

ANALISIS PREDIKSI DENGAN *MACHINE LEARNING* DAN DETERMINAN INVESTASI REGIONAL TINGKAT PROVINSI DI INDONESIA

Muhamad Ameer Noor, Rivai Geraldin Batubara, Kartiko Cokro Sewoyo, Bana Ali Fikri,
Reggina Ardalova
Direktorat Jenderal Perbendaharaan, Kementerian Keuangan

Abstract

In line with the pro-investment policy of the Indonesian government, investment is one of the keys in determining economic growth. This paper aims to predict investment and identify investment determinants, to provide input for the formulation of effective pro-investment policies. The study uses two models, namely machine learning and OLS regression, using data of 34 provinces in Indonesia during 2016-2020. Prediction power of machine learning is utilized to predict investment using panel data, while OLS regression with cross-section data is chosen to identify investment determinants. Results of Extra Trees Regressor model can predict Investment with an R^2 of 0.8428. The model also finds that GRDP, Government Expenditure, Distance to Economic Center, Population, Infrastructure, Health, and Natural Resources have high feature importance values in predicting investment. The results of the cross-section regression model confirmed that Capital Expenditure, Port Quality, Population and Oil Resources have a significant effect on investment in Indonesia

Abstrak

Sejalan dengan kebijakan pro-investasi pemerintah Indonesia, Investasi merupakan salah satu kunci utama dalam menentukan pertumbuhan ekonomi. Tulisan ini bertujuan untuk memprediksi investasi dan mengidentifikasi faktor-faktor yang mempengaruhi investasi, dalam rangka memberikan masukan untuk formulasi kebijakan pro-investasi yang lebih efektif. Penelitian menggunakan dua model, yakni *machine learning* dan regresi OLS, dengan menggunakan data 34 provinsi di Indonesia selama tahun 2016-2020. Potensi kapabilitas prediksi *machine learning* dimanfaatkan untuk memprediksi nilai Investasi dengan data panel, sedangkan regresi OLS dengan data *cross-section* digunakan untuk mengidentifikasi determinan investasi. Hasil model *Extra Trees Regressor* dapat memprediksi Investasi dengan R^2 sebesar 0,8428. Model tersebut juga menemukan variabel PDRB, Belanja Pemerintah, Jarak ke Pusat Ekonomi, Kependudukan, Infrastruktur, Kesehatan, dan Sumber Daya Alam memiliki nilai *feature importance* yang tinggi dalam memprediksi Investasi. Hasil model regresi *cross-section* mengkonfirmasi pengaruh signifikan dari Belanja Modal, Kualitas Pelabuhan, Jumlah Penduduk dan SDA Minyak terhadap investasi di Indonesia.

Keywords: determinan investasi, *extra trees regressor*, investasi regional, kebijakan investasi, *machine learning* dalam ekonomi, prediksi investasi

JEL Classification: E22, C53, H50

PENDAHULUAN

Investasi memiliki peranan yang penting dalam perekonomian suatu negara, sebagaimana diteliti dalam berbagai literatur studi seperti Anderson, D. (1990), Blomström et al. (1996), Suhendra & Anwar (2014), Chow, G. C. (2017), Gupta, K. (2021), dan Shabbir, et al. (2021).

Pentingnya investasi bagi pertumbuhan ekonomi disadari oleh Pemerintah Indonesia. Hal tersebut terbukti dengan berbagai kebijakan yang didorong untuk mempermudah investasi oleh Pemerintah Indonesia. Dari sisi kemudahan berbisnis (*Ease of Doing Business* / EoDB), Indonesia terus berbenah sebagaimana tercermin dalam peningkatan peringkat EoDB dari 120 di tahun 2014 menjadi 73 di tahun 2019 (World Bank, 2022). Sejalan dengan peningkatan kemudahan berbisnis tersebut, Pemerintah Indonesia juga mengimplementasikan sistem perizinan terintegrasi atau *online single submission* (OSS). Dari sisi perpajakan, insentif berupa *tax holiday*, *tax allowance*, serta penurunan tarif PPH Badan dari 25% menjadi 20% di tahun 2022.

Pada Tahun 2021, Direktorat Jenderal Perbendaharaan (DJPb), Kementerian Keuangan, mengemban amanat baru untuk mewujudkan Kementerian Keuangan sebagai *Regional Chief Economist* (RCE) atau penasihat ekonomi utama dalam hubungan yang sinergis dengan mitra-mitra pembangunan di daerah.

Dikaitkan dengan kebijakan umum pro-investasi pemerintah pusat, Peran DJPb sebagai RCE di daerah dapat diperluas dengan berfungsi sebagai ekonom yang dapat menangkap potensi-potensi

suatu daerah dan mendorong penanam modal untuk memanfaatkan potensi tersebut.

Untuk dapat mewujudkan fungsi tersebut, diperlukan analisis dua sisi yakni 1) analisis prediksi investasi dengan *machine learning* yang dapat menjadi dasar pemerintah dalam mengambil kebijakan tertentu dengan mempertimbangkan estimasi dampaknya terhadap penambahan investasi di suatu provinsi, dan 2) analisis determinan investasi regional yang dapat menangkap faktor-faktor yang paling krusial dalam mempengaruhi masuknya investasi pada suatu provinsi. Periode observasi yang digunakan dalam penelitian ini adalah menggunakan data tahunan selama 5 tahun dari tahun 2016-2020. Adapun data yang digunakan merupakan data regional dari 34 Provinsi.

Dalam konteks analisis prediksi investasi belum terdapat banyak penelitian terdahulu yang telah dilakukan, Adapun studi terdahulu terkait prediksi investasi di Indonesia oleh Budiono & Purba (2019) belum memanfaatkan model *machine learning*. Untuk keperluan prediksi, model *machine learning* berpotensi menawarkan kemampuan prediksi yang lebih tinggi dibandingkan model regresi pada umumnya, dengan *trade-off* berupa penurunan tingkat interpretasi dari model (Athey, S., 2018). Prediksi dengan *machine learning* memiliki banyak jenis model yang dapat digunakan. Untuk memilih model yang dapat menghasilkan akurasi prediksi terbaik, tulisan ini menggunakan Bahasa Pemrograman Python dan *Library* PyCaret untuk membandingkan akurasi dari 18 model *machine learning*. *Library* PyCaret dapat mengevaluasi dan menemukan model terbaik dengan

proses yang efisien (Mulpuru & Mishra, 2021).

Dalam konteks analisis determinan investasi, studi-studi terdahulu di Indonesia seperti Suhendra & Anwar (2014), Soekro & Widodo (2015), dan Fathia, et al. (2021) masih menggunakan variabel nasional dalam menentukan determinan investasi di Indonesia, sehingga belum menangkap perbedaan karakteristik 34 Provinsi di Indonesia yang masing-masing memiliki keunikan tersendiri. Disisi lain, studi determinan investasi regional di Rusia oleh Ledyeva, S. (2009), menawarkan model yang menarik untuk direplikasi, dengan menghasilkan determinan unik yang dapat mempengaruhi keputusan investor untuk lebih memilih berinvestasi di Provinsi tertentu.

