Pengembangan Aplikasi Berbasis Web dengan R Shiny untuk Data Clustering Menggunakan Algoritma PCA

by John Doe

Submission date: 26-Nov-2021 02:37AM (UTC+0700)
Submission ID: 1712757261
File name: Paper_Kolokium.pdf (1.51M)
Word count: 3060
Character count: 19218

Pengembangan Aplikasi Berbasis Web dengan R Shiny untuk Data Clustering Menggunakan Algoritma PCA

***Abstrak*—Data menjadi kebutuhan yang sangat penting di zaman sekarang ini. Apapun pekerjaannya, manusia pasti selalu membutuhkan data untuk memproses pekerjaannya menjadi lebih informatif dan deskriptif. Ketika mereka mengumpulkan data-data tersebut menjadi satu padu maka akan terbentuk informasi yang menjadi lebih kompleks yang sering disebut sebagai dataset. Dalam dataset inilah banyak cara dari mereka untuk dapat mengolah kumpulan dari data tersebut menjadi sebuah informasi yang berguna. Salah satu ilmu yang dapat memproses data tersebut adalah sains data. Sains data juga memiliki banyak cabang di dalamnya. Tetapi yang terutama di dalam sains data, penggunaan algoritma *machine learning* menjadi pembeda dengan ilmu data yang lainnya. Disini penelitian akan berfokus membangun sebuah aplikasi berbasis *web* menggunakan R Shiny dimana penggunaan algoritma akan berfokus kepada salah satu *machine learning* yang disebut PCA dimana metode yang paling sering digunakan untuk algoritma ini adalah *data clustering*. Sehingga nantinya aplikasi ini dapat berguna bagi para pengguna yang ingin memproses suatu data dengan algoritma PCA sebagai metode dari *data clustering*.**

Kata Kunci—Dataset, Machine Learning, PCA, Data Clustering, RShiny

I. PENDAHULUAN

Di era teknologi yang selalu berkembang setiap harinya, tidak henti-hentinya dunia selalu membutuhkan data dalam proses kerja kehidupan manusia di dalamnya. Data disini akan berperan sebagai hasil yang didapatkan untuk mendefinisikan suatu deskripsi dari sebuah objek atau kejadian [1]. 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. 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 cikal bakal dari pemrosesan suatu kesimpulan dari 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 pagi pihak yang membutuhkannya dengan data tersebut terkhusus bagi *data scientist* atau *data analyst* [2]. Mereka kemudian dapat menggunakan data-data hasil dari sains data tersebut menjadi ke dalam sebuah bentuk *machine learning*. *Machine learning* merupakan suatu bidang studi yang bisa dikatakan baru dan bersumber dari algoritma kuantum yang dinamakan *supervised* (terpandu) dan *unsupervised* (mandiri) [3]. Algoritma *supervised* adalah suatu pencarian algoritma yang bersumber dari contoh yang sudah tersedia (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 sedangkan algoritma *unsupervised* adalah data menyimpulkan fungsi dari beberapa contoh data *training,* disini algoritma mencoba menemukan struktur yang tersembunyi di dalam data sehingga menimbulkan adanya klasterisasi data atau biasa disebut *data clustering* [4],[5].

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 [6]. Dengan adanya PCA ini, penelitian ini akan mencoba membangun sebuah sistem aplikasi berbasis *web* dengan sebuah *tool* yang digunakan R Shiny. R Shiny merupakan aplikasi *web development* dengan basis bahasa pemrograman R dimana sangat sesuai untuk menganalisis dan memproses sebuah dataset dengan algoritma *machine learning* yang digunakan. Nantinya sistem dari aplikasi tersebut dapat memvisualisasikan dan memperhitungan dari hasil model dataset yang di-input oleh pengguna aplikasi menggunakan metode algoritma PCA.

Dari seluruh rincian yang sudah dijelaskan, penelitian ini akan berfokus untuk membangun sebuah aplikasi web menggunakan data klasterisasi dengan algoritma PCA dengan alasan bahwa algoritma ini sangat mudah dalam memproses sebuah perhitungan data. PCA sendiri memiliki metode klasterisasi dalam pemrosesan dan visualisasi datanya sehingga aplikasi berharap dapat digunakan oleh semua orang dari kalangan apapun seperti bisnis, sosial, manufaktur, dan yang lainnya untuk menggunakan aplikasi

ini dalam upaya mengetahui penyebaran klasterisasi data yang diinput seperti apa dan dapat menghasilkan suatu nilai ataupun kesimpulan yang didapatkan dengan menggunakan perhitungan metode algoritma PCA.

