

Proyek DAMI

by Lidia Ginting

Submission date: 06-Jan-2024 02:12AM (UTC+0700)

Submission ID: 2267094477

File name: Group8_Laporan.pdf (1.97M)

Word count: 3044

Character count: 24229

PROGRAM STUDI SARJANA SISTEM INFORMASI
PROPOSAL PROYEK AKHIR DATA MINING

**Prediksi Biaya Pasien pada Fasilitas Kesehatan Tingkat Pertama
(FKTP) menggunakan Algoritma Regresi Multi Linear**



OLEH:

12S20013 Lydia Tesalonika

12S20020 Wahyu Krisdangolyanti Simamora

12S20053 Andri Anjelia Hutapea

17
FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO
INSTITUT TEKNOLOGI DEL
LAGUBOTI
2023

6
DAFTAR ISI

DAFTAR ISI.....	ii
DAFTAR TABEL	iii
DAFTAR GAMBARiv
BAB I BUSINESS UNDERSTANDING	1
1.1 <i>Determine Business Objective</i>	1
1.2 <i>Determine Project Goal</i>	2
1.3 <i>Produce Project Plan</i>	2
BAB II DATA UNDERSTANDING	5
2.1 Data Collection.....	5
2.2 Data Description.....	5
2.3 Data Validation	7
BAB III DATA PREPARATION.....	11
3.1 Data selection	11
3.2 Data Cleaning	13
3.2.1 Missing and Duplicate Values	13
BAB IV BUILD MODEL.....	21
4.1 Select Modeling Technique.....	21
4.2 Model Building	21
BAB V MODEL EVALUATION	24
5.1 Result Evaluation	24
5.2 Result Review	24
BAB VI DEPLOYMENT	25
6.1 Deployment	25
6.2 Web Application	26
LAMPIRAN	27

DAFTAR TABEL

Tabel 1 Produce Project Plan	2
Tabel 2 Data Description	5
Tabel 3 Data Selection	11

DAFTAR GAMBAR

Gambar 1 Jenis Faskes	7
Gambar 2 Tipe Faskes	8
Gambar 3 Tingkat Layanan	8
Gambar 4 Nama Tindakan	9
Gambar 5 Biaya Tagih	9
Gambar 6 Biaya Verifikasi	10
Gambar 7 Hasil Run Data Selection	12
Gambar 8 Hasil Run Data Selection 2	12
Gambar 9 Hasil Run Data Selection 3	13
Gambar 10 Missing dan Duplicate Values	14
Gambar 11 Menampilkan Data dari Dataset.....	14
Gambar 12 Hasil Run setelah Menampilkan Data.....	15
Gambar 13 Hasil Run setelah Menampilkan Data 2	16
Gambar 14 Menghilangkan Baris Duplikat	16
Gambar 15 Mengganti Nilai yang Hilang (Missing Values)	16
Gambar 16 Hasil Mengganti Nilai yang Hilang (Missing Values).....	17
Gambar 17 Menghapus Baris dengan Nilai yang Hilang	18
Gambar 18 Hasil Run Menghapus Baris dengan Nilai yang Hilang	18
Gambar 19 Hasil Run Menghapus Baris dengan Nilai yang Hilang 2	18
Gambar 20 Menghapus Kolom	18
Gambar 21 Menyaring data dan Menampilkan Isi	19
Gambar 22 Hasil Menyaring data dan Menampilkan Isi	19
Gambar 23 Hasil Menyaring data dan Menampilkan Isi 2	20
Gambar 24 Menghapus Data	20
Gambar 25 Mengimpor pustaka-pustaka	21
Gambar 26 Read Data CSV	22
Gambar 27 View Data Information	22
Gambar 28 Describe the Data	22
Gambar 29 Check Null Data.....	22
Gambar 30 Variabel Independen	22
Gambar 31 Subset Training dan Testing	23

Gambar 32 Melatih Model dengan Testing	23
Gambar 33 Menghitung Skor (R-Squared).....	23
Gambar 34 Menghitung Akurasi dan Regresi Ensemel Gradient Boosting	23
Gambar 35 Result Evaluation	24
Gambar 36 Model.pkl	25
Gambar 37 FlashWebApp.....	25
Gambar 38 Code untuk Memanggil Model	25
Gambar 39 Run App	26
Gambar 40 Web Application Sebelum Input Data	26
Gambar 41 Web Application Setelah Input Data	26

BAB I

BUSINESS UNDERSTANDING

Business Understanding merupakan langkah pertama yang dilakukan dalam metodologi untuk memprediksi Biaya Pasien pada FKTP dengan Multi linear Regression Algorithm adalah untuk memahami bisnisnya. Bab ini akan menjelaskan pemahaman konten dan kebutuhan operasi data mining dari perspektif bisnis. Kegiatan ini meliputi Determine Business Objective, Determine Project Goal, dan Project Plan.

1.1 Determine Business Objective

Dalam proyek ini, Kami akan menyelidiki dan menganalisis sampel data yang berkaitan dengan Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) dalam rangka proyek ini. Tujuan utama BPJS Kesehatan adalah memberikan jaminan kesehatan kepada semua orang di Indonesia. Data yang akan kami analisis mencakup berbagai elemen terkait, seperti informasi tentang kepersertaan, FKTP, FKTRTL, FKTP non kapitasi, dan diagnosis sekunder. Dataset yang dipilih pada proyek ini adalah dataset FKTP-non kapitasi untuk tahun 2021.

BPJS (Badan Penyelenggara Jaminan Sosial), merupakan Badan Hukum Publik yang bertanggung jawab atas jaminan perlindungan dan pelayanan kesehatan bagi pengguna atau masyarakat. BPJS Kesehatan menghimpun dan menyimpan banyak data, termasuk keanggotaan peserta, klaim pengobatan, biaya layanan kesehatan, dan informasi lainnya. Analisis dataset ini dapat memberi tahu kita tentang tren kesehatan masyarakat, meningkatkan efisiensi sistem, dan membuat kebijakan yang lebih baik.

Metode Regresi adalah metode statistik yang digunakan untuk menganalisis hubungan antara dua atau lebih variabel. Hal yang dipelajari disebut variabel respon, sedangkan hal yang mempengaruhi variabel respon disebut variabel prediktor. Variabel respon adalah variabel ingin diketahui pengaruhnya terhadap variabel lain. Variabel respon diketahui sebagai variabel terikat, variabel dependent atau variabel keluaran. Sedangkan, variabel Prediktor adalah variabel yang memiliki pengaruh terhadap variabel respon. Variabel prediktor disebut sebagai variabel bebas, variabel independen atau variabel masukan.

Regresi Multi Linear adalah cara untuk memahami hubungan antara satu variabel yang ingin diprediksi dengan dua atau lebih variabel yang memungkinkannya. Misalnya, jika kita

ingin memprediksi harga rumah, kita bisa melibatkan variabel seperti luas tanah, jumlah kamar tidur, dan jarak ke pusat kota.

Algoritma ini mencari persamaan matematis yang paling baik mencocokkan pola hubungan di antara variabel-variabel tersebut. Tujuannya adalah menemukan seberapa besar setiap variabel berkontribusi terhadap hasil akhir. Dengan koefisien yang ditemukan, kita dapat membuat prediksi berdasarkan nilai-nilai variabel independen. Regresi Multi Linear sangat berguna untuk memahami dan memprediksi hubungan kompleks dalam berbagai bidang, termasuk ekonomi dan ilmu sosial.

