

**LAPORAN AKHIR  
EKSPLORASI DAN VISUALISASI DATA  
KELAS SD-A2**

**Analisis Pengaruh Keterlibatan Mahasiswa Teknologi Sains Data Universitas  
Airlangga dalam Masa Pemilihan Umum Tahun 2024**



**Oleh Kelompok 9 :**

Ilham Dicky Darmawan	164221023
Rizal Hidayatulloh	164221031
Arkan Syafiq At`taqy	164221062
Giovanni Eki Pamungkas	164221090
Salwa Sabian Raihana	164221113

**SEMESTER GASAL 2023/2024  
PROGRAM STUDI TEKNOLOGI SAINS DATA  
FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN  
UNIVERSITAS AIRLANGGA**

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>DAFTAR ISI</b>	<b>ii</b>
<b>DAFTAR TABEL</b>	<b>iii</b>
<b>DAFTAR GAMBAR</b>	<b>iv</b>
<b>BAB I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
<b>BAB II TINJAUAN PUSTAKA</b>	<b>3</b>
2.1 Tinjauan Pustaka Statistika	3
2.2 Tinjauan Pustaka Non Statistika	5
<b>BAB III METODOLOGI</b>	<b>7</b>
3.1 Sumber Data	7
3.2 Metode Data Pre-processing	7
3.3 Metode Analisis Data	8
3.4 Metode Visualisasi Data	8
<b>BAB IV ANALISIS DAN PEMBAHASAN</b>	<b>9</b>
4.1 Deskripsi Data	9
4.2 Data Pre-processing	10
4.3 Hasil Analisis (Prediksi/Klasifikasi/Klaster)	14
4.4 Visualisasi Data R-Shiny	15
<b>BAB V KESIMPULAN DAN SARAN</b>	<b>19</b>
5.1 Kesimpulan	19
5.2 Saran	20
<b>DAFTAR PUSTAKA</b>	<b>21</b>
<b>LAMPIRAN</b>	<b>22</b>

## DAFTAR TABEL

Tabel 1 : Variabel Penelitian	9
Tabel 2 : Statistika Deskriptif	10
Tabel 3 : Hasil Uji Grubbs, Lavarge, Cooks	12

## DAFTAR GAMBAR

Gambar 1 : Visualisasi Pengecekan Missing Values	10
Gambar 2 : Visualisasi Boxplot Cek Outlier	11
Gambar 3 : Visualisasi Boxplot Pembersihan Outlier	12
Gambar 4 : Encoding Data Kategorik	13
Gambar 5 : Reduksi Dimensi Data	13
Gambar 6 : Output Regresi Linear Berganda	14
Gambar 7 : Heatmap Korelasi Variabel	15
Gambar 8 : Scatter Plot	16
Gambar 9 : Pie Chart	16
Gambar 10 : Histogram & Density Plot	17
Gambar 11 : Box Plot	18

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pemilihan umum adalah wujud sistem politik yang demokratis dan ajang bagi masyarakat untuk menentukan wakil-wakil di pemerintahan. Momentum pemilihan umum seperti pemilihan presiden dan wakil presiden selalu menarik untuk dibahas. Hal ini memicu masyarakat untuk turut ramai memperbincangkan dan terlibat di dalamnya, termasuk memperbincangkan persoalan-persoalan politik, aktor-aktor politik, dan tantangan-tantangan politik.

Generasi milenial diproyeksi menjadi kelompok pemilih dengan proporsi terbesar mencapai 50% di Pemilu 2024. Oleh karena itu, perhatian terhadap keterlibatan politik rakyat dalam pemilu menjadi penting, karena tinggi rendahnya partisipasi ialah tanda serta indikasi penting berfungsinya demokrasi dan perwujudan kedaulatan rakyat. Selain itu, generasi milenial tidak hanya menjadi kelompok pemilih terbesar, tetapi juga menjadi kelompok yang paling aktif dalam menggunakan media sosial. Oleh karena itu, media sosial menjadi platform yang penting dalam kampanye politik, baik untuk menyebarkan informasi, mempengaruhi opini publik, maupun untuk mendengarkan aspirasi masyarakat.

Dalam konteks mahasiswa Teknologi Sains Data Universitas Airlangga, penelitian ini bertujuan untuk menganalisis pengaruh masa pemilihan umum tahun 2024 terhadap keterlibatan mereka. Penelitian ini dilakukan dengan metode kuisisioner online yang mencakup pertanyaan tentang Umur, Jenis Kelamin, Angkatan, Pemahaman, Akademis, Diskusi, Kepercayaan, Hubungan, Aktivitas, Sumber Informasi, dan Keterlibatan. Hasil dari penelitian ini diharapkan dapat memberikan gambaran tentang sejauh mana mahasiswa terlibat dalam pemilihan umum dan faktor-faktor apa saja yang mempengaruhinya.

## **1.2 Rumusan Masalah**

1. Bagaimana pengaruh keterlibatan mahasiswa teknologi sains data Universitas Airlangga berdasarkan faktor-faktor seperti umur, jenis kelamin, angkatan, pemahaman, akademis, diskusi, kepercayaan, hubungan, aktivitas, dan sumber informasi dalam konteks Pemilihan Umum 2024?
2. Apakah terdapat perbedaan signifikan dalam keterlibatan mahasiswa teknologi sains data Universitas Airlangga berdasarkan faktor-faktor seperti umur, jenis kelamin, angkatan, pemahaman, akademis, diskusi, kepercayaan, hubungan, aktivitas, dan sumber informasi dalam konteks Pemilihan Umum 2024?

## **1.3 Tujuan Penelitian**

1. Untuk mengetahui pengaruh keterlibatan mahasiswa teknologi sains data Universitas Airlangga berdasarkan faktor-faktor seperti umur, jenis kelamin, angkatan, pemahaman, akademis, diskusi, kepercayaan, hubungan, aktivitas, dan sumber informasi dalam konteks Pemilihan Umum 2024.
2. Untuk mengetahui apakah terdapat perbedaan signifikan dalam keterlibatan mahasiswa teknologi sains data Universitas Airlangga berdasarkan faktor-faktor seperti umur, jenis kelamin, angkatan, pemahaman, akademis, diskusi, kepercayaan, hubungan, aktivitas, dan sumber informasi dalam konteks Pemilihan Umum 2024.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Pustaka Statistika**

##### **2.1.1 Statistika**

Statistika adalah ilmu mengumpulkan, menata, menyajikan, menganalisis, dan menginterpretasikan data menjadi informasi untuk membantu pengambilan keputusan yang efektif. Istilah statistika dapat diartikan sebagai metode mengumpulkan, mengolah, menyajikan, menganalisis, dan menginterpretasikan data dalam bentuk angka (Dajan, 1995).

