

Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi

Alfanda Novebrian Maharadja^{1*}, Iqbal Maulana^{2*}, Budi Arif Dermawan^{3*}

* Teknik Informatika, Universitas Singaperbangsa Karawang

alfanda.maharadja17005@student.unsika.ac.id¹, iqbal.maulana@staff.unsika.ac.id², budi.arif@staff.unsika.ac.id³

Article Info

Article history:

Received 01-07-2021

Revised 23-07-2021

Accepted 23-07-2021

Keyword:

Koefisien Determinasi,
Korupsi,
Regresi Linear Berganda,
RMSE.

ABSTRACT

Tindak pidana korupsi merupakan kegiatan yang dapat mengakibatkan kerugian keuangan negara atau perekonomian negara serta dapat menghambat pembangunan nasional. Semenjak penindakan kasus korupsi 2013-2020, pada tahun 2014 merupakan angka tertinggi dalam jumlah kasus, yaitu sebanyak 629 kasus, sedangkan pada tahun 2020 negara mengalami kerugian tertinggi sebesar Rp. 18,6 Triliun. Adanya permasalahan tersebut perlu dilakukan kebijakan yang tepat sertaantisipasi dalam meminimalisir kerugian negara pada tahun selanjutnya. Oleh karena itu penelitian ini melakukan prediksi kerugian negara berdasarkan tindak pidana korupsi dengan menggunakan regresi linear berganda. Regresi linear berganda merupakan salah satu metode statistik yang digunakan untuk menelusuri pola hubungan antara variabel terikat dengan dua atau lebih variabel bebas. Pembelajaran regresi linear berganda dalam penelitian ini menghasilkan model regresi yang dimana menghasilkan nilai konstanta yaitu 284645.5891073216 serta nilai koefisien yaitu -139837.38007863 dan 363493.06049751. Kemudian penelitian ini melakukan pengukuran performa model regresi linear dengan kondisi pembagian data 80% untuk data *training* dan 20% untuk data *testing*. Dari kondisi pembagian data tersebut memperoleh nilai RMSE sebesar 8447373.485 untuk *data training* dan 9769609.026 untuk *data testing*. Sedangkan untuk nilai koefisien determinasi memperoleh nilai sebesar 0.579 untuk *data training* yang tingkat hubungan antar variabelnya cukup kuat dan 0.662 untuk *data testing* yang berarti tingkat hubungan antar variabelnya kuat. Dengan melakukan prediksi menggunakan metode regresi linear berganda dapat memberikan informasi yang membantu pemerintah dalam mengambil kebijakan yang tepat terhadap permasalahan kasus korupsi serta meminimalisir dan mengantisipasi kerugian negara yang lebih besar untuk tahun selanjutnya.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Korupsi merupakan bentuk/jenis kasus tindak pidana, dalam UU No. 31 Tahun 1999 yang telah diubah dengan UU No. 20 Tahun 2001 tentang Pemberantasan Tindak Pidana Korupsi dijelaskan bahwa tindakan pidana korupsi sangat merugikan keuangan negara atau perekonomian negara dan menghambat pembangunan nasional. Berbagai upaya pemerintah untuk memberantas korupsi telah dilakukan di setiap periode pemerintahan negara ini. Tindakan pemerintah dalam pemberantasan korupsi diantaranya, menerbitkan Keppres No. 28 Tahun 1967

tentang Pembentukan Tim Pemberantasan Korupsi, namun dalam pelaksanaannya tidak bisa dilakukan secara maksimal, bahkan bisa dikatakan hampir tidak berfungsi. Selanjutnya di saat pemerintahan Presiden Abdurrahman Wahid muncul Tap MPR Nomor XI/MPR/1998 tentang Pengelolaan Negara yang Bersih dan Bebas Korupsi, Kolusi, dan Nepotisme kemudian membentuk badan-badan negara untuk mendukung upaya pemberantasan korupsi. Pada masa kepemimpinan Presiden Megawati Soekarno Putri membentuk Komisi Pemberantasan Tindak Pidana Korupsi (KPTPK) yang merupakan cikal bakal Komisi Pemberantasan Korupsi (KPK), dan di era Presiden Susilo Bambang Yudhoyono, beliau menerbitkan Instruksi

Presiden Nomor 5 Tahun 2004, kemudian dilanjutkan dengan penyiapan Rencana Aksi Nasional Pemberantasan Korupsi (RAN) yang disusun Bappenas [1].

