

Penggunaan R Studio untuk Analisis Multivariat pada Pemuliaan Kopi

Miftahur Rizqi Akbar¹⁾

¹⁾Pusat Penelitian Kopi dan Kakao Indonesia, Jl. PB Sudirman 90 Jember 68118

Analisis multivariat memiliki peranan yang penting dalam bidang pemuliaan secara umum. Penggunaan analisis multivariat pada pemuliaan kopi akan memberikan pandangan bagi pemulia dalam menentukan varietas yang memiliki sifat keunggulan tertentu baik ketahanan hama dan penyakit maupun cita rasa yang baik. Identifikasi karakter-karakter yang berpengaruh dominan membuat proses seleksi pada pemuliaan tanaman kopi menjadi lebih cepat. Aplikasi R Studio memberikan keunggulan dalam analisis multivariat karena berbasis *open source* yang gratis dan dapat digunakan untuk berbagai metode analisis multivariat serta memiliki kemudahan dalam pengoperasian dan tampilan visual yang menarik.



Analisis multivariat adalah analisis statistik yang melibatkan sekumpulan variabel. Setiap variabel dapat memiliki hubungan maupun tidak dengan variabel yang lain. Terdapat perbedaan istilah antara variabel dan variat dimana variabel adalah karakteristik, sifat, simbol, atau atribut yang memiliki nilai¹⁾. Variat adalah suatu kombinasi linear dari variabel-variabel yang memiliki bobot empiris yang telah ditentukan. Berdasarkan skala pengukurannya, variabel dapat terbagi menjadi empat, yaitu variabel berskala nominal, ordinal, interval, dan rasio¹⁾. Analisis multivariat berguna untuk melakukan analisis variabel dan populasi yang besar sehingga kita dapat memperoleh kesimpulan dari data yang besar. Penggunaan analisis multivariat akan membantu dalam seleksi genotipe kopi yang berjumlah ratusan dengan karakteristik tanaman kopi yang berbeda.

Analisis multivariat pada awalnya dikembangkan oleh Francis Galton and Karl Pearson pada

akhir abad ke-19 pada kuantifikasi hubungan antara keturunan dan karakteristik tetuanya, serta pengembangan koefisien korelasi. Kemudian pada awal abad 20, Charles Spearman meletakkan pondasi analisis faktor (*factorial analysis*)²⁾. Analisis multivariat terus berkembang setelah ditemukan permasalahan di bidang ilmu *science* dan pada tahun 1930 Fisher mengembangkan analisis fungsi diskriminan linier (*linear discriminant function analysis*) untuk menyelesaikan permasalahan pada bidang taksonomi. Fisher mengenalkan analisis varian (*analysis of variance*) pada tahun 1920 kemudian mengembangkan analisis varian multivariat (*multivariate analysis of variance*) atau biasa disebut MANOVA.

Metode analisis multivariat belum banyak berkembang pada awal ditemukannya yang disebabkan oleh keterbatasan perkembangan teknologi komputer sehingga metode analisis multivariat hanya terbatas pada pengembangan rumus matematika dan riset multivariat. Akan

tetapi pada pertengahan abad ke-20, penggunaan komputer mulai berkembang sehingga metode baru analisis multivariat mulai ditemukan. Pada awal tahun abad ke-21, teknologi komputer telah maju dan relatif murah serta didukung dengan ketersediaan *software* statistik sehingga metode analisis multivariat dapat diimplementasikan, bahkan untuk set data yang besar seperti ilmu genetik, *imaging*, dan astronomi²⁾.