1.2 Determine Project Goal

Melalui analisis data BPJS Kesehatan, proyek ini bertujuan memberikan beberapa keuntungan yang signifikan. Pertama, analisis tersebut diharapkan dapat memberikan wawasan yang diperlukan untuk meningkatkan efisiensi dan efektivitas layanan kesehatan kepada peserta BPJS. Dengan begitu, upaya ini akan membantu optimalisasi layanan kesehatan sesuai dengan kebutuhan peserta. Selain itu, tujuan proyek ini adalah memberikan kontribusi penting dalam proses perencanaan kebijakan kesehatan.

Diharapkan analisis data mendalam ini akan menyediakan informasi yang relevan dan akurat untuk perumusan kebijakan yang lebih baik. Dengan pemahaman yang lebih baik tentang dinamika partisipasi dan kebutuhan layanan kesehatan, diharapkan kebijakan yang diimplementasikan dapat lebih sesuai dengan kebutuhan masyarakat yang dilayani oleh BPJS Kesehatan. Hal ini dapat memberikan gambaran mengenai faktor-faktor apa saja yang berkontribusi terhadap biaya perawatan pasien di FKTP.

Tujuan akhirnya adalah menghasilkan model prediksi yang dapat membantu pihak BPJS Kesehatan dalam merencanakan dan mengelola alokasi sumber daya dengan lebih efektif.
Dengan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi biaya perawatan, BPJS Kesehatan dapat mengoptimalkan pendanaan dan meningkatkan efisiensi layanan kesehatan di FKTP.

1.3 Produce Project Plan

Proses produksi pada proyek ini dapat dilihat pada tabel dibawah ini yang terdiri dari aktivitas, detail dan durasi pelaksanaan proyek, dapat dilihat pada tabel dibawah ini:

Tabel 1 Produce Project Plan

Aktivitas	Detail	Durasi
-----------	--------	--------

Pemilihan Kasus	Pemilihan kasus Pemilihan Algoritma	1 Hari 1 Hari
Business Understanding	Menentukan Objektif Bisnis	2 Hari
	Menentukan Tujuan Bisnis	2 Hari
	Membuat Rencana proyek	2 Hari
Data Understanding	Mengumpulkan Data	1 Hari
	Menelaah Data	2 Hari
	Memvalidasi Data	2 Hari
Data Preparation	Memilah Data	2 Hari
	Membersihkan Data	2 Hari
	Mengkonstruksi Data	2 Hari
	Menentukan Label Data	3 Hari
	Mengintegrasikan Data	5 hari
Modeling	Membangun Skenario Pengujian	5 Hari
	Membangun Model	3 Hari
Model Evaluation	Mengevaluasi Hasil Pemodelan	2 Hari
	Melakukan Review Proses Pemodelan	4 Hari
Deployment	Melakukan Deployment Model	1 Hari
	Membuat laporan akhir Proyek	-

Dalam pelaksanaan proyek penelitian, Python digunakan sebagai bahasa pemrograman utama. Python adalah bahasa pemrograman yang sangat populer dalam konteks Data Science, Machine Learning, dan Internet of Things (IoT). Python juga dikenal sebagai bahasa pemrograman yang dapat menjalankan sejumlah instruksi secara langsung dengan metode

orientasi objek (Object Oriented Programming) dan menggunakan semantik dinamis untuk meningkatkan tingkat keterbacaan syntax.

Proyek ini juga mengadopsi metode algoritma Multi Linear untuk mengevaluasi hubungan positif dan negatif antara variabel independen dan variabel dependen. Algoritma ini memungkinkan prediksi nilai variabel dependen berdasarkan perubahan nilai variabel independen. Cara kerja Multi Regresi melibatkan penyusunan grafik garis antara dua variabel data, x dan y. Variabel independen, x, terletak di sepanjang sumbu horizontal, dan tujuan utama analisis regresi ini adalah untuk memahami hubungan sebab-akibat antara variabel satu dengan variabel lainnya.

BAB II

DATA UNDERSTANDING

Pada tahapan *data understanding*/ pemahaman data, akan dijelaskan bahwa tahapan ini dimulai dengan pengumpulan data awal untuk mengetahui karakteristik dan variabel-variabel utama data. Ini memungkinkan tim untuk membuat hipotesis awal tentang informasi yang tersembunyi dan mendapatkan pemahaman awal tentang data.

2.1 Data Collection

Pengumpulan data merupakan tahap awal yang dilakukan guna memperoleh data yang akan digunakan. Oleh karena itu dataset yang akan digunakan untuk memprediksi biaya Pasien ke Fasilitas Kesehatan Tahap Pertama (FKTP) dengan Algoritma Regresi Multi Linear adalah dataset fktp-non kapitas tahun 2021 yaitu fktpnonkapitasi.dta.

2.2 Data Description

Dataset yang digunakan dalam melakukan prediksi biaya pasien ke FKTP adalah terdiri beberapa data dan 21 variabel. Berikut akan dibahas mengenai atribut yang tersedia pada dataset:

Tabel 2 Data Description

No.	Variabel	Label variabel	Deskripsi
1.	PSTV01	Nomor Peserta	Nomor identifikasi peserta yang bersifat unik dan telah dideidentifikasi untuk melindungi identitas peserta sebenarnya
2.	PSTV02	Nomor keluarga	Nomor yang mengidentifikasi kepala keluarga dalam sampel dan berfungsi sebagai penanda keluarga (peserta BPJS Kesehatan dalam satu keluarga memiliki nomor kepala keluarga yang sama)
3.	PSTV15	Bobot	Faktor pengali yang menggambarkan jumlah individu di dalam populasi diwakili oleh individu di dalam sampel
4.	PNK02	ID Kunjungan	Nomor identifikasi unik untuk menandakan setiap kunjungan FKTP oleh peserta

5.	PNK03	Tanggal kunjungan	Tanggal melakukan kunjungan
6.	PNK04	Tanggal tindakan	Tanggal melakukan tindakan
7.	PNK05	Tanggal pulang	Tanggal menyelesaikan kunjungan
8.	PNK06	Provinsi faskes	Provinsi tempat peserta mengakses fasilitas kesehatan tingkat pratama (FKTP) <small>27</small>
9.	PNK07	Kode Kab/Kota faskes	Kabupaten/kota tempat peserta mengakses fasilitas kesehatan tingkat pratama (FKTP) <small>1</small>
10.	PNK08	Kepemilikan faskes	Kepemilikan dari fasilitas kesehatan tingkat pratama (FKTP) tempat peserta berkunjung
11.	PNK09	Jenis faskes	Jenis dari fasilitas kesehatan tingkat pertama (FKTP) tempat peserta berkunjung <small>1</small>
12.	PNK10	Tipe faskes	Tipe dari fasilitas kesehatan tingkat pertama (FKTP) tempat peserta berkunjung <small>1</small>
13.	PNK11	Tingkat layanan	Tingkat layanan yang diterima peserta di FKTP
14.	PNK12	Segmen peserta	Segmen peserta saat mengakses FKTP <small>1</small>
15.	PNK13	Kode dan Nama diagnosis berdasarkan ICD-10 (3 digit)	Kode dan nama diagnosis berdasarkan 3 digit pertama kode ICD 10 yang diperoleh dari hasil input sistem informasi BPJS Kesehatan <small>1</small>
16.	PNK13A	Kode diagnosis berdasarkan ICD-10 (3 digit)	Kode dan nama diagnosis berdasarkan 3 digit pertama kode ICD 10 yang diperoleh dari hasil input sistem informasi BPJS Kesehatan
17.	PNK14	Kode diagnosis (3-5 digit)	Kode diagnosis menurut ICD 10 (jumlah digit tidak sama pada semua observasi dengan rentang 3-5 digit kode ICD 10)
18.	PNK15	Nama Diagnosis	Nama diagnosis yang terbaca oleh sistem informasi BPJS Kesehatan berdasarkan kode diagnosis yang ter-input dalam sistem
19.	PNK16	Nama Tindakan	Nama jenis tindakan yang dilakukan kepada pasien

