesuai harapan	Diterima
3.	Unduh <i>dataset</i> baru hasil <i>pre-processing</i>	Menekan button “download”	Hasil <i>dataset</i> baru sesuai dengan pilihan kolom variabel yang dipilih <i>user</i> dan dapat disimpan ke direktori <i>hardware</i>	Sesuai harapan	Diterima

			<i>user</i>		
PCA					
1.	Hasil seluruh data (profil, <i>summary</i> , dan histogram)	Memilih variabel dari <i>dataset</i> dan membuka halaman “Data PCA”	Seluruh hasil dapat ditampilkan dengan hasil data yang sesuai	Sesuai harapan	Diterima
2.	Unduh hasil data	Menekan <i>button</i> “download”	Seluruh hasil data dapat terunduh ke dalam direktori <i>hardware user</i>	Sesuai harapan	Diterima
3.	Hasil seluruh komputasi (<i>eigenvalue</i> , <i>coordinates</i> , <i>squared cosine</i> , <i>contribution</i> , <i>summary</i> , <i>model</i> , <i>predict</i>)	Memilih variabel dari <i>dataset</i> dan membuka halaman “Komputasi PCA”	Seluruh hasil dapat ditampilkan dengan komputasi menggunakan algoritma PCA	Sesuai harapan	Diterima
4.	Unduh hasil komputasi PCA	Menekan <i>button</i> “download”	Seluruh hasil komputasi dapat terunduh ke dalam direktori <i>hardware user</i>	Sesuai harapan	Diterima
5.	Hasil seluruh visualisasi (<i>screepplot</i> , <i>squared cosine plot</i> , <i>contribution plot</i> , <i>individuals plot</i> , <i>variables plot</i> , <i>biplot</i>)	Memilih variabel dari <i>dataset</i> dan membuka halaman “Visualisasi PCA”	Seluruh hasil dapat ditampilkan dengan visualisasi menggunakan algoritma PCA	Sesuai harapan	Diterima

6.	Unduh hasil visualisasi PCA	Menekan <i>button</i> “download”	Seluruh hasil visualisasi dapat terunduh ke dalam direktori <i>hardware user</i>	Sesuai harapan	Diterima
<i>Data Report</i>					
1.	Mengunduh hasil <i>data report</i>	Menekan <i>button</i> “download”	<i>Data report</i> dapat terunduh ke dalam direktori <i>hardware user</i>	Sesuai harapan	Diterima
2.	<i>Data report</i> terbaca	Membuka hasil unduhan <i>data report</i>	Tampilan <i>data report</i> sesuai dengan hasil keseluruhan tampilan aplikasi	Sesuai harapan	Diterima

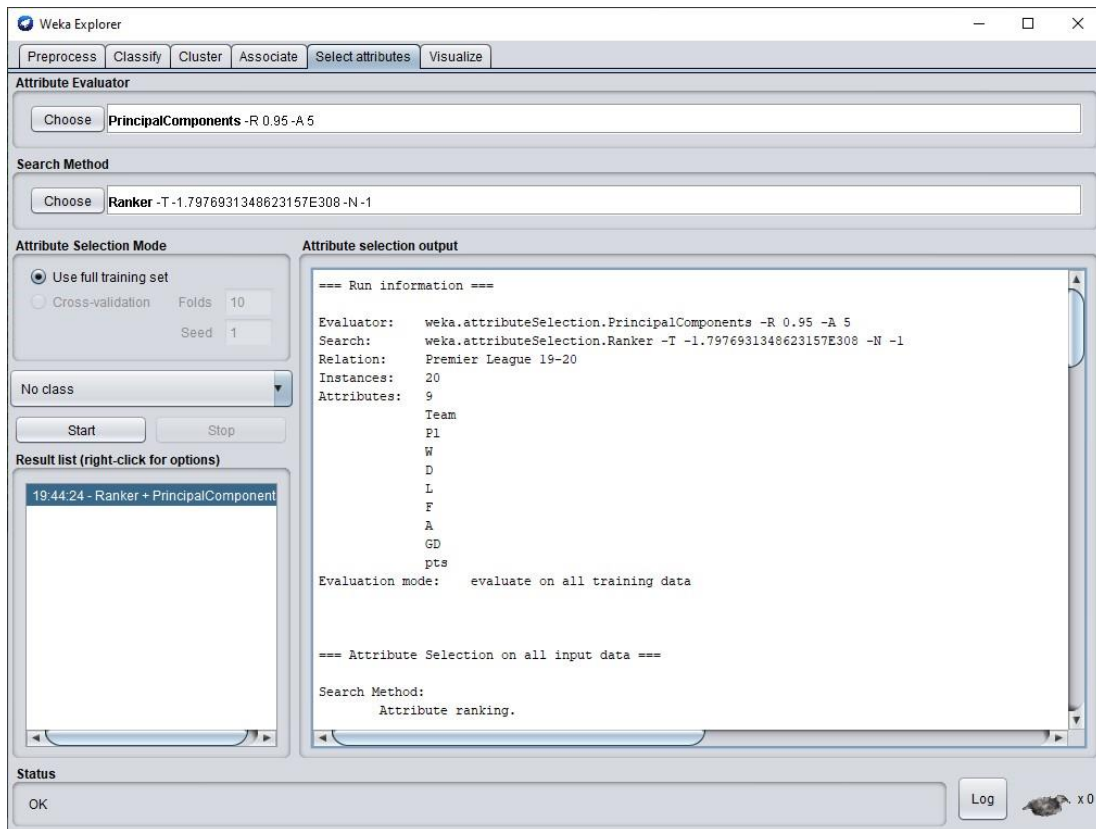
Black box testing yang dilakukan ini difokuskan terhadap *output* masing-masing unit pada aplikasi sehingga adanya kombinasi dengan *unit testing*. Pengujian yang sudah dilakukan terhadap aplikasi ini menghasilkan validitas 100% atau keseluruhan fungsional sudah memenuhi kebutuhan aplikasi yang diharapkan. Sehingga dari sisi pengembang, aplikasi ini sudah layak digunakan ke umum.