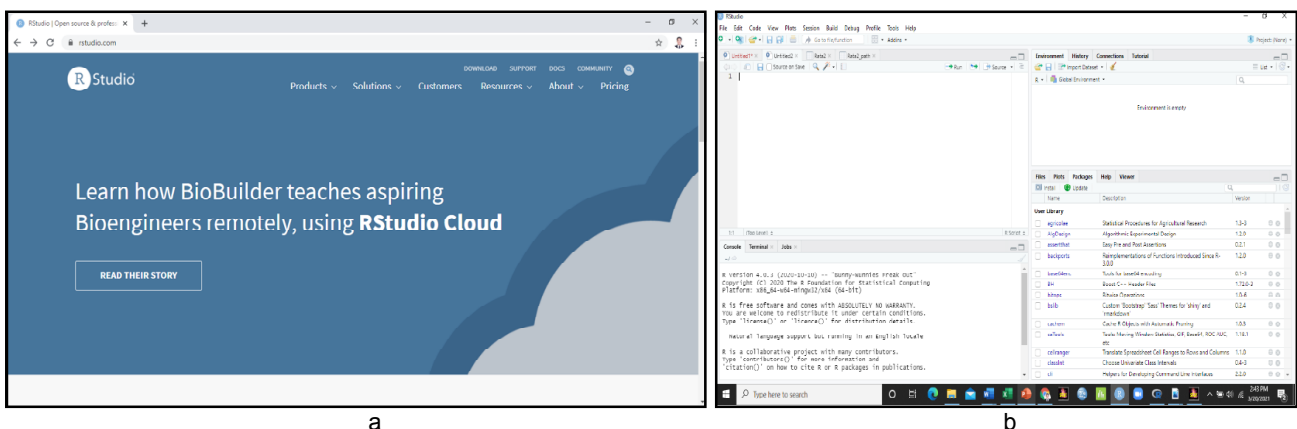
Analisis multivariat terdiri dari beberapa macam analisis, seperti *principal component analysis*, *multiple regression*, *discriminant analysis*, *multivariate analysis of variance*, *factor analysis*, *cluster analysis*, dan *multidimensional scaling*. Pada tulisan ini, penulis fokus pada pembahasan analisis komponen utama (*principal component analysis*).

Analisis komponen utama (*principal component analysis*) adalah analisis multivariat yang mereduksi dimensi set data yang terdiri dari variabel-variabel yang saling berkorelasi dengan mempertahankan keragaman sebesar mungkin pada set data yang baru³⁾. Hal tersebut dicapai dengan mentrans-formasi ke set variabel-variabel baru, *principal component* (PCs), yang tidak saling berkorelasi. Banyaknya komponen utama yang terbentuk sama dengan banyaknya variabel asli. Pereduksian (penyederhanaan) dimensi dilakukan dengan kriteria persentase keragaman data yang diterangkan oleh beberapa komponen utama pertama. Apabila beberapa komponen utama pertama telah menerangkan lebih dari 75% keragaman data asli, maka analisis cukup dilakukan sampai dengan komponen utama tersebut.

Analisis multivariat dapat dilakukan dengan berbagai software antara lain SPSS, SAS, MATLAB, dan R. Kelebihan software R adalah software ini bersifat *open source* dan gratis sehingga dapat digunakan secara luas. Keunggulan lainnya yaitu tersedianya *packages* lebih dari 10,000 *packages* hingga saat ini. Dengan jumlah *packages* yang banyak memungkinkan dapat dilakukan analisis data dengan berbagai metode. Saat ini *software* R telah tersedia versi yang lebih mudah yaitu R Studio. Instalasi R Studio dapat dilakukan dengan mengunjungi situs www.rstudio.com kemudian unduh dan instal aplikasi R Studio seperti ditunjukkan pada Gambar 1a dan 1b. Setelah melakukan instalasi, kita lanjutkan dengan menuliskan perintah untuk analisis multivariat.

Packages adalah sekumpulan kode yang memberikan fungsi baru pada R⁴⁾. Dalam analisis menggunakan R Studio, penggunaan *packages* dapat beragam disesuaikan dengan fitur masing-masing *packages*. Berikut disajikan contoh *raw data* tentang karakteristik kopi Robusta dan kopi Arabika (Tabel 1). Dari data tersebut akan dilakukan analisis komponen utama dengan menggunakan R Studio.