Namun berbagai upaya pemerintah dalam memberantas korupsi belum menunjukkan penanganan kongkret. Dapat dilihat pada Tabel 1 semenjak penindakan kasus korupsi 2013-2020, pada tahun 2014 merupakan angka tertinggi jumlah kasus, dengan 629 kasus. Sedangkan pada tahun 2020 negara mengalami kerugian tertinggi sebesar Rp. 18,6 Triliun [2].

TABEL I.
REKAPITULASI KERUGIAN NEGARA BERDASARKAN KASUS
TINDAK PIDANA KORUPSI TAHUN 2013-2020

Tahun	Jumlah Kasus	Jumlah Kerugian
2013	560	Rp6.011.269.014.296
2014	629	Rp7.183.496.886.861
2015	550	Rp3.107.725.943.808
2016	482	Rp1.450.117.648.234
2017	576	Rp6.562.929.686.822
2018	454	Rp5.645.728.426.504
2019	271	Rp8.405.040.149.122
2020	444	Rp18.615.067.880.559

Penelitian ini nantinya akan menerapkan metode regresi linear merupakan suatu algoritma yang digunakan untuk menelusuri pola hubungan antara variabel terikat dengan dua atau lebih variabel bebas [3]. Penelitian sebelumnya yang menerapkan metode regresi linear untuk melakukan perbandingan antara regresi linear, *backpropagation*, dan *fuzzy mamdani* yang dimana dalam penelitian tersebut menunjukkan bahwa korelasi dari regresi linear sangat bagus, yaitu 0,929 [4]. Selain itu penelitian lainnya, memprediksi harga saham dengan regresi linear berganda yang dimana dilakukan *hybrid* dengan *K-Means* dan *Moving Avarage* mampu menghasilkan *Root Mean Square Error* sebesar 122,831 [5]. Sebagai tambahan dalam prediksi harga sembako dengan menggunakan metode regresi linear berganda [6] menghasilkan nilai koefisien determinasi sebesar 0,842 yang dimana pengaruh variabel independen di dalam variabel dependen menghasilkan presentase 84,2%, sedangkan sisanya sebesar 15,8% dipengaruhi oleh variabel yang tidak dimasukkan ke dalam model penelitian.

Berdasarkan penelitian sebelumnya tersebut, maka penulis tertarik mengusulkan penelitian tentang penerapan metode regresi linear dalam membangun model prediksi kerugian negara berdasarkan kasus tindak pidana korupsi.

Dari penjelasan yang telah disebutkan di atas, pemetaan permasalahan perlu dilakukan agar Pemerintah dapat mengambil kebijakan yang tepat terhadap permasalahan kasus korupsi serta dapat mengantisipasi kerugian negara yang lebih besar untuk tahun selanjutnya.

II. METODE PENELITIAN

A. Studi Literatur

Studi literatur merupakan pengumpulan data dengan cara mempelajari dan memahami hal-hal yang berkaitan dengan penelitian sehingga sumber-sumber studi literatur yang digunakan sebagai studi pustaka. Hal-hal yang diperlukan dalam penelitian meliputi metode prediksi menggunakan algoritma regresi linear berganda.

B. Dataset

Dataset yang digunakan adalah data sekunder yang diperoleh secara langsung melalui instansi *Indonesia Corruption Watch* serta dapat diakses melalui *website* resmi acuan ICW yaitu <https://antikorupsi.org>. *Dataset* yang diperoleh merupakan data kerugian negara dari tahun 2013 – 2020.

C. Regresi Linear Berganda

Regresi linear berganda adalah salah satu cara untuk melakukan prediksi yang melibatkan dua atau lebih variabel yaitu variabel pemberi pengaruh dan variabel terpengaruh. Variabel-variabel tersebut bersifat saling terkait atau ada hubungan sebab akibat. Metode regresi akan menjelaskan hubungan antar variabel tersebut [7]. Regresi linear berganda memperlihatkan hubungan variabel dependent dan independent dengan grafik berupa garis lurus. Melalui perhitungan regresi linear akan menghasilkan persamaan yang dapat dijadikan acuan untuk memperkirakan nilai variabel dependent di waktu mendatang dengan memasukan nilai variabel independent ke dalam persamaan [7]. Adapun model persamaan regresi linear berganda adalah sebagai berikut :