4.6.2 *Software Testing*

Pengujian ini akan menjadi pengujian dengan dasar perbandingan dengan aplikasi lain dimana aplikasi tersebut berupa aplikasi berbasis *desktop* yang dinamakan WEKA dan juga aplikasi berbasis web yang lain dinamakan PCA Online. Pengujian dilakukan untuk mendapatkan hasil validasi dan verifikasi terhadap *software* aplikasi yang sudah dikembangkan sebelum aplikasi penelitian ini.

Pada pengujian ini fokus pengujian lebih kepada validasi dan verifikasi dari program yang sejenis sudah ada. Seperti yang sudah dijelaskan sebelumnya, *testing* ini akan menggunakan aplikasi WEKA sebagai bentuk komparasi perbandingan aplikasinya. Disini sampel *dataset* yang digunakan untuk pengujian adalah satu data saja yaitu **Premier League** karena disini dijadikan perbandingan hasil terhadap *software* bukan perbandingan hasil

terhadap *dataset*. Seperti yang terlihat pada Gambar 4.29 merupakan perbandingan dari aplikasi ini dengan aplikasi WEKA seperti apa hasilnya.



Gambar 4.29 Penggunaan aplikasi WEKA

Pada aplikasi WEKA yang terlihat pada Gambar 4.30 terhadap nilai yang dihasilkan dari aplikasi pada penelitian ini mendekati kesamaan. Dimana komputasi yang dihasilkan berbeda dimana pada aplikasi WEKA terdapat nilai *correlation matrix*, *eigenvectors*, dan *ranked attributes*. Sedangkan pada aplikasi penelitian ini terdapat nilai *model & predict PCA*.

Pada aplikasi WEKA terdapat perhitungan dan visualisasi yang didapatkan. Namun pada bagian visualisasi tidak terdapat visualisasi yang berhubungan konteks langsung terhadap komputasi PCA. Aplikasi WEKA hanya memperlihatkan visualisasi plot untuk sebaran nilai data dari *dataset* yang di-*input* oleh pengguna.

Tabel 4.4 Hasil pengujian *software*

No.	Perbandingan	Aplikasi ini	WEKA	PCA Online
1.	Bentuk aplikasi	Web	<i>Desktop</i>	Web
2.	Bahasa pemrograman	R	Java	JavaScript
3.	<i>Pre-processing</i>	<i>modeling</i>	<i>full-training</i>	<i>sampling</i>
4.	<i>Instance dataset</i>	<i>component</i>	<i>attribute</i>	<i>object</i>
5.	<i>Format dataset</i>	.txt, .csv	.arff, .arff.gz, .names, .data, .csv, .json, .json.gz	.txt, .csv

4.7 Pembahasan

Aplikasi ini akan menampilkan satu *output* utama yaitu pada halaman PCA yang memiliki *subunit* fungsi luaran yaitu data, visualisasi, dan komputasi. Bagian “data” disini berperan sebagai tampilan dari identitas *dataset* yang telah di-*input* sebelumnya oleh pengguna. Lalu *dataset* tersebut juga akan diproses dalam bentuk komputasi data dan hasil visualisasi *plot* dari algoritma PCA. Pengujian hasil menjadi standar bentuk aplikasi sebelum aplikasi dapat digunakan secara umum.