Langkah awal yang dilakukan untuk memulai analisis dengan R Studio adalah menyimpan file dalam bentuk “.csv” atau bentuk “.txt”. Kita memberikan nama file yaitu kopi2.csv. Penamaan file sesuai dengan keinginan dengan syarat masih menggunakan format yang sesuai. Selanjutnya kita membuat *syntax* atau sekumpulan perintah untuk menganalisis data kita.



Gambar 1. Situs download R Studio (a) dan tampilan setelah instalasi R Studio (b)

Tabel 1. Contoh data karakteristik kopi Arabika dan kopi Robusta

Genotipe	d. tajuk	p. ruas	p. daun	l. daun	b. 100	rendemen	p. buah	l. buah	potensi
Marogogipe	180	8.1	17.9	11.4	390	16	2.3	1.8	362.5
Mokka	176	2.6	3.4	2	36	12	0.6	0.4	172
Gayo 1	165	6.5	3.2	2.2	250	17.8	1.5	1.1	1875
Sigarar Utang	120	3.5	12	8	300	18.9	1.5	1.4	2000
Komasti	210	4.3	17.7	6.8	230	18	1.6	1.2	2100
USDA 762	185	4	12	8.2	187	17	1.4	1.1	2300
BP 42	220	5	29.3	12.2	200	22.9	2	1.6	1500
BP 358	240	7	29.1	12	180	21.4	1.6	1	2125
BP 534	180	6.5	18.4	6.6	200	21	1.5	1.4	2200
BP 936	175	6	22	6	300	20	2	1.5	2200
BP 939	178	6.6	22.4	10.2	320	21.1	1.3	1.5	1900
Sintaro 1	180	7	16	11	283	19.8	1.5	1.2	2125

```
#install
install.packages ("FactoMineR")

#load
library ("FactoMineR")

#install
install.packages ("factoextra")

#load
library ("factoextra")
```

Gambar 2. Listing instalasi packages untuk analisis komponen utama

```
library ("FactoMineR")
res.pca <- PCA(dat, graph = FALSE)

print (res.pca)

library ("factoextra")
eig.val <- get_eigenvalue(res.pca)
eig.val

var<- get_pca_var (res.pca)
var

print (var$contrib,9)
```

Gambar 3. Listing untuk penentuan nilai eigen dan nilai komponen utama

Beberapa packages dapat digunakan pada analisis komponen utama antara lain *FactoMwineR*, *ade4*, *stats*, *ca*, *MASS*, dan *ExPosition*. Kita akan menggunakan *FactoMineR package*⁵⁾ untuk

menganalisis komponen utama dan *factoextra package*⁶⁾ untuk melakukan visualisasi data dan interpretasi.

Tabel 2. Nilai eigen, persentase keragaman, dan keragaman kumulatif

	Nilai eigen	Persentase keragaman	Persentase keragaman kumulatif
Dim.1	4.6289	51.4329	51.4329
Dim.2	1.8044	20.0498	71.4828
Dim.3	1.1861	13.1799	84.6627
Dim.4	0.5890	6.5446	91.2073
Dim.5	0.3847	4.2750	95.4824
Dim.6	0.2279	2.5329	98.0153
Dim.7	0.0892	0.9921	99.0075
Dim.8	0.0556	0.6187	99.6262
Dim.9	0.0336	0.3737	100.0000

Tabel 3. Nilai komponen utama antar karakter

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
d.tajuk	2.319	36.2915	1.1487	6.3625	8.6839
p.ruas	10.8701	1.7396	3.9192	65.0343	6.4023
p.daun	14.7661	12.2329	6.8260	5.7573	0.0071
l.daun	13.8826	2.7731	3.9542	11.5034	36.6781
b.100	10.5251	24.8510	1.5852	0.2756	2.3713
rendemen	13.7813	5.5704	1.4797	0.0666	0.3920
p.buah	15.0751	3.8439	3.2657	0.3964	40.9199
l.buah	15.7363	9.7982	4.3645	6.7421	3.6475
potensi	3.0441	2.8989	6.1734	3.8614	0.8974

