



Prediksi Jumlah Target dan Realisasi Wajib Pajak Atas PBB – P2 Menggunakan Multi Regression, Regression Lasso, dan PCR

Jasman Pardede^{#1}, Maleakhi Ekklesia^{#2}

[#]Fakultas Teknologi Industri, Program Studi Informatika, Institut Teknologi Nasional Bandung
Jl. PH. H. Mustopa No. 23, Bandung

¹jasman@itenas.ac.id

²info2astrid@mhs.itenas.ac.id

Abstrak— Perubahan besaran Pajak Bumi dan Bangunan memberikan dampak bagi beberapa sektor maupun masyarakat di Kota Bandung, karena perubahan yang cukup signifikan dalam besaran Pajak Bumi dan Bangunan ini memberikan pengaruh pada kesadaran dan juga kepedulian masyarakat dalam membayar pajak. Terdapat beberapa penggunaan *machine learning* dalam penentuan pajak ini dimana salah satunya adalah dengan memprediksikan sebuah besaran Target dan Realisasi pada Pajak Bumi Bangunan, sehingga dilakukan sebuah penelitian dengan membandingkan metode *Multi Regression*, *Regression Lasso*, dan *PCR (Principle Component Regression)* untuk menentukan jumlah prediksi terbaik pada Target dan Realisasi penerimaan Pajak Bumi dan Bangunan di Kota Bandung. Hasil pengujian yang telah dilakukan, didapatkan bahwa *Regression Lasso* memiliki nilai keakuratan prediksi terbaik sebesar 79.08%. Hasil tersebut diakibatkan karena *Regression Lasso* dapat mengestimasi parameter, mengatasi permasalahan *multicolinearity* yang terjadi pada data yang digunakan dan menyeleksi variabel yang ada untuk mendapatkan nilai keakuratan prediksi terbaik. Sehingga dapat disimpulkan bahwa terdapat pengaruh positif pada variabel Ketetapan, Target, dan Realisasi yang saling berkaitan. Hal tersebut saling mempengaruhi besarnya, dimana ketetapan mempengaruhi jumlah target, dan jumlah target mempengaruhi jumlah realisasi pada model yang terbentuk, dan proses *pre – processing* membuat model *Regression Lasso* ini menjadi lebih baik.

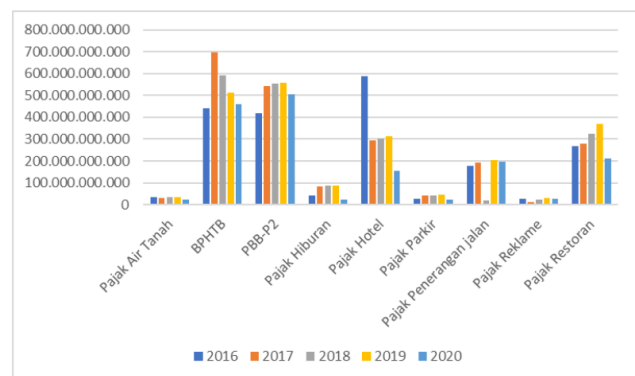
Kata kunci— Multi Regression, PCR, Regression Lasso, Mean Absolute Percentage Error, Pajak Bumi dan Bangunan

I. PENDAHULUAN

Pajak Bumi dan Bangunan (PBB) adalah sumber utama pendapatan pemerintah daerah yang digunakan untuk mendukung pembangunan dan layanan publik di tingkat lokal. Penting bagi pemerintah daerah untuk memprediksi jumlah target PBB secara akurat guna optimalisasi perencanaan keuangan dan pembangunan kota. PBB merupakan jenis pajak daerah diatur oleh pemerintah

daerah, mengalami perubahan setiap 4 tahun. Kota Bandung telah mengelola PBB - P2 sejak 2013 dengan 9 mata pajak dalam dua sistem pemungutan, yakni *self assessment* dan *official tax* [1].

Berdasarkan hasil penelitian terdahulu yang dilakukan oleh [2] didapatkan hasil dari *coefficient determination* pada analisis regresi yang dilakukan menggunakan Regresi Polinomial ini, mendapatkan nilai rata – rata diatas 90%. Perubahan Pajak Bumi dan Bangunan itu sendiri memberikan dampak bagi beberapa sektor maupun masyarakat di Kota Bandung, karena perubahan yang cukup signifikan dalam besaran Pajak Bumi dan Bangunan ini memberikan pengaruh pada kesadaran dan juga kepedulian masyarakat dalam membayar pajak yang dapat dilihat melalui Gambar 1 realisasi pajak daerah pada Kota Bandung tahun 2016- 2020, yang memperlihatkan perubahan yang signifikan pada besaran Pajak Bumi dan Bangunan – P2.