Dengan demikian, tulisan ini bertujuan untuk 1) Membandingkan dan memilih model *machine learning* dengan tingkat akurasi terbaik untuk memprediksi investasi regional Indonesia di tingkat Provinsi, 2) Memprediksi nilai investasi regional Indonesia di tingkat Provinsi, dan 3) mengidentifikasi determinan investasi regional Indonesia di tingkat Provinsi. Penelitian menggunakan pendekatan kuantitatif dengan model *machine learning* untuk prediksi investasi dan dengan model *ordinary least square* untuk identifikasi determinan investasi.

Mempertimbangkan potensi dampak positif Investasi dalam berbagai literatur, serta kebijakan pro-investasi pemerintah Indonesia, penelitian ini dapat bermanfaat dalam 1) menyediakan model *machine learning* untuk memprediksi nilai investasi pada setiap Provinsi di Indonesia 2) menyediakan *prototype* model simulasi bagi pengambil kebijakan untuk

memperkirakan dampak perubahan kebijakan terkait, terhadap perubahan nilai investasi pada suatu Provinsi, dan 3) menjadi dasar dalam diseminasi informasi yang lebih efektif pada calon investor, dengan menyajikan informasi-informasi yang berpotensi menjadi penentu keputusan investor berdasarkan signifikansi variabel independen dalam model regresi dan/atau besaran pengaruh variabel independen dalam model prediksi *machine learning*.

TINJAUAN LITERATUR

Literatur Tentang Dampak Investasi

Studi oleh Anderson, D. (1990) menunjukkan bahwa alokasi investasi yang efisien dalam meningkatkan output ekonomi memiliki dampak yang besar terhadap pertumbuhan ekonomi, sedangkan alokasi yang inefisien dapat menciptakan stagnasi atau penurunan pertumbuhan ekonomi. Studi oleh Blomström et al. (1996) juga mengemukakan hal serupa mengenai pentingnya efisiensi alokasi investasi. Sejalan dengan itu, Chow G. C. (2017) menemukan bahwa investasi berdampak positif terhadap ekonomi, meskipun di saat yang bersamaan ditemukan adanya potensi ekonomi yang hilang akibat alokasi investasi yang terlalu tersentralisasi pada masa awal pembangunan Tiongkok.

Dalam konteks Penanaman Dalam Negeri (PMDN) atau *Domestic Direct Investment* (DDI) dan Penanaman Modal Asing (PMA) atau *Foreign Direct Investment* (FDI), Shabbir, et al. (2021) menemukan bahwa keduanya berpengaruh positif terhadap pertumbuhan ekonomi di Pakistan. Sejalan dengan itu, studi oleh Gupta, K.

(2021) di Indonesia menunjukkan bahwa peningkatan modal atau investasi secara umum dapat mendorong pertumbuhan ekonomi.

Selain itu, literatur yang ditulis oleh Moran, et al. (2005), Heijman & Ophem (2007), dan Noor & Saputra (2020) menekankan bahwa investasi dengan perencanaan sektoral atau *transfer knowledge* yang baik juga berpotensi untuk mengurangi eksternalitas negatif ekonomi di suatu negara.

Secara umum, berbagai literatur tersebut menekankan dampak positif atau potensi dampak positif investasi di suatu negara. Mengingat kebijakan pro-investasi pemerintah Indonesia, analisis mengenai prediksi dan determinan nilai investasi dapat memberikan nilai tambah, sebagai bahan pertimbangan dalam memformulasikan kebijakan yang lebih efektif untuk mendorong kenaikan investasi.

Literatur Prediksi Investasi dan Prediksi dengan Machine Learning

Di Indonesia, literatur studi mengenai prediksi investasi belum begitu banyak. Studi terdahulu mengenai prediksi investasi di Indonesia oleh Budiono & Purba (2019) menggunakan regresi data panel dengan *random effect model*. Penelitian dilakukan pada 34 Provinsi di Indonesia dalam periode tahun 2010-2017. Model prediksi investasi disusun dengan menggunakan akses listrik dan kebijakan pemerintah pro-investasi sebagai prediktor. Keduanya berpengaruh signifikan terhadap nilai investasi di Indonesia.

Athey, S. (2018) membahas mengenai dampak model *machine learning* pada bidang keilmuan ekonomi. Menurut Athey, *machine learning* memiliki potensi tinggi untuk dimanfaatkan

dalam *modelling* sektor ekonomi. Hal ini didasari oleh potensi kekuatan prediksi yang lebih akurat dibandingkan model regresi umum. Namun demikian, kompleksitas *modelling* dalam *machine learning* juga menciptakan adanya *trade-off* dalam penggunaannya, yakni penurunan tingkat interpretasi hasil *modelling*.

Bahasa pemrograman Python yang bersifat *open source* memiliki pengembangan *tools machine learning* yang sangat beragam. Pada satu sisi, hal tersebut memperkaya pilihan model yang dapat memprediksi secara lebih akurat dengan *resource* yang tersedia dalam setiap studi kasus. Di sisi lain, hal tersebut juga memiliki dampak negatif, dimana proses membandingkan puluhan model terbaik dapat memakan waktu yang cukup panjang. Mulpuru & Mishra, (2021), dalam penelitiannya memanfaatkan *library* PyCaret dalam Python. *Library* tersebut dapat membandingkan akurasi dari 18 model *machine learning* secara serentak, sehingga proses pemilihan model terbaik menjadi lebih efisien. Hal tersebut mendasari pemanfaatan *library* PyCaret dalam proses pemilihan model prediksi dalam model ini.

Literatur Mengenai Determinan Investasi Nasional

Ledyaeva, S. (2009) melakukan pengujian terhadap determinan dan hubungan spasial terhadap aliran *Foreign Direct Investment (FDI)* di wilayah Rusia selama masa transisi tahun 1995-2005. Penelitian menggunakan model OLS dan SAR terhadap data cross-section dan panel. Ditemukan bahwa faktor penting dari aliran FDI ke wilayah di Rusia terkait dengan keberadaan kota besar dan pelabuhan,

sumber daya minyak dan gas, jarak ke pasar Eropa dan risiko politik dan legislatif. Penelitian menyimpulkan bahwa pemerintah harus berupaya menjaga agar tingkat pertumbuhan dan urbanisasi di setiap wilayah tetap terjaga sehingga menarik investasi. Peningkatan kualitas transportasi juga turut meningkatkan aliran investasi ke daerah tersebut dan sekitarnya. Hal ini akan meningkatkan akses ke pasar di Eropa dari wilayah-wilayah Rusia. Selanjutnya keberadaan sumber daya alam di suatu wilayah menjadi pendorong dan nilai tambah dalam menarik investasi ke wilayah yang dipromosikan. Penelitian ini memberikan sudut pandang menarik terkait investasi swasta di daerah. Namun demikian, penelitian ini masih terbatas pada skala nasional di Rusia sehingga diperlukan penelitian lebih lanjut untuk melakukan penelitian serupa di negara lain sehingga dapat memberikan gambaran lebih lengkap terkait determinan investasi swasta di suatu wilayah.

Hermes Niels dan Lensink Robert (2001) menguji pengaruh ekonomi makro dari kebijakan fiskal pada sampel tiga puluh negara kurang berkembang. Penelitian ini menguji keberadaan hubungan antara variabel kebijakan fiskal dan investasi. Secara eksplisit penelitian dipusatkan pada berbagai aspek dari kebijakan fiskal dan pengaruhnya terhadap investasi. Hasil penelitian menemukan bahwa belanja pemerintah yang berbeda memiliki efek berbeda terhadap investasi. Selanjutnya ditemukan bahwa efek dari belanja modal memiliki pengaruh positif terhadap investasi swasta hanya jika sudah melewati ambang batas tertentu. Belanja untuk gaji dan kesehatan akan mendorong investasi sampai tingkat

tertentu. Jika belanja dari kedua kategori tersebut melampaui ambang batas, maka tingkat investasi akan turun.