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (1)$$

$$a = \frac{\sum y - b(\sum x)}{n} \quad (2)$$

$$y = a + b_1x_1 + \dots + b_nx_n \quad (3)$$

D. Penguji Keakurasian

Keakurasian suatu prediksi ditentukan oleh seberapa besar penyimpangan atau kesalahan ini, yang terjadi antara data yang diprediksi dengan data yang sebenarnya atau data aktual [8], [5]. Kesalahan dalam perumusan sebuah prediksi tidak hanya disebabkan oleh unsur *error* tapi juga ketidakmampuan suatu model peramalan mengenali unsur yang lain dalam deret data yang mempengaruhi besarnya penyimpangan dalam prediksi [9]. Besarnya kesalahan atau penyimpangan ini dapat disebabkan oleh besarnya faktor yang tidak diduga (*outliers*) dimana tidak ada metode prediksi yang mampu menghasilkan prediksi yang akurat atau dapat juga disebabkan metode prediksi yang digunakan tidak dapat memprediksi dengan tepat

komponen tren, komponen musiman atau komponen siklus yang mungkin terdapat dalam deret data [9].

Adapun dalam menghitung besarnya kesalahan tersebut yaitu adalah *root mean square error* (RMSE). RMSE merupakan akar dari MSE serta menggunakan metode berbasis gradien. Hasil prediksi dinyatakan baik jika semakin rendah nilainya RMSE maka semakin baik prediksi yang dilakukan.

Adapun rumus dari pengujian RMSE tersebut sebagaimana ditampilkan pada persamaan (4), adalah sebagai berikut.

$$RMSE = \sqrt{\frac{1}{n} \sum (Y_t - Y'_t)^2} \tag{4}$$

Dalam suatu model pun perlu dilihat dari segi kecocokannya, dalam melihat kecocokan suatu model dapat menerapkan persamaan koefisien determinasi (R^2). Dinamakan koefisien determinasi karena $R^2 \times 100\%$ daripada variasi yang terjadi dalam variabel tak bebas Y dapat dijelaskan oleh variabel bebas X dengan adanya regresi linier Y atas X [10]. Besarnya nilai koefisien determinasi adalah berkisar $0 < R^2 < 1$. Artinya jika R^2 mendekati 1 maka dapat dikatakan pengaruh variabel bebas terhadap variabel terikat adalah sangat kuat. Berikut interpretasi koefisien determinasi [11].

TABEL II.
INTERPRETASI KOEFISIEN DETERMINASI

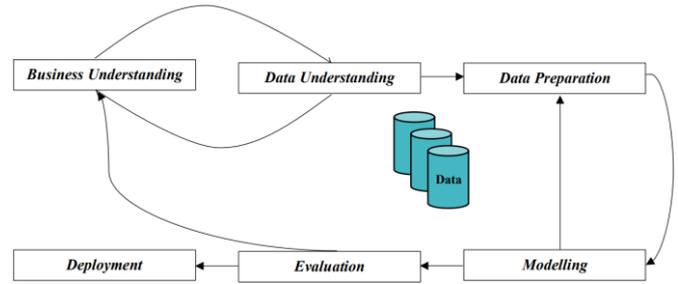
Interval Koefisien	Tingkat Hubungan
0,80 – 1,000	Sangat Kuat
0,60 – 0,799	Kuat
0,40 – 0,599	Cukup Kuat
0,20 – 0,399	Rendah
0,00 – 0,199	Sangat Rendah

Adapun rumus dari koefisien determinasi tersebut sebagaimana ditampilkan pada persamaan (5), adalah sebagai berikut.

$$R^2 = \frac{b_1(\sum x_1 y) + b_2(\sum x_2 y) + \dots + b_n(\sum x_n y)}{\sum y^2} \tag{5}$$

E. Metode Analisis Data

Metode analisis data yang digunakan pada penelitian yaitu *Cross-Industry Standard Process for Data Mining*.



Gambar 1. The CRISP-DM cycle [12].

Metodologi CRISP – DM merupakan standarisasi proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit [13]. Dapat dilihat Gambar 1 yang merupakan enam fase saling terkait yang digunakan untuk menggambarkan proses CRISP-DM [12], adapun penjelasan enam fase tersebut sebagai berikut.

