

## Pengembangan Paket R Untuk Analisis Diskriminan Berbasis *Graphical User Interface* Web Interaktif

Nur Isra<sup>1</sup>, Suwardi Annas<sup>2</sup>, Muhammad Kasim Aidid<sup>3</sup>

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Kata Kunci: Analisis Diskriminan Linear, *Graphical User Interface*, Paket R.

### Abstrak:

Penggunaan perangkat lunak berlisensi memerlukan biaya yang relatif mahal, dan sulitnya memperoleh perangkat lunak berlisensi menjadi salah satu penyebab meningkatnya penggunaan perangkat lunak bajakan. Salah satu upaya dalam mengurangi tingkat permasalahan perangkat lunak bajakan adalah melakukan pengembangan perangkat lunak yang memiliki lisensi publik bersifat *open source* seperti perangkat lunak R. Penelitian ini dilakukan untuk menyusun beberapa paket yang terdapat pada perangkat lunak R yang akan memudahkan pengguna dalam melakukan analisis statistika, khususnya untuk analisis diskriminan linear. Paket pendukung R tersebut yaitu paket R-*Shiny* yang mampu membuat tampilan berbasis *Graphical User Interface*. Pengembangan paket R dalam penelitian ini menggunakan metode *waterfall*. Paket ini bernama *Linear Discriminant Analysis Application* (LDA App). Berdasarkan pengujian yang dilakukan pada LDA App menunjukkan bahwa LDA App mampu menyelesaikan analisis statistika sesuai fungsinya. Perbandingan antara LDA App dan *software* statistika lainnya memiliki output yang sama, akurat, dan lebih efisien dalam melakukan analisis diskriminan linear.

### 1. Pendahuluan

Statistika adalah suatu metode yang digunakan dalam pengumpulan dan analisis data untuk memperoleh informasi yang bermanfaat. Statistika menyediakan prinsip dan metodologi untuk merancang proses pengumpulan data, meringkas dan menyajikan data yang telah diperoleh, menganalisis, menginterpretasikan data serta pengambilan keputusan yang efektif (Siagian & Sugiarto, 2006).

Dalam melakukan analisis dan pengolahan data yang baik, cepat, efisien, dan akurat maka diperlukan perangkat lunak (*software*) sebagai alat bantu. Perangkat lunak merupakan prasyarat penting dalam menganalisis dan mengolah data. Alat analisis yang efisien merupakan modal utama yang digunakan seorang peneliti agar memperoleh hasil analisis yang tepat. Dimana alat analisis ini biasanya dihitung secara manual yang tentunya membutuhkan waktu lama, sehingga perlu untuk melakukan proses analisis dan pengolahan data dengan menggunakan perangkat lunak yang dapat mempermudah dalam proses analisis data yang efisien.

Perkembangan ilmu statistika yang diiringi dengan perkembangan teknologi yang pesat mengakibatkan munculnya alat analisis baru sekaligus perbaikan dari alat analisis sebelumnya. Sebagian besar alat analisis ini adalah alat analisis kompleks yang memerlukan waktu lama untuk menghitung secara manual (Sriwahyuni dkk., 2018). Saat ini tersedia berbagai macam perangkat lunak (*software*) bantu untuk statistika, baik itu *software* berbayar seperti SAS, STATA, SmartPLS dan SPSS atau *software* yang berbasis *Open Source* seperti R.

Penggunaan perangkat lunak berlisensi memerlukan biaya yang relatif mahal. Hal tersebut menjadi kendala dalam pelaksanaan penelitian untuk proses analisis data. Sulitnya memperoleh perangkat lunak tersebut menjadi salah satu penyebab meningkatnya penggunaan perangkat lunak *illegal*. Penggunaan perangkat lunak yang tidak memiliki lisensi resmi sangat marak di Indonesia (Surahmat & Tenggono, 2018).

Berdasarkan laporan dari BSA (*Business Software Alliance*) tahun 2015 tercatat bahwa peredaran perangkat lunak bajakan mencapai 14,4 Triliun Rupiah dengan tingkat peredaran perangkat lunak bajakan mencapai 84% dari jumlah perangkat lunak yang beredar (Surahmat & Tenggono, 2018). Pada tahun 2017 penggunaan perangkat lunak bajakan di Indonesia hanya turun satu persen menjadi 83%. Salah satu solusi dalam mengatasi masalah penyediaan

<sup>1</sup> Corresponding author.

E-mail address: nurisra62@gmail.com



perangkat lunak tersebut yaitu dengan menggunakan perangkat lunak berbasis *open source*. Penggunaan perangkat lunak *open source* juga dapat menghemat pengeluaran dana (Gartina, 2009).

Adapun upaya dalam mengurangi tingkat permasalahan perangkat lunak bajakan adalah melakukan pengembangan perangkat lunak yang memiliki lisensi publik bersifat *open source*. Perangkat lunak R merupakan perangkat lunak statistika yang berbasis *open source* berlisensi publik dan berbasis pemrograman. Namun, tidak semua orang terbiasa untuk menggunakan program R tersebut.

Oleh karena itu, pengembangan paket dalam perangkat lunak R dengan antarmuka *user friendly* sangat diperlukan untuk memudahkan penggunaan R dalam menganalisis data. Salah satu paket pendukung R tersebut yaitu paket R-*Shiny* yang berbasis *Graphical User Interface (GUI) web* interaktif tanpa perlu melakukan instalasi suatu aplikasi. R-*Shiny* merupakan *toolkit* dari R yang dapat digunakan membuat program secara online yang memungkinkan R dapat diakses melalui *web* karena dilengkapi dengan *server interface*.