Implikasi dari penelitian ini yaitu bahwa pemerintah pada negara kurang berkembang harus waspada terhadap fakta bahwa kebijakan pendapatan dan belanja harus dilaksanakan secara hati-hati. Belanja modal yang terlalu sedikit di bawah ambang batas membutuhkan pengeluaran tambahan sehingga mendorong pengaruh positif dari belanja modal pemerintah terhadap investasi. Pemerintah di negara kurang berkembang harus memahami tingkat ambang batas dari masing-masing instrumen pemerintah untuk mendorong tingkat investasi swasta.

METODOLOGI PENELITIAN

Analisis Prediksi Investasi dengan *Machine Learning*

Variabel Investasi dapat direpresentasikan dalam berbagai bentuk seperti Penanaman Modal Dalam Negeri (PMDN), Penanaman Modal Asing (PMA), maupun komponen Penanaman Modal Tetap Bruto (PMTB) dalam Produk Domestik Bruto (PDB). Adapun tujuan utama dari penelitian ini adalah sebagai dasar pertimbangan pemerintah dalam menarik Investasi, oleh karena itu PMTB, yang perhitungannya mencakup kebijakan investasi dari pemerintah dalam bentuk Belanja Modal, tidak dipilih sebagai variabel independen.

Selanjutnya, analisis korelasi dilakukan antara PMDN dan PMA, untuk menentukan apakah diperlukan dua model prediksi untuk masing-masing jenis investasi tersebut. Koefisien korelasi sebesar 0.73 menunjukkan korelasi yang cukup tinggi antara PMDN

dan PMA pada periode data observasi tahun 2016-2020. Dengan demikian, penelitian ini menggunakan data gabungan PMA dan PMDN (*PrInv*) sebagai variabel independen.

Athey, S. (2018) menekankan bahwa model *machine learning* memiliki keunggulan dalam model prediksi semiparametrik atau ketika terdapat jumlah kovariat (variabel independen) yang relatif banyak dibandingkan jumlah sampel observasinya. Oleh karena itu, penelitian ini terlebih dahulu mengumpulkan 49 variabel independen potensial berdasarkan berbagai penelitian terdahulu yang telah dibahas dalam Bab Tinjauan Literatur. Untuk menyederhanakan penyajian, 49 variabel independen potensial tersebut dikelompokkan menjadi 12 kelompok besar variabel yang dapat dijabarkan dalam persamaan berikut:

$$\begin{aligned} PrInv = & \beta XG_1 + \beta XG_2 + \beta XG_3 + \beta XG_4 + \\ & \beta XG_5 + \beta XG_6 + \beta XG_7 + \beta XG_8 + \\ & \beta XG_9 + \beta XG_{10} + \beta XG_{11} + \beta XG_{12} \\ & + \beta XG_{13} \end{aligned}$$

Penjelasan dari masing-masing variabel tersebut adalah sebagai berikut:

PrInv: *Private Investment* atau Penanaman Modal Tetap Swasta, yang merupakan gabungan dari PMA dan PMDN, dalam satuan Rupiah.

β : Koefisien Determinasi Variabel Independen.

XG_1 : Variabel kelompok Belanja Pemerintah Konsolidasian yang terdiri dari variabel Belanja Pegawai, Belanja Barang, Belanja Modal, dan Gabungan Belanja-Belanja Pemerintah Lainnya selain Transfer, Subsidi, dan Pembayaran Hutang.

XG_2 : Variabel kelompok Ketenagakerjaan yang terdiri dari variabel Upah Minimum Regional Provinsi (UMP) yang merepresentasikan biaya operasional bagi penanam modal dan Umur Harapan Hidup (UHH) yang menjadi *proxy* kualitas kesehatan tenaga kerja.

XG_3 : Variabel kelompok Ukuran Pasar (*Market Size*) yang terdiri dari variabel Produk Domestik Regional Bruto (PDRB), PDRB per Kapita, Populasi, Net Ekspor, dan Kepadatan Penduduk.

XG_4 : Variabel kelompok Suku Bunga yang terdiri dari BI Rate/7RRR dan Fed Rate.

XG_5 : Variabel kelompok Infrastruktur yang terdiri dari Jumlah Pelabuhan, Jumlah Pelabuhan per 10 juta penduduk, Kualitas Infrastruktur Pelabuhan Indriastiwi, F. (2017), *Dummy Eksistensi Bandara Utama, Dummy Eksistensi Pelabuhan Utama*, dan Indeks Komposit Pembangunan Infrastruktur (Faradis & Afifah, 2020).

XG_6 : Variabel kelompok Jarak Ibukota Provinsi ke Pusat Ekonomi yang terdiri dari Variabel jarak Ibukota Provinsi ke Ibukota Indonesia, jarak Ibukota Provinsi ke 5 Ibukota Negara Penyumbang PMA Indonesia terbesar periode tahun 2017-2021 (Singapura, Tokyo, Beijing, Hong Kong, Seoul), dan ongkos kirim paket J&T dari Ibukota Indonesia ke Ibukota Provinsi.

XG_7 : Variabel kelompok *Favorable Investment Climate* yang terdiri

- dari Indeks Demokrasi Indonesia, *Corruption Perception Index* (CPI), Persentase Penyelesaian Tindak Pidana (*Crime Clearance Rate*), Risiko Kriminalitas, dan Peringkat *Ease of Doing Business* (EoDB).
- XG₈: Variabel *dummy* tahun Pandemi COVID-19.
- XG₉: Variabel *dummy* eksistensi kota metropolitan dalam provinsi.
- XG₁₀: Variabel kelompok kekayaan sumber daya alam yang terdiri dari variabel cadangan minyak, cadangan gas, cadangan batu bara, *dummy* provinsi kaya minyak, *dummy* provinsi kaya gas, dan *dummy* provinsi kaya batu bara.
- XG₁₁: Variabel inflasi.
- XG₁₂: Variabel kurs dollar AS terhadap Rupiah.

Selanjutnya, untuk menyederhanakan model prediksi, dilakukan eliminasi variabel independen potensial dengan menggunakan koefisien korelasi dan *feature importance*.

Koefisien korelasi yang digunakan adalah antara masing-masing variabel independen potensial dengan variabel *PrIInv*. Eliminasi dilakukan dengan *threshold* koefisien korelasi sama dengan atau lebih besar dari |0,2|. Koefisien korelasi keseluruhan yang terbentuk disajikan dalam bentuk *Heatmap* dari *library* Seaborn dalam Python yang disajikan dalam Lampiran III.