- 1) *Business Understanding*
Tahapan ini merupakan fase awal untuk mengetahui masalah yang diselesaikan untuk mencapai tujuan yang diinginkan.
- 2) *Data Understanding*
Pada tahap ini diperlukan untuk mempersiapkan data yang akan diolah agar dapat dimodelkan dengan memeriksa apakah data tersebut normal, lengkap dan konsisten sehingga dapat dimodelkan sesuai dengan metode data mining yang akan digunakan.
- 3) *Data Preparation*
Tahap ini mencakup semua aktivitas untuk membuat kumpulan data akhir. Kumpulan data yang dihasilkan biasanya berupa data mentah dan kurang berkualitas, misalnya nilai yang hilang, nilai masukan yang salah, dan ketidakkonsistenan. Akibatnya, data harus diproses terlebih dahulu. Proses pembersihan meliputi penghapusan data duplikat, pengisian / pemrosesan data yang hilang, koreksi data yang tidak konsisten, dan koreksi kesalahan ketik.
- 4) *Modelling*
Merupakan tahapan pemilihan teknik data mining dengan menentukan algoritma yang akan digunakan.. Dalam tahap ini, berbagai macam teknik pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis tertentu. Tahap pembuatan model juga mencakup penilaian dan analisa komparatif dari berbagai model yang dibangun.
- 5) *Evaluation*
Setelah tahap *modelling* selesai dilakukan, model tersebut harus dievaluasi untuk melihat kualitas dan efektivitas sebelum disebarakan untuk digunakan. Pada tahap ini juga ditentukan apakah model dapat mencapai

tujuan yang telah ditetapkan pada tahap pertama, apakah terdapat permasalahan penting dari penelitian yang tidak tertangani dengan baik, dan mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.

6) *Deployment*

Pada tahap terakhir, penelitian yang telah dilakukan akan dipresentasikan dalam bentuk laporan dan dapat direkomendasikan ke pihak yang bersangkutan.

III. HASIL DAN PEMBAHASAN

A. *Hasil Penelitian*

Penelitian ini menggunakan metode regresi linear berganda untuk memprediksi kerugian negara berdasarkan kasus tindak pidana korupsi yang diperoleh melalui lembaga *Indonesia Corruption Watch*.

1) *Bussiness Understanding*.

Tahap ini, terdapat permasalahan kasus tindak pidana korupsi yang mengakibatkan kerugian negara. Sehingga penelitian ini bertujuan untuk membangun model prediksi kerugian negara berdasarkan kasus tindak pidana korupsi sebagai bagian dari kajian kerentanan korupsi agar dapat membantu pemerintah dalam menentukan kebijakan yang tepat terhadap permasalahan kasus korupsi.

2) *Data Understanding*

Tahap ini meliputi, pencarian data lalu diidentifikasi dan dilakukan pemahaman terhadap data. Pada tahap ini data yang digunakan adalah data kuantitatif, dimana jenis data yang dapat diukur, dihitung serta dapat dideskripsikan dengan menggunakan angka.

Tahun	Jenis Kasus	Jumlah Kasus	Jumlah Kerugian	
0	2013	1	2	6800000000
1	2014	1	6	2381000000
2	2015	1	1	140000000
3	2016	1	2	1600653653
4	2017	1	0	0
...
107	2016	14	0	0
108	2017	14	0	0
109	2018	14	0	0
110	2019	14	0	0
111	2020	14	5	1691500000000

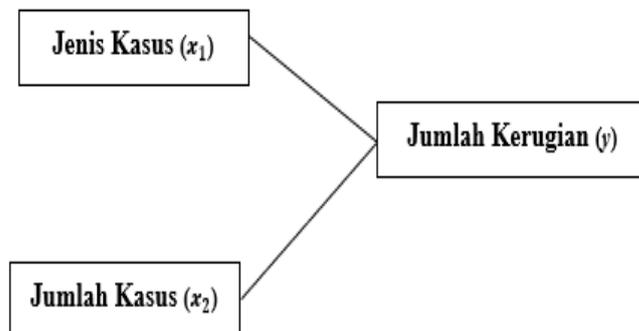
Gambar 2. Data Kerugian Negara Periode 2013 sampai 2020.

Dari Gambar 2, dapat dijelaskan bahwa data tersebut merupakan data sekunder yakni perolehan dari lembaga *Indonesia Corruption Watch*. Kondisi data bersifat kuantitatif karena informasi didalamnya berupa angka atau bilangan.