Kebutuhan akan penggunaan metode statistika multivariat saat ini semakin meningkat. Hal tersebut disebabkan karena pada umumnya permasalahan yang ada lebih kompleks, yang mana melibatkan lebih dari satu variabel. Oleh sebab itu, diperlukan analisis yang lebih komprehensif dari metode statistika univariat. Statistika univariat hanya memperhatikan satu variabel terikat, adapun statistika multivariat melibatkan lebih dari satu variabel terikat (Johnson & Wichern, 2002). Statistika multivariat terbagi atas dua metode, yaitu model dependensi dan model interdependensi. Salah satu contoh metode statistika multivariat model dependensi yaitu Analisis Diskriminan.

Analisis diskriminan merupakan suatu metode statistika multivariat yang digunakan untuk mengelompokkan sejumlah objek ke dalam beberapa kelompok berdasarkan beberapa variabel, sehingga setiap objek menjadi anggota dari salah satu kelompok dan tidak ada objek yang menjadi lebih dari satu kelompok (Johnson & Wichern, 2002).

Berdasarkan hal tersebut, maka perlu adanya pengembangan paket R dalam analisis diskriminan berbasis *Graphical User Interface (GUI) web* interaktif yang dapat diakses melalui *web browser* sehingga proses analisis data dapat dilakukan dengan mudah. Beberapa penelitian yang telah melakukan pengembangan paket R, yaitu (Utomo, Ahmar, & Aidid, 2018) dengan Pengembangan Paket R untuk Analisis *Time Series* dengan *Graphical User Interface (GUI)*, dan (Paisal, Satyahadewi & Perdana, 2021) dengan Pengembangan Aplikasi Statistika Berbasis *Web* Interaktif untuk Analisis Uji- T. Adapun penelitian sebelumnya yang telah melakukan penelitian menggunakan metode analisis diskriminan, yaitu (Umam, 2018) dengan Analisis Diskriminan melalui metode Fisher terhadap Mahasiswa Hukum dalam Memilih Profesi, dan (Astuti dkk., 2021) dengan Pendekatan Analisis Diskriminan pada Pembelajaran Daring di Masa Pandemi Covid-19.

## 2. Kajian Pustaka

### 2.1 Rekayasa Perangkat Lunak

Rekayasa Perangkat Lunak merupakan disiplin ilmu yang berfokus pada semua aspek dalam menghasilkan sebuah perangkat lunak. Dalam menerapkan rekayasa perangkat lunak, pengembang menggunakan pendekatan yang sistematis dan terorganisir, menggunakan *tools* dan teknik yang bergantung pada pemecahan masalah, batasan sistem, serta *resources* lain yang ada (Sommerville, 2011). Metode yang digunakan dalam pengembangan perangkat lunak yaitu metode *waterfall*. Tahapan dalam pengembangan perangkat lunak menurut Sommerville, yaitu analisis kebutuhan dan definisi, perancangan sistem dan perangkat lunak, implementasi dan pengujian unit, integrasi dan pengujian sistem, serta operasi dan pemeliharaan.

### 2.2 Perangkat Lunak R

Perangkat lunak R adalah salah satu perangkat lunak yang mengolah data serta menganalisis statistik berbasis bahasa pemrograman yang bersifat *open source*, mengalami perkembangan pesat mengikuti kebutuhan analisis data. Perangkat lunak R memiliki banyak fungsi, umumnya divisualisasikan dalam jendela window atau disimpan dalam berbagai format misalnya jpg, png, bmp, ps, emf, pictex dan xfig. R dapat digunakan untuk mengekstrak atau pengambilan data, manajemen data (*organizing*), penampilan data (*visualizing*), pemodelan data (*modeling*) (Chamber, 2008).

### 2.3 Paket Shiny

*Shiny* adalah paket R pada bahasa pemrograman R yang dapat mengembangkan perangkat lunak berbasis *web* dengan menggunakan paket *Shiny Web Framework* yang memudahkan dalam pembuatan aplikasi *web* interaktif langsung dari R (RStudio, 2020). *Shiny* merupakan program dalam R yang mampu membuat *Graphical User Interface* yang berbasis *web* interaktif. *Web GUI* yang dimaksud merupakan *web* yang dapat berinteraksi atau berkomunikasi dengan R berupa mengirim informasi atau permintaan dalam bentuk menu grafis ke R serta menampilkan kembali hasilnya pada laman *web* sebagaimana luaran yang dihasilkan oleh R baik berupa teks atau grafik (Wijayanti dkk., 2020). Ada dua kelompok komponen pada program R *Shiny*, yaitu *user interface* dan *server* (Tirta, 2014).

### 2.4 Analisis Diskriminan

Analisis Diskriminan merupakan metode statistika untuk mengelompokkan sejumlah objek ke dalam sejumlah kelompok berdasarkan beberapa peubah yang membentuk fungsi diskriminan, sedemikian sehingga setiap objek menjadi anggota dari salah satu kelompok. Tujuan analisis diskriminan pada dasarnya untuk mengetahui peubah yang dapat membedakan anggota populasi. Analisis ini dilakukan dengan membentuk kombinasi linear dari sejumlah peubah, sehingga nilai diskriminan dari setiap objek yang terlihat merupakan petunjuk pada objek tersebut untuk menjadi anggota pada satu kelompok tertentu. Apabila banyaknya kelompok yang dilibatkan lebih dari dua, pada umumnya diperlukan lebih dari satu fungsi diskriminan untuk pegelompokan objek-objek tersebut (Tiro, Sukarna, Aswi, 2010).

### 2.5 Fungsi Diskriminan

Fungsi diskriminan adalah kombinasi linear dari peubah-peubah yang diamati yang akan menghasilkan cara terbaik dalam pemisahan kelompok tersebut. Fungsi ini akan memberikan nilai yang sedekat mungkin dalam kelompok dan sejauh mungkin antar kelompok. Fungsi ini digunakan untuk menjelaskan perbedaan antar kelompok dan juga untuk pengklasifikasian (Annas & Irwan, 2015). Fungsi diskriminan linear Fisher untuk kasus dua kelompok, dapat ditulis sebagai berikut:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p \tag{1}$$

di mana:

$Y$  menyatakan skor diskriminan

$b_p$  menyatakan pembobot diskriminan untuk variabel ke-p

$X_i$  menyatakan variabel bebas ke-p.