Feature importance menunjukkan seberapa besar pengaruh suatu variabel independen secara relatif dibandingkan variabel independen lainnya dalam model prediksi *machine learning tree-based*. Model *preliminary* yang

menggunakan seluruh variabel independen potensial dibuat menggunakan *library* PyCaret untuk mencari model *tree-based* terbaik. Selanjutnya, *library* Yellowbrick dalam Python digunakan untuk menghasilkan *feature importance* dalam model tersebut, sebagaimana disajikan dalam Lampiran IV. *Feature importance* turut dipertimbangkan dalam proses eliminasi variabel independen potensial untuk mencegah tereliminasi variabel independen potensial yang memiliki pengaruh besar dalam estimasi semiparametrik, namun koefisien korelasinya dengan variabel *PrIInv* kecil. Berdasarkan proses eliminasi tersebut, 49 variabel independen potensial dieliminasi menjadi 20 variabel independen, sehingga persamaan yang digunakan untuk prediksi menjadi sebagai berikut:

$$\begin{aligned} PrIInv = & \beta_{51G} + \beta_{52G} + \beta_{53G} + \beta_{OtEG} \\ & + \beta_{UHH} + \beta_{UMP-1} + \beta_{PDRB-1} \\ & + \beta_{J_Pen} + \beta_{K_Pen} + \beta_{PortQ} + \\ & \beta_{DMap} + \beta_{InfraI} + \beta_{DistCp} + \\ & \beta_{DistSG} + \beta_{CrmCP} + \beta_{CrmRi} + \\ & \beta_{DMet} + \beta_{DOil} + \beta_{DNG} + \\ & \beta_{DCoal} \end{aligned}$$

Penjelasan dari masing-masing variabel tersebut adalah sebagai berikut:

- 51G: Belanja Pegawai konsolidasian Pusat dan Daerah per Provinsi dalam Rupiah.
- 52G: Belanja Barang konsolidasian Pusat dan Daerah per Provinsi dalam Rupiah.
- 53G: Belanja Modal konsolidasian Pusat dan Daerah per Provinsi dalam Rupiah.
- OtEG: Gabungan Belanja-Belanja konsolidasian Pemerintah Lainnya selain Transfer, Subsidi,

dan Pembayaran Hutang per Provinsi dalam Rupiah.

UHH: Umur Harapan Hidup per Provinsi dalam Satuan Tahun.

UMP-1: Upah Minimum Provinsi dalam Rupiah dengan *lag* periode 1 tahun.

PDRB-1: Produk Domestik Regional Bruto dalam Rupiah dengan *lag* periode 1 tahun.

J_Pen: Jumlah Penduduk dalam suatu Provinsi

K_Pen: Tingkat Kepadatan Penduduk dalam Suatu Provinsi

PortQ: Kualitas pelabuhan di suatu Provinsi berdasarkan penelitian Indriastiwi, F. (2017) dalam satuan indeks.

DMAP: Variabel *Dummy* eksistensi Bandara Utama di suatu Provinsi.

Infralx: Indeks Komposit Pembangunan Infrastruktur di Suatu Provinsi berdasarkan penelitian Faradis & Afifah (2020) dalam satuan indeks.

DistCp: Jarak Ibukota Provinsi ke Ibukota Indonesia (Jakarta) dalam satuan kilometer.

DistSG: Jarak Ibukota Provinsi ke SIngapura dalam satuan kilometer

CrmCP: *Crime Clearance Rate* atau persentase tindak pidana yang berhasil diselesaikan dibandingkan jumlah tindak pidana yang dilaporkan

CrmRi: Risiko Penduduk Terkena Tindak Pidana (Per 100.000 Penduduk)

DMet: Variabel *Dummy* Provinsi yang memiliki Metropolitan berdasarkan Definisi Peraturan Pemerintah Nomor 17 Tahun 2017.

DOil: Variabel *Dummy* Provinsi Kaya Cadangan Minyak (*Proven Reserve*) dengan *Threshold* cadangan minyak lebih besar dari 100 MMBSTB.

DNG: Variabel *Dummy* Provinsi Kaya Cadangan Gas (*Proven Reserve*) dengan *Threshold* cadangan gas lebih besar dari 900 KM³.

DCoal: Variabel *Dummy* Provinsi Kaya Cadangan Batu Bara (*Proven Reserve*) dengan *Threshold* cadangan batu bara lebih besar dari 3.000 Juta Ton.

Penjelasan lebih lengkap mengenai variabel-variabel yang digunakan dalam model *machine learning* tersebut disajikan dalam tabel deskripsi data pada Lampiran V.

Selanjutnya, model tersebut ditransformasi lebih lanjut dengan memberikan variabel *dummy* bagi masing-masing Provinsi yang dimaksudkan untuk menangkap keunikan karakteristik prediksi Investasi di masing-masing Provinsi yang belum mampu digambarkan oleh 20 variabel independen yang dipilih. Proses transformasi data tersebut dilakukan dengan menggunakan fitur OneHotEncoder dalam Python.

Berdasarkan hasil transformasi data tersebut, dilakukan komparasi model dengan menggunakan *library* PyCaret. Komparasi dimaksudkan untuk memilih 1 model terbaik dari 18 model prediksi *machine learning* yang tersedia dalam PyCaret. Kriteria yang digunakan untuk memilih model terbaik adalah model dengan R² terbesar.

Setelah model prediksi terbaik dipilih, dilakukan *hyperparameter tuning* terhadap model terbaik tersebut untuk menyempurnakan akurasi prediksi yang

dihasilkan model .Model *machine learning* yang telah di-*tuning* akan dievaluasi berdasarkan nilai R^2 dari model tersebut. Proses *tuning* dan finalisasi model prediksi dilakukan tenang *library* Scikit. Selain itu, analisis *feature importance* dengan *Library* Yellowbrick kembali digunakan untuk melihat variabel-variabel apa saja yang paling dominan dalam mempengaruhi nilai Investasi di suatu Provinsi.

Analisis Determinan Investasi

Analisis determinan investasi dilakukan dengan menggunakan regresi *Ordinary Least Square* (OLS) dengan aplikasi Eviews versi 10. Persamaan yang dibentuk secara umum mengadopsi persamaan yang digunakan dalam analisis prediksi dengan *machine learning*. Namun demikian, berbeda dengan model *machine learning*, model regresi OLS diharuskan untuk lolos uji asumsi klasik agar keandalan model terjaga. Untuk itu, diperlukan beberapa transformasi terhadap model dan data yang digunakan.

Transformasi model dilakukan dengan mengeliminasi beberapa variabel independen yang berdasarkan penilaian *Centered Variance Inflation Factors* (VIF) awal terindikasi saling berkorelasi. Variabel-variabel yang dieliminasi berdasarkan kriteria tersebut meliputi 51G, 52G, OtEG, PDRB-1, K_Pen, DistCp, dan DMet.

Transformasi data dilakukan karena ditemukannya indikasi heteroskedastisitas dalam uji awal dengan metode *Breusch-Pagan-Godfrey*. Hal tersebut diduga terjadi karena adanya fluktuasi nilai investasi di beberapa tahun yang tidak seragam terjadi antar provinsi. Oleh karena itu,

data panel ditransformasi menjadi data *cross-section*.

Transformasi menjadi *cross-section* dilakukan dengan mengakumulasi data selama 5 tahun (2016-2020) untuk variabel yang satuan datanya dapat diakumulasi seperti *PrInv* dan 53G. Untuk variabel yang satuan datanya tidak dapat diakumulasi seperti *UHH*, *UMP-1*, *J_Pen*, *CrmCP*, dan *CrmRi* transformasi dilakukan dengan meratakan nilai data selama 5 tahun (2016-2020). Adapun, untuk variabel yang datanya bersifat statis seperti *PortQ*, *DMap*, *Infralx*, *DistSG*, *DOil*, *DNG*, dan, *DCoal*, data yang digunakan adalah nilai pada satu tahun.