II. TINJAUAN PUSTAKA

A. *Principal Component Analysis*

Principal Component Analysis atau disebut PCA adalah salah satu machine learning yang dikategorikan ke dalam algoritma unsupervised learning. PCA sendiri merupakan algoritma yang memiliki banyak cara membentuk dasar untuk analisis data yang bersifat multivariat [7]. Algoritma juga PCA memiliki beberapa tujuan di dalam menganalisis sebuah data antara lain :

1) *Modeling*
2) *Simplification*
3) *Dimensionality Reduction*

B. *Data Clustering*

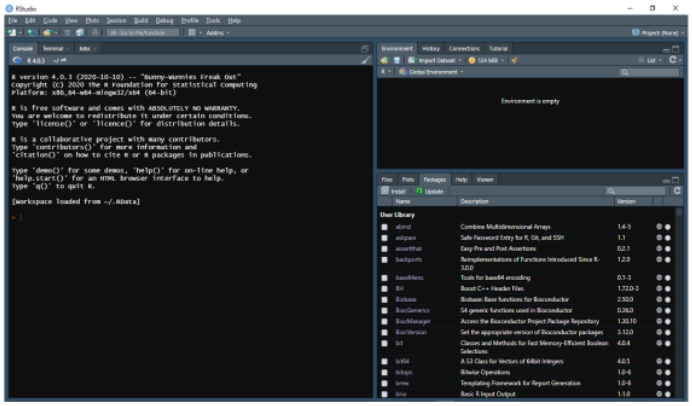
Data clustering atau klasterisasi data merupakan salah satu metode yang akan membagi suatu data ke dalam kelompok-kelompok kecil dengan objek yang sama atau serupa. Setiap kelompok disebut suatu klaster yang terdiri dari beberapa objek yang sama diantara klaster mereka sendiri namun berbeda dengan klaster objek yang lain [8].

C. *Bahasa Pemrograman R*

R merupakan salah satu bahasa pemrograman yang menjadi sarana pengolahan analisis data yang interaktif. Terlebih bahasa ini berkembang dengan cepat dan telah memiliki banyak *packages*. Namun R merupakan bahasa yang rata-rata berumur pendek dan dikembangkan untuk satu tujuan analisis yang digunakan seperti dengan penelitian ini [9].

D. *R Shiny*

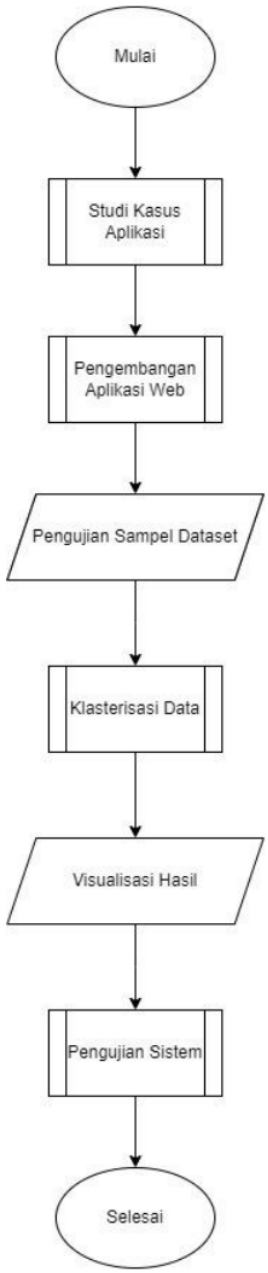
Shiny merupakan suatu *framework* aplikasi pengembangan *web* yang mana kebutuhan *environment* aplikasi tersebut menggunakan bahasa pemrograman R yang dikembangkan oleh RStudio [10]. Shiny merupakan aplikasi yang mudah dan reliabel penggunaanya sebagai pengembangan *web* karena langsung terintegrasi dengan bahasa R. Dengan aplikasi ini, sistem yang dibangun diharapkan akan menjadi lebih dinamis dan mudah digunakan kedepannya bagi para *user*.



Gambar 1. Aplikasi RStudio

Gambar 2. *Packages* RStudio

III. METODE PERANCANGAN SISTEM



Gambar 3. *Flowchart* perancangan sistem

Gambar 3 merepresentasikan langkah-langkah yang akan dibentuk dalam merancang sebuah sistem hingga menjadi produk sebuah aplikasi *web* menggunakan R Shiny.

Langkah-langkah dari *flowchart* yang bersifat kredensial berawal dari langkah pengujian sampel dataset hingga pengujian sistem. Jika proses tersebut berjalan sesuai alur maka secara otomatis pengembangan aplikasi juga akan bekerja secara baik nantinya.