Gambar. 1 Grafik Realisasi Pajak Daerah di Kota Bandung

Jika dilihat dari penelitian pendahuluan yang telah dilakukan oleh [3], penelitian tersebut mengeksplorasi Pajak Bumi dan Bangunan di Kabupaten Bandung Barat menggunakan metode dari jaringan syaraf tiruan, yaitu *Backpropagation* dengan beberapa variabel yang mampu

menentukan jumlah prediksi realisasi penerimaan Pajak Bumi dan Bangunan tersebut adalah Jumlah Perkotaan, Jumlah Pedesaan, Jumlah Wajib Pajak, Luas Wilayah, NJOP, dan NJOTKP. Penelitian tersebut melakukan analisis pada data sebanyak 495 yang diambil pada tahun 2015 – 2017. Penelitian tersebut menghasilkan nilai akurasi 74% dengan *learning rate* 0.2, nilai *epoch* sebesar 150, dan nilai *Mean Square Error* sebesar 0.001.

Maka dari itu, penelitian ini bertujuan untuk menerapkan metode Multi Regresi, Regresi Lasso, dan PCR (*Principle Component Regression*) dalam memprediksi jumlah target dan realisasi PBB – P2 di Kota Bandung pada tahun 2020 – 2022. Diperlukan model regresi karena metode ini dapat mengatasi masalah multikolinearitas dan memberikan kemampuan penyeleksi fitur pada variabel prediktor yang tidak dapat diselesaikan dengan cara konvensional.

Penelitian ini mengeksplorasi perbandingan kinerja tiga algoritma unggulan, yakni Multi Regresi, Regresi Lasso, dan PCR dalam mengungkap faktor – faktor yang memengaruhi perubahan besaran Pajak Bumi dan Bangunan (PBB) di Kota Bandung. Alasan di balik pemilihan ketiga algoritma ini adalah kemampuan Regresi Lasso dalam menangani tantangan multikolinearitas dan seleksi fitur, sedangkan PCR diakui dapat mengatasi permasalahan kolinearitas dan kompleksitas model. Diharapkan, perbandingan ini akan memberikan wawasan yang bermanfaat, menjadi landasan bagi pemerintah daerah dalam merumuskan kebijakan yang lebih efektif untuk meningkatkan penerimaan PBB. Sebagai hasilnya, judul penelitian diusulkan sebagai "Prediksi Jumlah Target dan Realisasi Wajib Pajak Atas PBB – P2 Menggunakan Multi Regression, Regression Lasso, dan PCR".

II. TINJAUAN PUSTAKA

A. Machine Learning

Penelitian ini hakikatnya termasuk pada *machine learning*, seperti yang dijelaskan dalam penelitian sebelumnya [4]. *Machine learning* merupakan suatu teknik untuk menginferensi data dengan pendekatan matematis, di mana metode ini digunakan untuk membentuk model matematis yang merepresentasikan pola-pola data dari hubungan di setiap data. Fokus utama inferensi pada *machine learning* adalah menjelaskan hubungan antar variabel yang direpresentasikan dalam bentuk matematis, dimana pada hal ini Pajak Bumi dan Bangunan menjadi fokus utama dalam mengetahui hubungan variabel yang mempengaruhi besaran Target maupun Realisasi yang terjadi.

B. Multi Regression

Berdasarkan bidang *machine learning* terdapat beberapa metode regresi yang dapat digunakan dalam melakukan prediksi, salah satunya adalah *multi regression*. Menurut penelitian yang telah dilakukan pada [5] *multi regression* merupakan suatu model regresi yang

menggunakan lebih dari satu *variable* bebas atau sebagai predictor, persamaan dasar tersebut dapat dituliskan sebagai berikut,

$$Y = \alpha + \beta X_1 + \beta X_2 + \dots + \beta X_n + e. \quad (1)$$

C. Principle Component Regression (PCR)

Dalam penggunaan *variable* yang cukup banyak pada penelitian ini, dilakukan proses *selection feature* dan diterapkan pada proses regresi melalui metode *principle component regression*. Menurut penelitian [6] metode *principle component regression* menerapkan proses dari *principle component analysis* yang digunakan untuk mengidentifikasi pola data dengan mereduksi dimensi data menjadi dimensi yang lebih kecil dengan mempertahankan informasi yang ada. Terdapat beberapa proses pencarian nilai dari PCA ini sebagai berikut :

1. Menentukan matriks fitur terlebih dahulu.
2. Mencari matriks varians dari kovarian data.
3. Mencari nilai dari *eigen* dan *vector eigen*.
4. Hasil reduksi dimensi ditemukan dengan *vector eigen* yang bersesuaian dengan nilai *eigen* terbesar.