##### **2.1.2 Sampling**

Sampling merupakan proses dalam menyeleksi observasi (sampel) untuk menyediakan deskripsi yang cukup dan menguatkan inferensi dari suatu populasi. Teknik sampling adalah merupakan teknik pengambilan sampel untuk menentukan sampel yang akan digunakan dalam penelitian (Garalka, 2019).

##### **2.1.3 Populasi dan Sampel**

Populasi dan sampel merupakan elemen yang penting dalam sebuah penelitian. Populasi merupakan keseluruhan unit yang akan dijadikan sebagai objek pengamatan. Sedangkan sampel merupakan bagian dari populasi yang terpilih untuk mewakili sebuah populasi (Shukla, 2020). Karakteristik sampel yang baik adalah ukuran sampel yang sesuai, unit sampel terpilih dapat mewakili populasi, pengambilan sampel efektif atau menghemat waktu, biaya, dan tenaga, serta sampel harus dipilih secara adil tanpa ada bias. Adanya sampel sangat penting untuk mempercepat waktu penelitian dan menghemat biaya.

##### **2.1.4 Outlier**

Outlier adalah data pengamatan yang tidak konsisten pada seriesnya. Efek kejadian tersebut dapat dihitung dengan model intervensi jika waktu dan penyebab diketahui. Deteksi outlier pada multivariat tidaklah mudah, masalah menjadi kompleks ketika terdapat dua atau lebih outlier yang berasal dari variable yang lebih dari dua (Rousseeuw & Van Zomeren, 1990).

### **2.1.5 Missing Value**

*Missing Value* adalah nilai yang ingin diperoleh selama pengumpulan data, tetapi tidak didapatkan karena berbagai alasan. *Missing Value* muncul karena responden tidak menjawab semua pertanyaan pada kuisioner pada saat input data, kesalahan pengukuran yang salah, beberapa data anonim. terdapat 3 masalah terkait *missing value* menurut Luengo,j.(2011), yaitu (1) hilangnya efisiensi, (2) Komplikasi dalam menangani dan menganalisis data, (3) bias yang dihasilkan dari perbedaan antara data yang hilang dan data yang lengkap.

### **2.1.6 Data Preprocessing**

Data preprocessing adalah tahapan untuk menghilangkan beberapa permasalahan yang bisa mengganggu pemrosesan data, data preprocessing merupakan langkah paling awal sebelum melakukan data mining (Garcia 2015). Data yang dikumpulkan dalam kumpulan data kemungkinan besar mengandung kesalahan seperti nilai yang hilang, ketidakkonsistenan, dan nilai nilai yang tidak diinginkan. Garcia (2015), langkah-langkah preprocessing data sebelum data mining: (1) Menentukan gambaran umum data untuk mengidentifikasi kesalahan. (2) Integrasi data untuk menghindari redundansi dan inkonsistensi. (3) Pembersihan data untuk menghapus atau memperbaiki data kotor. (4) Transformasi data untuk membuat atribut baru.

### **2.1.7 Analisis Data**

Analisis data adalah proses menguraikan, memeriksa, dan menafsirkan data untuk mengidentifikasi pola, hubungan, atau informasi yang dapat digunakan untuk mengambil keputusan atau menarik kesimpulan. Ini melibatkan penggunaan alat statistik, metode matematika dan teknik komputer untuk mengatur, membersihkan, memeriksa dan menafsirkan data. Tujuannya adalah untuk memperoleh pemahaman yang lebih mendalam terhadap suatu fenomena atau permasalahan berdasarkan bukti-bukti yang terkandung dalam materi, Hastono (2001). Dengan analisis data yang tepat, kita dapat mengambil keputusan yang lebih baik di berbagai bidang seperti bisnis, sains, kesehatan, dan banyak bidang lainnya.



### **2.1.8 Regresi Linear Berganda**

Analisis regresi adalah hubungan yang didapat dan dinyatakan dalam bentuk persamaan matematik yang menyatakan hubungan fungsional antar variabel. Regresi Linier Berganda adalah model probabilitas yang menyatakan hubungan linear antara lebih dari satu variabel independen dan memiliki satu variabel dependen (Dedi, 2015).

### **2.1.9 R-shiny**

Shiny adalah sebuah paket R yang memudahkan Anda untuk membuat aplikasi web interaktif (aplikasi) langsung dari R. R Shiny memiliki fitur fitur atau package untuk membuat web sederhana. R Shiny bekerja secara *real time*, tiap perubahan inputan akan tampil langsung secara otomatis.

### **2.1.10 Visualisasi Data**

Visualisasi data adalah teknik mengkomunikasikan data atau informasi dengan membuat objek visual, seperti titik, garis atau batang dalam sebuah grafik. Bertujuan untuk mengkomunikasikan informasi secara jelas dan efisien kepada pembaca (Friedman, 2008). Viegas dan Wattenberg (2011), berpendapat bahwa visualisasi yang ideal tidak hanya mengkomunikasikan secara jelas tetapi menstimulasi atensi dan keterlibatan audiens.

## **2.2 Tinjauan Pustaka Non Statistika**

### **2.2.1 Pemilu**

Menurut Sigit Pamungkas (2009) pemilu adalah arena kompetisi untuk mengisi jabatan politik di pemerintahan yang didasarkan pada pilihan formal dari warga negara yang memiliki syarat. Sarbaini (2015) berpendapat pemilu adalah mekanisme utama yang harus terdapat di dalam tahapan penyelenggaraan negara yang menganut sistem demokrasi. Pemilu dianggap bentuk paling nyata dari demokrasi itu sendiri yang berarti kedaulatan dari rakyat dan untuk rakyat.

### **2.2.2 Keterlibatan Mahasiswa**

Keterlibatan mahasiswa dalam pemilu memiliki dampak penting dalam dinamika demokrasi. Menurut penelitian oleh Anugrah et al. (2018), mahasiswa dianggap sebagai agen perubahan dalam masyarakat, dan

partisipasi mereka dalam pemilu dapat mempengaruhi arah dan kebijakan politik. Keterlibatan mahasiswa tidak hanya mencakup hak pilih mereka tetapi juga melibatkan Pemahaman, Akademis, Diskusi, Kepercayaan, Hubungan, hingga aktivitas kesehariannya.