20.	PNK17	Biaya tagih	Biaya yang ditagihkan fasilitas kesehatan untuk setiap ID kunjunga
21.	PNK18	Biaya verifikasi	Biaya yang diverifikasi BPJS Kesehatan untuk setiap nomor ID Kunjungan

Dataset fktp-non kapitasi akan melalui sebuah proses untuk melakukan peninjauan dan analisis data guna memperoleh pemahaman dasar mengenai data tersebut yang disebut sebagai *Exploratory Data Analysis (EDA)*. Adapun tujuan dilakukannya EDA ini adalah untuk memahami struktur data dan menentukan variabel yang relevan untuk digunakan pada proyek.

9

2.3 Data Validation

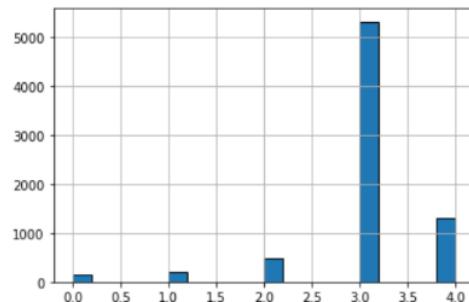
Pada tahap ini, data yang akan digunakan divalidasi untuk menghindari kesalahan atau masalah input yang tidak memiliki nilai. Validasi data merupakan proses memastikan bahwa data yang digunakan dalam analisis atau model memiliki kualitas yang baik dan dapat diandalkan. Terdapat 6 variabel yang digunakan dalam program ini yaitu 5 variabel bebas dan 1 variabel terikat

1. Variabel PNK09 (Jenis Faskes)

```
1 df['PNK09'].describe()
```

	PNK09
count	7453.000000
mean	2.993023
std	0.735172
min	0.000000
25%	3.000000
50%	3.000000
75%	3.000000
max	4.000000
Name:	PNK09, dtype: float64

```
1 df['PNK09'].hist(edgecolor='black', bins=20);
```

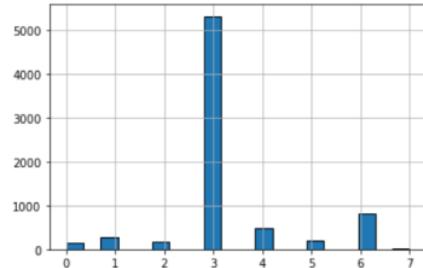


Gambar 1 Jenis Faskes

2. Variabel PNK10 (Tipe Faskes)

```
1 df['PNK10'].describe()
count    7453.000000
mean      3.293707
std       1.214627
min       0.000000
25%      3.000000
50%      3.000000
75%      3.000000
max      7.000000
Name: PNK10, dtype: float64
```

```
1 df['PNK10'].hist(edgecolor='black', bins=20);
```

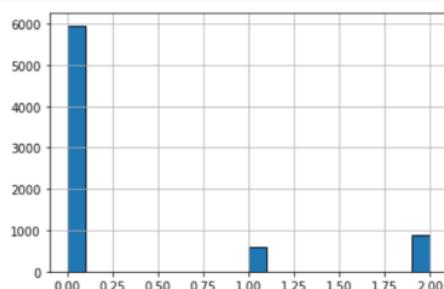


Gambar 2 Tipe Faskes

3. Variabel PNK11 (Tingkat Layanan)

```
1 df['PNK11'].describe()
count    7453.000000
mean      0.320542
std       0.675984
min       0.000000
25%      0.000000
50%      0.000000
75%      0.000000
max      2.000000
Name: PNK11, dtype: float64
```

```
1 df['PNK11'].hist(edgecolor='black', bins=20);
```



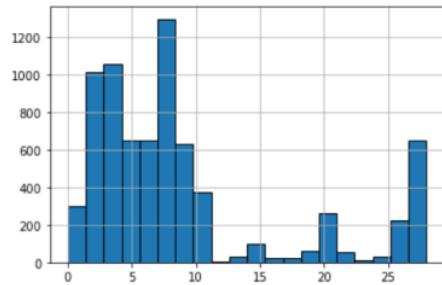
Gambar 3 Tingkat Layanan

4. Variabel PNK16 (Nama Tindakan)

```
1 df['PNK16'].describe()
```

```
count    7453.000000
mean     9.141151
std      8.045675
min     0.000000
25%     4.000000
50%     7.000000
75%     9.000000
max    28.000000
Name: PNK16, dtype: float64
```

```
1 df['PNK16'].hist(edgecolor='black', bins=20);
```



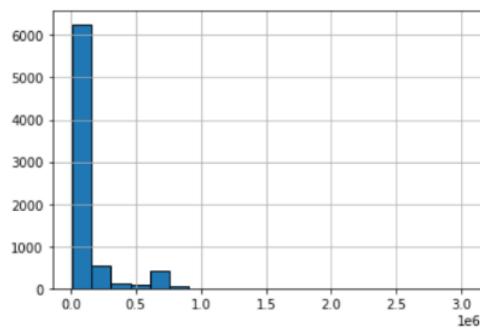
Gambar 4 Nama Tindakan

5. Variabel PNK17 (Biaya Tagih)

```
1 df['PNK17'].describe()
```

```
count    7.453000e+03
mean     1.099373e+05
std      1.818357e+05
min     1.000000e+04
25%     3.000000e+04
50%     4.500000e+04
75%     1.200000e+05
max    3.000000e+06
Name: PNK17, dtype: float64
```

```
1 df['PNK17'].hist(edgecolor='black', bins=20);
```



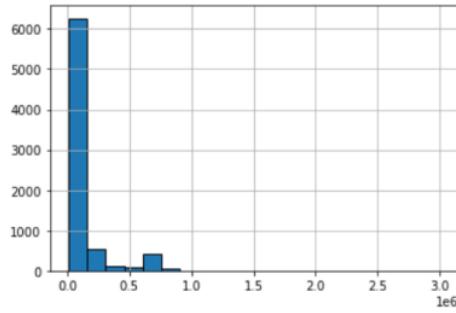
Gambar 5 Biaya Tagih

6. Variabel PNK18 (Biaya Verifikasi)

```
1 df['PNK18'].describe()
```

```
count    7.453000e+03
mean     1.099373e+05
std      1.818357e+05
min      1.000000e+04
25%      3.000000e+04
50%      4.500000e+04
75%      1.200000e+05
max      3.000000e+06
Name: PNK18, dtype: float64
```

```
1 df['PNK18'].hist(edgecolor='black', bins=20);
```



Gambar 6 Biaya Verifikasi

BAB III

DATA PREPARATION

Pada tahapan *data preparation* tim akan melakukan *data preparation* dengan melakukan langkah-langkah ¹⁴ untuk memastikan bahwa kualitas baik pada data. Proses dalam *data preparation* ini terdiri dari *data selection*, *data cleaning*, *data transformation*, dan *data reduction*.