D. Regression Lasso (Least Absolute Shrinkage and Selection Operator)

Selanjutnya, terdapat beberapa metode seleksi fitur yang dapat diimplementasikan melalui pendekatan Lasso. Menurut hasil penelitian yang dirinci dalam karya [7], Lasso adalah metode yang efektif dalam merampingkan koefisien dengan menginterpretasikan angka yang mendekati nol dan melakukan seleksi variabel independen, menghasilkan model dengan variabel terbaik. Estimasi parameter regresi Lasso dilakukan dengan menggunakan persamaan (2).

$$\sum_{i=1}^n (y_i - \sum_{j=1}^k x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^k |\beta_j|. \quad (2)$$

E. Pajak Bumi dan Bangunan – P2

Pajak bumi dan bangunan merupakan suatu biaya pajak yang perlu disetorkan atas keberadaan tanah dan bangunan. Menurut penelitian [8] sifat dari pajak bumi dan bangunan ini adalah kebendaan atau objek, sehingga tarifnya ditentukan dari keadaan objek bangunan yang ada. Dasar hukum yang mengatur pajak bumi dan bangunan ini diatur pada Undang – Undang (UU) No. 12 Tahun 1994, dan No.28 Tahun 2009. Besaran dari PBB – P2 ini sendiri dapat ditentukan dengan PBB – P2 = Tarif PBB – (P2 x Dasar Pengenaan Pajak).

F. Correlation Pearson

Dalam mengatasi adanya fitur yang cukup banyak pada penelitian ini, dilakukan proses dari *selection feature* dengan menghitung *correlation pearson*. Menurut [9] *correlation pearson* merupakan suatu proses dalam mengukur hubungan antara satu atau lebih dari *variable* dalam interval (-1 hingga 0), hubungan dari *correlation pearson* ini menunjukkan nilai hubungan dari masing – masing *variable* pada hubungan yang positif maupun *negative*. Dalam menghitung nilai dari *correlation*

pearson dapat diukur melalui persamaan (3)

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum (x - \bar{x})^2} \sqrt{n \sum (y - \bar{y})^2}} \quad (3)$$

III. METODOLOGI PENELITIAN

A. Objek Penelitian (Dataset)

Dataset yang digunakan dalam proses penelitian ini didapatkan melalui Badan Pendapatan Daerah (Bapenda) Kota Bandung yang telah diolah pada tahun 2020 – 2022. Dataset ini terdiri dari 6 buah fitur yang terdiri dari Wilayah, Tahun, Jumlah SPPT, Ketetapan, Jumlah Target, dan Jumlah Realisasi. Dari seluruh fitur yang terdapat pada dataset ini, 5 fitur tersebut bersifat numerik, dan 1 fitur yaitu Wilayah adalah fitur string, sehingga data tersebut perlu diubah kedalam bentuk *numerical*.

B. Pre – Processing Data

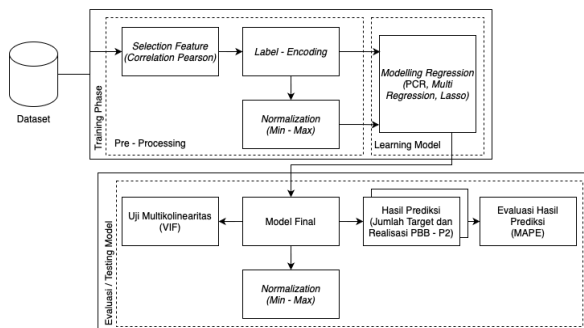
Menurut penelitian yang dilakukan oleh [10] proses dari *pre – processing* data adalah tahapan untuk menghilangkan maupun memperbaiki permasalahan yang ada pada data, dalam pembentukan model *machine learning* dapat dilakukan dengan konsisten dan baik. Metode *pre – processing* yang dilakukan untuk mengolah data yang akan dibentuk kedalam model regresi ini dilakukan dengan beberapa tahapan yaitu :

1. Menentukan *feature* terbaik pada model dengan *selection feature* dalam menentukan kinerja terbaik menggunakan *correlation pearson*.
2. Mengubah data pada *feature* Wilayah menjadi *numerical* dengan proses metode *encoding* dengan metode *label – encoding* (Alphabetical).
3. Karena data – data pada *dataset* memiliki rentang nilai yang berbeda, maka dilakukan proses *normalization* data dengan mengubah seluruh data pada masing – masing fitur menjadi interval 0 hingga 1 [11] dengan *min – max normalization*. *Min – Max Normalization* ini dilakukan dengan persamaan (4).

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min}) \quad (4)$$

C. Tahapan Penelitian

Dalam proses pengolahan tersebut, ditampilkan pada Gambar 2 proses dari *pre – processing* dan penelitian ini.