### **2.2.3 Dinamika Pemilu**

Dalam memahami dinamika pemilu, banyak penelitian telah mengkaji berbagai aspek yang mencakup perubahan sosial, politik, dan budaya yang terjadi selama proses pemilihan umum. Menurut Norris (2015), pemilu tidak hanya merupakan proses mekanis untuk memilih pemerintahan, tetapi juga mencerminkan dinamika dalam masyarakat yang lebih luas. Beberapa aspek yang dapat dianalisis dalam konteks dinamika pemilu melibatkan partisipasi pemilih, peran media massa, dan strategi kampanye politik.

### **2.2.4 Pemilu Pada Mahasiswa**

Dalam konteks pemilu, pengaruhnya pada keterlibatan mahasiswa Teknologi Sains Data di Universitas Airlangga dapat menjadi aspek menarik. Pemilu, seperti yang dinyatakan oleh Anugrah et al. (2018), dapat memotivasi mahasiswa untuk aktif terlibat dalam diskusi politik dan inisiatif terkait. Sebagai agen perubahan, mahasiswa dapat menggunakan pengetahuan teknologi mereka untuk mengkaji isu-isu yang muncul selama kampanye. Analisis terhadap pengaruh Pemilihan Umum Tahun 2024 dapat memberikan pemahaman tentang bagaimana proses politik mempengaruhi keterlibatan mahasiswa Teknologi Sains Data dalam dinamika teknologi dan data.

### **2.2.5 Partisipasi & Sumber Informasi**

Sumber informasi kini tidak hanya terbatas hanya sebuah berita koran maupun televisi. sumber informasi sekarang bisa melalui media sosial. perkembangan yang begitu pesat membuat sosial media menjadi sebuah kunci akan perkembangan informasi yang kian pesat, terlebih informasi pemilu yang akan dilaksanakan. \*\*Partisipasi dan Sumber Informasi Mahasiswa TSD dalam Pemilu 2024. Penelitian ini mengamati partisipasi dan sumber informasi mahasiswa Teknologi Sains Data (TSD) Universitas Airlangga selama Pemilu 2024. Fokusnya adalah Pemahaman, Akademis, Diskusi, Kepercayaan, Hubungan, Aktivitas, Sumber Informasi, dan Keterlibatan mahasiswa TSD.

## **BAB III**

### **METODOLOGI**

#### **3.1 Sumber Data**

Dalam penelitian ini, data yang digunakan adalah data primer yang diperoleh melalui kuesioner Google Form yang diisi oleh mahasiswa program studi Teknologi Sains Data. Jumlah total sampel yang digunakan dalam penelitian ini adalah 80 mahasiswa dengan populasi mahasiswa Teknologi Sains Data. Dataset ini berisi 11 variabel yaitu variabel Umur, Jenis\_kelamin, Angkatan, Pemahaman, Akademis, Diskusi, Keterlibatan, Kepercayaan, Hubungan, Aktivitas, dan Sumber\_informasi.

#### **3.2 Metode Data Pre-processing**

Dalam Data Pre-processing terdapat empat tahapan, yakni *checking & handling missing values*, *checking & handling outlier*, transformasi data, dan reduksi dimensi. Adapun hal pertama yang dilakukan adalah pemeriksaan terhadap nilai yang hilang dalam dataset. Hasilnya menunjukkan bahwa dataset yang kami gunakan tidak mengandung missing values. Selanjutnya, dilakukan pengecekan outlier dengan menggunakan visualisasi boxplot, serta uji statistik seperti Grubbs, leverage, dan Cook's distance yang menunjukkan adanya outlier pada variabel umur, pemahaman, dan hubungan. Yang kemudian untuk penanganan memilih metode IQR (Interquartile Range) dengan menghapus data yang berada di luar batas atas dan batas bawah berdasarkan IQR.

Selanjutnya, melakukan transformasi data menggunakan metode one-hot encoding. Variabel kategorikal seperti Jenis Kelamin, Angkatan, dan Sumber Informasi diubah menjadi representasi numerik dengan nilai 1 dan 0. Proses terakhir adalah reduksi dimensi menggunakan metode backward elimination. Melalui metode backward elimination, kami menghilangkan variabel yang tidak signifikan. Secara keseluruhan dalam melakukan data pre-processing ini kami melakukan penanganan missing values, outlier, transformasi data, dan reduksi dimensi.

### 3.3 Metode Analisis Data

Setelah menyelesaikan proses Data Pre-processing, langkah selanjutnya dalam analisis kami menerapkan analisis regresi linear berganda untuk memahami pengaruh dan perbedaan signifikan antara variabel dependen (tingkat keterlibatan) dan satu atau lebih variabel independen, yaitu Umur, Jenis Kelamin, Angkatan, Pemahaman, Akademis, Diskusi, Kepercayaan, Hubungan, Aktivitas, dan Sumber Informasi. Dalam proses ini, variabel dependen, yaitu tingkat keterlibatan, menjadi fokus untuk memahami kontribusi variabel independen.

### 3.3 Metode Visualisasi Data

Dalam metode visualisasi data, digunakan beberapa teknik untuk membantu memahami dan menggambarkan dataset yang telah dipre-process. Pertama, membuat heatmap korelasi untuk menunjukkan hubungan antar variabel. Selanjutnya, menyajikan distribusi variabel menggunakan pie chart, scatter plot, histogram dan density plot, serta boxplot untuk memberikan gambaran visual yang jelas tentang pola dan karakteristik data. Variabel yang diplot dipilih berdasarkan kepentingan analisis, seperti 'Pemahaman', 'Akademis', 'Kepercayaan', 'Hubungan', 'Keterlibatan', 'Jenis\_KelaminLaki.Laki', 'AngkatanDevaskara', 'AngkatanVidyadatum', dan 'Sumber\_InformasiTV'. Visualisasi ini membantu dalam mengidentifikasi tren, outliers, serta hubungan antar variabel.