Aturan klasifikasi analisis diskriminan berdasarkan peluang posterior terbesar. Misalkan  $\pi_k$  adalah peluang prior dari kelompok  $k, k=1, \dots, m$ , dan  $f_k(x)$  adalah fungsi sebaran bersyarat  $x$  pada kelompok  $k$ . Peluang posterior suatu pengamatan  $x$  berasal dari kelompok  $k$ , adalah (Johnson & Wichern, 2002):

$$P(Y = k|x) = \frac{f_k(x)\pi_k}{\sum_{i=1}^m f_i(x)\pi_i} \tag{2}$$

nilai  $\pi_k$  dapat diduga sebagai

$$\pi_k = \frac{\text{banyaknya amatan dari kelas } k}{\text{banyaknya seluruh amatan}}$$

Pada *linear discriminant analysis*,  $f_k(x)$  diasumsikan berupa sebaran multivariate normal

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)} \tag{3}$$

dimana,

$p$  merupakan dimensi (banyaknya peubah)

$\mu_k$  merupakan vektor rata-rata

$\Sigma_k$  merupakan matriks ragam-peragam

Suatu observasi dengan karakteristik  $x$  akan diklasifikasikan sebagai anggota kelompok 0 jika  $P(Y = 0|x) > P(Y = 1|x)$ .

Dalam analisis diskriminan, nilai akurasi hasil klasifikasi dapat dihitung dengan menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk melakukan analisis seberapa baik *classifier* dalam mengenali data dari kelas yang berbeda. Tabel *confusion matrix* dapat dilihat pada tabel 1 berikut:

**Tabel 1 Confusion Matrix**

Aktual	Prediksi	
	Positif	Negatif
Positif	True Positive (TP)	False Positive (FP)
Negatif	False Negative (FN)	True Negative (TN)

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

### 2.6 Uji Normalitas

Analisis diskriminan harus memenuhi asumsi normalitas. Tujuan dari uji normalitas untuk mengetahui apakah distribusi sebuah data mengikuti atau mendekati distribusi normal. Uji normalitas dilakukan pada setiap variabel dengan logika bahwa jika secara individual masing-masing variabel memenuhi asumsi normalitas, maka secara bersama-sama (multivariat) variabel tersebut juga dianggap memenuhi asumsi normalitas (Santoso, 2017).

Pengujian asumsi tersebut adalah *Shapiro-Wilks' Test*. Uji Statistik Shapiro-Wilk didasarkan pada suatu sampel acak berukuran  $n, x_1, x_2, \dots, x_n$  yang didefinisikan sebagai:

$$W_x = \frac{\bar{\sigma}_x^2}{s_x^2} \quad (5)$$

Uji ini akan tolak  $H_0$  dengan suatu ukuran taraf  $\alpha$  jika  $W_x < k_\alpha$  dengan  $k_\alpha$  merupakan persentil  $100\alpha\%$  dari distribusi  $W_x$  (Villasenor Alva & Estrada, 2009).

## 2.7 Uji Kehomogenan Matriks Ragam-peragam

Uji kehomogenan matriks ragam-peragam juga dibutuhkan dalam uji signifikan pembeda peubah dan fungsi diskriminan. Prosedur pada uji Box M ini menggunakan pendekatan sebaran khi-kuadrat dengan derajat bebas  $\frac{1}{2}(g-1)p(p+1)$  (Morrison, 1990).

Statistik uji yang digunakan adalah Box'M, yaitu:

$$M = (\sum_i (n_i - 1) \ln |S_{pooled}| - \sum_i [(n_i - 1) \ln |S_i|]) \quad (6)$$

$$C = (1 - \mu)M$$

dimana

$k$  : banyak kelompok

$|S_{pooled}|$  : matriks kovarian gabungan dalam kelompok

$S_i$  : matriks kovarian kelompok ke-i

Daerah kritis yaitu Tolak  $H_0$  jika  $C > \chi^2_{p(p+1)(g-1)/2}(\alpha)$ , yang artinya dapat disimpulkan bahwa matriks varian kovarian antar kelompok tidak homogen (Morrison, 1990). Selain itu, juga dapat dibuktikan dengan p-value (nilai probabilitas)  $< \alpha(0,05)$  maka tolak  $H_0$  yang artinya matriks varian kovarian antar kelompok tidak sama atau tidak homogen.

## 2.8 Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk menguji apakah adanya korelasi antara variabel bebas (independent) (Santosa, 2005). Secara teori, idealnya dalam analisis diskriminan tidak ada multikolinearitas antar variabel. Asumsi yang harus terpenuhi adalah tidak ada korelasi antara variabel bebas. Jika variabel bebas mempunyai korelasi yang kuat, maka dikatakan terjadi multikolinieritas.

Pengecekan multikolinearitas dapat dilihat dengan bantuan matriks korelasi. Nilai korelasi untuk multikolinearitas adalah sampai sebesar 0,8, dengan kriteria uji, tolak  $H_0$  jika nilai korelasi antar variabel ( $r$ )  $< 0,8$ . Apabila ada dua variabel memiliki nilai korelasi  $r > 0,8$  maka variabel tersebut berkorelasi sehingga terdapat multikolinearitas antar variabel. Nilai korelasinya dihitung dengan menggunakan Korelasi Pearson (Yendra & Desvina, 2019). Namun, jika hanya sebagian kecil variabel bebas yang berkorelasi, maka asumsi terjadi multikolinieritas dapat diabaikan dan proses analisis diskriminan tetap dilanjutkan (Rizkiana & Hendikawati, 2016). Analisis diskriminan tidak terlalu sensitif dengan pelanggaran asumsi ini, kecuali pelanggarannya bersifat ekstrim (Hair dkk.,1987).