Gambar 2. Alur penelitian

Tahapan penelitian yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Peneliti mengambil data penerimaan Pajak Bumi dan Bangunan P2 di Kota Bandung yang berasal dari Badan Pendapatan Daerah (BAPENDA).
2. Target yang ditentukan pada pembentukan model data tersebut adalah Jumlah Target, dan Jumlah Realisasi Penerimaan Pajak Bumi dan Bangunan.
3. Setelah itu data yang ada pada dataset yang akan digunakan, dilakukan proses *encoding* pada data *categorical* atau *non – numeric* menjadi *numeric*, lalu melakukan *scaling* data pada interval 0 hingga 1.
4. Peneliti kemudian membandingkan metode *Multi Regression*, *Regression Lasso*, dan *Principle Component Regression* (PCR).
5. Mencari 4 model setiap metode regresi dengan studi kasus sebagai berikut:
 - **Model A** adalah model *Multi Regression* dengan *variable dependent* Jumlah Target pada wilayah Kecamatan.
 - **Model B** adalah model *Multi Regression* dengan *variable dependent* Jumlah Realisasi pada wilayah Kecamatan.
 - **Model C** adalah model *Multi Regression* dengan *variable dependent* Jumlah Target pada wilayah Kelurahan.
 - **Model D** adalah model *Multi Regression* dengan *variable dependent* Jumlah Realisasi pada wilayah Kelurahan.
 - **Model E** adalah model *Regression Lasso* dengan *variable dependent* Jumlah Target pada wilayah Kecamatan.
 - **Model F** adalah model *Regression Lasso* dengan *variable dependent* Jumlah Realisasi pada wilayah Kecamatan.
 - **Model G** adalah model *Regression Lasso* dengan *variable dependent* Jumlah Target pada wilayah Kelurahan.
 - **Model H** adalah model *Regression Lasso* dengan *variable dependent* Jumlah Realisasi pada wilayah Kelurahan.
 - **Model I** adalah model PCR dengan *variable dependent* Jumlah Target pada wilayah Kecamatan.
 - **Model J** adalah model PCR dengan *variable dependent* Jumlah Realisasi pada wilayah Kecamatan.
 - **Model K** adalah model PCR dengan *variable dependent* Jumlah Target pada wilayah Kelurahan.
 - **Model L** adalah model PCR dengan *variable dependent* Jumlah Realisasi pada wilayah Kelurahan.
6. Mengukur hasil akurasi hasil prediksi regresi ini pada setiap model dengan menggunakan *Mean Absolute Percentage Error* dan *R² Coefficient of Determination*.

IV. HASIL DAN PEMBAHASAN

A. Uji Coefficient dan Keakuratan Prediksi

Dari rancangan penelitian yang dilakukan, untuk mengetahui apa saja variable yang memiliki pengaruh pada pembentukan model dilakukan pencarian variable dengan metode *correlation pearson* untuk proses dari *selection feature* yang diperlihatkan pada Tabel I.

TABEL I
Matriks Correlation Pearson

	Tahun	SPPT	Ketetapan	Target	Realisasi
Tahun	1	-0.13	0.012	0.239	-0.03
SPPT	-0.13	1	0.293	0.257	0.202
Ketetapan	0.012	0.293	1	0.95	0.965
Target	0.239	0.257	0.95	1	0.909
Realisasi	-0.03	0.202	0.965	0.909	1

Pada Tabel I hasil pada masing – masing korelasi pada setiap fitur yang digunakan pada dataset ini, dengan diasumsikan bahwa nilai dari *threshold* tersebut adalah ≥ 0.25 . Sehingga hasil dari *selection feature* pada pembentukan model regresi ini menyisakan Wilayah, Jumlah SPPT, Ketetapan, Jumlah Target, dan Realisasi. Namun dilakukan beberapa pengujian dengan 4 buah studi kasus yang dilakukan untuk mengetahui seberapa besar hubungan fitur yang digunakan pada masing – masing model regresi yang akan dibentuk, dilakukan pengujian pembentukan model, dimana terdapat beberapa *feature* yang akan digunakan pada masing – masing studi kasus tersebut dengan studi kasus sebagai berikut :

1. **Studi Kasus 1** – Menggunakan seluruh fitur yang ada pada dataset.
2. **Studi Kasus 2** – Menjadikan Jumlah Target sebagai *variable independent* pada model prediksi Jumlah Realisasi dengan *variable dependent* yang digunakan yaitu Wilayah, Jumlah SPPT, dan Ketetapan.
3. **Studi Kasus 3** – Menggunakan fitur dari hasil *selection feature* tanpa menggunakan Wilayah, dengan fitur yang digunakan yaitu, Jumlah SPPT, dan Ketetapan.
4. **Studi Kasus 4** – Menggunakan fitur dari hasil *selection feature* yaitu, Jumlah SPPT, Ketetapan, dan Wilayah.

Telah diperlihatkan pada Tabel II hingga Tabel V *variable – variable* yang digunakan pada masing – masing studi kasus pengujian ini.