## BAB IV

### ANALISIS DAN PEMBAHASAN

#### 4.1 Deskripsi Data

Dalam penelitian ini, variabel yang digunakan mencakup Umur, Jenis Kelamin, Angkatan, Pemahaman (1-10), Akademis (1-10), Diskusi (1-10), Kepercayaan (1-10), Hubungan (1-10), Aktivitas (1-10), Sumber Informasi, dan Keterlibatan (1-10). Skala 1-10 menunjukkan tingkat yang lebih besar. Sampel yang digunakan berjumlah 80 sampel. Variabel dependen pada data adalah keterlibatan.

Variabel	Tipe Data	Keterangan
Umur	Numerik	Usia responden dalam tahun
Jenis_Kelamin	Kategorik	Jenis kelamin responden (Laki-Laki dan Perempuan)
Angkatan	Kategorik	Nama angkatan responden (Euclid, Archimedes, Devaskara, dan Vidyadatum)
Pemahaman	Numerik	Seberapa besar tingkat pemahaman responden terhadap pemilu (1-10)
Akademis	Numerik	Seberapa besar pengaruh masa pemilu terhadap keseharian akademis mahasiswa TSD (1-10)
Diskusi	Numerik	Seberapa sering responden terlibat diskusi politik dengan teman-teman (1-10)
Kepercayaan	Numerik	Seberapa besar tingkat kepercayaan informasi politik yang responden peroleh selama masa pemilu 2024 (1-10)
Hubungan	Numerik	Seberapa besar pengaruh masa pemilu terhadap hubungan pertemanan responden (1-10)
Aktivitas	Numerik	Sejauh mana responden merasa perlu melibatkan diri dalam aktivitas politik sebagai mahasiswa TSD (1-10)
Sumber_Informasi	Kategorik	Dari mana responden mendapatkan informasi tentang pemilu (TV, Berita Digital, Media Sosial, dan Koran)
Keterlibatan	Numerik	Sejauh mana responden terlibat dalam proses Pemilu 2024 (1-10)

Tabel 1. Variabel Penelitian

Description: df [8 x 11]

Variabel <chr>	Mean <dbl>	Median <dbl>	Q1 <dbl>	Q3 <dbl>	Variance <dbl>	SD <dbl>	Range <dbl>	Sum <dbl>	Min <dbl>	Max <dbl>
Umur	19.4625	19	19.00	20	1.492247	1.221576	5	1557	17	22
Pemahaman	6.7000	7	6.00	8	3.301266	1.816939	9	536	1	10
Akademis	3.8875	4	2.00	5	5.063133	2.250141	8	311	1	9
Diskusi	4.6250	4	3.00	7	5.553797	2.356650	9	370	1	10
Kepercayaan	5.5000	6	4.00	7	4.177215	2.043824	8	440	1	9
Hubungan	2.7875	2	1.00	4	5.359335	2.315024	8	223	1	9
Aktivitas	3.4875	3	1.75	5	5.670728	2.381329	9	279	1	10
Keterlibatan	4.8250	5	3.00	7	4.880380	2.209158	9	386	1	10

8 rows

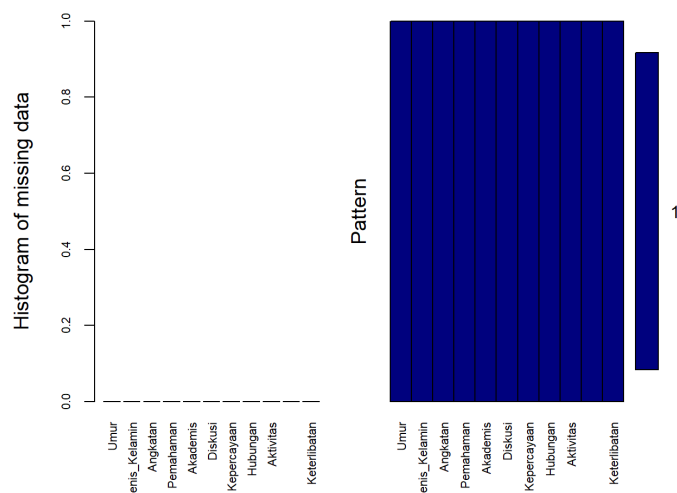
Tabel 2. Statistika Deskriptif

Output diatas menampilkan perhitungan statistik secara deskriptif dimana menampilkan seluruh variabel yang digunakan serta menghitung mean, median, Q1, Q2, Q3, Variansi, Standar Deviasi, Jumlah, Nilai Min, Nilai Max pada tiap variabel yang ada.

## 4.2 Data Pre-processing

### a. Checking & handling missing values.

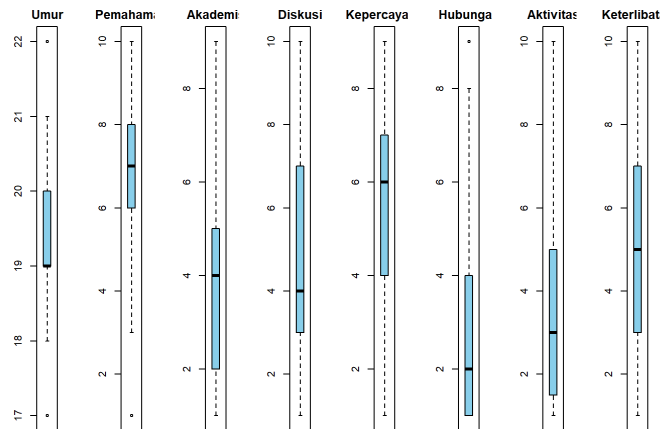
Memeriksa dan menangani nilai yang hilang merupakan langkah awal yang penting dalam analisis statistik. Kami menggunakan sintaks `colSums(is.na(data1))` untuk mengecek keberadaan data yang hilang atau missing values menggunakan R. Berdasarkan hasil visualisasi menggunakan histogram, dapat disimpulkan bahwa dataset yang kami ambil tidak mengandung missing values atau data yang hilang.



Gambar 1. Visualisasi Pengecekan Missing Values

b. *Checking & handling outlier.*

Setelah melakukan pengecekan missing values atau data yang hilang, kami melanjutkan dengan pengecekan outlier. Kami menggunakan visualisasi boxplot, uji Grubbs, uji leverage dan uji cook's distance untuk melihat apakah terdapat outlier pada dataset yang kami analisis dan uji uji dilakukan untuk penanganan outlier yang sesuai model.