TABEL III.
DESKRIPSI ATRIBUT

Atribut	Keterangan
Tahun	Menjelaskan data kerugian negara berdasarkan tahun.
Jenis Kasus	Jenis kasus yang mengakibatkan kerugian negara, dengan kondisi terbagi menjadi 14 jenis kasus, adapun jenis kasusnya yaitu, sebagai berikut : 1 = Anggaran Ganda 2 = Kegiatan / Proyek Fiktif 3 = Laporan Fiktif 4 = Mark Down 5 = Mark Up 6 = Pemerasan 7 = Penggelapan 8 = Penyalahgunaan Anggaran 9 = Penyalahgunaan Wewenang 10 = Pematangan / Penyunatan 11 = Pungutan Liar 12 = Suap / Gratifikasi 13 = Pencucian Uang 14 = Manipulasi Saham
Jumlah Kasus	Menjelaskan data kerugian negara berdasarkan jumlah kasus.
Jumlah Kerugian	Data jumlah kerugian yang diperoleh dari lembaga <i>Indonesia Corruption Watch</i> .

Berdasarkan kondisi data yang telah diperoleh maka penentuan atribut yang akan digunakan sebagai variabel independent adalah Jenis kasus dan Jumlah Kasus sedangkan untuk variabel dependen adalah Jumlah Kasus.



Gambar 3. Penentuan Variabel Independen dan Dependen

3) *Data Preparation*

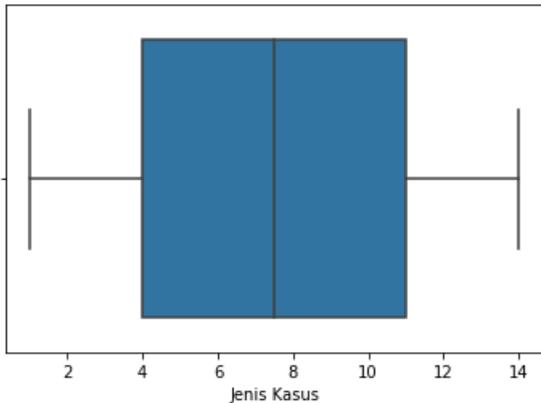
Ada beberapa tahap dalam *data preparation* yang dilakukan, berikut adalah langkah-langkah yang dilakukan.

- Pada tahap *Select Data* dilakukan proses untuk memilih data yang dibutuhkan untuk tujuan permodelan. Data yang digunakan adalah data pada Gambar 2.
- Selanjutnya pada tahap *cleaning data* dilakukan pemeriksaan *missing value* pada masing-masing atribut dengan menggunakan perintah sebagai berikut :

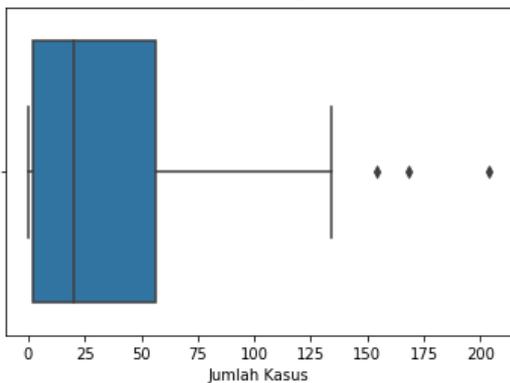
TABEL IV.
HASIL PEMERIKSAAN *MISSING VALUE*

Atribut	Jumlah <i>Missing Value</i>
Jenis Kasus	0
Jumlah Kasus	3
Jumlah Kerugian	11

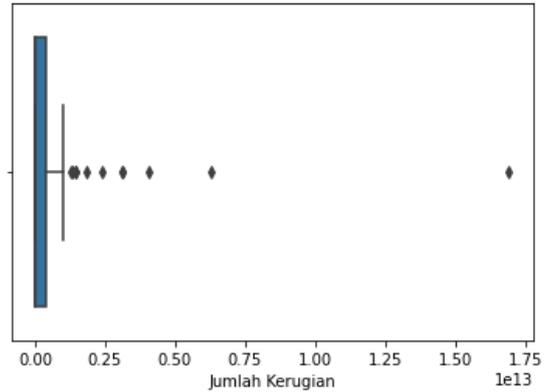
Pada Tabel IV dapat dilihat bahwa dari masing-masing atribut tidak di temukan *missing value* pada data tersebut. Selanjutnya dalam proses *cleaning data* perlu diketahui penyebaran suatu variabel pada masing-masing atribut dengan menggunakan perintah `sns.boxplot`.