4.7.1 Data PCA

Bagian ini akan mendeskripsikan *dataset* yang telah di-*input* oleh pengguna. Terdapat tiga unit fungsi yang akan ditampilkan pada bagian ini, yaitu :

1. Profil *dataset*

Menjelaskan identitas dari *dataset* seperti baris dan kolom yang dimiliki oleh *dataset* serta tipe data yang dimiliki oleh tiap frekuensi tiap nilai dari datanya.

2. *Summary dataset*

Berupa rangkuman statistik dari *dataset* yang di-*input* oleh pengguna dimana terdapat nilai *mean*, median, nilai kuartil, dan standar deviasi. Selain itu ditambahkan juga nilai dari profil *dataset* sebelumnya seperti baris, kolom, tipe data, dan yang terakhir adanya tipe variabel dari *dataset* yang digunakan.

3. Histogram *dataset*

Menggambarkan sebuah grafik dengan bentuk diagram batang dari nilai tiap frekuensi dari kolom variabel *dataset*.

4.7.2 Komputasi PCA

Komputasi PCA merupakan bagian interpretasi penerapan algoritma PCA dalam kalkulasi pada *dataset* yang digunakan. Terdapat tujuh unit fungsi *output* pada *tool* aplikasi ini yaitu :

1. *Eigenvalue* : Nilai faktor variabel yang dijelaskan oleh masing-masing PC.
2. *Coordinates* : Nilai komponen yang menggambarkan titik asal (koordinat) pada tiap variabel pada PC.
3. *Squared cosine* : Nilai komponen terhadap kuadrat jarak observasi (kurva kosinus) ke *coordinates* pada PC.
4. *Contribution* : Nilai komponen yang diperoleh dari perbandingan nilai komponen \cos^2 dengan *eigenvalue* pada PC.
5. *Summary* : Rangkuman nilai statistik PCA pada tiap dimensi PC.
6. *Model & Predict* : Sebaran data yang telah terstandarisasi pada tiap dimensi PC dan Sebaran data baru yang telah dianalisis dari hasil tiap dimensi PC.

Lalu perhitungan dari komputasi PCA ini bersifat konstan dan sistematis sehingga jika terdapat satu kesalahan pada salah satu komputasi, maka hasil lainnya akan memengaruhi terlebih pada komputasi bagian *model* PCA. Mengapa demikian ? Karena nilai *model* PCA menjadi *parent data* untuk fitur komputasi aplikasi selanjutnya berlaku juga ke bagian visualisasi ke depannya.

4.7.3 Visualisasi PCA

Visualisasi pada PCA menjadi hasil representatif dari nilai komputasi sebelumnya dimana sebaran hasil tersebut dikeluarkan dalam bentuk *plotting*. *Output* pada visualisasi akan diberikan sebanyak enam plot dengan sebagai berikut :

1. *Screeplot* : mewakilkan nilai varians kumulatif PC dari nilai *eigenvalue*
2. *Squared cosine plot* : mewakilkan nilai *squared cosine* ke dalam diagram batang
3. *Contribution plot* : mewakilkan nilai *contribution* ke dalam diagram batang.
4. *Individuals plot* : menggambarkan nilai *squared cosine* ke dalam sebaran *plot* PCA
5. *Variables plot* : menggambarkan nilai *contribution* ke tiap variabel *plot* PCA
6. *Biplot* : menggambarkan gabungan *individuals & variables plot* PCA

4.7.4 Data Report PCA

Pada bagian ini akan menjelaskan suatu ikhtisar dari hasil analisis data yang didapatkan. Rangkuman analisis yang didapatkan itu dirangkum dari awal proses khususnya pada bagian

data hingga pada bagian visualisasi. Seperti yang terlihat pada Gambar 4.32 & Gambar 4.33 sebagai contoh hasil *data report*.

Hasil *data report* menggunakan algoritma PCA

PCA

23-August-2022

Data

Profil Dataset

Menjelaskan identitas dari dataset seperti baris dan kolom yang dimiliki oleh dataset serta tipe data yang dimiliki oleh tiap frekuensi tiap nilai dari datanya.

```

Rows: 20
Columns: 7
$ W <int> 32, 26, 18, 20, 18, 16, 15, 14, 14, 15, 15, 15, 11, 11, 9, 10, 9, -
$ D <int> 3, 3, 12, 6, 9, 11, 14, 14, 12, 9, 7, 10, 11, 10, 14, 9, 8, 7, 10, -
$ L <int> 3, 9, 8, 12, 12, 11, 9, 10, 12, 14, 16, 15, 10, 17, 15, 19, 21, 22, -
$ F <int> 85, 102, 66, 69, 67, 81, 51, 56, 39, 43, 51, 44, 38, 31, 39, 49, 4, -
$ A <int> 23, 35, 36, 54, 41, 47, 40, 48, 39, 50, 48, 58, 58, 50, 54, 62, 67, -
$ GD <int> 52, 67, 30, 15, 20, 14, 11, 8, 0, -7, -9, -12, -20, -10, -15, -13, -
$ pts <int> 99, 81, 66, 66, 62, 59, 59, 56, 54, 54, 52, 49, 44, 43, 41, 39, 35, -
    
```