Gambar 2. Visualisasi Boxplot Cek Outlier

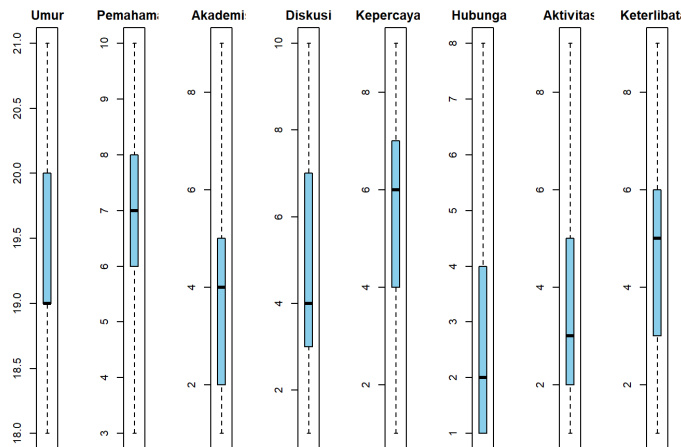
Pada hasil boxplot, terlihat bahwa terdapat outlier pada variabel umur, pemahaman, dan hubungan. Sedangkan dengan menggunakan uji grubbs dan uji leverage didapatkan kesimpulan tolak  $H_0$  pada setiap variabel berarti bahwa terdapat setidaknya satu nilai yang secara signifikan berbeda dari nilai-nilai lain dalam sampel. Hasil uji cook's distance adalah gagal tolak  $H_0$ , tidak ada bukti statistik yang cukup kuat untuk menyatakan bahwa ada observasi yang secara signifikan mempengaruhi model statistik, yang terangkum dalam tabel di bawah.

Variable	Uji Grubbs' Min	Uji Grubbs' Max	Uji Lavarge	Uji Cook's
Umur	Tolak $H_0$	Tolak $H_0$	Tolak $H_0$	Gagal Tolak $H_0$
Pemahaman	Tolak $H_0$	Tolak $H_0$	Tolak $H_0$	Gagal Tolak $H_0$
Akademis	Gagal Tolak $H_0$	Tolak $H_0$	Tolak $H_0$	Gagal Tolak $H_0$

Diskusi	Gagal Tolak H0	Tolak H0	Tolak H0	Gagal Tolak H0
Keterlibatan	Gagal Tolak H0	Tolak H0	Tolak H0	Gagal Tolak H0
Kepercayaan	Tolak H0	Gagal Tolak H0	Tolak H0	Gagal Tolak H0
Hubungan	Gagal Tolak H0	Tolak H0	Tolak H0	Gagal Tolak H0
Aktivitas	Gagal Tolak H0	Tolak H0	Tolak H0	Gagal Tolak H0

Tabel 3. Hasil Uji Grubbs, Lavarge, Cooks

Untuk mengatasi outlier dan berdasarkan uji yang telah dilakukan, kami melakukan penghapusan outlier dengan menggunakan metode IQR (Interquartile Range). Metode ini melibatkan perhitungan kuartil pertama (Q1) dan kuartil ketiga (Q3), kemudian menentukan batas atas dan batas bawah berdasarkan IQR. Data yang berada di luar kisaran ini dianggap sebagai outlier dan dihapus. Gambar boxplot setelah pembersihan outlier seperti di bawah ini.



Gambar 3. Visualisasi Boxplot Pembersihan Outlier

### c. *Transforming the data.*

Transformasi data adalah suatu proses di mana representasi data diubah menjadi bentuk lain untuk memenuhi asumsi atau kebutuhan tertentu. Dalam penelitian ini, kami menerapkan metode one-hot encoding sebagai bentuk transformasi data. Proses encoding dilakukan dengan menggunakan fungsi



'model.matrix()' untuk mengubah variabel Jenis Kelamin, Angkatan, dan Sumber Informasi yang sebelumnya bersifat kategorikal menjadi representasi numerik dengan nilai 1 dan 0. Adapun penamaan variabel kategorikal setelah encoding adalah sebagai berikut:

- Jenis\_kelamin menjadi Jenis\_kelaminLaki-Laki.
- Angkatan menjadi Angkatan\_Devaskara, Angkatan\_Euclid, Angkatan\_Vidyadatum.
- Sumber\_informasi menjadi Sumber\_Informasi\_Media\_Sosial, Sumber\_Informasi\_TV.

Jenis_KelaminLaki.Laki	Angkatan_Devaskara	Angkatan_Euclid	Angkatan_Vidyadatum	\
1	1	0	0	
0	1	0	0	
1	0	0	0	
1	0	0	0	
0				
	Sumber_Informasi_Media_Sosial	Sumber_Informasi_TV		
		1	0	
		0	0	
		1	0	
		1	0	
		1	0	

Gambar 4. Encoding Data Kategorik

d. *Dimensionality reduction.*

Setelah dilakukan proses backward elimination, model terbaik yang dihasilkan memiliki nilai Adjusted R-Squared sebesar 0.519. Ini berarti sekitar 51.9% variabilitas dalam respons dapat dijelaskan oleh kombinasi variabel prediktor yang tersisa, yaitu jenis kelamin (Laki-Laki), angkatan (Devaskara dan Vidyadatum), pemahaman, akademis, kepercayaan, hubungan, dan sumber informasi dari TV."

```
[1] "Variables: 1 | Adj R-Squared: 0.51963211894961"
[2] "Variables: Jenis_KelaminLaki.Laki | Adj R-Squared: 0.51963211894961"
[3] "Variables: AngkatanDevaskara | Adj R-Squared: 0.51963211894961"
[4] "Variables: AngkatanVidyadatum | Adj R-Squared: 0.51963211894961"
[5] "Variables: Pemahaman | Adj R-Squared: 0.51963211894961"
[6] "Variables: Akademis | Adj R-Squared: 0.51963211894961"
[7] "Variables: Kepercayaan | Adj R-Squared: 0.51963211894961"
[8] "Variables: Hubungan | Adj R-Squared: 0.51963211894961"
[9] "Variables: Sumber_InformasiTV | Adj R-Squared: 0.51963211894961"
```

Gambar 5. Reduksi Dimensi Data

### 4.3 Hasil Analisis

Dalam penelitian ini kami menggunakan Analisis Regresi Linear Berganda, berikut adalah hasil analisisnya :

*H0 : Tidak Terdapat perbedaan secara signifikan*

*H1 : Terdapat perbedaan secara signifikan*

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.38474 -0.87466 -0.02951  0.97907  3.08923