²⁶ 3.1 Data selection

Dalam tahapan ini akan dipilih data yang dibutuhkan. Maka sebelumnya kita harus membaca dan menganalisis data yang telah tersedia yaitu dataset TB20152021_fktpnonkapitasi.dta. Data yang digunakan adalah data FKTP non kapitasi berdasarkan karakteristik fasilitas kesehatan tahun 2021. Dan untuk memperoleh data khusus tahun 2021 akan dilakukan data cleaning pada tahapan proses selanjutnya.

Tabel 3 Data Selection

```
df = pd.read_stata("TB20152021_fktpnonkapitasi.dta")
#melihat dataset
print(df)
```

	PSTV01	PSTV02	PSTV15	PNK02	PNK03	\
0	93858078	93216423	7.659537	183920215Y000376	2015-02-26	
1	93747649	346217457	1.800196	19500915Y000074	2015-09-21	
2	359887820	72989971	0.959155	250630919P000299	2019-09-26	
3	84126594	84126594	1.110887	326360919P001086	2019-09-09	
4	87558937	62126532	32.451832	252721019P001142	2019-10-14	
...	
36488	107435122	107435122	17.946516	108680818Y000770	2018-08-06	
36489	447273237	435380768	2.330930	122070319Y000676	2019-03-09	
36490	80686271	426470275	1.154869	296600118Y000277	2018-01-20	
36491	93887676	93887676	3.204735	296600517Y000031	2017-05-03	
36492	93887676	93887676	3.204735	296600816Y000004	2016-08-01	
	PNK04	PNK05	PNK06	PNK07	PNK08	\
0	2015-02-26	2015-02-27	SULAWESI SELATAN	BARRU		
1	2015-09-23	2015-09-23	KALIMANTAN SELATAN	TAPIN		
2	2019-09-29	2019-09-29	PAPUA	JAYAPURA		
3	2019-09-12	2019-09-12	JAWA TENGAH	REMBANG		
4	2019-10-17	2019-10-17	JAWA TENGAH	KEBUMEN		
...	
36488	2018-08-06	2018-08-06	SULAWESI SELATAN	WAJO		
36489	2019-03-09	2019-03-09	JAWA TIMUR	TUBAN		
36490	2018-01-20	2018-01-20	PAPUA BARAT	RAJA AMPAT		
36491	2017-05-03	2017-05-03	PAPUA BARAT	RAJA AMPAT		
36492	2016-08-01	2016-08-01	PAPUA BARAT	RAJA AMPAT		

Gambar 7 Hasil Run Data Selection

	PNK08	...	PNK10	PNK11	PNK12	\
0	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
1	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
2	TNI AD	...	KLINIK RAWAT INAP	RITP	PPU	
3	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
4	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
...	
36488	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RJTP	PBI APBN	
36489	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RJTP	PBPU	
36490	PEMERINTAH KABUPATEN/KOTA	...	NON RAWAT INAP	RJTP	PBI APBD	
36491	PEMERINTAH KABUPATEN/KOTA	...	NON RAWAT INAP	RJTP	PBI APBD	
36492	PEMERINTAH KABUPATEN/KOTA	...	NON RAWAT INAP	RJTP	PBI APBD	
	PNK13	PNK13A	PNK14	\
0		NaN	9999			
1		NaN	9999			
2	A01 Typhoid and paratyphoid fevers	A01	A010			
3	A01 Typhoid and paratyphoid fevers	A01	A01			
4	A01 Typhoid and paratyphoid fevers	A01	A01			
...	
36488	Z97 Presence of other devices	Z97	Z975			
36489	Z97 Presence of other devices	Z97	Z975			
36490	Z97 Presence of other devices	Z97	Z975			
36491	Z97 Presence of other devices	Z97	Z975			
36492	Z97 Presence of other devices	Z97	Z975			

Gambar 8 Hasil Run Data Selection 2

```

          PNK15 \
0           9999
1           9999
2           Typhoid fever
3           Typhoid and paratyphoid fevers
4           Typhoid and paratyphoid fevers
...
36488 Presence of (intrauterine) contraceptive device
36489 Presence of (intrauterine) contraceptive device
36490 Presence of (intrauterine) contraceptive device
36491 Presence of (intrauterine) contraceptive device
36492 Presence of (intrauterine) contraceptive device

          PNK16   PNK17   PNK18
0       Evakuasi medis / Ambulans Darat 120000 120000
1       Rawat Inap di R. Perawatan Biasa 240000 240000
2       Rawat Inap di R. Perawatan Biasa 600000 600000
3       Rawat Inap di R. Perawatan Biasa 480000 480000
4       Rawat Inap di R. Perawatan Biasa 600000 600000
...
36488 Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36489 Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36490           Pelayanan KB : Suntik    15000 15000
36491           Pelayanan KB : Suntik    15000 15000
36492           Pelayanan KB : Suntik    15000 15000

[36493 rows x 21 columns]

```

Gambar 9 Hasil Run Data Selection 3

3.2 Data Cleaning

Untuk menghasilkan data yang berkualitas, dibutuhkan *cleaning data* untuk membersihkan data. Pada tahapan ini, 3 proses utama yang dilakukan adalah

- Mengisi nilai-nilai yang hilang
- Memeriksa dan menghapus data duplicate
- Menghapus data tahun 2015-2019 karena data yang digunakan adalah data tahun 2021

3.2.1 Missing and Duplicate Values

Pada dataset terdapat missing values yaitu nilai ‘NaN’ dan akan dilakukan cleaning data duplicate values pada dataset.

		PNK08	...	PNK10	PNK11	PNK12	\
0	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
1	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
2	TNI AD	...	KLINIK	RAWAT INAP	RITP	PPU	
3	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
4	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
...		
36488	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RJTP	PBI APBN	
36489	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RJTP	PBPU	
36490	PEMERINTAH KABUPATEN/KOTA	...		NON RAWAT INAP	RJTP	PBI APBD	
36491	PEMERINTAH KABUPATEN/KOTA	...		NON RAWAT INAP	RJTP	PBI APBD	
36492	PEMERINTAH KABUPATEN/KOTA	...		NON RAWAT INAP	RJTP	PBI APBD	
				PNK13	PNK13A	PNK14	\
0				Nan	Nan	9999	
1				Nan	Nan	9999	
2	A01 Typhoid and paratyphoid fevers			A01	A010		
3	A01 Typhoid and paratyphoid fevers			A01	A01		
4	A01 Typhoid and paratyphoid fevers			A01	A01		
...				
36488	Z97 Presence of other devices			Z97	Z975		
36489	Z97 Presence of other devices			Z97	Z975		
36490	Z97 Presence of other devices			Z97	Z975		
36491	Z97 Presence of other devices			Z97	Z975		
36492	Z97 Presence of other devices			Z97	Z975		

Gambar 10 Missing dan Duplicate Values

Pada tahapan ini, proses cleaning akan menampilkan data dari dataset dan selanjutnya akan menghapus data duplikat,

```
import pandas as pd

# Ganti 'nama_file.xlsx' dengan nama file sebenarnya
nama_file_excel = 'fktpnonkapitasi.xlsx'

# Membaca file Excel
dataframe = pd.read_excel(nama_file_excel)

# Menampilkan isi dataframe
print(dataframe)
```