Dengan demikian, persamaan yang digunakan dalam regresi OLS *cross-section* untuk analisis determinan investasi adalah sebagai berikut.

$$PrInv = C + \beta_{53G} + \beta_{UHH} + \beta_{UMP-1} + \beta_{J_Pen} + \beta_{PortQ} + \beta_{DMap} + \beta_{Infralx} + \beta_{DistSG} + \beta_{CrmCP} + \beta_{CrmRi} + \beta_{DOil} + \beta_{DNG} + \beta_{DCoal}$$

Penjelasan lebih lengkap mengenai variabel-variabel yang digunakan dalam model regresi OLS tersebut disajikan dalam tabel deskripsi data pada Lampiran VI.

HASIL PENELITIAN DAN PEMBAHASAN

Analisis Prediksi Investasi dengan *Machine Learning*

Untuk menghasilkan model dengan kinerja prediksi terbaik, *Library* PyCaret digunakan untuk membandingkan 18 model regresi *Machine Learning*. Model-model tersebut meliputi *Extra Trees Regressor*, *CatBoost Regressor*, *Random*

Forest Regressor, Gradient Boosting Regressor, Extreme Gradient Boosting, Orthogonal Matching Pursuit, Light Gradient Boosting Machine, Decision Tree Regressor, AdaBoost Regressor, K-Neighbors Regressor, Huber Regressor, Lasso Least Angle Regression, Lasso Regression, Bayesian Ridge, Passive Aggressive Regressor, Linear Regression, Ridge Regression, dan Elastic Net.

Hasil komparasi kinerja prediksi 18 model dengan PyCaret menunjukkan bahwa model *Extra Trees Regressor* menghasilkan akurasi prediksi terbaik dalam setiap kriteria akurasi, yang meliputi R^2 , *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Root Mean Squared Logarithmic Error* (RMSLE), dan *Mean Absolute Percentage Error* (MAPE). Tabel perbandingan kinerja prediksi 18 model tersebut disajikan dalam Lampiran VII.

Model *Extra Trees Regressor* merupakan model pengembangan dari model *Decision Tree Regressor*. *Decision Tree* membangun model regresi atau klasifikasi dalam bentuk struktur pohon. Kumpulan data dipecah menjadi himpunan bagian yang lebih kecil sementara pada saat yang sama *Decision Tree* terkait dikembangkan secara bertahap. Hasil akhir dari model berupa pohon dengan *decision node* dan *leaf node*. Sebuah *decision node* memiliki dua atau lebih cabang yang masing-masing mewakili nilai untuk atribut yang diuji. *Leaf node* mewakili keputusan atas target angka tertentu. *Decision node* paling atas dalam *decision tree* yang sesuai dengan prediktor terbaik disebut *root node*.

Extra Trees Regressor dikembangkan dari model *decision trees* dengan membangun sejumlah besar model

decision trees independen yang dihasilkan secara acak dengan menggunakan seluruh data set dalam setiap *decision tree*. Hasil prediksi dari model-model *decision tree* tersebut digunakan secara bersamaan untuk menghasilkan nilai prediksi yang lebih akurat dalam *Extra Trees Regressor* (Geurts, et al., 2006).

Selanjutnya, model *Extra Trees* yang telah dipilih berdasarkan komparasi R^2 18 model dalam PyCaret, diuji coba untuk memprediksi variabel *Prilnv* dengan menggunakan *library* Scikit, yang secara umum lebih *robust* dibandingkan PyCaret. Hasil uji coba awal model menghasilkan nilai R^2 sebesar 0,8428.

Selanjutnya, untuk meningkatkan akurasi prediksi model *Extra Trees*, dilakukan *hyperparameter tuning* dengan menggunakan *library* Scikit. Beberapa *hyperparameter* yang diuji coba meliputi *n_estimators* (5, 20, 24, 25, 30, 50, 100, dan 200), *min_samples_leaf* (1 dan 2), *min_samples_split* (2 dan 3), *criterion* ('mse' dan 'mae'), *bootstrap* (True dan False), *max_depth* (None dan 27), serta *max_features* ('auto' dan 'log2'),

Penjelasan untuk masing-masing *hyperparameter* tersebut adalah sebagai berikut:

1. *n_estimators*: jumlah *trees* yang digunakan dalam model berbasis pohon.
2. *min_samples_leaf*: jumlah minimum sampel yang dibutuhkan untuk membentuk suatu *leaf node*.
3. *min_samples_split*: jumlah minimum sampel yang dibutuhkan untuk memecah sebuah *node* internal.
4. *criterion*: Fungsi untuk mengukur kualitas sebuah pecahan *node* dan menentukan jenis *decision tree* yang

digunakan. criterion 'mae' menggunakan *mean absolute error* sebagai ukuran, sedangkan 'mse' menggunakan *mean squared error*.

5. bootstrap: menentukan apakah data sampel (*bootstrap*) tau dataset utuh yang digunakan dalam membentuk masing-masing pohon dalam model *tree based*..
6. max_depth: jumlah maksimum kedalaman sebuah *decision tree*. Apabila kriteria *None* dipilih, maka jumlah maksimum tidak dibatasi.
7. max_features: jumlah maksimum *features* (variabel independen yang berfungsi sebagai input dalam *decision tree*) yang digunakan dalam menentukan jumlah pecahan *node* terbaik. Kriteria 'auto' berarti jumlah maksimum *features* sama dengan jumlah *features* yang ada, sedangkan kriteria 'log' berarti jumlah maksimum *features* sama dengan logaritma dari jumlah *features* yang ada.

Berdasarkan hasil *hyperparameter tuning* tersebut, model *Extra Trees* dengan R^2 terbaik menggunakan *hyperparameter* *n_estimators* 25, *min_samples_leaf* 1, *min_samples_split* 2, criterion 'mse', bootstrap False, max_depth 27, dan max_features 'auto'. Hasil model *Extra Trees Regressor* final tersebut memiliki nilai R^2 sebesar 0,8428. Dengan demikian, 84,28% dari variasi *Prilnv* di 34 Provinsi Indonesia dapat dijelaskan dengan model prediksi tersebut.

Model *machine learning Extra Trees Regressor* juga dapat mengidentifikasi variabel-variabel independen yang memiliki *feature importance* tinggi atau pengaruh yang besar terhadap hasil prediksi. Variabel-variabel tersebut secara berturut-turut dari *feature*

importance tertinggi adalah *PDRB-1*, *51G*, *DistCp*, *J_Pen*, *PortQ*, *53G*, *52G*, *K_Pen*, *Infralx*, *UHH*, *CrmCP*, *CrmRi*, *DistSg*, *DOil*, *OtEG*, dan *DCoal*. Hasil identifikasi *Feature Importance* tersebut disajikan dalam lampiran VIII.

Signifikansi *relative importance* variabel lag *PDRB* (*PDRB-1*), Jumlah Penduduk (*J_Pen*), Kepadatan Penduduk (*K_Pen*), Belanja Pegawai (*51G*), Belanja Barang (*52G*), dan Belanja Pemerintah Lainnya (*OtEG*) dapat dianalogikan dengan kategori variabel Ukuran Pasar atau *Market Size* dalam penelitian investasi regional Ledyeva, S. (2009) di Rusia dan variabel *PDB* dalam penelitian Fathia, et al. (2021). Enam variabel tersebut, baik secara langsung maupun tidak dapat mempengaruhi ukuran pasar atau jumlah permintaan barang dan jasa di suatu Provinsi. Keputusan penanam modal untuk mempertimbangkan ukuran pasar dalam memilih lokasi investasi juga sejalan dengan teori *least cost* Alfred Weber, dimana pembangunan suatu industri dapat mempertimbangkan lokasi produksi yang berdekatan dengan potensi pasar yang lebih besar.