A. Studi Kasus Aplikasi

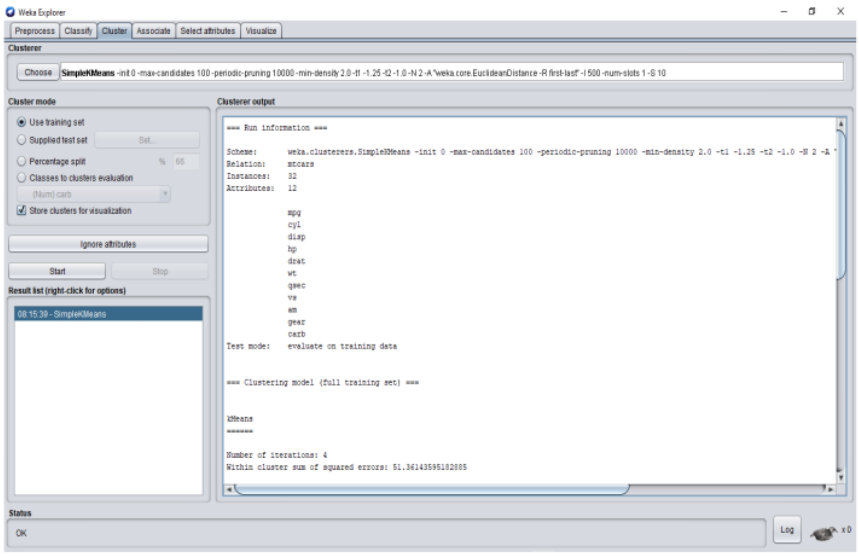
Disini perancangan sistem akan dimulai dengan mempelajari dan membandingkan aplikasi sejenis dengan basis sains data dan *machine learning* yang sudah ada untuk digunakan secara publik.

Perbandingan untuk studi kasus aplikasi ini hanya menggunakan salah satu aplikasi dengan fungsi aplikasi yang berhubungan dengan data, yaitu WEKA. Waikato Environment for Knowledge Analysis atau yang disingkat WEKA adalah aplikasi analisis untuk kebutuhan *data mining* tentunya dengan menggunakan algoritma *machine learning* yang juga terdapat di dalamnya. Para pengguna yang menggunakan aplikasi ini bertujuan untuk menggali informasi terhadap dataset yang dimilikinya. Banyak pilihan algoritma model yang ingin dijalankan baik dengan algoritma *supervised* maupun *unsupervised*. Tetapi sesuai dengan kebutuhan penelitian disini, studi kasus akan menggali algoritma *machine learning* menggunakan di aplikasi ini yang bertujuan sebagai *data clustering*.

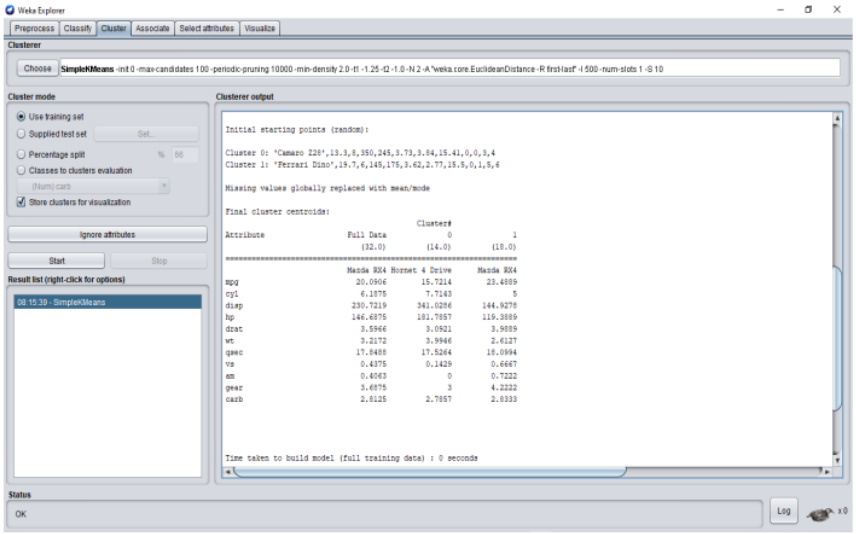
Gambar 4. Aplikasi Weka

Gambar 4 menjelaskan aplikasi WEKA yang masih dalam bentuk GUI. Studi kasus untuk aplikasi WEKA disini memilih pilihan "explorer" karena hanya disini pilihan yang bisa memilih metode klasterisasi pada suatu dataset.

Dalam aplikasi ini tidak ada algoritma PCA sebagai algoritma untuk klasterisasi datanya. Namun disini pilihan yang mendekati untuk klasterisasi data menggunakan algoritma K-Means. PCA dan K-Means merupakan algoritma yang identik sebagai algoritma untuk metode klasterisasi data karena mereka memiliki tujuan yang sama yaitu *dimensional reduction* [11]. Hal inilah yang menjadi alasan studi kasus aplikasi untuk perbandingan klasterisasi dengan algoritma yang lain. Dan berikut hasil dari klasterisasi dari dataset yang digunakan yaitu mtcars dimana ini merupakan kumpulan 32 *data entry* dari merk mobil dengan format dataset csv menggunakan algoritma K-Means.