Summary Data

Berupa rangkuman statistik dari dataset yang di-input oleh pengguna dimana terdapat nilai mean, median, nilai kuartil, dan standar deviasi. Selain itu ditambahkan juga nilai dari profil dataset sebelumnya seperti baris, kolom, tipe data, dan yang terakhir adanya tipe variabel dari dataset yang digunakan.

W	D	L	F
Min.: 5.00	Min.: 3.00	Min.: 3.00	Min.: 26.00
1st Qu.: 9.75	1st Qu.: 7.00	1st Qu.: 10.75	1st Qu.: 39.00
Median: 14.00	Median: 9.50	Median: 14.50	Median: 46.50
Mean: 14.40	Mean: 9.20	Mean: 14.40	Mean: 51.70
3rd Qu.: 18.50	3rd Qu.: 11.25	3rd Qu.: 17.50	3rd Qu.: 62.25
Max.: 32.00	Max.: 14.00	Max.: 27.00	Max.: 102.00

A	GD	pts
Min.: 23.00	Min.: -49.00	Min.: 22.00
1st Qu.: 48.75	1st Qu.: -19.25	1st Qu.: 40.50
Median: 52.00	Median: -8.00	Median: 53.00
Mean: 51.70	Mean: 9.00	Mean: 52.40
3rd Qu.: 60.50	3rd Qu.: 14.25	3rd Qu.: 59.75
Max.: 75.00	Max.: 67.00	Max.: 99.00

Data summary

Name	summary_dataset
Number of rows	20
Number of columns	7
Column type frequency:	
numeric	7

Gambar 4.32 Contoh hasil *Data Report* (1)

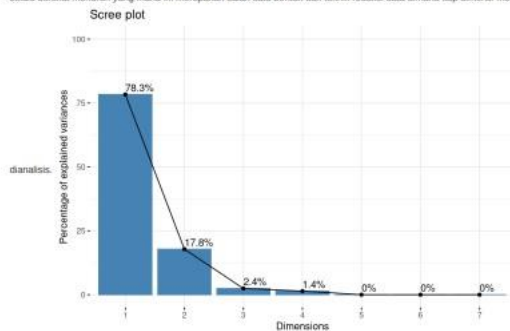
	PC1
Standard deviation	8.535e-17
Proportion of Variance	9.000e+00
Cumulative Proportion	1.000e+00

Predict Data tidak dapat ditampilkan pada data report karena data tersebut merupakan data table hasil prediksi dari data model

Visualisasi

Screeplot PCA

Screeplot sebagai gambaran plot dari nilai eigenvalues. Hasil kumulatif dari tiap dimensi akan berjumlah 100 % dimana dari awal dimensi akan selalu bersifat menurun yang mana ini merupakan salah satu bentuk dari teknik reduksi data dimana tiap dimensi mewakili nilai-nilai yang akan



Individuals Plot PCA

Individuals plot sebagai gambaran plot yang didapatkan dari sebaran nilai squared cosine yang digambarkan menjadi dimensi plot ini (Individuals



Gambar 4.33 Contoh hasil *Data Report* (2)

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini mengangkat dari hasil pengembangan aplikasi pengolahan data berbasis web dengan R Shiny yang dapat melakukan analisis data menggunakan algoritma PCA. Kemudian tentu dari hasil aplikasi ini nantinya dapat digunakan oleh pengguna yang ingin menganalisis suatu data yang dimilikinya dengan menggunakan algoritma PCA. Pengguna akan disuguhkan dengan aplikasi menghasilkan suatu identitas *dataset*, hasil komputasi dan visualisasi *dataset* dengan PCA, dan yang terakhir sebuah *data report* yang berupa dokumen dengan mencakup keseluruhan isi hasil rangkuman aplikasi. Aplikasi dalam penelitian ini juga diuji menggunakan dua pendekatan pengujian dengan metode *black box testing* dan juga *software testing*. Hal ini dilakukan sebagai tolak ukur terhadap aplikasi yang layak digunakan menjadi aplikasi analisis data berbasis web sebelum dapat dipublikasikan ke umum.