Signifikansi *relative importance* variabel yang berkaitan dengan jarak terhadap pusat ekonomi juga ditemukan dalam penelitian Ledyeva, S. (2009). Variabel jarak antara Ibukota Provinsi dengan Ibukota Indonesia (*DistCp*) dan jarak antara Ibukota Provinsi dengan Singapura (*DistSG*) dapat menjadi data *proxy* yang menggambarkan potensi perdagangan dan aliran modal suatu Provinsi dengan daerah lain yang memiliki tingkat perekonomian besar. Adapun berdasarkan data BKPM (2022), Singapura merupakan negara penyumbang PMA terbesar ke Indonesia.

Variabel terkait Infrastruktur, yakni *PortQ* dan *Infralx* juga memiliki *relative importance* yang cukup tinggi sebagai prediktor dalam model. Di antara dua variabel tersebut, variabel kualitas pelabuhan memiliki nilai *relative importance* yang lebih tinggi. Hal tersebut mengindikasikan bahwa adanya pelabuhan yang baik memiliki daya tarik investasi yang lebih tinggi. Hal tersebut juga sejalan dengan karakteristik Indonesia sebagai negara kepulauan dan karakteristik perdagangan internasional yang sebagian besar dilakukan melalui transportasi laut. Di sisi lain, variabel yang merepresentasikan eksistensi bandara utama (*DMap*), tidak memiliki *relative importance* yang signifikan.

Signifikansi *relative importance* variabel Usia Harapan Hidup (*UHH*) yang merupakan variabel kesehatan dapat dilihat dari perspektif ketenagakerjaan. Pekerja yang sehat cenderung lebih produktif, serta memerlukan lebih sedikit cuti sakit dan biaya pengobatan. Dikaitkan dengan teori *least cost* Alfred Weber, hal tersebut dapat menjadi salah satu pertimbangan penanam modal dari segi biaya tenaga kerja.

Variabel *crime completion rate* (*CrmCp*) merepresentasikan kinerja penegak hukum, sedangkan variabel persentase penduduk terkenda tindak pidana (*CrmRi*) merepresentasikan tingkat keamanan di masyarakat. Signifikansi *relative importance* kedua variabel tersebut mengindikasikan bahwa risiko kelembagaan berupa tingkat keamanan di suatu Provinsi juga dapat mempengaruhi rasa keamanan penanam modal untuk berinvestasi di Provinsi tersebut. Variabel serupa yang merepresentasikan risiko kelembagaan, yakni risiko politik dan risiko legislatif,

dalam penelitian Ledyeva, S. (2009) juga terbukti signifikan dalam mempengaruhi tingkat investasi di Rusia.

Terakhir, signifikansi *relative importance* variabel *dummy* provinsi dengan cadangan minyak tinggi (*DOil*) dan variabel *dummy* provinsi dengan cadangan batu bara tinggi (*DCoal*) menunjukkan bahwa kekayaan sumber daya alam suatu provinsi dapat menjadi daya tarik bagi penanam modal untuk berinvestasi dalam bidang eksplorasi dan ekstraksi Sumber Daya Alam. Hal serupa juga ditemukan oleh Ledyeva, S. (2009) di Rusia melalui signifikansi variabel *dummy* Sakhalin (Provinsi Kaya SDA) dan variabel index minyak dan gas. Di sisi lain, tidak signifikannya variabel *dummy* provinsi dengan cadangan gas tinggi (*DGas*) juga dapat menjadi dasar penelitian lebih lanjut atau evaluasi kebijakan investasi untuk menelusuri indikasi kurang tertariknya investor untuk berinvestasi di bidang tersebut.

Analisis Determinan Investasi

Berdasarkan hasil regresi OLS untuk analisis determinan, didapati nilai koefisien determinasi (R^2) dari regresi Ordinary Least Square adalah 0.8362 yang berarti bahwa 83,62% variasi dari *Prilnv* dapat dijelaskan oleh seluruh variabel bebas. Sementara itu 16.38% variasi dari *Prilnv* dijelaskan oleh variabel lain yang tidak diikutsertakan dalam model.

Selanjutnya, untuk memastikan keandalan model, dilakukan uji asumsi klasik. Adapun untuk modelling yang berbasis data *cross-section*, uji yang dilakukan meliputi uji normalitas, uji multikolinearitas, dan uji heteroskedastisitas.

Nilai Prob. dari uji normalitas dengan Jarque-Bera sebesar 0.721141 (72.11%) lebih besar dari $\alpha=5\%$ sehingga dapat disimpulkan bahwa residual dari model persamaan memiliki distribusi normal. Untuk pengujian multikolinearitas, digunakan nilai *Centered Variance Inflation Factors* (VIF). Berdasarkan nilai *Centered VIF* tersebut, tidak ada yang lebih besar dari 10 sehingga dapat disimpulkan bahwa tidak terdapat permasalahan multikolinearitas dalam model. Selanjutnya, Uji Heteroskedastisitas dilakukan dengan menggunakan metode Breusch-Pagan-Godfrey. Nilai Prob. chi-square adalah 0,1964 (19,64%) lebih besar dari $\alpha=5\%$, sehingga dapat disimpulkan tidak terdapat permasalahan heteroskedastisitas dalam model. Hasil model regresi dan uji asumsi klasik dengan Eviews disajikan pada Lampiran IX sampai dengan XII.

Setelah melalui Uji Asumsi Klasik, dilakukan penilaian atas signifikan F dari model. Nilai *Prob. F* dari model adalah sebesar 0.0000 (0,00%) lebih kecil dari $\alpha=5\%$, sehingga dapat disimpulkan bahwa seluruh variabel bebas secara bersama-sama berpengaruh signifikan secara statistik terhadap variabel *Prilnv*. Analisis determinan dilakukan dengan melihat signifikansi masing-masing variabel independen terhadap *Prilnv*. Hal tersebut dapat dilakukan dengan Uji t. Dalam model determinan tersebut, variabel bebas yang berpengaruh signifikan secara individu terhadap variabel *Prilnv* pada $\alpha=5\%$ adalah variabel *53G*, *PortQ*, *J_pen* dan *Doil*. Oleh karena itu, dapat disimpulkan bahwa dengan pengujian data *cross-section* pada 34 provinsi di Indonesia selama tahun 2016 sampai tahun 2020 (diakumulasi), tingkat investasi swasta

pada setiap provinsi dipengaruhi secara signifikan oleh Belanja Modal Konsolidasi Pemerintah Pusat dan Pemerintah Daerah, Kualitas Pelabuhan, Jumlah Penduduk dan Tingkat Cadangan Minyak pada tiap provinsi. Berdasarkan penelitian Hermes Niels dan Lensink Robert (2001), belanja modal pemerintah memiliki pengaruh positif pada investasi swasta setelah tingkat belanja tertentu. Pemerintah harus merancang jenis dan tingkat belanja secara hati-hati agar pengaruhnya terhadap tingkat investasi menjadi positif. Belanja modal pemerintah mendorong tingkat produktivitas suatu daerah sehingga hal ini dapat menjadi pendorong bagi swasta untuk menanam investasi baru maupun meningkatkan investasi yang sudah ada sebelumnya.