## 3. Metodologi penelitian

### 3.1 Jenis Penelitian

Jenis penelitian ini adalah penelitian *research and development*. Penelitian ini dilakukan dengan membuat pengembangan paket R berbasis *Graphical User Interface web* interaktif agar lebih memudahkan pengguna dalam melakukan analisis diskriminan.

### 3.2 Sumber Data

Sumber data yang digunakan pada penelitian ini merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) yaitu data IPM (Indeks Pembangunan Manusia) Provinsi Sulawesi Utara tahun 2021 (<https://www.bps.go.id/>).

### 3.3 Teknik Analisis Data

Teknik analisis data yang dilakukan dalam penelitian ini terdiri atas dua bagian, yaitu:

#### 3.3.1 Pengembangan Paket R

Pengembangan paket R berbasis *Graphical User Interface* ini digunakan untuk melakukan analisis diskriminan. Pengembangan perangkat lunak tersebut mengikuti tahapan (Sommerville, 2011) dengan menggunakan metode *waterfall*. Metode *waterfall* dianggap lebih terstruktur dalam perancangan dan pengembangan sistem. Metode *waterfall* mempunyai tahapan utama yang mencerminkan aktifitas pengembangan dasar (Sommerville, 2011). Adapun tahapan metode *waterfall* menurut Sommerville sebagai berikut:

1. Analisis kebutuhan sistem
2. Perancangan sistem dan perangkat lunak
3. Implementasi dan pengujian unit
4. Integrasi dan pengujian sistem
5. Operasi dan pemeliharaan

### 3.3.2 Analisis Diskriminan

Adapun teknik analisis data pada analisis diskriminan, yaitu:

1. Menentukan objek serta variabel yang akan dianalisis
2. Melakukan uji asumsi (uji normalitas, uji homogenitas, dan uji multikolinearitas)
3. Megestimasi fungsi diskriminan
4. Menghitung tingkat ketepatan klasifikasi hasil prediksi fungsi diskriminan
5. Melakukan validasi fungsi pengelompokan
6. Menginterpretasikan hasil

## 4. Hasil dan Pembahasan

### 4.1 Analisis Kebutuhan Sistem

Analisis diskriminan merupakan metode statistika multivariat untuk mengelompokkan sejumlah objek ke dalam beberapa kelompok, sehingga setiap objek menjadi anggota dari salah satu kelompok. Analisis diskriminan juga dapat melihat tingkat keakuratan pengelompokan yang terbentuk.

Pada penelitian ini, paket yang dikembangkan yaitu *Linear Discriminant Analysis Application* (LDA App) yang menyediakan tampilan antarmuka web interaktif yang bersifat *user friendly* dengan R sehingga memudahkan pengguna dalam melakukan analisis diskriminan. *Linear Discriminant Analysis Application* (LDA App) digunakan untuk melakukan analisis diskriminan linear yang menyediakan:

1. Statistika Deskriptif (*Group Means*)
2. Uji Asumsi (Uji Normalitas, Uji Homogenitas, dan Uji Multikolinearitas)
3. Fungsi Diskriminan
4. Prediksi dari Klasifikasi
5. Akurasi Klasifikasi Analisis Diskriminan

Ada beberapa paket penunjang untuk pengembangan *Linear Discriminant Analysis Application* (LDA App), dimana paket tersebut sudah terdapat pada *software* R. Adapun paket yang digunakan yaitu paket *shiny*, *shinythemes*, *shinycssloaders*, *rstatix*, *mvShapiroTest*, *MASS*, dan *tidyverse*.

### 4.2 Perancangan Sistem

Pada tahap perancangan sistem dilakukan dengan membuat diagram alir terlebih dahulu. Pada tahap ini mencakup pembuatan sistem mulai dari tahap input hingga tahap output. Diagram alir dibuat dengan tujuan agar proses perancangan sistem yang akan dibuat lebih terarah sehingga sistem dapat berjalan dengan baik. Ada dua komponen dalam membentuk program dengan *shiny*, yaitu *ui.r* dan *server.r*. *user-interface* ini digunakan untuk mendesain tampilan dari program yang akan dibuat dan *server* merupakan inti dari program yang akan melakukan simulasi berbagai analisis data dan selanjutnya mengirim hasilnya ke bagian output.

### 4.3 Implementasi Sistem

Pada tahapan implementasi sistem ini digunakan paket *shiny* sebagai tampilan *Graphical User Interface* (GUI) yang memudahkan pengguna dalam melakukan analisis diskriminan. *Linear Discriminant Analysis Application* (LDA App) ini terdiri atas beberapa tab menu yaitu menu *upload* data, pemilihan variabel, hasil (deskriptif, uji asumsi, hasil analisis diskriminan), bantuan dan tentang.

#### 4.3.1 Menu Data

Menu data merupakan menu pertama pada tampilan aplikasi tersebut. Tab menu data ini berfungsi untuk melakukan input data yang akan dianalisis. Adapun tipe data yang dapat diinput yaitu berupa file dengan format *.csv* (*comma separated value*), *.xlsx* (*microsoft excel worksheet*), atau *.txt* (*text document*).



Gambar 1 Menu Data LDA App

#### 4.3.2 Menu Pemilihan Variabel

Menu pemilihan variabel digunakan untuk memilih variabel terikat (*dependent variable*) dan variabel bebas (*independent variable*) yang akan dianalisis.