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.98017    0.99462  -1.991  0.051213 .
X1              NA           NA      NA      NA
Jenis_KelaminLaki.Laki  0.69468    0.36487    1.904  0.061890 .
AngkatanDevaskara    1.48659    0.47783    3.111  0.002890 **
AngkatanVidyadatum    1.25084    0.52416    2.386  0.020299 *
Pemahaman          0.28376    0.11682    2.429  0.018256 *
Akademis           0.22359    0.09465    2.362  0.021536 *
Kepercayaan         0.34509    0.09551    3.613  0.000633 ***
Hubungan           0.17677    0.11544    1.531  0.131120
Sumber_InformasiTV    1.20754    0.67863    1.779  0.080418 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.397 on 58 degrees of freedom
Multiple R-squared:  0.5779,    Adjusted R-squared:  0.5196
F-statistic: 9.924 on 8 and 58 DF,  p-value: 1.417e-08
```

Gambar 6. Output Regresi Linear Berganda

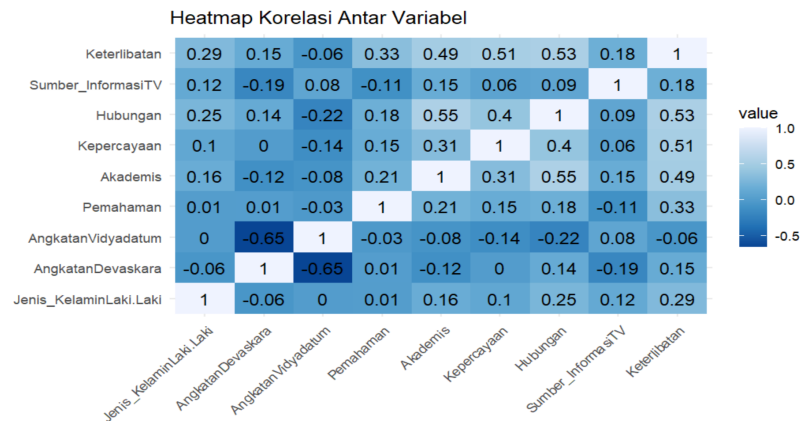
Dalam analisis regresi linear berganda, diperoleh persamaan:

- keterlibatan = -1.98 + 0.69 (Jenis\_KelaminLaki.Laki) + 1.48 (AngkatanDevaskara) + 1.25 (AngkatanVidyadatum) + 0.28 (Pemahaman) + 0.22 (Akademis) + 0.34 (Kepercayaan) + 0.17 (Hubungan) + 1.20 (Sumber\_InformasiTV).
- Nilai p-value <  $\alpha$  (0.05) menunjukkan terdapat perbedaan yang signifikan antara variabel prediktor 'AngkatanDevaskara', 'AngkatanVidyadatum', 'Pemahaman', 'Akademis', dan 'Kepercayaan' terhadap variabel respons 'Keterlibatan'.
- Nilai p-value >  $\alpha$  (0.05) menunjukkan tidak terdapat perbedaan yang signifikan antara 'Jenis\_KelaminLaki.Laki', 'Hubungan' dan 'Sumber\_InformasiTV' terhadap variabel respons 'Keterlibatan'.
- Model analisis regresi linear berganda memiliki R-Square sebesar 0.578 dan Adjusted R-Square sebesar 0.52, yang menunjukkan bahwa model mampu menjelaskan sekitar 57,8% dan 52% dari variasi dalam variabel target (keterlibatan) setelah mempertimbangkan jumlah variabel prediktor yang digunakan dalam model.

- Secara keseluruhan, kedua model menunjukkan kinerja yang cukup baik dalam menjelaskan hubungan antara variabel prediktor dengan variabel respon.

## 4.4 Visualisasi Data R-Shiny

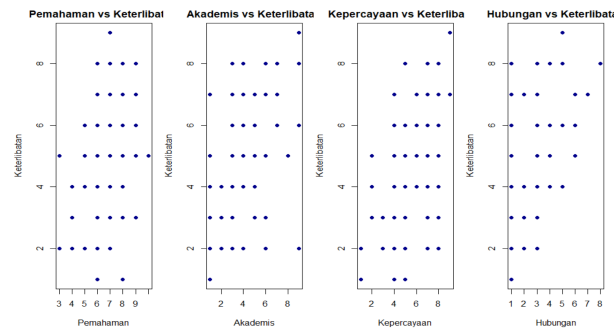
### 4.4.1 Heatmap Correlation



Gambar 7. Heatmap Korelasi Variabel

Berdasarkan hasil visualisasi Heatmap di atas menggambarkan nilai korelasi antar variabel, dengan intensitas warna mencerminkan kekuatan korelasi. Semakin gelap warnanya, semakin kuat korelasinya. Korelasi terkuat antara variabel prediktor dan respon dari heatmap diatas adalah variabel 'Hubungan' dan 'Keterlibatan', yaitu dengan nilai korelasi sebesar 0.53. Selanjutnya, korelasi terkuat kedua antara variabel prediktor dan respon dari heatmap diatas adalah variabel 'Kepercayaan' dan 'Keterlibatan', yaitu dengan nilai korelasi sebesar 0.51. Terakhir, korelasi terkuat ketiga antara variabel prediktor dan respon dari heatmap diatas adalah variabel 'Akademis' dan 'Keterlibatan', yaitu dengan nilai korelasi sebesar 0.49.

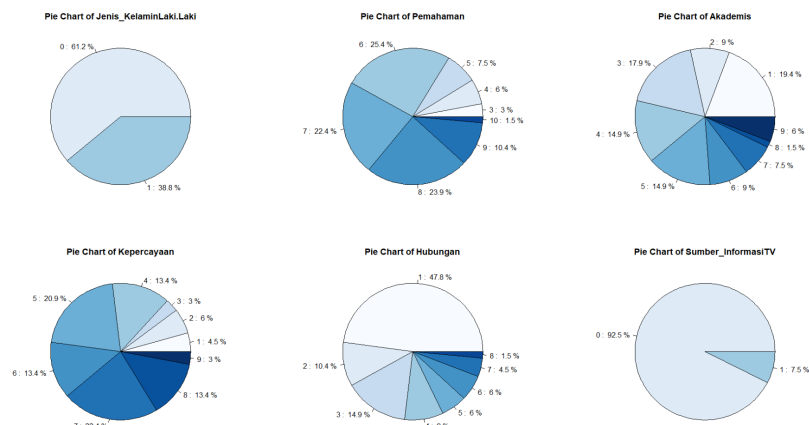
#### 4.4.2 Scatter Plot



Gambar 8. Scatter Plot

Berdasarkan hasil visualisasi Scatter Plot di atas menunjukkan bahwa pada variabel ‘Pemahaman’, ‘Akademis’, ‘Kepercayaan’, dan ‘Hubungan’ terdapat hubungan positif dengan variabel ‘keterlibatan’. Interpretasi dari scatter plot tersebut adalah bahwa semakin tinggi tingkat pemahaman, akademis, kepercayaan, dan hubungan yang dimiliki mahasiswa/i TSD, semakin tinggi juga tingkat keterlibatannya dalam suatu aktivitas atau situasi, seperti dalam pemilu. Korelasi positif ini menandakan bahwa variabel-variabel tersebut secara signifikan berkontribusi terhadap tingkat keterlibatan mahasiswa/i TSD dalam pemilu tahun 2024. Oleh karena itu, dapat disimpulkan bahwa pemahaman, akademik, kepercayaan, dan hubungan memiliki pengaruh yang penting dalam meningkatkan keterlibatan mahasiswa/i TSD dalam pemilu tahun 2024.