Gambar 5. Pendistribusian Data pada Atribut Jenis Kasus



Gambar 6. Pendistribusian Data pada Atribut Jumlah Kasus



Gambar 7. Pendistribusian Data pada Atribut Jumlah Kerugian

Berdasarkan visualisasi penyebaran suatu variabel bahwa atribut *Jenis Kasus* berdistribusi normal sedangkan untuk atribut *Jumlah Kasus* dan *Jumlah Kerugian* mengalami *outlier*. Sebagai berikut perintah untuk mengetahui kuantitas data yang mengalami *outlier*.

```
#Melihat missing value
df.isnull().sum()
```

TABEL V.
KUANTITAS DATA YANG MENGALAMI *OUTLIER*

Atribut	Jumlah <i>Outlier</i>
Jenis Kasus	0
Jumlah Kasus	3
Jumlah Kerugian	11

Telah diketahui bahwa penyebaran data suatu variabel pada atribut *Jumlah Kasus* yang mengalami *outlier* sebanyak 3 dan *Jumlah Kerugian* sebanyak 11. Dalam menangani data *outlier* diterapkan perintah `dropna(axis = 0)` untuk menghapus baris yang mengalami *outlier* serta dilakukan penghapusan nilai 0 pada atribut *Jumlah Kasus* dan *Jumlah kerugian*.

Sebagai berikut perintah untuk menghapus baris yang mengalami *outlier* pada dataset.

```
# menghapus outlier
df = df.dropna(axis = 0)
```

Selanjutnya melakukan penanganan untuk menghapus nilai 0 pada atribut jumlah kasus dan jumlah kerugian, sebagai berikut perintah yang dilakukan untuk menghapus nilai 0 :

```
# menghapus nilai 0
index_names = df[df['Jumlah Kasus'] == 0].index
df.drop(index_names, inplace = True)

# menghapus nilai 0
index_names = df[df['Jumlah Kerugian'] == 0].index
df.drop(index_names, inplace = True)
```

Setelah dilakukan penanganan kasus data *outlier* tahapan selanjutnya adalah melakukan standarisasi data. Sebelumnya pada data kerugian negara yang diperoleh akan dilakukan konversi nilai kerugian untuk menyederhanakan nominal angka kerugian yang terkandung dalam atribut Jumlah Kerugian dengan kondisi Rupiah menjadi USD, adapun perintah yang dilakukan sebagai berikut.

```
# konversi nilai Jumlah Kerugian
df['Jumlah Kerugian'] = (df['Jumlah Kerugian']
/ 14500).astype('int64').round()
df.head()
```

Adapun tahapan dalam melakukan standarisasi data bertujuan mengubah data sedemikian rupa dan menjadikan data lebih stabil, untuk melakukan standarisasi pada dataset digunakan *library* pada *Python* yaitu `StandardScaler()`.

TABEL VI.
HASIL STANDARISASI DATA

-1.4361659 3e+00	-1.429720 66e+00	9.7955351 1e-01	-8.22674953 e-01
-1.0033488 0e+00	-1.429720 66e+00	-8.733514 06e-01	-8.40796363 e-01
-5.7053167 0e-01	-1.429720 66e+00	-1.006104 04e+00	-8.49986189 e-01
...
-1.0033488 0e+00	1.6641011 0e+00	-7.671493 00e-01	-8.13932456 e-01
-5.7053167 0e-01	1.6641011 0e+00	-9.530029 85e-01	-8.50068245 e-01
-1.3771454 1e-01	2.2828654 5e+00	-1.006104 04e+00	-8.28826121 e-01

4) Modelling. Pada tahap *modelling* dilakukan penambangan data pada dataset yang telah melalui tahap *data preparation*. Dengan kondisi dataset yang sudah siap digunakan sebelumnya akan dilakukan pembagian data, adapun kondisi pembagian data yang dilakukan adalah 80% untuk data *training* dan 20% untuk data *testing*. Kemudian langkah selanjutnya akan dilakukan pembelajaran dengan metode regresi linear berganda. Sebagai berikut fungsi untuk memperoleh nilai konstanta dan koefisien dalam suatu model:

```
#Mengetahui nilai a (konstanta) dan
b(koefesien)
ml = LinearRegression()
persamaan = ml.fit(X,y)
```

Berdasarkan dari hasil dari pembelajaran regresi linear berganda menghasilkan model regresi sebagai berikut.

$$Y = 284645.5891073216 + (-139837.38007863) x_1 + -20978.0813428 x_2 + (363493.06049751) x_3$$

Model regresi linear berganda tersebut kemudian diterapkan pada data *testing*. Hasilnya seperti diperlihatkan pada Tabel VI.