Gambar 2 Menu Pemilihan Variabel

#### 4.3.3 Menu Hasil

##### 1. Statistika Deskriptif

Menu statistika deskriptif berfungsi untuk menampilkan nilai statistik data yaitu rerata kelompok (*group means*).



Gambar 3 Menu Statistika Deskriptif

##### 2. Uji Asumsi

Menu uji asumsi berfungsi untuk melakukan uji asumsi sehingga diketahui apakah data sudah memenuhi uji asumsi atau tidak. Menu uji asumsi terdiri dari uji normalitas, uji kehomogenan matriks ragam-peragam, dan uji multikolinearitas.

##### a) Uji normalitas

Uji normalitas dilakukan untuk melihat apakah variabel bebas (*independent variable*) berdistribusi normal atau tidak. Data yang tidak berdistribusi normal akan menyebabkan masalah pada ketepatan fungsi diskriminan.



Gambar 4 Menu Uji Normalitas

##### b) Uji kehomogenan matriks ragam-peragam

Uji kehomogenan matriks ragam-peragam dilakukan untuk menguji varians dari setiap variabel dengan menggunakan uji *Box's M*. Asumsi yang terpenuhi adalah matriks varian kovarian antar kelompok relatif sama (homogen).



**Gambar 5** Menu Uji Kehomogenan

c) Uji multikolinearitas

Uji multikolinearitas dilakukan untuk menguji apakah terdapat korelasi yang kuat antar variabel bebas (*independent variable*). Asumsi yang terpenuhi adalah tidak ada korelasi antar variabel bebas (*independent variable*).



**Gambar 6** Menu Uji Multikolinearitas

### 3. Hasil Analisis Diskriminan

Menu hasil analisis diskriminan berfungsi untuk mengetahui hasil estimasi fungsi diskriminan, prediksi klasifikasi dan akurasi untuk memperoleh kesimpulan.



**Gambar 7** Menu Hasil Analisis Diskriminan

a) Fungsi Diskriminan

Fungsi diskriminan merupakan kombinasi linear dari peubah yang diamati yang akan menghasilkan cara terbaik dalam pemisahan kelompok.

b) Prediksi Klasifikasi

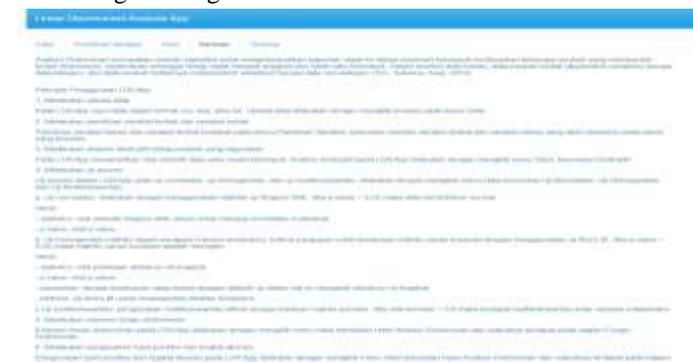
Prediksi klasifikasi pada LDA App dilakukan berdasarkan nilai dari score diskriminan dan peluang posterior.

c) Tingkat Akurasi

Nilai akurasi hasil klasifikasi pada LDA App dihitung dengan menggunakan *confusion matrix*.

#### 4.3.4 Menu Bantuan

Menu bantuan berfungsi untuk membantu pengguna dalam melakukan analisis diskriminan. Pada menu bantuan ini terdapat deskripsi singkat dan langkah-langkah dalam melakukan analisis diskriminan.



**Gambar 8** Menu Bantuan

#### 4.3.5 Menu Tentang

Menu tentang merupakan menu tambahan yang berisi tentang informasi pengembang dan bentuk pencapaian telah selesainya sistem.



Gambar 9 Menu Tentang

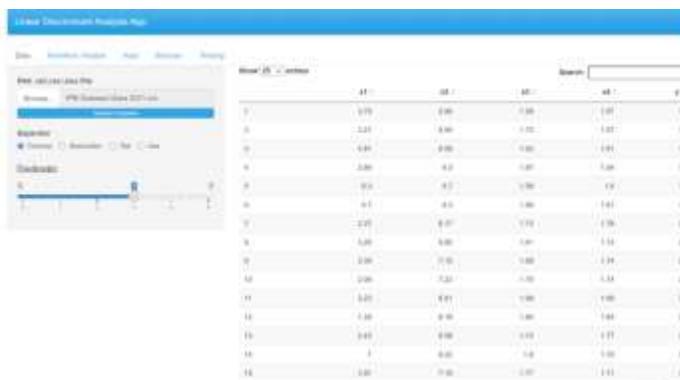
#### 4.4 Integrasi Sistem

Pada tahap integrasi sistem, *Linear Discriminant Analysis Application* dilakukan integrasi dari keseluruhan fungsi-fungsi perangkat. Data yang digunakan untuk melakukan pengujian pada perangkat LDA App merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS). Pada tahap pengujian dilakukan dengan membandingkan hasil output dari LDA App dengan perangkat lunak lainnya, yaitu dengan membandingkan hasil *output* dari *R Studio* dengan *Linear Discriminant Analysis Application* (LDA App).

#### 4.5 Pengujian

Pada tahap pengujian, dilakukan pengujian LDA App menggunakan contoh kasus dengan tujuan untuk melihat hasil dari perangkat yang telah dibuat. Pengujian dilakukan dengan menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) yaitu data IPM (Indeks Pembangunan Manusia) Provinsi Sulawesi Utara tahun 2021. Berikut penggunaan LDA App pada data tersebut sebagai berikut:

1. Melakukan *upload* data.



Gambar 10 Penginputan Data pada LDA App

Gambar 10 pada LDA App menunjukkan bahwa pada menu data akan ditampilkan data yang telah diinput.