TABEL III
VARIABEL STUDI KASUS 1

STUDI KASUS 1				
X ₁	X ₂	X ₃	X ₄	Y
Wilayah	Tahun	SPPT	Ketetapan	Realisasi
X ₁	X ₂	X ₃	X ₄	Y
Wilayah	Tahun	SPPT	Ketetapan	Target

TABEL IIIII
VARIABEL STUDI KASUS 2

STUDI KASUS 2				
X ₁	X ₂	X ₃	X ₄	Y
Wilayah	Target	SPPT	Ketetapan	Realisasi
X ₁	X ₂	X ₃	Y	
Wilayah	SPPT	Ketetapan	Target	

TABEL IVV
VARIABEL STUDI KASUS 3

STUDI KASUS 3		
X ₁	X ₂	Y
SPPT	Ketetapan	Realisasi
X ₁	X ₂	Y
SPPT	Ketetapan	Target

TABEL V
VARIABEL STUDI KASUS 4

STUDI KASUS 4			
X ₁	X ₂	X ₃	Y
Wilayah	SPPT	Ketetapan	Realisasi
X ₁	X ₂	X ₃	Y
Wilayah	SPPT	Ketetapan	Target

Dilakukan pengujian dalam menghitung nilai *coefficient determination*, dan *Mean Absolute Percentage Error*. Pada penelitian [12] *R Square* atau *Coefficient Determination* ini merupakan suatu teknik maupun analisis mengenai koefisien pada pembentukan model yang menjelaskan seberapa jauh data *dependent* dapat dijelaskan oleh data *independent*. Lalu menurut penelitian yang dilakukan oleh [13] *Mean Absolute Percentage Error* merupakan hasil dari nilai rata – rata perbandingan antara nilai prediksi dan nilai realisasi (*real*).

Sehingga hasil evaluasi *Mean Absolute Percentage Error* ini dapat mengevaluasi seberapa akurat pemodelan yang dibentuk. Lalu berikut diuji pada masing – masing studi kasus yang telah dilakukan sebagai berikut :

TABEL VI
UJI COEFFICIENT DETERMINATION DENGAN PRE - PROCESSING

MODEL	STUDI KASUS 1	STUDI KASUS 2	STUDI KASUS 3	STUDI KASUS 4
MODEL A	0.744	0.767	0.767	0.767
MODEL B	0.547	0.629	0.579	0.5653
MODEL C	0.955	0.904	0.903	0.904
MODEL D	0.857	0.87	0.861	0.8605
MODEL E	0.744	0.767	0.768	0.7676
MODEL F	0.547	0.629	0.579	0.5653
MODEL G	0.88	0.885	0.886	0.8859
MODEL H	0.857	0.87	0.861	0.8605
MODEL I	0.744	0.767	0.768	0.767
MODEL J	0.547	0.629	0.579	0.5653
MODEL K	0.88	0.885	0.886	0.8859
MODEL L	0.857	0.87	0.579	0.8605
MEAN	0.763	0.789	0.75	0.771

TABEL VII
UJI MEAN ABSOLUTE PERCENTAGE ERROR DENGAN PRE - PROCESSING

MODEL	STUDI KASUS 1	STUDI KASUS 2	STUDI KASUS 3	STUDI KASUS 4
MODEL A	26.8	25.65	25.63	25.65
MODEL B	21.86	19.37	21.13	21.47
MODEL C	46.2	16.83	16.84	16.83
MODEL D	14.78	14.08	14.53	14.56
MODEL E	26.8	25.65	25.63	25.65
MODEL F	21.86	19.32	21.13	21.472
MODEL G	22.88	22.11	22.09	22.11
MODEL H	14.78	14.08	14.53	14.56
MODEL I	26.8	25.65	25.63	25.65
MODEL J	22.8	19.32	21.13	21.47
MODEL K	22.88	22.11	22.09	22.11
MODEL L	20.86	14.08	14.53	14.56
MEAN	24.1	19.8	20.4	20.5

Hasil penelitian yang dilakukan pada Tabel VI hingga Tabel VII telah didapatkan hasil nilai keakuratan prediksi melalui perhitungan *Mean Absoulte Percentage Error* dan *Coefficient Determination* dengan hasil sebagai berikut :

- Hasil rata – rata nilai akurasi pada pemodelan Jumlah Target pada Wilayah Kecamatan dengan menggunakan *Multi Regression* adalah 74,06% sedangkan dengan menggunakan *Regression Lasso* adalah 74,07% dan bila menggunakan PCR adalah 74,0675%.
- Hasil rata – rata nilai akurasi pada pemodelan Jumlah Realisasi pada Wilayah Kecamatan dengan menggunakan *Multi Regression* adalah 79,042% sedangkan dengan menggunakan *Regression Lasso* adalah 79,0545% dan bila menggunakan PCR adalah 78,82%.
- Hasil rata – rata nilai akurasi pada pemodelan Jumlah Target pada Wilayah Kelurahan dengan menggunakan *Multi Regression* adalah 75,825% sedangkan dengan menggunakan *Regression Lasso* adalah 77,7025% dan bila menggunakan PCR adalah 77,7%.
- Hasil rata – rata nilai akurasi pada pemodelan Jumlah Realisasi pada Wilayah Kelurahan dengan menggunakan *Multi Regression* adalah 85,5125% sedangkan dengan menggunakan *Regression Lasso* adalah 85,52% dan bila menggunakan PCR adalah 83,9925%.