5.2 Saran

Penelitian ini masih jauh di bawah kata sempurna yang dengan kata lain hasil dari penelitian ini yang berupa aplikasi terdapat beberapa saran yang mungkin bisa menjadi acuan untuk penelitian ini bisa dikembangkan lebih lanjut. Saran-saran tersebut berupa :

1. Aplikasi yang dikembangkan dengan algoritma PCA bisa ditambahkan dengan perhitungan yang belum terdapat pada aplikasi dan juga hasil bentuk visualisasinya dikarenakan terbatasnya peneliti untuk mengeksplorasi persamaan yang terdapat pada algoritma PCA.
2. Aplikasi dibuat lebih dinamis dengan kata lain terdapat *storing* pada aplikasi yang bisa digunakan dengan struktur *database* serta *user* tidak hanya memilih variabel dan mengunduh suatu hasil tetapi ditambahkan hal-hal yang lebih membuat aplikasi menjadi ramah terhadap *user*.
3. Adanya fungsi *catch error* pada aplikasi yang menampilkan *error* dimana saat ini *error* yang ditampilkan masih mengikuti aturan *default* mengikuti dari R Studio.

DAFTAR PUSTAKA

- Abdi, H., & Williams, L. J. (2010). Principal Component Analysis. *Wiley Online Library*, 2. www.utdallas.edu/
- Adhisyanda Aditya, M., Dicky Mulyana, R., Putu Eka, I., & Rheno Widiyanto, S. (2020). Penggabungan Teknologi Untuk Analisa Data Berbasis Data Science. *seminar-id.com*, 51–56. <http://seminar-id.com/prosiding/index.php/sainteks/article/view/404>
- Boutsidis, C., Mahoney, M. W., & Drineas, P. (2008). Unsupervised feature selection for principal components analysis. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 61–69. <https://doi.org/10.1145/1401890.1401903>
- Chang, X., Nie, F., Yang, Y., Zhang, C., & Huang, H. (2016). Convex sparse PCA for unsupervised feature learning. *ACM Transactions on Knowledge Discovery from Data*, 11(1). <https://doi.org/10.1145/2910585>
- Darningsih, B. E., Dinara, M., Quraisyin, F. N. R. N. D. K. D. T. N. S. H. A. C. D., Moertijoso, B., Camelia, T. H. R. S. A. F. R. A., Surokim, & Julijanti, Y. R. C. S. M. W. T. H. (2016). RISET KOMUNIKASI: Strategi Praktis Bagi Peneliti Pemula. In *komunikasi.trunojoyo.ac.id*. <http://komunikasi.trunojoyo.ac.id/wp-content/uploads/2016/01/BUKU-RISET-KOM-2016.pdf#page=138>
- Donalek, C. (2011). Supervised and unsupervised learning. *sites.astro.caltech.edu*, 199. https://sites.astro.caltech.edu/~george/aybi199/Donalek_classif1.pdf
- Dy, J., & Brodley, C. (2004). Feature selection for unsupervised learning. *jmlr.org*, 5, 845–889. <https://www.jmlr.org/papers/volume5/dy04a/dy04a.pdf?ref=https://githubhelp.com>
- Feldman, D., Schmidt, M., & Sohler, C. (2013). Turning Big data into tiny data: Constant-size coresets for k-means, PCA and projective clustering. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, 1434–1453. <https://doi.org/10.1137/1.9781611973105.103>
- Ghojogh, B., & Crowley, M. (2019). Unsupervised and supervised principal component analysis: Tutorial. *arxiv.org*. <https://arxiv.org/abs/1906.03148>
- Irmansyah, F. (2003). Pengantar Database. *staff.blog.ui.ac.id*. <https://staff.blog.ui.ac.id/r-suti/files/2010/03/pengantardb.pdf>
- Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *science.org*. <https://doi.org/10.1126/science.aac4520>