TABEL VII.
PREDIKSI KERUGIAN NEGARA DENGAN REGRESI LINEAR BERGANDA

No.	Actual	Predicted
56	48548601	45189960.0
31	10344	927719.0
32	46583525	34711596.0
73	4952057	12433629.0
81	615993	3688817.0
34	31445467	37983033.0
66	68401745	37172135.0
58	55404772	48824891.0
38	11883409	15082971.0
5	190344	990654.0
39	37217442	23079818.0
51	14153827	45210939.0
37	37322918	27805228.0
61	31397904	24834349.0

5) *Evaluation* : Pada tahap evaluasi ini, dilakukan proses pengujian kualitas dan efektivitas. Dalam menguji ketepatan dan keakuratan model prediksi kerugian negara berdasarkan kasus tindak pidana korupsi diterapkan proses perhitungan nilai *root mean square error* serta perhitungan untuk mengetahui kecocokan suatu model dengan menggunakan persamaan koefisien determinasi. Berikut hasil nilai RMSE dan koefisien determinasi.

TABEL VIII.
PERFORMA MODEL

Model	RMSE	R ²
Training	8447373.485	0.579
Testing	9769609.026	0.662

Berdasarkan Tabel VII, adapun perolehan performa model regresi linear berganda dalam memprediksi kerugian negara berdasarkan kasus tindak pidana korupsi menghasilkan

tingkat akurasi dengan nilai *root mean square error* sebesar 8447373.485 pada model *training* dan 9769609.026 pada model *testing*. Sedangkan perolehan nilai koefisien determinasi pada model *training* yaitu sebesar 0.579 yang dimana memiliki tingkat hubungan cukup kuat karena kemampuan variabel bebas dalam menjelaskan varians dari variabel terikatnya adalah sebesar 57,9% dan terdapat 42,1% varians variabel terikat yang dijelaskan oleh faktor lain dan untuk model *testing* menghasilkan nilai koefisien 0.662 yang dimana memiliki tingkat hubungan kuat karena kemampuan variabel bebas dalam menjelaskan varians dari variabel terikatnya adalah sebesar 66,2% dan terdapat 33,8% varians variabel terikat yang dijelaskan oleh faktor lain.

6) *Deployment* : *Deployment* merupakan tahapan terakhir dalam pembuatan laporan hasil data mining. Dimana laporan yang berisikan mengenai pengetahuan yang diperoleh atau pengenalan pola pada data dalam proses *data mining*. Dari penelitian ini, dihasilkan suatu pola *data mining* yang terdapat data *training* dan data *testing*. Dalam proses pengujian dalam menguji tingkat keakurasian menggunakan bahasa pemrograman *Python* didapatkan pengetahuan baru bahwa regresi linear dapat melakukan prediksi kerugian negara berdasarkan kasus tindak pidana korupsi.

B. Pembahasan

Penelitian ini membahas terkait penerapan metode regresi linear berganda terhadap pengolahan data kerugian negara berdasarkan kasus tindak pidana korupsi. Metodologi yang dilakukan adalah CRISP – DM (*Cross-Industry Standard Process for Data Mining*). Dataset yang digunakan yaitu tren penindakan kasus korupsi tahun 2013 sampai dengan 2020. Diperoleh pembelajaran regresi linear berganda menghasilkan model regresi yang dimana menghasilkan nilai konstanta yaitu 284645.5891073216 serta nilai koefisien yaitu -139837.38007863 dan 363493.06049751. Kemudian dilakukan pengukuran performa model regresi linear dengan kondisi pembagian data 80%:20%. Dari kondisi pembagian data tersebut diperoleh nilai RMSE sebesar 8447373.485 untuk *data training* dan 9769609.026 untuk *data testing*. Sedangkan untuk nilai koefisien determinasi diperoleh nilai sebesar 0.579 untuk *data training* yang tingkat hubungan antar variabelnya cukup kuat yang dimana kondisi tersebut menjelaskan varians dari variabel terikatnya adalah sebesar 57,9% dan terdapat 42,1% varians variabel terikat yang dijelaskan oleh faktor lain, sedangkan untuk *data testing* memperoleh nilai koefisien determinasi 0.662 yang berarti tingkat hubungan antar variabelnya kuat. Dimana kondisi tersebut menjelaskan bahwa kemampuan variabel bebas dalam menjelaskan varian dari variabel terikatnya adalah sebesar 66,2% dan terdapat 33,8% varians variabel terikat yang dijelaskan oleh faktor lain.