Sehingga jika dilihat dari hasil pengujian tersebut dapat dikatakan bahwa *Regression Lasso* memiliki nilai rata – rata keakuratan prediksi lebih tinggi dibandingkan dengan *Multi Regression* dan PCR. Hal tersebut diakibatkan karena *Regression Lasso* menyeleksi nilai peubah dan pendugaan koefisien regresi dengan melakukan pendekatan galat yang mendekati nol atau hingga nol.

Oleh karena itu nilai dari keakuratan prediksi pada model *Regression Lasso* ini memiliki model yang lebih baik. Lalu dari pengujian yang telah dilakukan pada Tabel VI hingga VII, didapatkan bahwa *feature* terbaik adalah pada Studi Kasus 2 dengan nilai rata – rata R^2 terbesar yaitu 0.789. Untuk melihat pengaruh *pre – processing* pada pembentukan model regresi ini, dilakukan uji hasil analisis yang sama tanpa *pre – processing* yang dilakukan pada Tabel VIII hingga IX :

TABEL VIII
UJI COEFFICIENT DETERMINATION TANPA PRE - PROCESSING

MODEL	STUDI KASUS 1	STUDI KASUS 2	STUDI KASUS 3	STUDI KASUS 4
MODEL A	0.968	0.863	0.863	0.8632
MODEL B	0.956	0.95	0.947	0.9482
MODEL C	0.955	0.904	0.903	0.9042
MODEL D	0.943	0.941	0.94	0.9406
MODEL E	0.968	0.863	0.863	0.8632
MODEL F	0.956	0.951	0.947	0.9482
MODEL G	0.955	0.904	0.903	0.9042
MODEL H	0.943	0.941	0.9404	0.9406
MODEL I	0.968	0.863	0.863	0.863
MODEL J	0.956	0.951	0.947	0.9482
MODEL K	0.955	0.904	0.903	0.9042
MODEL L	0.943	0.941	0.94	0.9406
MEAN	0.955	0.914	0.913	0.914

TABEL IX
UJI MEAN ABSOLUTE PERCENTAGE ERROR TANPA PRE - PROCESSING

MODEL	STUDI KASUS 1	STUDI KASUS 2	STUDI KASUS 3	STUDI KASUS 4
MODEL A	76.01	23.11	23.16	23.11
MODEL B	14.55	14.44	14.89	15.11
MODEL C	63.37	20.62	20.7	20.62
MODEL D	14.42	14.78	14.74	14.71
MODEL E	76.01	23.11	23.16	23.11
MODEL F	14.55	14.44	14.89	15.11
MODEL G	63.37	20.62	20.7	20.62
MODEL H	14.42	14.78	14.74	14.714
MODEL I	76.01	23.11	23.16	23.11
MODEL J	14.55	14.44	14.89	15.11
MODEL K	63.37	20.62	20.709	20.62
MODEL L	29.53	25.46	14.74	14.714
MEAN	43.34	19.12	18.37	18.38

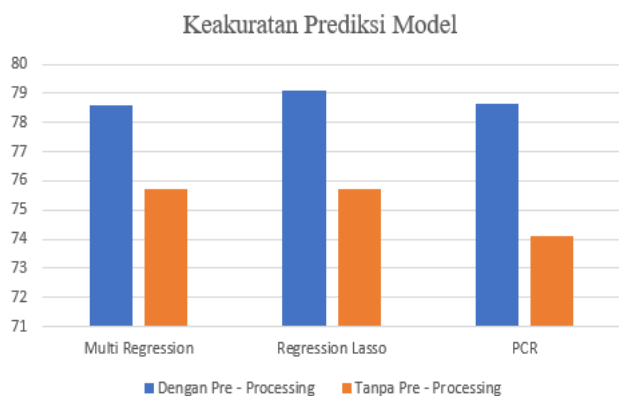
Hasil penelitian yang dilakukan pada Tabel VIII hingga Tabel IX telah didapatkan hasil nilai keakuratan prediksi melalui perhitungan *Mean Absoulte Percentage Error* dan *Coefficient Determination* dengan hasil sebagai berikut :

- Hasil rata – rata nilai akurasi pada pemodelan Jumlah Target pada Wilayah Kecamatan dengan menggunakan *Multi Regression* adalah 63,65%

sedangkan dengan menggunakan *Regression Lasso* adalah 63,66% dan bila menggunakan PCR adalah 63,6525 %.