- Kassambara, A. (2018). *Machine learning essentials: Practical guide in R*.
<https://books.google.com/books?hl=en&lr=&id=745QDwAAQBAJ&oi=fnd&pg=PP2&dq=Essentials+in+R:&ots=5ENtsPT0Lt&sig=TSuvqG5cqSGtslmHVhRLS5PPKhI>
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. In *Informatica* (Vol. 31).
https://books.google.com/books?hl=id&lr=&id=vLiTXDHR_sYC&oi=fnd&pg=PA3&dq=supervised+machine+learning&ots=CZmvAz0Hhgk&sig=oHuY5j1CD37vBArdU_GdK7UJFVA
- Kuchta, J. (2016). Completeness and Consistency of the System Requirement Specification. *mostwiedzy.pl*. <https://doi.org/10.15439/2016F468>
- Lloyd, S., Mohseni, M., & Rebentrost, P. (2013). Quantum algorithms for supervised and unsupervised machine learning. In *arxiv.org*. <https://arxiv.org/abs/1307.0411>
- Lucca, G., & Penta, M. Di. (2003). Considering browser interaction in web application testing. *ieeexplore.ieee.org*. <https://doi.org/10.1109/WSE.2003.1234011>
- M Ringnér. (2008). What is principal component analysis? *nature.com*.
<https://www.nature.com/articles/nbt0308-303>
- Mitchell, T. (1997). *Machine learning*.
<https://profs.info.uaic.ro/~ciortuz/SLIDES/2017s/ml0.pdf>
- Mutiara, A. B., & Refianti, R. (2018). MACHINE LEARNING KUANTUM UNTUK SAINS DATA Penerbit Gunadarma. In *amutiara.staff.gunadarma.ac.id*. Gunadarma.
<http://amutiara.staff.gunadarma.ac.id/Publications/files/4333/bukuMachineLearningKuantum.pdf>
- Nasution, M. (2002). Pangkalan data: konsep dan definisi. *researchgate.net*.
<https://doi.org/10.13140/RG.2.2.14380.69762/1>
- Nasution, M. (2019). Sains data. *researchgate.net*.
<https://doi.org/10.13140/RG.2.2.21816.49924/2>
- Pan, J. (1999). Software testing. *sci.brooklyn.cuny.edu*.
<http://www.sci.brooklyn.cuny.edu/~sklar/teaching/s08/cis20.2/papers/software-testing.pdf>
- Potter, G., Wong, J., Alcaraz, I., & In, P. C. (2016). Web application teaching tools for statistics using R and shiny. *escholarship.org*, 9, 1. <https://escholarship.org/uc/item/00d4q8cp>
- Rahayu, G. (2017). Principal Component Analysis untuk Dimensi Reduksi Data Clustering

- Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (Vol. 0, Nomor 0). <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/3265>
- Ridgway, J., Mccusker, S., Pead, D., & Futurelab, B. (2007). Doing a literature review. *cambridge.org*. <https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/doing-a-literature-review/00B62000B6760AB78E1BD27E32A94C9F>
- Rijali, A. (2008). A. *Metode Penelitian*. 17(33). <http://sc.syekhnurjati.ac.id/esscamp/risetmhs/BAB31414112081.pdf>
- Rijali, A. (2018). Analisis data kualitatif. *jurnal.uin-antasari.ac.id*, 17. <http://jurnal.uin-antasari.ac.id/index.php/alhadharah/article/view/2374>
- Sharma, R., Alam, M., & A, R. (2012). K-means clustering in spatial data mining using weka interface. *Citeseer*. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.4365&rep=rep1&type=pdf>
- Shen, Y., Abubakar, M., Liu, H., & Energies, F. H. (2019). Power quality disturbance monitoring and classification based on improved PCA and convolution neural network for wind-grid distribution systems. *mdpi.com*, 12(7). <https://doi.org/10.3390/en12071280>
- Sievert, C. (2020). *Interactive web-based data visualization with R, plotly, and shiny*. <https://books.google.com/books?hl=en&lr=&id=7zPNDwAAQBAJ&oi=fnd&pg=PT9&dq=r+shiny&ots=eZM6Rj86jn&sig=T76G2svQsn68n0USyXDg4TYRuNo>
- Tadi, M. (2018). PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM ALGORITMA K-MEANS UNTUK MENENTUKAN CENTROID PADA CLUSTERING. *KONSTANTA*, 1(1). <https://journal.iainkudus.ac.id/index.php/Konstanta/article/view/3543>
- Verma, A., Khatana, A., & Eng, S. C. (2017). A comparative study of black box testing and white box testing. *researchgate.net*. <https://doi.org/10.26438/ijcse/v5i12.301304>
- Wold, S., Esbensen, K., Laboratory, P. G.-C. and intelligent, & 1987, U. (1987). Principal component analysis. *Elsevier*, 2, 37–52. <https://www.sciencedirect.com/science/article/pii/0169743987800849>
- Yaseen, M., Ibrahim, N., & Mustapha, A. (2019). Requirements prioritization and using iteration model for successful implementation of requirements. *researchgate.net*, 10(1). <https://doi.org/10.14569/IJACSA.2019.0100115>