Gambar 5. Keterangan dataset mtcars menggunakan WEKA dengan algoritma K-Means



Gambar 6. Hasil klasterisasi dataset mtcars menggunakan WEKA dengan algoritma K-Means

B. Pengembangan Aplikasi Web

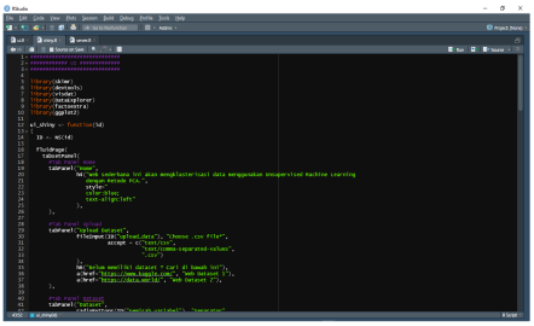
Proses selanjutnya setelah melakukan studi kasus pada aplikasi yang sejenis, maka penelitian dilanjutkan dengan membangun sebuah aplikasi *web* menggunakan R Shiny. Pada pembuatan *web* ini, sistem membutuhkan *script code* dengan format R sebanyak 3 jenis *script*. Penjelasan tiap *script* akan dijelaskan sebagai berikut :

1. *UI*

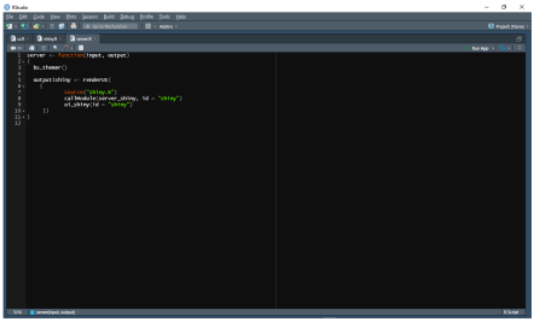
   Sesuai dengan namanya, pada *script* ini menjadi penunjang utama *user interface* (UI) untuk tampilan *web* dari sistem aplikasi ini. Penggunaan bahasa dalam *script* ini tidak hanya menggunakan bahasa R tapi juga dengan HTML. Di dalam *script* ini terdapat tema untuk kustomisasi tampilan yang diinginkan, gambar, dan judul sebagai profil tampilan di aplikasi nantinya. Berikut tampilan dari *code* di *script* UI yang disajikan pada gambar di bawah.

Gambar 7. UI

2. *Shiny*
   Untuk shiny sendiri merupakan isi seluruh *code* PCA dari awal hingga akhir termasuk *page* yang digunakan dan *library* yang harus disediakan. Sehingga isi dari *script code* disini menjadi penentu untuk isi tampilan dari aplikasi ini.
3. *Server*
   Server disini berfungsi sebagai konektor antara UI dan shiny untuk aplikasi *web*-nya. *Script* ini membuat UI menjadi *source* yang harus ditampilkan.



Gambar 8. Shiny



Gambar 9. Server

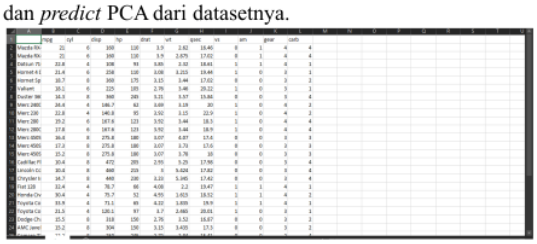
Gambar 8 dan 9 menjelaskan *script code* dari shiny dan server. Untuk *script code* shiny bukan menjadi opsi utama untuk penamannya melainkan bebas dengan pilihan masing-masing. Di dalam shiny terdapat isi *code* dari perhitungan algoritma PCA, klasterisasi, hingga visualisasi data yang menjadi *script* utama dalam pengembangan aplikasi ini.

C. *Pengujian Sampel Dataset*

Sampel dataset yang digunakan disini bersumber dari *R dataset packages* yang sudah tersedia di RStudio bernama *factoextra*. Namun dikarenakan dataset yang digunakan disini harus diinput secara manual ke aplikasi *web* dengan *soft file* yang tersedia (menggunakan format csv) maka sampel yang digunakan disini di-*download* terlebih dahulu yang didapatkan dari *public code* di Github. Untuk sampelnya sendiri masih sama digunakan seperti pada studi kasus aplikasi sebelumnya yaitu dataset dari mtcars yang berisikan 32 tipe merk mobil dengan 11 variabel di dalamnya.