Gambar 11 Menampilkan Data dari Dataset

	PSTV01	PSTV02	PSTV15	PNK02	PNK03	\
0	93858078	93216423	7.659537e+07	183920215Y000376	2015-02-26	
1	93747649	346217457	1.800196e+07	19500915Y000074	2015-09-21	
2	359887820	72989971	9.591551e-01	250630919P000299	2019-09-26	
3	84126594	84126594	1.110887e+07	326360919P001086	2019-09-09	
4	87558937	62126532	3.245183e+06	252721019P001142	2019-10-14	
...	
36488	107435122	107435122	1.794652e+07	108680818Y000770	2018-08-06	
36489	447273237	435380768	2.330938e+07	122070319Y000676	2019-03-09	
36490	80686271	426470275	1.154869e+07	296600118Y000277	2018-01-20	
36491	93887676	93887676	3.204735e+06	296600517Y000031	2017-05-03	
36492	93887676	93887676	3.204735e+06	296600816Y000004	2016-08-01	
	PNK04	PNK05	PNK06	PNK07	PNK08	\
0	2015-02-26	2015-02-27	SULAWESI SELATAN	BARRU		
1	2015-09-23	2015-09-23	KALIMANTAN SELATAN	TAPIN		
2	2019-09-29	2019-09-29		PAPUA	JAYAPURA	
3	2019-09-12	2019-09-12	JAWA TENGAH	REMBANG		
4	2019-10-17	2019-10-17	JAWA TENGAH	KEBUMEN		
...	
36488	2018-08-06	2018-08-06	SULAWESI SELATAN	WAJO		
36489	2019-03-09	2019-03-09	JAWA TIMUR	TUBAN		
36490	2018-01-20	2018-01-20	PAPUA BARAT	RAJA AMPAT		
36491	2017-05-03	2017-05-03	PAPUA BARAT	RAJA AMPAT		
36492	2016-08-01	2016-08-01	PAPUA BARAT	RAJA AMPAT		
	PNK08	...	PNK10	PNK11	PNK12	\
0	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
1	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
2	TNI AD	...	KLINIK RAWAT INAP	RITP	PPU	
3	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
4	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RITP	PBI APBN	
...	
36488	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RJTP	PBI APBN	
36489	PEMERINTAH KABUPATEN/KOTA	...	RAWAT INAP	RJTP	PBPU	
36490	PEMERINTAH KABUPATEN/KOTA	...	NON RAWAT INAP	RJTP	PBI APBD	
36491	PEMERINTAH KABUPATEN/KOTA	...	NON RAWAT INAP	RJTP	PBI APBD	
36492	PEMERINTAH KABUPATEN/KOTA	...	NON RAWAT INAP	RJTP	PBI APBD	
	PNK13	PNK13A	PNK14	\		
0		NaN	NaN	9999		
1		NaN	NaN	9999		
2	A01 Typhoid and paratyphoid fevers		A01	A010		
3	A01 Typhoid and paratyphoid fevers		A01	A01		
4	A01 Typhoid and paratyphoid fevers		A01	A01		
...		
36488	Z97 Presence of other devices		Z97	Z975		
36489	Z97 Presence of other devices		Z97	Z975		
36490	Z97 Presence of other devices		Z97	Z975		
36491	Z97 Presence of other devices		Z97	Z975		
36492	Z97 Presence of other devices		Z97	Z975		

Gambar 12 Hasil Run setelah Menampilkan Data

```

          PNK15 \
0           9999
1           9999
2             Typhoid fever
3       Typhoid and paratyphoid fevers
4       Typhoid and paratyphoid fevers
...
36488 Presence of (intrauterine) contraceptive device
36489 Presence of (intrauterine) contraceptive device
36490 Presence of (intrauterine) contraceptive device
36491 Presence of (intrauterine) contraceptive device
36492 Presence of (intrauterine) contraceptive device

          PNK16   PNK17   PNK18
0   Evakuasi medis / Ambulans Darat 120000 120000
1   Rawat Inap di R. Perawatan Biasa 240000 240000
2   Rawat Inap di R. Perawatan Biasa 600000 600000
3   Rawat Inap di R. Perawatan Biasa 480000 480000
4   Rawat Inap di R. Perawatan Biasa 600000 600000
...
36488 Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36489 Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36490             Pelayanan KB : Suntik    15000 15000
36491             Pelayanan KB : Suntik    15000 15000
36492             Pelayanan KB : Suntik    15000 15000

[36493 rows x 21 columns]

```

Gambar 13 Hasil Run setelah Menampilkan Data 2

```
# Menghilangkan Baris Duplikat
dataframe.drop_duplicates(inplace=True)
```

Gambar 14 Menghilangkan Baris Duplikat

```
# Mengganti Nilai yang Hilang (Missing Values)
nilai_tertentu = 0 # Ganti dengan nilai yang sesuai
dataframe.fillna(nilai_tertentu, inplace=True)

# Menampilkan isi dataframe setelah penggantian nilai yang hilang
print(dataframe)
```

Gambar 15 Mengganti Nilai yang Hilang (Missing Values)

		PNK08	...	PNK10	PNK11	PNK12	\
0	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
1	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
2		TNI AD	...	KLINIK	RAWAT INAP	RITP	PPU
3	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
4	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RITP	PBI APBN	
...		
36488	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RJTP	PBI APBN	
36489	PEMERINTAH KABUPATEN/KOTA	...		RAWAT INAP	RJTP	PBPU	
36490	PEMERINTAH KABUPATEN/KOTA	...		NON RAWAT INAP	RJTP	PBI APBD	
36491	PEMERINTAH KABUPATEN/KOTA	...		NON RAWAT INAP	RJTP	PBI APBD	
36492	PEMERINTAH KABUPATEN/KOTA	...		NON RAWAT INAP	RJTP	PBI APBD	
		PNK13	PNK13A	PNK14	\		
0		0	0	9999			
1		0	0	9999			
2	A01 Typhoid and paratyphoid fevers		A01	A010			
3	A01 Typhoid and paratyphoid fevers		A01	A01			
4	A01 Typhoid and paratyphoid fevers		A01	A01			
...				
36488	Z97 Presence of other devices		Z97	Z975			
36489	Z97 Presence of other devices		Z97	Z975			
36490	Z97 Presence of other devices		Z97	Z975			
36491	Z97 Presence of other devices		Z97	Z975			
36492	Z97 Presence of other devices		Z97	Z975			
		PNK15	\				
0		9999					
1		9999					
2		Typhoid fever					
3		Typhoid and paratyphoid fevers					
4		Typhoid and paratyphoid fevers					
...		...					
36488	Presence of (intrauterine) contraceptive device						
36489	Presence of (intrauterine) contraceptive device						
36490	Presence of (intrauterine) contraceptive device						
36491	Presence of (intrauterine) contraceptive device						
36492	Presence of (intrauterine) contraceptive device						
		PNK16	PNK17	PNK18			
0	Evakuasi medis / Ambulans Darat	120000	120000				
1	Rawat Inap di R. Perawatan Biasa	240000	240000				
2	Rawat Inap di R. Perawatan Biasa	600000	600000				
3	Rawat Inap di R. Perawatan Biasa	480000	480000				
4	Rawat Inap di R. Perawatan Biasa	600000	600000				
...				
36488	Pelayanan KB : Pemasangan IUD / Implant	100000	100000				
36489	Pelayanan KB : Pemasangan IUD / Implant	100000	100000				
36490	Pelayanan KB : Suntik	15000	15000				
36491	Pelayanan KB : Suntik	15000	15000				
36492	Pelayanan KB : Suntik	15000	15000				

[36006 rows x 21 columns]

Gambar 16 Hasil Mengganti Nilai yang Hilang (Missing Values)

Selanjutnya adalah dengan menghapus data dengan nilai yang hilang:

```
# Menghapus Baris atau Kolom dengan Nilai yang Hilang
dataframe.dropna(axis=0) # untuk menghapus baris
dataframe.dropna(axis=1) # untuk menghapus kolom
```