LAMPIRAN

Dataset Premier League 19-20

Team	Pl	W	D	L	F	A	GD	pts
Liverpool	38	32	3	3	85	33	52	99
Manchester City	38	26	3	9	102	35	67	81
Manchester United	38	18	12	8	66	36	30	66
Chelsea	38	20	6	12	69	54	15	66
Leicester City	38	18	8	12	67	41	26	62
Tottenham Hotspur	38	16	11	11	61	47	14	59
Wolverhampton Wanderers	38	15	14	9	51	40	11	59
Arsenal	38	14	14	10	56	48	8	56
Sheffield United	38	14	12	12	39	39	0	54
Burnley	38	15	9	14	43	50	-7	54
Southampton	38	15	7	16	51	60	-9	52
Everton	38	13	10	15	44	56	-12	49
Newcastle United	38	11	11	16	38	58	-20	44
Crystal Palace	38	11	10	17	31	50	-19	43
Brighton and Hove Albion	38	9	14	15	39	54	-15	41
West Ham United	38	10	9	19	49	62	-13	39
Aston Villa	38	9	8	21	41	67	-26	35
Bournemouth	38	9	7	22	40	65	-25	34
Watford	38	8	10	20	36	64	-28	34
Norwich City	38	5	6	27	26	75	-49	21

Dataset COVID-19 di Indonesia

PROVINSI	DIRAWAT	TERKONFIRMASI	SEMBUH	MENINGGAL
DKI Jakarta	662	865688	851418	13608
Jawa Barat	475	708889	693658	14756
Jawa Tengah	1012	486934	455634	30288
Jawa Timur	77	400099	370276	29746
Kalimantan Timur	18	158339	152869	5452
Daerah Istimewa Yogyakarta	94	157008	151645	5269
Banten	51	132848	130102	2695
Riau	331	128936	124488	4117
Bali	55	114397	110282	4060
Sulawesi Selatan	29	110007	107737	2241
Sumatera Utara	47	106128	103186	2895

Sumatera Barat	6	89873	87713	2154
Kalimantan Selatan	11	69957	67554	2392
Nusa Tenggara Timur	52	64263	62863	1348
Sumatera Selatan	12	59988	56895	3081
Kepulauan Riau	427	54308	52133	1748
Kepulauan Bangka Belitung	12	52362	50888	1462
Lampung	517	49752	45350	3885
Sulawesi Tengah	41	47258	45611	1606
Kalimantan Tengah	46	45661	44183	1432
Kalimantan Barat	4	41676	40608	1064
Aceh	3	38431	36362	2066
Kalimantan Utara	18	35948	35119	811
Sulawesi Utara	112	34803	33629	1062
Papua	110	34376	33706	560
Jambi	13	29792	28996	783
Nusa Tenggara Barat	70	27810	26922	818
Papua Barat	99	23684	23228	357
Bengkulu	1	23106	22632	473
Sulawesi Tenggara	6	20173	19639	528
Maluku	5	14596	14326	265
Sulawesi Barat	3	12368	12018	347
Maluku Utara	0	12105	11802	303
Gorontalo	0	11849	11388	461

Dataset Iris

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
05.01	03.05	01.04	00.02	setosa
04.09	3	01.04	00.02	setosa
04.07	03.02	01.03	00.02	setosa
04.06	03.01	01.05	00.02	setosa
5	03.06	01.04	00.02	setosa
05.04	03.09	01.07	00.04	setosa
04.06	03.04	01.04	00.03	setosa

5	03.04	01.05	00.02	setosa
04.04	02.09	01.04	00.02	setosa
04.09	03.01	01.05	00.01	setosa
05.04	03.07	01.05	00.02	setosa
04.08	03.04	01.06	00.02	setosa
04.08	3	01.04	00.01	setosa
04.03	3	01.01	00.01	setosa
05.08	4	01.02	00.02	setosa
05.07	04.04	01.05	00.04	setosa
05.04	03.09	01.03	00.04	setosa
05.01	03.05	01.04	00.03	setosa
05.07	03.08	01.07	00.03	setosa
05.01	03.08	01.05	00.03	setosa
05.04	03.04	01.07	00.02	setosa
05.01	03.07	01.05	00.04	setosa
04.06	03.06	1	00.02	setosa
05.01	03.03	01.07	00.05	setosa
04.08	03.04	01.09	00.02	setosa
5	3	01.06	00.02	setosa
5	03.04	01.06	00.04	setosa
05.02	03.05	01.05	00.02	setosa
05.02	03.04	01.04	00.02	setosa
04.07	03.02	01.06	00.02	setosa
04.08	03.01	01.06	00.02	setosa
05.04	03.04	01.05	00.04	setosa
05.02	04.01	01.05	00.01	setosa
05.05	04.02	01.04	00.02	setosa
04.09	03.01	01.05	00.02	setosa
5	03.02	01.02	00.02	setosa
05.05	03.05	01.03	00.02	setosa
04.09	03.06	01.04	00.01	setosa
04.04	3	01.03	00.02	setosa
05.01	03.04	01.05	00.02	setosa