D. *Klasterisasi Data*

Tahap ini menjadi kunci utama dalam pengembangan aplikasi *web* ini. Dalam proses ini akan melakukan pembuatan *code* untuk perhitungan dari algoritma PCA di dalam aplikasi menggunakan bahasa R. Untuk perhitungannya akan menggunakan *data model*, *summary,* dan *predict* PCA dari datasetnya.



Gambar 10. Dataset mtcars

Gambar 10 menunjukkan bahwa nama menjadi baris utama yang akan ditampilkan nantinya ke dalam plot visualisasi. Sebelas variabel lainnya menjadi perhitungan klasterisasi ke dalam *data model* untuk datasetnya.

E. *Visualisasi Hasil*

Setelah melakukan perhitungan dan klasterisasi, maka akan dilakukan penyebaran klasterisasi datanya dengan cara memvisualisasikan hasilnya menggunakan teknik *plotting*. Disini visualisasi menggunakan enam plot yaitu *scree plot*, *cos2 plot, individuals plot, contrib plot, variables plot,* dan *biplot* yang dihasilkan dari *data model*.

F. *Pengujian Sistem*

Proses terakhir dari rancangan penelitian ini adalah pengujian sistem. Pengujian sistem aplikasi ini harus dilakukan berulang kali hingga sempurna dari langkah pengujian sampel dataset hingga visualisasi hasilnya. Jika semua proses tersebut berjalan dengan lancar maka pengujian sistem aplikasi dinyatakan berhasil dan dapat berlanjut ke arah pembuatan aplikasinya.

IV. IMPLEMENTASI

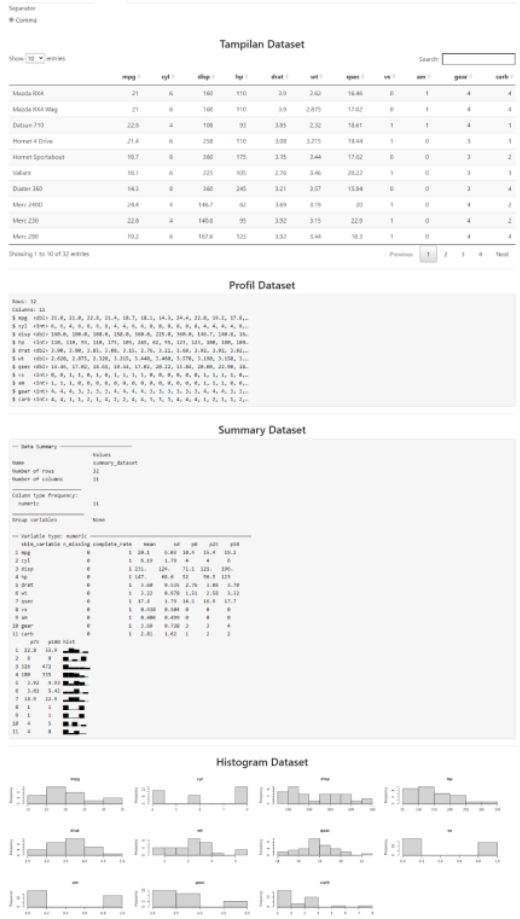
A. *Tampilan Aplikasi*

Tampilan aplikasi ini akan menjelaskan keseluruhan fitur yang bisa digunakan oleh *user* dan judul dari aplikasi sebagai penjelasan fungsi dari aplikasi *web* ini. Aplikasi ini juga sudah di-*hosting* menggunakan shinyapps. Tidak lupa pilihan tema yang disajikan juga beragam sehingga dapat menyesuaikan pilihan tampilan tema yang diinginkan oleh pengguna aplikasi.

Gambar 11. Tampilan aplikasi

B. Dataset

Profil Dataset akan merincikan bentuk datasetnya apakah numerik atau string. Statistik dari dataset ini juga menjadi profil dataset. Selain itu variabel dari dataset ini juga disajikan dalam bentuk histogram (diagram batang).