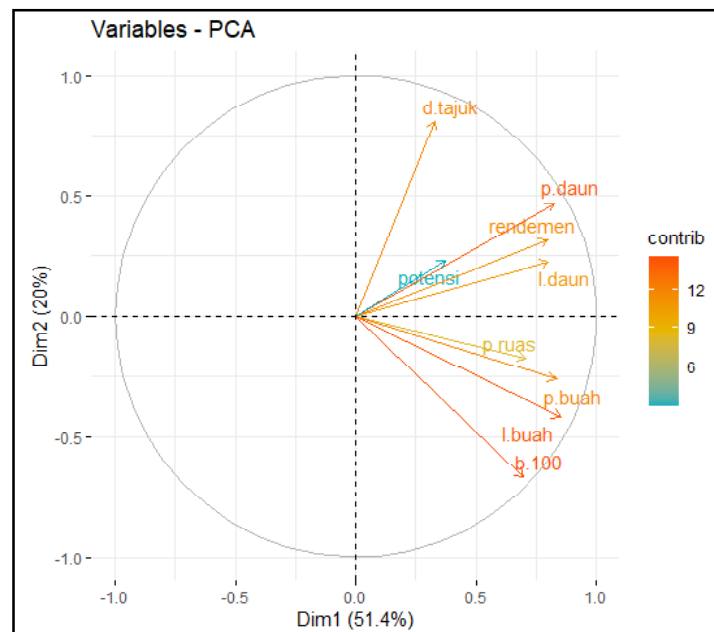
```

Library ("corrplot")
corrplot (var$contrib, is.corr = FALSE)

fviz_pca_var (res.pca,
col.var = "contrib", gradient.cols = c
("#00AFBB", "#E7B800",
"#FC4E07"),
repel = TRUE # Avoid text overlapping)

ind <- get_pca_ind (res.pca)
ind
    
```

a

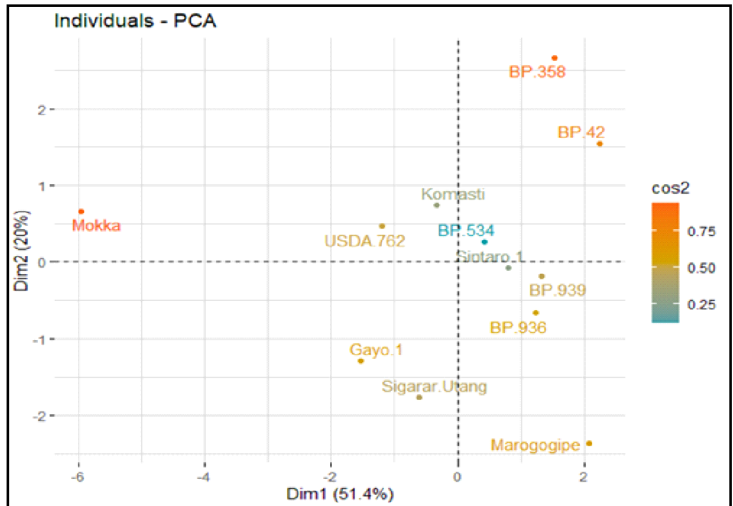


b

Gambar 4. Listing visualisasi PCA berdasarkan karakter (a) dan output analisis (b)

```
fviz_pca_ind(res.pca, col.ind = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800",
    "#FC4E07"),
  repel = TRUE

# Avoid text overlapping
(slow if many points)
```