Gambar 17 Menghapus Baris dengan Nilai yang Hilang

	PSTV01	PSTV02	PSTV15	PNK02	PNK03	PNK04	PNK05	PNK06	PNK07	PNK08	...	PNK10	PNK11	PNK12
0	93858078	93216423	7.659537e+07	183920215Y000376	2015-02-26	2015-02-26	2015-02-27	SULAWESI SELATAN	BARRU	PEMERINTAH KABUPATEN/KOTA	—	RAWAT INAP	RJTP	P APB
1	93747649	346217457	1.800195e+07	19500915Y000074	2015-09-21	2015-09-23	2015-09-23	KALIMANTAN SELATAN	TAPIN	PEMERINTAH KABUPATEN/KOTA	—	RAWAT INAP	RJTP	P APB
2	359887820	72989971	9.591551e-01	250630919P000299	2019-09-26	2019-09-29	2019-09-29	PAPUA	JAYAPURA	TNI AD	—	KLINIK RAIWAT INAP	RJTP	PP
3	84126594	84126594	1.110687e+07	326360919P001086	2019-09-09	2019-09-12	2019-09-12	JAWA TENGAH	REMBANG	PEMERINTAH KABUPATEN/KOTA	—	RAWAT INAP	RJTP	P APB
4	87558937	62126532	3.245183e+06	252721019P001142	2019-10-14	2019-10-17	2019-10-17	JAWA TENGAH	KEBUMEN	PEMERINTAH KABUPATEN/KOTA	—	RAWAT INAP	RJTP	P APB
...

Gambar 18 Hasil Run Menghapus Baris dengan Nilai yang Hilang

36488	107435122	107435122	1.794652e+07	108680818Y000770	2018-09-06	2018-08-06	2018-08-06	SULAWESI SELATAN	WAJO	PEMERINTAH KABUPATEN/KOTA	—	RAWAT INAP	RJTP	P APB
36489	447273237	435380768	2.330830e+07	122070319Y000676	2019-03-09	2019-03-09	2019-03-09	JAWA TIMUR	TUBAN	PEMERINTAH KABUPATEN/KOTA	—	RAWAT INAP	RJTP	P APB
36490	80688271	426470275	1.154869e+07	296600118Y000277	2018-01-20	2018-01-20	2018-01-20	PAPUA BARAT	RAJA AMPAT	PEMERINTAH KABUPATEN/KOTA	—	NON RAWAT INAP	RJTP	P APB
36491	93887676	93887676	3.204735e+06	296600517Y000031	2017-05-03	2017-05-03	2017-05-03	PAPUA BARAT	RAJA AMPAT	PEMERINTAH KABUPATEN/KOTA	—	NON RAWAT INAP	RJTP	P APB
36492	93887676	93887676	3.204735e+06	296600816Y000004	2016-08-01	2016-08-01	2016-08-01	PAPUA BARAT	RAJA AMPAT	PEMERINTAH KABUPATEN/KOTA	—	NON RAWAT INAP	RJTP	P APB

36006 rows × 21 columns

Gambar 19 Hasil Run Menghapus Baris dengan Nilai yang Hilang 2

Berikut code untuk mengubah type data yang berisi `tanggal-bulan-tahun` menjadi type data *datetime*

```
# Mengubah kolom PNK03, PNK04, dan PNK05 menjadi tipe data datetime
dataframe['PNK03'] = pd.to_datetime(dataframe['PNK03'], errors='coerce')
dataframe['PNK04'] = pd.to_datetime(dataframe['PNK04'], errors='coerce')
dataframe['PNK05'] = pd.to_datetime(dataframe['PNK05'], errors='coerce')
```

Gambar 20 Menghapus Kolom

Code untuk menampilkan data yang hanya memiliki tahun 2021, karena sample yang digunakan adalah data dengan tahun 2021.

```

# Menyaring data hanya untuk tahun 2021
dataframe_2021 = dataframe[dataframe['PNK03'].dt.year == 2021]
dataframe_2021 = dataframe[dataframe['PNK04'].dt.year == 2021]
dataframe_2021 = dataframe[dataframe['PNK05'].dt.year == 2021]

# Menampilkan isi dataframe setelah penyaringan
print(dataframe_2021)

```

Gambar 21 Menyaring data dan Menampilkan Isi

	PSTV01	PSTV02	PSTV15	PNK02	PNK03	\
16	380665968	79105174	77371845.0	326441221P000568	2021-12-12	
21	2829328	82385224	16625354.0	310450121P000350	2021-01-24	
22	405151945	405151945	11816345.0	203490121P000812	2021-01-26	
73	48109427	48109427	8762647.0	89411220P001121	2020-12-31	
78	84623344	84623344	63545923.0	107380421P001224	2021-04-21	
...
36470	27651694	325386495	4735886.0	129021021P000482	2021-10-19	
36471	413554295	362828666	59456315.0	277220121P000017	2021-01-04	
36472	91793281	91793281	11153269.0	344640121P001554	2021-01-14	
36473	47870233	329612503	10557176.0	16910121P002189	2021-01-19	
36483	88069454	88069454	4764698.0	424091121P002796	2021-11-26	
	PNK04	PNK05	PNK06	PNK07	\	
16	2021-12-14	2021-12-14	JAWA TENGAH	REMBANG		
21	2021-01-26	2021-01-26	RIAU	KOTA DUMAI		
22	2021-01-29	2021-01-29	NUSA TENGGARA BARAT	LOMBOK TENGAH		
73	2021-01-05	2021-01-05	JAWA TIMUR	LAMONGAN		
78	2021-04-24	2021-04-24	NUSA TENGGARA BARAT	LOMBOK BARAT		
...	
36470	2021-10-19	2021-10-19	SUMATERA UTARA	LABUHAN BATU UTARA		
36471	2021-01-04	2021-01-04	SULAWESI SELATAN	SOPPENG		
36472	2021-01-14	2021-01-14	SUMATERA BARAT	PESISIR SELATAN		
36473	2021-01-19	2021-01-19	NUSA TENGGARA BARAT	KOTA BIMA		
36483	2021-11-26	2021-11-26	BANTEN	LEBAK		
	PNK08	...	PNK10	PNK11	PNK12	\
16	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RITP	PPU
21	SWASTA	...	KLINIK	RAWAT	INAP	PPU
22	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RITP	PBI APBN
73	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RITP	PPU
78	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RITP	PBI APBN
...	
36470	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RJTP	PBI APBD
36471	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RJTP	PBI APBD
36472	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RJTP	PBPU
36473	PEMERINTAH KABUPATEN/KOTA	...	RAWAT	INAP	RJTP	PBI APBD
36483	SWASTA	...	KLINIK	NON RAWAT	INAP	PPU
	PNK13	PNK13A	PNK14	\		
16	A01 Typhoid and paratyphoid fevers	A01	A01			
21	A01 Typhoid and paratyphoid fevers	A01	A010			
22	A01 Typhoid and paratyphoid fevers	A01	A010			
73	A01 Typhoid and paratyphoid fevers	A01	A010			
78	A01 Typhoid and paratyphoid fevers	A01	A010			
...	
36470	Z96 Presence of other functional implants	Z96	Z967			
36471	Z96 Presence of other functional implants	Z96	Z96			
36472	Z97 Presence of other devices	Z97	Z975			
36473	Z97 Presence of other devices	Z97	Z975			
36483	Z97 Presence of other devices	Z97	Z975			