Selanjutnya, signifikansi variabel kualitas pelabuhan terhadap investasi juga ditemukan dalam penelitian Wekesa, et al. (2016), dimana investasi perbaikan kualitas pelabuhan mempengaruhi tingkat investasi asing di Kenya. Sementara itu, signifikansi variabel jumlah penduduk dan variabel cadangan minyak terhadap investasi asing, juga ditemukan dalam penelitian Ledyeva (2009) di Rusia.

KESIMPULAN DAN SARAN

Kesimpulan:

Penelitian ini membandingkan kinerja prediksi 18 model *machine learning* untuk menghasilkan model prediksi Investasi terbaik. Berdasarkan komparasi tersebut, model *Extra Trees Regressor* menghasilkan akurasi prediksi terbaik dengan nilai R^2 sebesar 0,8428.

Selanjutnya, model *Extra Trees Regressor* tersebut juga menemukan bahwa

variabel-variabel terkait *market size*, jarak terhadap pusat ekonomi, Infrastruktur, kesehatan, kriminalitas, dan kekayaan sumber daya alam berpengaruh signifikan sebagai prediktor investasi provinsi-provinsi di Indonesia. Hal tersebut mengindikasikan bahwa variabel-variabel tersebut turut menjadi salah satu faktor penentu keputusan penanam modal untuk berinvestasi di suatu Provinsi.

Dalam konteks pengaruh variabel kekayaan sumber daya alam terhadap investasi, ditemukan bahwa variabel kekayaan minyak dan variabel kekayaan batu bara memiliki pengaruh yang signifikan, sedangkan kekayaan gas, tidak memiliki pengaruh yang signifikan. Hal tersebut mengindikasikan diperlukannya penelitian lebih lanjut atau evaluasi kebijakan yang berkaitan dengan investasi gas di Indonesia.

Sejalan dengan nilai *relative importance* yang signifikan dalam model prediksi *machine learning*, beberapa variabel juga memiliki pengaruh yang signifikan dalam model determinan dengan regresi OLS. Variabel-variabel tersebut meliputi Belanja Modal Pemerintah (53G), Kualitas Pelabuhan (*PortQ*), Jumlah Penduduk (*J_Pen*) dan *Dummy* Provinsi dengan cadangan Minyak besar (*DOil*).

Saran:

Dalam membuat kebijakan untuk meningkatkan investasi di Indonesia, pemerintah dapat mempertimbangkan untuk membuat kebijakan yang berkaitan dengan variabel-variabel prediktor yang berpengaruh tinggi dalam model *machine learning* dan variabel-variabel determinan yang signifikan dalam model regresi. Dengan berfokus pada indikator dalam variabel

tersebut, kebijakan pro-investasi dapat lebih efektif dalam mendorong peningkatan investasi karena sesuai dengan faktor-faktor yang menjadi penentu keputusan investasi di suatu Provinsi oleh penanam modal. Untuk itu, kebijakan yang dapat ditempuh pemerintah antara lain adalah dengan pembangunan Infrastruktur,

peningkatan akses dan kualitas kesehatan, serta peningkatan kualitas penegakan hukum di setiap Provinsi.

Khusus untuk infrastruktur, dari 3 variabel independen terkait infrastruktur yang digunakan, variabel kualitas pelabuhan (*PortQ*) memiliki *relative importance* yang paling tinggi dalam model *machine learning*. Hal tersebut mengindikasikan bahwa pembangunan infrastruktur pelabuhan berkapasitas tinggi yang efektif dalam menekan biaya logistik, berpotensi untuk menghasilkan daya tarik investasi yang paling efektif.

Selain itu, kebijakan mengenai izin, perpajakan, dan royalti di bidang SDA juga dapat dievaluasi untuk meningkatkan minat penanam modal dalam berinvestasi di sektor tersebut. Urgensi untuk melakukan evaluasi kebijakan investasi di bidang SDA tersebut juga cukup tinggi mengingat adanya indikasi bahwa investor kurang berminat dalam berinvestasi di sektor SDA gas, dalam model *machine learning*. Signifikansi *relative importance* variabel jarak Ibukota Provinsi ke Ibukota Indonesia (*DistCp*) mengindikasikan potensi dampak positif yang menarik di tengah-tengah rencana pemerintah untuk memindahkan Ibukota Negara ke daerah Kalimantan Timur. Sebagai Daerah yang terletak di tengah-tengah Indonesia, maka *DistCp* dari Ibukota Negara ke masing-masing Ibukota Provinsi akan lebih merata. Untuk itu,

pemerintah sebaiknya melakukan perencanaan yang matang agar Ibukota baru tersebut dapat mendekati fungsi Jakarta sebagai *hub* ekonomi dan transportasi, guna memaksimalkan potensi pemerataan Investasi yang dapat terjadi.

Adapun variabel independen yang berkaitan dengan *market size* juga memiliki *relative importance* yang signifikan sebagai prediktor model *machine learning*. Namun demikian, variabel-variabel tersebut cenderung bersifat *given* sehingga sulit untuk dilakukan intervensi oleh pengambil kebijakan.

Dalam konteks yang lebih teknis operasional, model prediksi *machine learning* dalam penelitian ini dapat dilakukan *deployment* sebagai *tools* simulasi untuk pengambil kebijakan. Simulasi tersebut dapat membantu pengambil kebijakan untuk melihat proyeksi dampak kebijakan yang akan ditempuh, terhadap kenaikan investasi di suatu provinsi. Sebagai contoh, *mock-up* aplikasi simulasi prediksi tersebut disajikan dalam Lampiran I.

Selain itu, variabel-variabel yang signifikan, baik dalam model *machine learning* maupun model regresi, dapat dituangkan ke dalam *dashboard* yang dapat diakses calon penanam modal sebagaimana dicontohkan dalam Lampiran II. *Dashboard* tersebut dapat mendukung pemerintah dalam mempromosikan potensi-potensi setiap Provinsi di Indonesia kepada calon penanam modal. Pada akhirnya, hal tersebut dapat mendorong penanam modal untuk berinvestasi di suatu Provinsi.

Dalam konteks yang lebih mikro, yakni perwujudan DJPb sebagai *Regional Chief Economist*, unit vertikal DJPb dapat

memanfaatkan temuan variabel-variabel penting dalam penelitian ini, beserta *tools* berupa *dashboard* dan *deployment* simulasi dampak perubahan suatu variabel terhadap investasi. Ketiga hal tersebut dapat memperkuat strategi komunikasi unit vertikal DJPb dalam melaksanakan promosi investasi regional dan kerja sama pembangunan dengan *stakeholders* pemerintah di daerah.

IMPLIKASI DAN KETERBATASAN

Penelitian ini menggabungkan data PMA dan PMDN sebagai variabel dependen untuk menyederhanakan model analisis, berdasarkan koefisien korelasi yang cukup tinggi antara kedua variabel tersebut (0,73). Namun demikian, nilai koefisien korelasi tersebut juga berimplikasi bahwa PMA dan PMDN tidak sepenuhnya memiliki karakteristik yang sama. Untuk itu, penelitian selanjutnya dapat membuat dua model yang berbeda untuk masing-masing PMA dan PMDN. Pengembangan tersebut berpotensi menghasilkan nilai prediksi yang lebih tinggi, serta menangkap perbedaan pertimbangan investasi yang berpotensi berbeda antara penanam modal asing dan domestik.