2. Hasil rata – rata nilai akurasi pada pemodelan Jumlah Realisasi pada Wilayah Kecamatan dengan menggunakan *Multi Regression* adalah 85,252% sedangkan dengan menggunakan *Regression Lasso* adalah 85,252% dan bila menggunakan PCR adalah 85,2 %.
3. Hasil rata – rata nilai akurasi pada pemodelan Jumlah Target pada Wilayah Kelurahan dengan menggunakan *Multi Regression* adalah 68,672 % sedangkan dengan menggunakan *Regression Lasso* adalah 63,6% dan bila menggunakan PCR adalah 68,67025%.
4. Hasil rata – rata nilai akurasi pada pemodelan Jumlah Realisasi pada Wilayah Kelurahan dengan menggunakan *Multi Regression* adalah 85,3375% sedangkan dengan menggunakan *Regression Lasso* adalah 85,3% dan bila menggunakan PCR adalah 78,889%.

Lalu pada pengujian *pre – processing* ini didapatkan bahwa *feature* terbaik ada pada Studi Kasus 3 dengan nilai rata – rata keakuratan prediksi sebesar 81,626%. Sehingga berdasarkan pengujian tersebut didapatkan bahwa hasil keakuratan prediksi pada masing – masing model ditampilkan nilai dari rata – rata keakuratan prediksi setiap model yang ditampilkan pada Gambar 3.



Gambar 3. Grafik keakuratan prediksi model

Gambar 3 memperlihatkan bahwa hasil dari model dengan *pre – processing* memiliki nilai keakuratan yang lebih tinggi dibandingkan dengan tanpa *pre – processing* pada pembentukan model regresi yang dilakukan.

B. Uji Multikolinearitas

Untuk menguji keberadaan korelasi yang tinggi antara dua atau lebih *variable independent* yang digunakan dalam pembuatan model ini, dilakukan pengujian multikolinearitas. Multikolinearitas [14] ini terjadi ketika terdapat hubungan linear yang kuat antara *variable –*

variable independent dalam model yang dibentuk, multikolenaritas ini dapat diketahui melalui persamaan (5).

$$(VIF) = 1 / (1 - R^2) \tag{5}$$

Melalui persamaan (5) tersebut dilakukan pengujian dalam menguji multikolinearitas yang terjadi pada masing – masing studi kasus yang dilakukan pada Tabel II hingga V. Sehingga berikut merupakan hasil pengujian yang dilakukan :

TABEL X
HASIL UJI MULTIKOLINEARITAS

VARIABLE	KECAMATAN	KELURAHAN
Wilayah	1.061	1.006
Tahun	20.34	8.69
Jumlah SPPT	1.47	1.24
Ketetapan	33.26	31.36
Realisasi	20.4	17.17
Target	7.73	10.64

Diperlihatkan pada Tabel X hasil perhitungan *variance infcaltion factor* masing – masing *variable predictor* pada model. Menurut penelitian [15] jika nilai VIF > 10, hal tersebut diasumsikan bahwa terdapat korelasi yang tinggi antara *variable predictor* yang diberikan dan *variable predictor* lainnya pada model yang dibentuk, sehingga hal ini dapat dikatakan bahwa terdapat permasalahan multikolinearitas pada model yang dibentuk.

V. KESIMPULAN

Berdasarkan hasil uji korelasi menggunakan *correlation pearson* didapatkan korelasi antara Ketetapan terhadap Target sebesar 0.95, Ketetapan terhadap Realisasi sebesar 0.96, dan Target terhadap Realisasi 0.909. Hasil ini menunjukkan bahwa besaran Ketetapan akan menentukan jumlah dari Target maupun Realisasi yang terpenuhi dari jumlah penerimaan Pajak Bumi dan Bangunan – P2 pada Kecamatan maupun Kelurahan yang ada di Kota Bandung. Semakin tinggi besaran Ketetapan yang ditentukan, maka dapat diprediksikan bahwa besaran Target maupun Realisasi juga bertambah.

Untuk menentukan *variable* apa saja yang mempengaruhi jumlah Target maupun Realisasi pada penerimaan Pajak Bumi dan Bangunan – P2 di Kota Bandung ini, didapatkan bahwa *variable* yang memiliki nilai kecocokan model terbaik adalah pada Studi Kasus 2 yaitu dengan nilai rata – rata *coefficient determination* sebesar 80,14% dengan proses *pre – processing*, dan Studi Kasus 3 dengan nilai rata – rata *coefficient determination* sebesar 81,626% tanpa proses *pre – processing*. Sehingga berdasarkan penelitian ini, didapatkan bahwa *pre – processing* tidak menentukan sebuah model tersebut menjadi lebih baik, *pre – processing* perlu memperhatikan karakteristik data yang digunakan dan model apa yang akan diterapkan pada data itu sendiri.