Gambar 22 Hasil Menyaring data dan Menampilkan Isi

```

          PNK15 \
16           Typhoid and paratyphoid fevers
21           Typhoid fever
22           Typhoid fever
73           Typhoid fever
78           Typhoid fever
...
36470   Presence of other bone and tendon implants
36471   Presence of other functional implants
36472 Presence of (intrauterine) contraceptive device
36473 Presence of (intrauterine) contraceptive device
36483 Presence of (intrauterine) contraceptive device

          PNK16   PNK17   PNK18
16     Rawat Inap di R. Perawatan Biasa 320000 320000
21     Rawat Inap di R. Perawatan Biasa 240000 240000
22     Rawat Inap di R. Perawatan Biasa 540000 540000
73     Rawat Inap di R. Perawatan Biasa 900000 900000
78     Rawat Inap di R. Perawatan Biasa 540000 540000
...
36470   Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36471   Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36472   Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36473   Pelayanan KB : Pemasangan IUD / Implant 100000 100000
36483   Pelayanan KB : Pemasangan IUD / Implant 100000 100000

```

[7453 rows x 21 columns]

Gambar 23 Hasil Menyaring data dan Menampilkan Isi 2

19
Menghapus data dengan tahun 2015, 2016, 2017, 2018, 2019 dan 2020.

```

# Get the years to be deleted
delete_years = [2015, 2016, 2017, 2018, 2019, 2020]

# Create masks for each column
mask_PNK03 = dataframe["PNK03"].isin(delete_years)
mask_PNK04 = dataframe["PNK04"].isin(delete_years)
mask_PNK05 = dataframe["PNK05"].isin(delete_years)

# Combine masks
mask = mask_PNK03 | mask_PNK04 | mask_PNK05

# Delete data
df = dataframe[~mask]

```

Gambar 24 Menghapus Data

BAB IV

BUILD MODEL

Pada bab sebelumnya, tim telah mempersiapkan data untuk dapat digunakan dalam membangun model.

4.1 Select Modeling Technique

Pada proses ini, tim akan membangun model Regresi, yaitu Regresi Multi Linear. Alasan tim menerapkan algoritma ini adalah untuk membantu tim dalam membuat sebuah keputusan. Banyak algoritma yang dapat melakukan Prediksi, namun tidak semua algoritma memiliki performansi yang sama karena dipengaruhi oleh kualitas data yang dimiliki. Untuk itu tim akan memilih model Regresi yang ⁴ kinerjanya lebih baik dari pada regresi lainnya.

Metode Regresi adalah metode statistik yang digunakan untuk menganalisis hubungan antara dua atau lebih variabel. Hal yang dipelajari disebut variabel respon, sedangkan hal yang mempengaruhi variabel respon disebut variabel prediktor. Variabel respon adalah variabel ingin diketahui pengaruhnya terhadap variabel lain. Variabel respon diketahui sebagai variabel terikat, variabel dependent atau variabel keluaran. Sedangkan, variabel Prediktor adalah variabel yang memiliki pengaruh terhadap variabel respon. Variabel prediktor disebut sebagai variabel bebas, variabel independen atau variabel masukan.

4.2 Model Building

Pada sub bab ini akan menggambarkan bagaimana tahapan pembangunan model, yang mana ini digunakan untuk mencari hasil dari prediksi biaya pasien pada FKTP. Berikut merupakan proses membangun model:

1. Mengimpor pustaka-pustaka yang dibutuhkan dalam analisis data dan pembangunan model

```
import pandas as pd
import numpy as np
import seaborn as sns
%matplotlib inline
import matplotlib.pyplot as plt
import math
import sklearn
from sklearn.model_selection import train_test_split
```

Gambar 25 Mengimpor pustaka-pustaka

2. Membaca data dari file CSV yang digunakan yaitu fktpnonkapitasi.csv ke dalam sebuah DataFrame menggunakan pustaka Pandas ('pd').

```
#Read the data, show the head  
df = pd.read_csv('fktpnonkapitasi.csv')  
df.head()
```

Gambar 26 Read Data CSV

3. Menampilkan informasi umum tentang dataset: jumlah baris dan kolom, tipe data setiap kolom.

```
# view data information  
df.info()
```

Gambar 27 View Data Information

4. Menyajikan ringkasan statistik deskriptif dari dataset

```
# describe the data  
df.describe()
```

Gambar 28 Describe the Data

5. Memeriksa nilai null(nilai kosong) dalam setiap kolom.

```
# Check null data  
df.isna().sum()
```

Gambar 29 Check Null Data

6. Memisahkan fitur (variabel independen) dan variabel target(variabel dependen) dari dataset. Variabel x berisi atribut yang digunakan untuk melakukan prediksi atau analisis, sedangkan variabel y berisi nilai yang ingin diprediksi atau dianalisis. Variabel m akan menyimpan jumlah baris data yang dapat digunakan.

```
X = df[['PNK09', 'PNK10', 'PNK11', 'PNK16', 'PNK17']]  
y = df['PNK18']  
m = len(y)
```

Gambar 30 Variabel Independen

7. Membagi data menjadi subset pelatihan (training) dan pengujian (testing)

```
x_train , x_test , y_train , y_test = train_test_split(X , y , test_size = 0.10,random_state =2)
```

Gambar 31 Subset Training dan Testing

8. Melatih model dengan data pelatihan, setelah proses pelatihan selesai maka model regresi linear akan digunakan untuk melakukan prediksi pada data baru atau dievaluasi kinerjanya pada data pengujian.

```
1 from sklearn.preprocessing import LabelEncoder  
2  
3 label_encoder = LabelEncoder()  
4 df['PNK16'] = label_encoder.fit_transform(df['PNK16'])  
5  
6 # Lanjutkan dengan persiapan regresi |  
7 X = df[['PNK09', 'PNK10', 'PNK11', 'PNK16', 'PNK17']]  
8 y = df['PNK18']  
9 x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.10, random_state=2)  
10  
11 reg = LinearRegression()  
12 reg.fit(x_train, y_train)  
13  
LinearRegression()
```

Gambar 32 Melatih Model dengan Testing

9. Menghitung skor atau koefisien determinasi (R-Squared) pada data pengujian (x_test dan y_test)

```
1 reg.score(x_test,y_test)  
1.0
```

Gambar 33 Menghitung Skor (R-Squared)

10. Menggunakan model regresi ensemble Gradient Boosting dari scikit-learn untuk meningkatkan kinerja dan akurasi prediksi.

```
1 import sklearn.ensemble as ensemble  
2 import pandas as pd  
3  
4 clf = ensemble.GradientBoostingRegressor(loss='squared_error')  
5  
1 clf.fit(x_train, y_train)  
GradientBoostingRegressor()  
  
1 clf.score(x_test,y_test)  
0.9999912113192935
```

Gambar 34 Menghitung Akurasi dan Regresi Ensemel Gradient Boosting

BAB V

MODEL EVALUATION

Pada bab ini akan menjelaskan model evaluation yang meliputi Result Evaluation, dan Result Review.