2. Melakukan pemilihan variabel terikat dan variabel bebas.

Pemilihan variabel terikat dan variabel bebas terdapat pada menu pemilihan variabel, kemudian memilih variabel terikat dan variabel bebas yang akan dianalisis pada panel yang disediakan.



Gambar 11 Pemilihan Variabel pada LDA App

Gambar 11 pada LDA App menunjukkan bahwa pada menu pemilihan variabel menampilkan variabel terikat (*dependent variable*) dan variabel bebas (*independent variable*) yang akan dianalisis.

3. Melakukan analisis deskriptif data.  
 Pada LDA App ini menampilkan nilai statistik data yaitu rerata kelompok (*group means*).



**Gambar 12** Statistika Deskriptif pada LDA App

Gambar 12 pada LDA App menunjukkan bahwa pada menu deskriptif menampilkan statistik deskriptif data yaitu rerata kelompok (*group means*).

**Tabel 2** Hasil perbandingan LDA App dan R Studio

Group Means	LDA App	R Studio
X <sub>1</sub>	1: 4,245	1: 4,245
	2: 2,448	2: 2,448
X <sub>2</sub>	1: 6,468	1: 6,468
	2: 7,420	2: 7,420
X <sub>3</sub>	1: 1,625	1: 1,625
	2: 1,733	2: 1,733
X <sub>4</sub>	1: 1,600	1: 1,600
	2: 1,751	2: 1,751

Berdasarkan Tabel 2 menunjukkan bahwa hasil dari *group means* antara LDA App dan R Studio sama. Hasil *Group Means* yaitu untuk x<sub>1</sub> kelompok 1 sebesar 4,245 dan kelompok 2 sebesar 2,448; untuk x<sub>2</sub> kelompok 1 sebesar 6,468 dan kelompok 2 sebesar 7,42; untuk x<sub>3</sub> kelompok 1 sebesar 1,625 dan kelompok 2 sebesar 1,733; dan untuk x<sub>4</sub> kelompok 1 sebesar 1,6 dan kelompok 2 sebesar 1,751.

4. Melakukan uji asumsi pada data untuk mengetahui apakah data tersebut memenuhi uji asumsi atau tidak.
  - a) Uji normalitas

Pada uji normalitas akan dibandingkan hasil *p-value* dari LDA App dan R Studio.



**Gambar 13** Uji Normalitas pada LDA App

Gambar 13 pada LDA App menunjukkan bahwa pada menu uji normalitas menampilkan hasil dari uji normalitas yang terdiri dari *statistics of shapiro-wilk* dan *p-value* dari variabel bebas (*independent variable*).

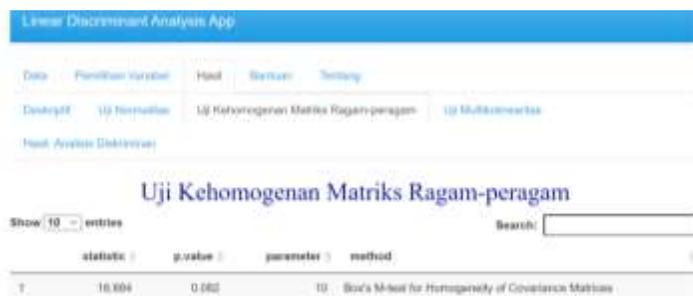
**Tabel 3** Hasil perbandingan LDA App dan R Studio

Uji Asumsi	LDA App	R Studio
	p-value	p-value
Normalitas	$x_1 : 0,653$	$x_1 : 0,653$
	$x_2 : 0,131$	$x_2 : 0,131$
	$x_3 : 0,707$	$x_3 : 0,707$
	$x_4 : 0,544$	$x_4 : 0,544$

Berdasarkan Tabel 3 menunjukkan bahwa hasil nilai *p-value* antara LDA App dan R Studio sama. Hasil yang diperoleh yaitu nilai *p-value* untuk  $x_1$  sebesar 0,653; untuk  $x_2$  sebesar 0,131; untuk  $x_3$  sebesar 0,707; dan untuk  $x_4$  sebesar 0,544.

b) Uji kehomogenan matriks ragam-peragam

Pada uji kehomogenan matriks ragam-peragam akan dibandingkan hasil *p-value* dari LDA App dan R Studio.



**Gambar 14** Uji Kehomogenan pada LDA App

Gambar 14 pada LDA App menunjukkan bahwa pada menu uji kehomogenan matriks ragam-peragam menampilkan hasil dari uji kehomogenan matriks ragam-peragam yang terdiri dari nilai statistik uji, *p-value*, dan parameter.

**Tabel 4** Hasil perbandingan LDA App dan R Studio

Uji Asumsi	LDA App	R Studio
	p-value	p-value
Kehomogenan	0,082	0,082

Berdasarkan Tabel 4 menunjukkan bahwa hasil nilai *p-value* antara LDA App dan R Studio sama. Hasil uji asumsi kehomogenan pada LDA App dan R Studio diperoleh nilai *p-value* sebesar 0,082.

c) Uji multikolinearitas

Pada uji multikolinearitas akan dibandingkan hasil nilai korelasi antar variabel dari LDA App dan R Studio.



**Gambar 15** Uji Multikolinearitas pada LDA App

Gambar 15 pada LDA App menunjukkan bahwa pada menu uji multikolinearitas menampilkan hasil dari uji multikolinearitas yaitu nilai korelasi antar variabel bebas.