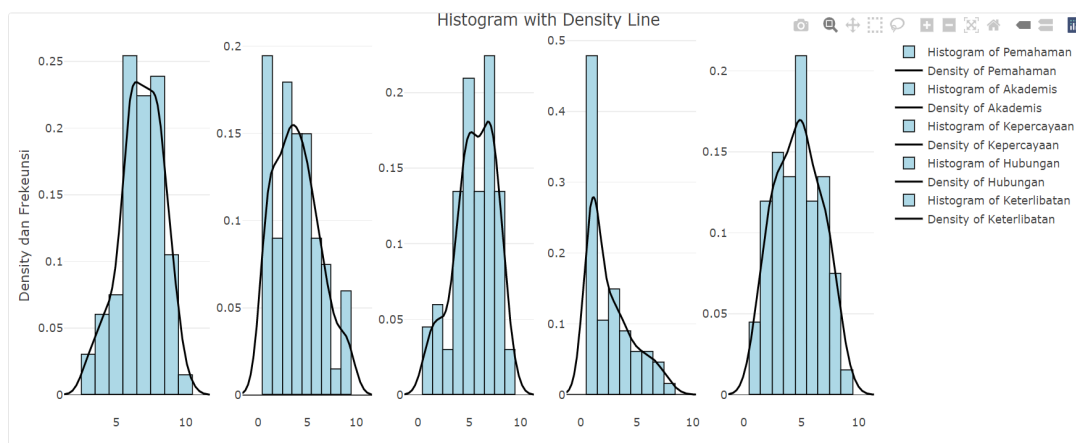
#### 4.4.3 Pie Chart



Gambar 9. Pie Chart

Berdasarkan hasil visualisasi Pie Chart di atas terkait partisipasi mahasiswa/i TSD dalam pemilu tahun 2024, terlihat bahwa mayoritas responden adalah perempuan(0) sebesar 61.2%, sementara partisipasi laki-laki(1) mencapai 38.8%. Variabel 'AngkatanDevaskara', terlihat bahwa sebanyak 53.7% mahasiswa/i TSD berasal dari angkatan Devaskara (2022), sementara 'Selain Angkatan Devaskara' hanya menyumbang 46.3%. Begitu pula, dalam visualisasi pie chart untuk variabel 'AngkatanVidyadatum', sekitar 73.1% responden berasal dari angkatan selain Vidyadatum (2023), dengan hanya 26.9% berasal dari Angkatan Vidyadatum. Pemahaman mahasiswa/i TSD mengenai pemilu tergolong baik, dengan sebagian besar memberikan nilai tinggi (25.4% nilai 6, 23.9% nilai 8, dan 22.4% nilai 7). Dalam konteks akademis, pemilu tidak tampak berpengaruh signifikan. Meskipun sekitar 20.9% responden memiliki kepercayaan yang cukup terhadap informasi pemilu tahun 2024, dampaknya terhadap hubungan pertemanan terlihat minim. Mayoritas responden lebih mengandalkan sumber informasi selain TV (92.5%) dalam memperoleh informasi mengenai pemilu .data yang ada .

#### 4.4.4 Histogram & Density Plot

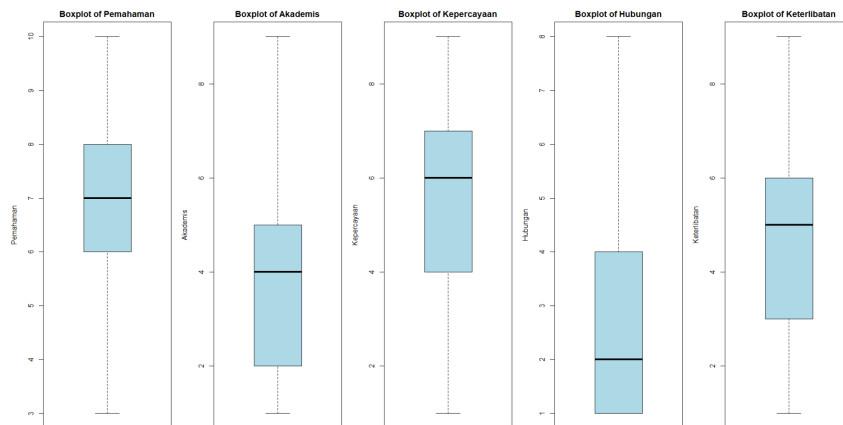


Gambar 10. Histogram & Density Plot

Berdasarkan hasil visualisasi Histogram dan Density Plot diatas, dapat disimpulkan bahwa sebagian besar mahasiswa/i TSD memiliki tingkat pemahaman yang cukup baik terkait pemilu, pemilu tidak berpengaruh secara signifikan terhadap prestasi akademik mahasiswa/i TSD, sebagian besar

mahasiswa/i TSD memiliki tingkat kepercayaan yang cukup tinggi terhadap informasi politik yang berkaitan dengan pemilu tahun 2024, pengaruh pemilu terhadap hubungan pertemanan mahasiswa/i TSD cenderung rendah, sehingga pemilu tidak secara signifikan mempengaruhi hubungan pertemanan mahasiswa/i TSD, dan yang terakhir keterlibatan mahasiswa/i TSD di pemilu cenderung cukup.

#### 4.4.5 Box Plot



Gambar 11. Box Plot

Berdasarkan hasil visualisasi Box Plot diatas, dapat disimpulkan bahwa pemahaman mahasiswa/i TSD terhadap pemilu cenderung baik, terkonsentrasi di sekitar nilai median, pemilu tidak memiliki pengaruh signifikan terhadap prestasi akademis mahasiswa/i TSD, mahasiswa/i TSD memiliki tingkat kepercayaan yang cukup terhadap informasi politik terkait pemilu 2024, dengan sebagian besar data berkumpul di sekitar nilai median, pemilu sama sekali tidak mempengaruhi hubungan pertemanan mahasiswa/i TSD, karena sebagian besar data berkumpul di sekitar nilai median yang rendah, dan yang terakhir keterlibatan mahasiswa/i TSD di pemilu cenderung cukup, terkonsentrasi di sekitar nilai median.

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil Analisis Pengaruh Keterlibatan Mahasiswa Teknologi Sains Data Universitas Airlangga dalam Masa Pemilihan Umum Tahun 2024 dengan mengambil responden dari 80 mahasiswa/i, dapat disimpulkan bahwa :

##### **1. Visualisasi Data**

Dari hasil analisis eksplorasi data yang dilakukan dengan menyajikan hubungan antar variabel didapatkan bahwa variabel respon berdistribusi normal. Berdasarkan nilai korelasi, dapat diketahui bahwa variabel prediktor yang berkorelasi atau berpengaruh terhadap variabel respons (keterlibatan) adalah hubungan (0.53), kepercayaan (0.51), dan akademis (0.49).