5	03.05	01.03	00.03	setosa
04.05	02.03	01.03	00.03	setosa
04.04	03.02	01.03	00.02	setosa
5	03.05	01.06	00.06	setosa
05.01	03.08	01.09	00.04	setosa
04.08	3	01.04	00.03	setosa
05.01	03.08	01.06	00.02	setosa
04.06	03.02	01.04	00.02	setosa
05.03	03.07	01.05	00.02	setosa
5	03.03	01.04	00.02	setosa
7	03.02	04.07	01.04	versicolor
06.04	03.02	04.05	01.05	versicolor
06.09	03.01	04.09	01.05	versicolor
05.05	02.03	4	01.03	versicolor
06.05	02.08	04.06	01.05	versicolor
05.07	02.08	04.05	01.03	versicolor
06.03	03.03	04.07	01.06	versicolor
04.09	02.04	03.03	1	versicolor
06.06	02.09	04.06	01.03	versicolor
05.02	02.07	03.09	01.04	versicolor
5	2	03.05	1	versicolor
05.09	3	04.02	01.05	versicolor
6	02.02	4	1	versicolor
06.01	02.09	04.07	01.04	versicolor
05.06	02.09	03.06	01.03	versicolor
06.07	03.01	04.04	01.04	versicolor
05.06	3	04.05	01.05	versicolor
05.08	02.07	04.01	1	versicolor
06.02	02.02	04.05	01.05	versicolor
05.06	02.05	03.09	01.01	versicolor
05.09	03.02	04.08	01.08	versicolor
06.01	02.08	4	01.03	versicolor
06.03	02.05	04.09	01.05	versicolor

06.01	02.08	04.07	01.02	versicolor
06.04	02.09	04.03	01.03	versicolor
06.06	3	04.04	01.04	versicolor
06.08	02.08	04.08	01.04	versicolor
06.07	3	5	01.07	versicolor
6	02.09	04.05	01.05	versicolor
05.07	02.06	03.05	1	versicolor
05.05	02.04	03.08	01.01	versicolor
05.05	02.04	03.07	1	versicolor
05.08	02.07	03.09	01.02	versicolor
6	02.07	05.01	01.06	versicolor
05.04	3	04.05	01.05	versicolor
6	03.04	04.05	01.06	versicolor
06.07	03.01	04.07	01.05	versicolor
06.03	02.03	04.04	01.03	versicolor
05.06	3	04.01	01.03	versicolor
05.05	02.05	4	01.03	versicolor
05.05	02.06	04.04	01.02	versicolor
06.01	3	04.06	01.04	versicolor
05.08	02.06	4	01.02	versicolor
5	02.03	03.03	1	versicolor
05.06	02.07	04.02	01.03	versicolor
05.07	3	04.02	01.02	versicolor
05.07	02.09	04.02	01.03	versicolor
06.02	02.09	04.03	01.03	versicolor
05.01	02.05	3	01.01	versicolor
05.07	02.08	04.01	01.03	versicolor
06.03	03.03	6	02.05	virginica
05.08	02.07	05.01	01.09	virginica
07.01	3	05.09	02.01	virginica
06.03	02.09	05.06	01.08	virginica
06.05	3	05.08	02.02	virginica
07.06	3	06.06	02.01	virginica

04.09	02.05	04.05	01.07	virginica
07.03	02.09	06.03	01.08	virginica
06.07	02.05	05.08	01.08	virginica
07.02	03.06	06.01	02.05	virginica
06.05	03.02	05.01	2	virginica
06.04	02.07	05.03	01.09	virginica
06.08	3	05.05	02.01	virginica
05.07	02.05	5	2	virginica
05.08	02.08	05.01	02.04	virginica
06.04	03.02	05.03	02.03	virginica
06.05	3	05.05	01.08	virginica
07.07	03.08	06.07	02.02	virginica
07.07	02.06	06.09	02.03	virginica
6	02.02	5	01.05	virginica
06.09	03.02	05.07	02.03	virginica
05.06	02.08	04.09	2	virginica
07.07	02.08	06.07	2	virginica
06.03	02.07	04.09	01.08	virginica
06.07	03.03	05.07	02.01	virginica
07.02	03.02	6	01.08	virginica
06.02	02.08	04.08	01.08	virginica
06.01	3	04.09	01.08	virginica
06.04	02.08	05.06	02.01	virginica
07.02	3	05.08	01.06	virginica
07.04	02.08	06.01	01.09	virginica
07.09	03.08	06.04	2	virginica
06.04	02.08	05.06	02.02	virginica
06.03	02.08	05.01	01.05	virginica
06.03	03.04	05.06	02.04	virginica
06.04	03.01	05.05	01.08	virginica
6	3	04.08	01.08	virginica