**Tabel 5** Hasil perbandingan LDA App dan R Studio

Uji Asumsi	LDA App	R Studio
Multikolinearitas	$x_1-x_2$ : -0,7	$x_1-x_2$ : -0,7
	$x_1-x_3$ : -0,8	$x_1-x_3$ : -0,8
	$x_1-x_4$ : -0,6	$x_1-x_4$ : -0,6
	$x_2-x_3$ : 0,8	$x_2-x_3$ : 0,8
	$x_2-x_4$ : 0,7	$x_2-x_4$ : 0,7
	$x_3-x_4$ : 0,6	$x_3-x_4$ : 0,6

Berdasarkan Tabel 5 menunjukkan bahwa hasil nilai korelasi antara LDA App dan R Studio sama. Hasil uji asumsi multikolinearitas pada LDA App dan R Studio diperoleh nilai korelasi antar variabel bebas, yaitu untuk nilai korelasi antara  $x_1$  dan  $x_2$  sebesar -0,7; nilai korelasi antara  $x_1$  dan  $x_3$  sebesar -0,8; nilai korelasi antara  $x_1$  dan  $x_4$  sebesar -0,6; nilai korelasi antara  $x_2$  dan  $x_3$  sebesar 0,8; nilai korelasi antara  $x_2$  dan  $x_4$  sebesar 0,7; dan nilai korelasi antara  $x_3$  dan  $x_4$  sebesar 0,6.

5. Setelah dilakukan uji asumsi, selanjutnya mengestimasi fungsi diskriminan.



**Gambar 16** Estimasi Fungsi Diskriminan pada LDA App

Gambar 16 pada LDA App menunjukkan bahwa pada menu Hasil Analisis Diskriminan bagian fungsi diskriminan menampilkan hasil dari estimasi fungsi diskriminan.

**Tabel 6** Hasil perbandingan LDA App dan R Studio

Hasil Analisis	LDA App	R Studio
Estimasi Fungsi Diskriminan	$b_1$ : -0,161	$b_1$ : -0,161
	$b_2$ : -0,136	$b_2$ : -0,136
	$b_3$ : 4,104	$b_3$ : 4,104
	$b_4$ : 20,858	$b_4$ : 20,858

Berdasarkan Tabel 6 menunjukkan bahwa hasil nilai estimasi fungsi diskriminan antara LDA App dan R Studio sama. Hasil dari pemodelan diperoleh fungsi diskriminan :

$$Y = -0,161(x_1) - 0,136(x_2) + 4,104(x_3) + 20,852(x_4)$$

6. Setelah mengestimasi fungsi diskriminan, selanjutnya yaitu melakukan pengecekan hasil prediksi klasifikasi dan tingkat akurasi.

a) Prediksi Klasifikasi

Prediksi klasifikasi pada LDA App dilakukan berdasarkan nilai dari score diskriminan dan peluang posterior.



**Gambar 17** Prediksi Klasifikasi pada LDA App

Gambar 17 pada LDA App menunjukkan bahwa pada menu Hasil Analisis Diskriminan bagian prediksi klasifikasi menampilkan hasil dari prediksi klasifikasi berdasarkan fungsi diskriminan.

**Tabel 7** Hasil Prediksi Klasifikasi pada LDA App

Aktual	Prediksi	
	1	2
1	6	0
2	0	9

Selanjutnya, melihat hasil output prediksi klasifikasi pada R *Studio* untuk memastikan output pada LDA App dan R *Studio* sama.

**Tabel 8** Hasil Prediksi Klasifikasi pada R *Studio*

Aktual	Prediksi	
	1	2
1	6	0
2	0	9

Berdasarkan Tabel 7 dan Tabel 8 menunjukkan bahwa hasil prediksi klasifikasi antara LDA App dan R *Studio* sama. Hasil prediksi klasifikasi diperoleh bahwa jumlah kelompok 1 yang berhasil dikelompokkan dengan benar berjumlah 6 data, dan kelompok 2 berjumlah 9 data.

b) Akurasi

Nilai akurasi hasil klasifikasi pada LDA App dihitung dengan menggunakan *confusion matrix*. Perhitungan tingkat akurasi klasifikasi terdapat pada persamaan 4.



**Gambar 18** Tingkat Akurasi pada LDA App

Gambar 18 pada LDA App menunjukkan bahwa pada menu Hasil Analisis Diskriminan bagian akurasi menampilkan hasil dari tingkat akurasi.

**Tabel 9** Hasil perbandingan LDA App dan R *Studio*

Hasil Analisis	LDA App	R Studio
Tingkat Akurasi	100 %	100 %

Berdasarkan Tabel 9 menunjukkan bahwa hasil tingkat akurasi antara LDA App dan R Studio sama. Hasil tingkat akurasi diperoleh dengan menggunakan persamaan 4, yaitu:

$$Akurasi = \frac{6 + 9}{6 + 0 + 9 + 0} = \frac{15}{15} = 1 \times 100 = 100\%$$

#### 4.6 Operasi dan Pemeliharaan

Pada tahap operasi dan pemeliharaan, LDA App sudah dapat digunakan oleh pengguna, agar perangkat lunak dan paket lain yang digunakan pada perangkat tersebut tetap berjalan dengan baik maka dilakukan pemeliharaan terhadap perangkat lunak tersebut.

#### 4.7 Batasan Sistem

LDA App merupakan perangkat lunak yang dikembangkan dengan paket *Shiny* dan paket penunjang lainnya. Adapun batasan-batasan yang ada pada perangkat, yaitu:

1. Memiliki ketergantungan pada perangkat lunak R
2. Hanya bisa digunakan untuk analisis diskriminan linear, yaitu:
  - a. Statistika Deskriptif (*Group Means*)
  - b. Uji Asumsi (Uji Normalitas, Uji Homogenitas, dan Uji Multikolinearitas)
  - c. Estimasi Fungsi Diskriminan
  - d. Prediksi dari Klasifikasi
  - e. Akurasi

### 5. Kesimpulan dan Saran

#### 5.1 Kesimpulan

*Linear Discriminant Analysis Application* (LDA App) merupakan paket yang dikembangkan dengan salah satu paket pendukung R yaitu paket R-*Shiny* berbasis *Graphical User Interface web* interaktif yang digunakan untuk analisis diskriminan linear. Pada LDA App terdapat beberapa menu yaitu menu data, pemilihan variabel, statistika deskriptif, uji asumsi, hasil analisis diskriminan, bantuan, dan tentang. Pengujian yang dilakukan pada perangkat menunjukkan bahwa LDA App mampu menyelesaikan analisis statistika yang telah ditentukan sesuai dengan fungsinya. Perbandingan antara LDA App dan *software* statistika lainnya memiliki output yang sama, akurat, dan lebih efisien dalam melakukan analisis diskriminan linear. Paket ini dipublikasikan melalui *shinyapps.io* sehingga memudahkan pengguna untuk mengakses melalui *web browser*.