5.1 Result Evaluation

Pada sub-bab ini akan diterangkan result/ hasil dan evaluation dari model yang sudah dibuat.

```
1 from sklearn import metrics
2 import numpy as np
3
4 # Assuming y_pred is the predictions made by your model
5
6 meanAbErr = metrics.mean_absolute_error(y_test, y_pred)
7 meanSqErr = metrics.mean_squared_error(y_test, y_pred)
8 rootMeanSqErr = np.sqrt(metrics.mean_squared_error(y_test, y_pred))
9
10 print('Mean Absolute Error:', meanAbErr)
11 print('Mean Square Error:', meanSqErr)
12 print('Root Mean Square Error:', rootMeanSqErr)
13 print("MAPE :", (np.abs(y_test - y_pred) / y_test).mean())
14
```

Mean Absolute Error: 106.36560227296526
Mean Square Error: 278557.30485152313
Root Mean Square Error: 527.7852829053905
MAPE : 0.0013264673249713027

Gambar 35 Result Evaluation

5.2 Result Review

Hasil yang didapatkan dari hasil evaluasi, yaitu dimana, Evaluasi yang harus dipenuhi untuk Regression Problem adalah MAE < 900 dan MAPE < 70%.

- 1) MAE (Mean Absolute Error) dari program ini adalah 106.36560227296526 yang berarti bahwa telah **memenuhi** evaluasi atau lebih kecil dari 900.
- 2) MAPE (Mean Absolute Percentage Error) dari program ini adalah 0.0013264673249713027 atau setara dengan 14% artinya untuk hasil MAPE ini telah **memenuhi** evaluasi atau lebih kecil 70%

BAB VI

DEPLOYMENT

Tahap keenam ini akan dilakukan *cost prediction* deployment. Pada bab ini kelompok kami akan melakukan *cost prediction* deployment dan akan di jelaskan pengerjaan dari deployment pada lokal komputer.

6.1 Deployment

Pada fase plan deployment ini, model yang telah terbentuk pada fase modeling akan digunakan sesuai dengan tujuan data mining yang dibutuhkan. Adapun tahapan untuk melakukan deployment sebagai berikut:

1. Membuat model dalam bentuk file pickle (model.pkl) dan menyimpannya di local komputer dengan format model.pkl dengan code:

```
1 import pickle  
2 pickle.dump(reg,open('model.pkl','wb'))
```

Gambar 36 Model.pkl



Gambar 37 FlashWebApp

2. Membuat code untuk memanggil model. Code dibawah akan disimpan dalam format .py dan berfungsi untuk menghubungkan dengan template htmlnya. Berikut codenya:

```
In [4]: 1 from flask import Flask, request, render_template  
2 import pickle  
3 import numpy as np  
4  
5 app = Flask(__name__)  
6  
7 # Memuat model  
8 model_file = open('model.pkl', 'rb')  
9 model = pickle.load(model_file, encoding='bytes')  
10  
11 # Fungsi untuk merender halaman utama  
12 @app.route('/')  
13 def index():  
14     return render_template('Web Application.html', cases=0)  
15  
16 @app.route('/predict', methods=['POST'])  
17 def predict():  
18     PNK09, PNK10, PNK11, PNK16, PNK17, PNK18 = [x for x in request.form.values()]  
19  
20     data = []  
21  
22     data.append(object(PNK09))  
23     data.append(object(PNK10))  
24     data.append(object(PNK11))  
25     data.append(object(PNK16))  
26     data.append(int64(PNK17))  
27     data.append(int64(PNK18))  
28  
29     prediction = model.predict([data])  
30     output = object(prediction[0])  
31     output = int64(prediction[0])  
32     return render_template('Web Application.html', case=output, PNK09=PNK09, PNK10=PNK10, PNK11=PNK11, PNK16=PNK16, PNK17=PNK17, PNK18=PNK18)
```

Gambar 38 Code untuk Memanggil Model

3. Menjalankan code berikut untuk mendapatkan URL dari deployment.

```
In [*]: 1 | !run app.py
* Serving Flask app "app" (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
* Debug mode: on
* Restarting with watchdog (windowsapi)
* Restarting with watchdog (windowsapi)
```

Gambar 39 Run App

Pada tahap terakhir saat melakukan *deployment* di local komputer tidak berhasil, ditandai dengan ketidakmunculan hasil yaitu <http://127.0.0.1:5000/> (tidak dapat running).

6.2 Web Application

Pada bagian web application dibawah, akan ditampilkan tampilan front end dari Prediksi Biaya Pasien pada FKTP dengan Algoritma Regresi Multi linear, web application berhasil di deploy dan dapat dilihat pada gambar di bawah ini.

Projek Penambangan Data
Predicting Patient Cost to Facultas Kesehatan Tahap Pertama(FKTP) with Linear Regression

Prediksi Biaya Pasien untuk Tahap Pertama (FKTP)

Jenis Faskes

Tipe Faskes

Nama Tindakan

Tingkat Layanan

RTRP

Tampilan Biaya

Biaya Pasien

Total Biaya: Rp 0

Gambar 40 Web Application Sebelum Input Data

Projek Penambangan Data
Predicting Patient Cost to Facultas Kesehatan Tahap Pertama(FKTP) with Linear Regression

Prediksi Biaya Pasien untuk Tahap Pertama (FKTP)

Jenis Faskes

Publikitas

Tipe Faskes

resep obat

Nama Tindakan

Evaluasi medis

Tingkat Layanan

RTRP

Tampilan Biaya

Biaya Pasien

Total Biaya: Rp 114.000

Gambar 41 Web Application Setelah Input Data

LAMPIRAN

1. Link Github: [lydiaharianja/Group-8_Project-Data-Mining_Case-3-Using-Regression
\(github.com\)](https://github.com/lydiaharianja/Group-8_Project-Data-Mining_Case-3-Using-Regression)

Proyek DAMI

ORIGINALITY REPORT

20%

SIMILARITY INDEX

13%

INTERNET SOURCES

5%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Pelita Harapan Student Paper	6%
2	repo.itera.ac.id Internet Source	2%
3	ejurnal.undana.ac.id Internet Source	1%
4	es.scribd.com Internet Source	1%
5	Submitted to Universitas Negeri Semarang Student Paper	1%
6	repository.upi.edu Internet Source	1%
7	core.ac.uk Internet Source	1%
8	geograf.id Internet Source	1%
9	repository.uin-suska.ac.id Internet Source	1%

10	sipora.polije.ac.id Internet Source	<1 %
11	users.stat.umn.edu Internet Source	<1 %
12	Submitted to Institute of Technology Blanchardstown Student Paper	<1 %
13	repositori.usu.ac.id Internet Source	<1 %
14	docobook.com Internet Source	<1 %
15	pt.scribd.com Internet Source	<1 %
16	id.stpatrickschoolwhitelake.org Internet Source	<1 %
17	j-cup.org Internet Source	<1 %
18	jurnal.harianregional.com Internet Source	<1 %
19	nxbom.treasure-gnss.eu Internet Source	<1 %
20	peraturan.bpk.go.id Internet Source	<1 %
21	repository.trisakti.ac.id	

Internet Source

<1 %

22 www.slideshare.net

Internet Source

<1 %

23 id.123dok.com

Internet Source

<1 %

24 repository.uinsu.ac.id

Internet Source

<1 %

25 repository.unikom.ac.id

Internet Source

<1 %

26 Ahmad Habib Husaini, Rini Mayasari, Susilawati Susilawati. "Pengelompokan Ulasan Aplikasi PeduliLindungi dengan Algoritma K-Medoids", INTECOMS: Journal of Information Technology and Computer Science, 2022

Publication

<1 %

27 Nina Rahmadiliyani, Muhammad Khalid Fahrizal Adha. "Pelaksanaan Rujukan Rawat Jalan Pelayanan Kesehatan Tingkat Pertama Pasien BPJS di UPT Puskesmas Martapura Barat", Jurnal Kesehatan Indonesia, 2019

Publication

<1 %

Exclude bibliography On