Gambar 12. Deskripsi dataset mtcars

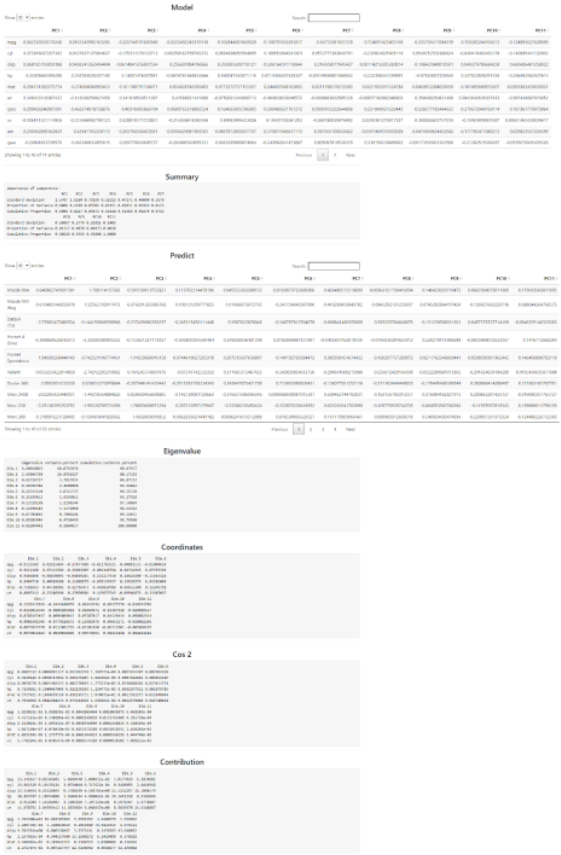
Pada gambar 12 terlihat terdapat tampilan dataset, profil dataset, hingga histogram dataset untuk nilai tiap variabelnya. Penyajian datasetnya disajikan dengan pemisah *comma* sebagai fungsi penyajian dataset dalam bentuk tabel (*dataframe*).

C. Perhitungan PCA

PCA akan diperhitungkan dengan tiga fungsi yang akan menjadi perhitungan klasterisasi datanya yaitu :

1. *Data Model* PCA
2. *Summary* PCA
3. *Predict* PCA
4. *Eigenvalue* PCA
5. *Coord* PCA
6. *Cos 2* PCA
7. *Contrib* PCA

Tampilan dari perhitungan PCA sesuai dengan fungsi yang dimasukkan di dalam *script code* dapat dilihat sesuai gambar di bawah ini.



Gambar 13. Perhitungan PCA

D. Visualisasi PCA

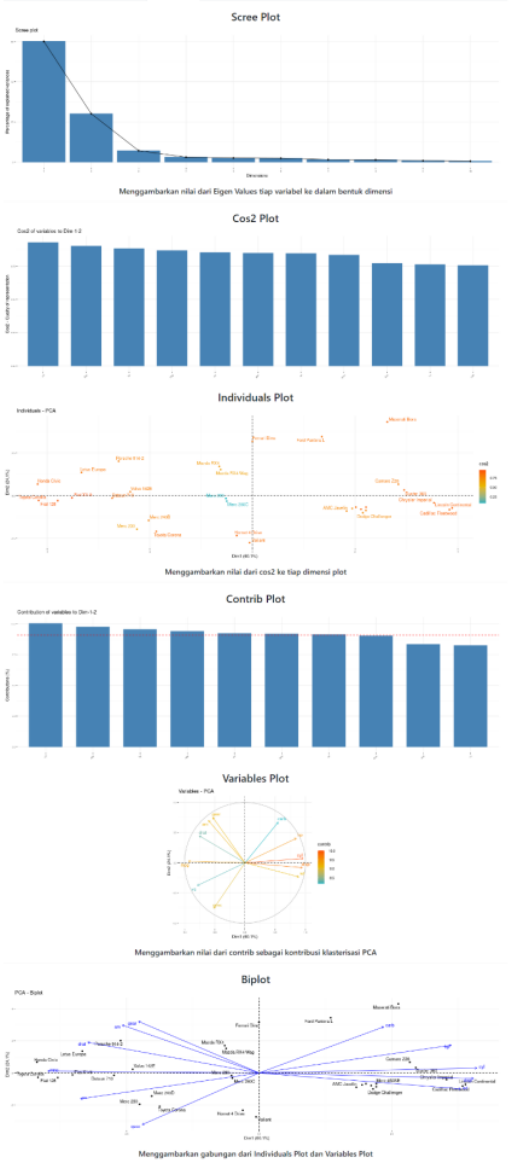
Setelah perhitungan dilakukan, maka visualisasi dalam aplikasi ini juga harus ditampilkan sebagai informasi yang akan didapatkan bagi para pengguna aplikasi. Enam plot yang disajikan akan mewakilkan informasi dari klasterisasi data menggunakan algoritma PCA dari perhitungan *model*.

Enam plot yang akan divisualisasikan ke dalam aplikasi ini yaitu :

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

Plot-plot ini akan mewakilkan klasterisasi setelah perhitungan PCA yang kompleks. Para pengguna yang tidak mengetahui algoritma PCA akan disuguhkan visualisasi ini tanpa harus mengetahui perhitungan tersebut secara keseluruhan. Sehingga para pengguna akan tetap dapat menarik informasi ini menjadi sebuah kesimpulan apa yang

didapatkan dari informasi dataset yang digunakan. Gambar di bawah merincikan bentuk dari visualisasi PCA.