#### 5.2 Saran

Saran untuk penelitian selanjutnya yaitu:

1. Mengembangkan paket R untuk analisis diskriminan lainnya, seperti analisis diskriminan kuadratik.
2. Melakukan inovasi terkait dengan pengembangan paket statistika lainnya dan menambah fitur yang lebih menarik sehingga lebih memudahkan pengguna dalam melakukan analisis, seperti fitur download dari hasil analisis.

### Referensi

Annas, S., & Irwan. (2015). Penerapan Analisis Diskriminan Dalam Pengelompokan Desa Miskin Di Kabupaten Wajo. *Scientific Pinsi*, 1(1), 34–43.

- Astuti, C. C., Wiguna, A., & Ariyanti, N. (2021). Pendekatan Analisis Diskriminan Pada Pembelajaran Daring Di Masa Pandemi Covid-19. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 15(1), 137–146.
- Badan Pusat Statistik. (2021). *Indeks Pembangunan Manusia*. BPS: Jakarta.
- Chamber, J. (2008). *Software for Data Analysis, Springer Statistical and Computing*. New York: Springer Verlag New York.
- Gartina, D. (2009). Penggunaan Software Open Source Dalam Mendukung Kegiatan Penelitian dan Administrasi Perkantoran. *Informatika Pertanian*, 18(1), 45–62.
- Hair Jr, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis 7 Edition*. New Jersey: Prentice Hall.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis, Fifth Edition*. New Jersey: Prentice Hall.
- Morrison, D.F. (1990). *Multivariate Statistical Methods Third Edition*. USA : Mc Graw Hill Inc.
- Paisal, Satyahadewi, N., & Perdana, H. (2021). Pengembangan Aplikasi Statistika Berbasis Web Interaktif Untuk Analisis Uji- T. *Jurnal Penelitian Ilmiah*, 10(3), 331–340.
- Rizkiana, P., & Hendikawati, P. (2016). Klasifikasi Tingkat Kepuasan Pelanggan Dengan Analisis Diskriminan. *Jurnal MIPA*, 38(1), 89–100.
- RStudio. (2020). *Shiny From RStudio*. Retrieved from Shiny: <https://shiny.rstudio.com/>
- Santosa PBA. (2005). *Analisis Statistik dengan MS. Excel dan SPSS*. Yogyakarta: Andi.
- Santoso, S. (2017). Statistik Multivariat dengan SPSS. Jakarta: PT. Elex Media Komputindo.
- Sarvina, Y. (2017). Pemanfaatan Software Open Source “ R ” Untuk Penelitian Agroklimat “ R ” Open Source Software for Agroclimate Research. *Informatika Pertanian*, 26(1), 23–30.
- Siagian, D., & Sugiarto. (2006). *Metode Statistika untuk Bisnis dan Ekonomi*. Jakarta: PT. Gramedia Pustaka Utama.
- Sommerville, I. (2011). *Software Engineering (9th ed.; Boston, Ed.)*. Massachusetts: Pearson Education.
- Sriwahyuni, A. A., Annas, S., & Ahmar, A. S. (2018). Pengembangan Paket Analisis Regresi Berbasis Web. *Seminar Nasional Variansi (Venue Artikulasi-Riset, Inovasi, Resonansi-Teori, Dan Aplikasi Statistika)*, 1, 31–45.
- Surahmat, S., & Tenggono, A. (2018). Evaluasi Penggunaan Aplikasi Office Berbasis Open Source Pada SMKN Kota Palembang Dengan Pendekatan Technology Acceptance Model. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 7(2), 98–103.
- Tiro, M. A., Sukarna, & Aswi. (2010). *Statistika Deskriptif Peubah Banyak*. Makassar: Andira Publisher.
- Tirta, I. M. (2014). Pengembangan E-Modul Statistika Terintegrasi dan Dinamik dengan R-shiny dan mathJax. *Prosiding Seminar Nasional Matematika, Universitas Jember*, 1(1), 223–232.
- Umam, K. (2018). Analisis Diskriminan Melalui Metode Fisher Terhadap Mahasiswa Hukum Dalam Memilih Profesi. *Penelitian Multidisiplin*, 01(01), 91–100.
- Utomo, A. T., Ahmar, A. S., & Aidid, M. K. (2017). *Pengembangan Paket R untuk Analisis Time Series Dengan Graphical User Interface ( GUI )*. Eprints Universitas Negeri Makassar.
- Villasenor Alva, J. A., & Estrada, E. G. (2009). A generalization of Shapiro-Wilk’s test for multivariate normality. *Communications in Statistics - Theory and Methods*, 38(11), 1870–1883.
- Wijayanti, D., Sugito, S., & Yasin, H. (2020). Analisis Model Antrean Non-Poisson Dan Ukuran Kinerja Sistem Berbasis Gui Web Interaktif Menggunakan R-Shiny (Studi Kasus: Bus di Terminal Penggaron Kota Semarang). *Jurnal Gaussian*, 9(4), 444–453.
- Yendra, R., & Desvina, A. P. (2019). Analisis Diskriminan Fisher Untuk Klasifikasi Risiko Kredit. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*, 11(1), 478–481.