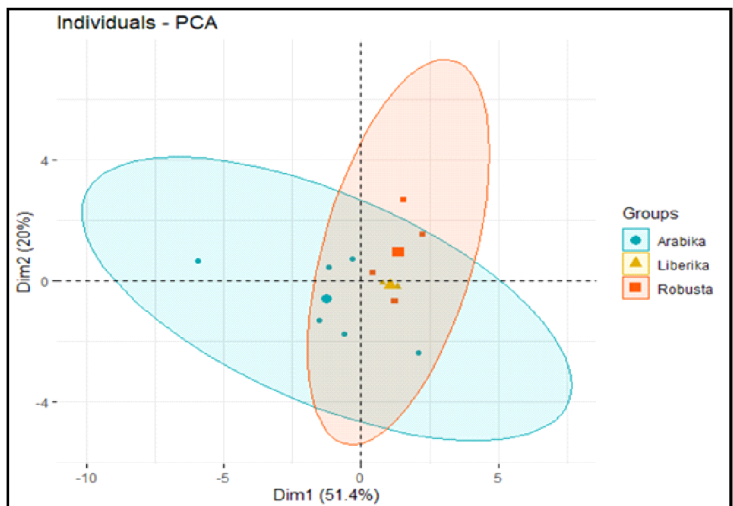
a

b

Gambar 5. Listing visualisasi PCA berdasarkan individu (a) dan output analisis (b)

```
# The variable Species (index = 10)
dihilangkan
dat4.pca <- PCA(dat4[,-10], graph = FALSE)

fviz_pca_ind(dat4.pca,
  label = "none", #hide individual labels
  habillage = dat4$Species, #color by
  groups palette = c("#00AFBB",
    "#E7B800", "#FC4E07"),
  addEllipses = TRUE # Concentration
  ellipses)
```



a

b

Gambar 6. Listing visualisasi PCA pengelompokan individu (a) dan output analisis (b)

Hasil pada Tabel 2. menunjukkan persentase kumulatif untuk 3 komponen sebesar 84.66% artinya dari 3 komponen analisis sudah dapat menggambarkan data berdasarkan analisis komponen utama. Tabel 3. menunjukkan nilai komponen pada masing-masing karakter. Nilai ini juga dapat digunakan dalam pembobotan karakter dalam pemilihan karakter penting dalam penentuan indeks seleksi⁷⁾. Hasil analisis komponen dapat kita visualisasikan berdasarkan karakter (Gambar 4.) dan individu (Gambar 5.). Pengelompokan dengan analisis komponen utama untuk individu disajikan pada Gambar 6.

Penutup

Aplikasi R Studio dapat melakukan analisis komponen utama dengan output data yang tidak berbeda dengan aplikasi statistik lain dan memiliki tampilan visual yang menarik, serta dapat dieksplorasi lebih mendalam, meskipun memiliki kekurangan pengoperasiannya yang sedikit lebih sulit. Berdasarkan keunggulan di atas, aplikasi R Studio memiliki manfaat yang besar untuk digunakan pada analisis multivariat untuk menunjang pemuliaan tanaman kopi.

Sumber Pustaka

- 1) Mattjik, A.A. & I.M. Sumertajaya (2013). Perancangan Percobaan dengan Aplikasi SAS dan Minitab, IPB Press, Bogor.
- 2) Kassambara, A. (2017). Practical Guide to Principal Component Method in R. STHDA, Merseille.
- 3) Jolliffe, I.T. (2002). Principal Component Analysis, Second Edition. Springer-Verlag New York, New York.
- 4) Barton, P. (2019). Learn the basics of statistical computing. <https://www.freecodecamp.org>.
- 5) Huson, F.; J. Josse; S. Le & J. Mazet (2017). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R packages version 1.36.
- 6) Kassambara, A. & F. Mundt (2017). Factoextra: Extract and Visualize The Results of Multivariate Data Analysis. R package version 1.0.5.9999.
- 7) Akbar, M.R.; B.S. Purwoko; I.S. Dewi & W.B. Suwarno (2018). Agronomic and drought tolerance evaluation of doubled haploid rice breeding lines derived from anther culture. *SABRAO Journal Breeding Genet.* 50(2), 115-128.