Gambar 14. Visualisasi PCA

V. PEMBAHASAN

A. *Dataset*

Dataset akan disajikan rincian deskripsi dari isi datasetnya. Mulai dari banyaknya baris, banyaknya kolom (sebagai nilai variabel), nilai statistik, dan bentuk tipe data dari isi kolom variabel apakah numerik atau string. Sebelum menggunakan profil dataset ini, di dalam R dikhususkan untuk membutuhkan tiga *library* yang harus tersedia *dplyr*, *skimr*, dan *visdat*.

Library "dplyr" akan menjabarkan profil dataset yang (bentuk tipe data tiap variabel). Lalu *library "skimr"* menjelaskan nilai statistik dari variabel dataset yang digunakan. Terakhir, *"visdat"* sebagai *library* yang membuat fungsi histogram pada variabel dataset.

Terdapat syarat untuk dataset yang dapat digunakan untuk perhitungan algoritma PCA sebagai berikut :

1. Dataset memiliki nilai yang bersifat multivariat.
2. Isi dari kolom data yang diperhitungkan tidak membentuk suatu grup kecil atau bisa dikatakan bervariasi satu sama lain (tidak duplikat). Seperti contoh di atas menggunakan dataset mtcars, nama mobil yang akan diklasteriasi tidak boleh ada yang sama.
3. Kolom variabel tidak boleh ada yang bersifat string dikarenakan perhitungan ini merupakan perhitungan analisis yang bersifat numerik (*unsupervised*).
4. Nilai dari baris utama tidak boleh bernilai *null*. Dengan contoh dataset di atas, nama mobil tidak boleh ada yang *null* walaupun tetap memiliki nilai tiap variabelnya.

B. *Perhitungan PCA*

Model PCA menjadi tolak ukur perhitungan dari algoritma PCA kedepannya. Nantinya ini menjadi perhitungan utama ke dalam algoritma PCA (*summary, predict, eigenvalue, cos2,* dan *contrib*) hingga ke visualisasi datasetnya. *Library* yang dibutuhkan untuk mengaplikasikan PCA ke dalam R adalah *factoextra*. Fungsi ini sebagai syarat utama jika ingin menambahkan algoritma PCA ke dalam RStudio.

Model PCA menjadi dasar dari algoritma PCA dimana teknik yang digunakan dalam perhitungannya berdasarkan prinsip dari *dimensionality reduction* pada tiap nilai variabelnya.

```
output$hasil_modelPCA <- DT::renderDT(
  {
    hasil_modelPCA <- dataset()

    pcaModel <- prcomp(hasil_modelPCA, scale. = TRUE, center = TRUE)
    pcaModel$rotation
  })
```

Gambar 15. *Script code* model PCA

Pada gambar di atas menunjukkan *script code* dari penggunaan perhitungan model PCA di dalam R. Disini perhitungan menggunakan fungsi yang sederhana yang dinamakan "*prcomp*". Di dalam fungsi tersebut terdapat variabel "*hasil_modelPCA*" yang menggambarkan matriks numerik dari dataset yang digunakan dan *scale* sebagai deskripsi dari nilai variabel bersifat logis atau tidak (bergantung pada nilai *true/false*) yang menunjukkan apakah variabel harus diskalakan agar memiliki varians unit sebelum PCA dilakukan.

Selanjutnya akan dilakukan proses *summary* dari *model* PCA yang sudah diperhitungkan. Disini *summary* PCA menjelaskan nilai statistik varians yang didapatkan dari nilai *model* seperti *standard deviation*, *proportion of variance*, dan *cumulative proportion*.

Perhitungan yang selanjutnya merupakan *predict* PCA yang didapatkan juga dari nilai *model* PCA. Pada perhitungan ini akan menganalisa nilai dari tiap satuan individu pada data (mobil) dan tambahan variabel dari informasi perhitungan *model* sebelumnya.

Terakhir, terdapat perhitungan *eigenvalues* yang menjadi dasar utama dalam perhitungan dimensi tiap varians, *coord* yang menjelaskan koordinat dari tiap variabel, *cos 2* sebagai kualitas representasi nilai untuk variabel di dalam grafik visualisasi (*individuals plot*), dan *contrib* yang menjadikan fungsi yang berisi kontribusi (dalam persen) dari variabel ke PCA (*variables plot*). Untuk nilai *cos2* didapatkan dari hasil perhitungan nilai *coord* dan *contrib* didapatkan dari hasil perhitungan nilai *cos2*.