Dari beberapa pengujian yang dilakukan, maka dapat disimpulkan bahwa dari ketiga metode regresi tersebut, *Regression Lasso* dengan proses *pre – processing* dengan

nilai keakuratan prediksi sebesar 79,08 memiliki kinerja model yang cukup baik dibandingkan dengan *Multi Regression* dan PCR. Hal ini karena adanya permasalahan multikolinearitas yang terjadi pada data yang digunakan, dan hal tersebut mampu diatasi oleh *Regression Lasso* dengan menyeleksi koefisien *variable* yang tidak penting dan akan memilih *variable* yang saling berkorelasi tinggi dengan proses penyusutan koefisien melalui galat hingga menyusut nol pada masing – masing koefisien model regresi yang terbentuk.

REFERENSI

- [1] B. Setiawan, M. R. R and A. Delis, “Analisis penerimaan pajak bumi dan bangunan-P2 serta pengaruhnya terhadap pendapatan daerah Kabupaten / Kota di Provinsi Jambi,” *Jurnal Paradigma Ekonomika*, vol. XV, no. 2, pp. 184 - 186, 2020.
- [2] M. Firdaus, M. Hafiyusholeh and S. Widodo, “Prediksi Nilai Jual Objek Pajak (NJOP) Tanah di Kabupaten Gresik Menggunakan Regresi Polinomial,” *Jurnal Matematika*, vol. I, no. 1, pp. 82 - 99, 2020.
- [3] R. U. Kuswana, G. Abdillah and A. Komarudin, “Prediksi Realisasi Penerimaan Pajak Bumi dan Bangunan di Pemerintah Kabupaten Bandung Barat Menggunakan Jaringan Syaraf Tiruan Backpropagation,” *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, Vols. -, no. -, pp. 580 - 585, 2019.
- [4] I. M. Faiza, Gunawan and W. Adriani, “Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir,” *Jurnal Minfo Polgan*, vol. XI, no. 2, pp. 59 - 60, 2022.
- [5] R. A. Sahulata, “Analisa Multi Regresi pada Pengukuran Kepercayaan, Keamanan dan Kualitas Layanan Pembelian Online,” *Seminar Nasional Sistem Informasi dan Teknik Informatika*, Vols. -, no. -, p. 675, 2019.
- [6] Kurniawati and B. Rikumahu, “egresi Principal Component Analysis Untuk Menentukan Faktor-Faktor Yang Berpengaruh Terhadap Return Saham (Studi pada Sub Sektor Telekomunikasi yang Terdaftar di Bursa Efek Indonesia Periode 2007-2016),” *Festival Riset Ilmiah Manajemen dan Akutansi (FRIMA)*, Vols. -, no. 1, pp. 560 - 577, 2018.
- [7] M. Robbani, F. Agustiani and N. Herrhyanto, “Regresi Least Absolute Shrinkage and Selection Operator (Lasso) pada Kasus Inflasi Di Indonesia Tahun 2014-2017,” *Jurnal EurekaMatika*, vol. VII, no. 2, pp. 5 - 6, 2019.
- [8] A. Rahman, “Pengaruh Kesadaran Wajib Pajak, Tingkat Pendidikan, dan Pendapatan Terhadap Kepatuhan Membayar Pajak Bumi dan Bangunan,” *Jurnal Akutansi*, vol. VI, no. 1, pp. 2 - 5, 2018.
- [9] F. Jabnabillah and N. Margina, “Analisis Korelasi Pearson Dalam Menentukan Hubungan Antara Motivasi Belajar Dengan Kemandirian Belajar Pada Pembelajaran Daring,” *Jurnal Sintak*, vol. I, no. 1, pp. 15 - 16, 2022.
- [10] Zelaya and C. V. Gonzales, “Zelaya, Carlos Vladimiro González. "Towards explaining the effects of data preprocessing on machine learning," IEEE 35th international conference on data engineering (ICDE), 2019.
- [11] D. A. Nasution, H. H. Khotimah and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *Journal of Computer Engineering, System and Science*, vol. IV, no. 1, p. 79, 2019.
- [12] I. Nabillah and I. Ranggadara, “Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut,” *Journal of Information System*, vol. V, no. 2, p. 252, 2020.
- [13] D. A. S. F, L. Rochmawati and I. Sonhaji, “Koefisien Korelasi (R) Dan Koefisien Determinasi (R2),” *Jurnal Penelitian Politeknik Penerbangan Surabaya*, vol. V, no. 4, pp. 289 - 296, 2020.
- [14] M. Sriningsih, D. Hatidja and J. D.Prang, “Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras Di Provinsi Sulut” *Jurnal Ilmiah Sains*, vol. XVIII, no. 2, p. 19, 2018.
- [15] W. R. Anggraeni, N. N. Debatara and S. W. Rizki, “Estimasi Parameter Regresi Ridge untuk Mengatasi Multikolinearitas,” *Bimaster*, vol. VII, no. 4, pp. 296 - 297, 2018.