IV. KESIMPULAN

Penerapan metode regresi linear berganda dalam memprediksi kerugian negara berdasarkan kasus tindak pidana korupsi didapatkan hasil pembelajaran regresi linear berganda dengan nilai konstanta yaitu 284645.5891073216 serta nilai koefisien yaitu -139837.38007863 dan 363493.06049751. Kemudian dilakukan pengukuran performa model regresi linear dengan kondisi pembagian data 80%:20% dimana terdiri dari kondisi 52 baris untuk data latih dan 14 baris untuk data uji. Dari kondisi pembagian data tersebut diperoleh nilai RMSE sebesar 8447373.485 untuk *data training* dan 9769609.026 untuk *data testing*. Sedangkan untuk nilai koefisien determinasi diperoleh nilai sebesar 0.579 untuk *data training* yang tingkat hubungan antar variabelnya cukup kuat dan 0.662 untuk *data testing* yang berarti tingkat hubungan antar variabelnya kuat.

Diharapkan dengan adanya penelitian ini model prediksi tersebut dapat menjadi bagian dari kajian kerentanan korupsi serta dapat serta dapat dilakukan pemetaan permasalahan agar Pemerintah dapat mengambil kebijakan yang tepat terhadap permasalahan kasus korupsi serta meminimalisir dan mengantisipasi kerugian negara yang lebih besar untuk tahun selanjutnya berdasarkan variabel-variabel yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] TPBPAK (Tim Penulis Buku Pendidikan Anti Korupsi), Pendidikan Anti Korupsi untuk Perguruan Tinggi, Jakarta: Kementerian Pendidikan dan Kebudayaan RI, 2011.
- [2] ICW, "Tren Penindakan Kasus Korupsi 2020," 2020. [Online]. Available: <https://www.antikorupsi.org/sites/default/files/dokumen/210101-Tren%20Penindakan%20Korupsi%20Tahun%202020.pdf>. [Accessed 2020].
- [3] T. N. Padilah and R. I. Adam, "Analisis Regresi Linier Berganda Dalam Estimasi," *Fibonacci : Jurnal Pendidikan Matematika dan Matematika*, vol. 5 No.2, pp. 117-128, 2019.
- [4] N. Nafi'iyah, "Perbandingan Regresi Linear, Backpropagation Dan," *SENIATI*, no. 2085-4218, pp. B.291-B.296, 2016.
- [5] A. Izzah, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier," *KINETIK*, Vols. 2, No. 3, no. 2503-2259, pp. 141-150, 2017.
- [6] D. A. Desi Novianti, "Pengujian Aplikasi E-Farmer Dalam Perhitungan Keuntungan Dengan Metode Blackbox Testingboundary Value Analysis," *JUNIF*, vol. 1 No. 2, pp. 76-81, 2020.
- [7] S. Susanto and D. Suryadi, *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*, Yogyakarta: Andi Offset, 2010.
- [8] Susiana, "Analisis Peramalan Penjualan Minyak Kelapa Sawit (MKS) pada PT Perkebunan Nusantara IV Unit

- Kebun Pabat," *Karismatika*, Vols. 53, No. 9, p. 1689–1699, 2019.
- [9] A. Pranata, M. Akbar Hsb, T. Akhdansyah, and S. Anwar, "Penerapan metode pemulusan eksponensial ganda dan tripel untuk meramalkan kunjungan wisatawan mancanegara ke Indonesia," *Jurnal Data Anal*, vol. 1 No.1, pp. 32-14, 2018.
- [10] Sujana, *Metode Statistik*, Bandung: Tarsito, 2001.
- [11] Sugiyono, *Metode Penelitian Kombinasi*, Bandung: CV Alfabeta, 2009.
- [12] Maimon, Oded, and Lior, *Data Mining and Knowledge Discovery Handbook*, Springer, New York: Springer, 2010.
- [13] D. T. Larose, C. D. Larose, *Discovering knowledge in data: an introduction to data mining (Vol. 4)*, 2014.