Nilai *cos2* berasal dari :

$$var.coord^2$$

sedangkan nilai *contrib* didapatkan dari :

$$\frac{var.cos2\ x\ 100}{total.cos2}$$

C. *Visualisasi PCA*

Langkah terakhir setelah perhitungan analisis PCA selesai maka akan dilakukan klasterisasi data yang akan divisualisasikan menggunakan plot sebagai informasi yang akan didapatkan.

Di dalam R, dibutuhkan dua *packages* sebagai fungsi menampilkan plot dan visualisasi dengan algoritma PCA yaitu *factoextra, corrplot,* dan *ggplot*. Untuk pembahasan tiap plot akan dijelaskan sebagai berikut.

1. *Scree plot* : menggambarkan nilai dari *eigenvalue* tiap variabel ke dalam bentuk dimensi
2. *Cos2 plot* : menggambarkan nilai dari *cos2*
3. *Individuals plot* : menggambarkan nilai dari individu yang dijelaskan dari fungsi *cos2* digunakan untuk memperkirakan kualitas representasi tiap dimensinya plotnya. (dijelaskan sesuai warna dari tiap plot).
4. *Contrib plot* : menggambarkan nilai dari *contrib*
5. *Variables plot* : menggambarkan nilai dari variabel yang dijelaskan dari fungsi *contrib* sebagai kontribusi dengan PCA yang akan diklasterisasi variabelnya. (dijelaskan sesuai warna dan arah panah).
6. *Biplot* : menggambarkan gabungan dari *individuals plot* dan *variables plot* (dijelaskan dengan arah panah yang mengarah ke plot).

Biplot menjadi informasi utama dalam visualisasi ini sebagai bentuk klasterisasi datanya dari tiap individu dan variabel membentuk sebuah klaster-klaster.

VI. KESIMPULAN

Dari penelitian dan sistem aplikasi *web* yang telah dibangun, maka terdapat kesimpulan sebagai berikut :

1. Aplikasi dapat menjalankan analisis menggunakan algoritma PCA dari sampel dataset yang telah digunakan.
2. Hasil perhitungan dari algoritma PCA sangat berpengaruh pada jenis dataset yang akan digunakan.
3. Visualisasi data merupakan hasil informasi dari klasterisasi yang diproses menggunakan algoritma PCA.

DAFTAR PUSTAKA

[1] F. Irmansyah, "Pengantar Database", 2003, bll 1–13.

[2] M. K. M. Nasution, "Sains Data", University of Sumatera Utara, 2019.

[3] A. B. Mutiara, R. Reanti, en D, "MACHINE LEARNING KUANTUM UNTUK SAINS DATA Penerbit Gunadarma, amutiara".

[4] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica (Ljubljana) Oct, vol 2007, bll 249–268.

[5] Lloyd, S., Mohseni, M., & Rebentrost, P. (n.d.). "Quantum algorithms for supervised and unsupervised machine learning", In arxiv.org. Retrieved June 27, 2021, from https://arxiv.org/abs/1307.0411.

[6] H. Abdi en L. J. Williams, "Principal component analysis", Wiley Interdiscip. Rev. Comput. Stat., vol 2, no 4, bll 433–459, Jul 2010.

[7] A. Directions, "Principal Component Analysis (PCA)", 2007, bll 1–12.

[8] O. A. Abbas, "Comparisons between data clustering algorithms", International Arab Journal of Information Technology (IAJIT), vol 5, no 3, 2008.

[9] Drs. A.P. Hardhono, M.Ed., Ph.D, and M.Kom. Dr. Imas Sukaesih Sitanggang, S.Si. n.d. "Pengenalan Dan Instalasi Perangkat Lunak Dan Lingkungan Pemrograman R", 2014, 1–29.

[10] J. Doi, G. Potter, J. Wong, I. Alcaraz, en P. Chi, "Web application teaching tools for statistics using R and shiny", Technology Innovations in Statistics Education, vol 9, no 1, 2016.

[11] Y. Liang, M. Balcan, en V. Kanchanapally, "Distributed PCA and k -Means Clustering", 2013, bll 1–8.

Pengembangan Aplikasi Berbasis Web dengan R Shiny untuk Data Clustering Menggunakan Algoritma PCA

ORIGINALITY REPORT

3%	3%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	fadilradit9.blogspot.com Internet Source	1%
2	www.scribd.com Internet Source	<1%
3	ARI YUNUS HENDRAWAN. "PENINGKATAN KINERJA ALGORITMA K MEANS DENGAN MENGGUNAKAN PARTICLE SWARM OPTIMIZATION DALAM PENGELOMPOKAN DATA PENYEDIAAN AKSES", Electro Luceat, 2020 Publication	<1%
4	ahlikuncijabodetabek.com Internet Source	<1%
5	text-id.123dok.com Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		