##### **2. Analisis Data**

###### **a. Signifikansi**

Secara keseluruhan, kita dapat menyimpulkan bahwa beberapa variabel prediktor, yaitu 'Jenis\_KelaminLaki.Laki', 'Pemahaman', 'Akademis', 'Kepercayaan', 'Hubungan', dan 'Sumber\_InformasiTV', berkontribusi secara signifikan terhadap perbedaan dalam variabel respons 'Keterlibatan', sementara 'AngkatanDevaskara' dan 'AngkatanVidyadatum' tidak memberikan kontribusi yang signifikan.

###### **b. Model**

Kebalikan Model Analisis Regresi Linier Berganda memiliki R-Square dan Adjusted R-Square sebesar 0.578 dan 0.520, yang menunjukkan bahwa model mampu menjelaskan (memberikan pengaruh) sekitar 57.8%, sedangkan 42.2% dijelaskan oleh variabel lain yang tidak termasuk dalam perhitungan dari variasi dalam variabel target (keterlibatan).

## 5.2 Saran

Berdasarkan hasil analisis dan kesimpulan diatas, peneliti menyarankan ketika data yang dianalisis memiliki outlier, penggunaan metode regresi linier konvensional seperti Ordinary Least Squares (OLS) dapat menjadi rentan terhadap dampak outlier tersebut. Dalam hal ini, penggunaan metode regresi yang lebih tahan terhadap outlier, seperti Robust Regression, bisa menjadi pilihan yang lebih baik. Disamping itu, untuk meningkatkan nilai R-squared, disarankan untuk menambahkan variabel yang memiliki validitas yang lebih tinggi dalam model regresi. Selain itu bisa dengan melakukan uji asumsi, peneliti dapat memastikan bahwa analisis regresi linier dilakukan dengan benar dan hasilnya dapat diinterpretasikan dengan kepercayaan yang tinggi.



## DAFTAR PUSTAKA

KPU Biro Advokasi Hukum Dan Penyelesain Sengketa (2022). *Dinamika Hukum Pemilu.KPU*

RESEARCH ARTICLE ANALISIS DAMPAK PENUNDAAN PEMILU 2024.  
<https://journal.uns.ac.id/Sovereignty/article/download/224/145>.

Sarbaini (2015). *Demokratisasi Dan Kebebasan Memilih Warga Negara Dalam Pemilihan Umum*

Vitaly Friedman (2008) "Data Visualization and Infographics" in: *Graphics, Monday Inspiration*, January 14th, 2008.

Kaiser, J. (2014). Dealing with Missing Values in Data. *Journal Of Systems Integration* (1804-2724), 5(1).

Hastono, S. P. (2001). *Analisis data*. Depok: Fakultas Kesehatan Masyarakat Universitas Indonesia.

Wardhani, P. S. N. (2018). Partisipasi politik pemilih pemula dalam pemilihan umum. *Jupiiis: Jurnal Pendidikan Ilmu-Ilmu Sosial*, 10(1), 57-62.

Rousseeuw, P.J., Van Zomeren, B.C., 1990, Un masking Multivariate Outliers and Leverage Points, *Journal of The American Statistical Association*, Vol. 85, No. 411, pp : 633-651.

Herdiani, E. T., & Ilyas, N. (2017). Modifikasi Penaksir Robust dalam Pelabelan Outlier Multivariat. *Jurnal Matematika, Statistika dan Komputasi*, 14(1), 46-53.

Budiarti, L., Tarno, T., & Warsito, B. (2013). Analisis Intervensi Dan Deteksi Outlier Pada Data Wisatawan Domestik (Studi Kasus Di Daerah Istimewa YOGYAKARTA). *Jurnal Gaussian*, 2(1), 39-48.

Sule.N (2022).R SHINY APLIKASI WEB INTERAKTIF MODERN.algortima.  
<https://algorit.ma/blog/r-shiny-adalah-2022/>

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72, pp. 59-139). Cham, Switzerland: Springer International Publishing.

Pahlevi, I. (2016). *Dinamika sistem pemilu masa transisi di indonesia*. *Jurnal Politica Dinamika Masalah Politik Dalam Negeri dan Hubungan Internasional*, 5(2).

## LAMPIRAN

1. Link kode R-Shiny, R Markdown, dan Data Penelitian = <http://bit.ly/Folder-Project-EVD-Kelompok9>
2. R Markdown = [https://bit.ly/Rmarkdown\\_Kelompok9](https://bit.ly/Rmarkdown_Kelompok9)
3. Desktop Rshiny = <http://bit.ly/DesktopRshiny>
4. Mobile Rshiny = <https://bit.ly/MobileRshiny>