Penelitian ini juga baru menggunakan lag sebesar 1 periode dalam variabel independen *PDRB-1* dan *UMP-1* dalam *modelling*. Dalam praktiknya, terdapat berbagai variabel independen yang dampaknya terhadap investasi berpotensi terdistribusi dalam beberapa periode *lag*. Hal tersebut juga sejalan dengan karakteristik investasi fisik yang bersifat jangka panjang, sehingga memerlukan waktu yang cukup panjang

pula bagi seorang penanam modal, sebelum memutuskan untuk berinvestasi atas potensi-potensi yang dilihatnya. Oleh karena itu, penyempurnaan *modelling* berikutnya juga dapat jenis-jenis model atau data yang mempertimbangkan distribusi *lag* dari variabel independen.

Dalam konteks model prediksi, kebermanfaatan model akan lebih optimal dengan mempertimbangkan selisih waktu ketersediaan data antara variabel independen dan variabel dependen. Seluruh data variabel independen perlu dipastikan sudah tersedia, sebelum data variabel dependen yang akan diprediksi tersedia. Dengan demikian, model akan terus optimal untuk digunakan dalam prediksi nilai variabel dependen di periode berikutnya. Data-data dalam penelitian ini masih cenderung menggunakan data dari *timeline* yang sama. Oleh karena itu, penyempurnaan model ke depannya juga dapat mempertimbangkan *timeline* ketersediaan data tersebut.

Beberapa data yang digunakan sebagai variabel independen seperti kualitas infrastruktur pelabuhan (PortQ) dan indeks komposit infrastruktur (InfraIx) masih bersifat statis karena adanya keterbatasan ketersediaan data. Periode data yang diteliti juga baru selama 5 periode karena adanya keterbatasan di beberapa data. Selain itu, variabel-variabel yang diduga dapat mempengaruhi investasi seperti perubahan kebijakan pro-investasi pemerintah, negara sumber investasi, kualitas kelembagaan nasional maupun daerah juga belum diperhitungkan. Untuk mengoptimalkan potensi prediksi dan *feature importance* dari model *machine learning* dalam menghasilkan masukan kebijakan pro-investasi,

penyempurnaan model selanjutnya dapat memperkaya data yang digunakan baik dari jumlah sampel maupun jumlah variabel yang lebih akurat

REFERENSI

- Anderson, D. (1990). Investment and economic growth. *World Development*, 18(8), 1057-1079.
- Athey, S. (2018). The impact of machine learning on economics. *The economics of artificial intelligence: An agenda*, 507-547.
- Blomström, M., Lipsey, R. E., & Zejan, M. (1996). Is fixed investment the key to economic growth?. *The Quarterly Journal of Economics*, 111(1), 269-276.
- Budiono, S., & Purba, J. T. (2019). Data panel model solution in forecasting investments through energy electricity and government policy in Indonesia. In *Proceedings of the International Conference on Industrial Engineering and Operations Management* (pp. 1132-1132).
- Chow, G. C. (2017). *Capital formation and economic growth in China* (pp. 1186-1221). Brill.
- Faradis, R., & Afifah, U. N. (2020). Indeks Komposit Pembangunan Infrastruktur Provinsi-Provinsi di Indonesia. *Jurnal Ekonomi dan Pembangunan Indonesia*, 20(1), 33-55.
- Fathia, N., Silvia, V., & Gunawan, E. (2021). Analysis of Foreign Investment Determinants in Indonesia. *Economics Development Analysis Journal*, 10(3), 338-350.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3-42.
- Gupta, K. (2021). The Importance of Financial Liberalisation for Economic

Growth: The Case of Indonesia. *Bulletin of Indonesian Economic Studies*, 57(2), 175-201.

Heijman, W., & Ophem, J. V. (2007). Abatement of Commuting's Negative Externalities by Regional Investment in Houses and Buildings. In *Regional Externalities* (pp. 245-254). Springer, Berlin, Heidelberg.

Niels, H., & Robert, L. (2001). Fiscal policy and private investment in less developed countries. The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki.

Indriastiwi, F. (2017). Identifikasi Fasilitas 24 Pelabuhan di Indonesia Menggunakan Analisis Cluster dan Analysis Hierarchy Process. *Jurnal Penelitian Transportasi Laut*, 19(1), 25-39.

Mulpuru, V., & Mishra, N. (2021). In silico prediction of fraction unbound in human plasma from chemical fingerprint using automated machine learning. *ACS omega*, 6(10), 6791-6797.

Moran, T., Graham, E. M., & Blomström, M. (Eds.). (2005). *Does foreign direct investment promote development?*. Columbia University Press.

Noor, M. A., & Saputra, P. M. A. (2020). Emisi Karbon dan Produk Domestik Bruto: Investigasi Hipotesis Environmental Kuznets Curve (EKC) pada Negara Berpendapatan Menengah di Kawasan ASEAN. *Jurnal Wilayah dan Lingkungan*, 8(3), 230-246.

Peraturan Pemerintah Republik Indonesia Nomor 17 Tahun 2017. *Sinkronisasi Proses Perencanaan dan Penganggaran Pembangunan Nasional*. 24 Mei 2017. Lembaran Negara Republik Indonesia Tahun 2017 Nomor 105. Jakarta.

Shabbir, M. S., Bashir, M., Abbasi, H. M., Yahya, G., & Abbasi, B. A. (2021).

Effect of domestic and foreign private investment on economic growth of Pakistan. *Transnational Corporations Review*, 13(4), 437-449.


Soekro, S. R., & Widodo, T. (2015). *Mapping and determinants of intra-ASEAN Foreign Direct Investment (FDI): Indonesia Case Study* (No. WP/12/2015).

Suhendra, I., & Anwar, C. J. (2014). Determinants of private investment and the effects on economic growth in Indonesia. *GSTF Journal on Business Review (GBR)*, 3(3), 1-6.

Wekesa, C. T., Wawire, N. H., & Kosimbei, G. (2016). Effects of infrastructure development on foreign direct investment in Kenya. *Journal of Infrastructure Development*, 8(2), 93-110.

World Bank. (2022). Ease of Doing Business Rank - Indonesia. <https://data.worldbank.org/indicator/IC.BUS.EASE.XQ?end=2019&locations=ID&start=2012&view=bar>


Lampiran I - Mock-up Aplikasi Simulasi Prediksi Investasi Provinsi



SEDAP
Setidjen Data Analytics Pioneers

Pilih Menu:

- ☐ Analisis
- ☒ Prediksi



Prediksi Jumlah Investasi Agregat (Dalam dan Luar Negeri)

Pilih Provinsi

Aceh

Aceh

Berikut adalah data yang terisi secara otomatis untuk provinsi : Aceh

Jarak Ibukota Provinsi ke Ibukota Negara (dalam km)

1830

Jarak Ibukota Provinsi ke Singapura (dalam km)

1059

Keberadaan Bandara Utama (data menurut BPS); 0 = tidak punya 1 = punya

0

Kualitas Infrastruktur Pelabuhan

0.32

Indeks Komposit Pembangunan Infrastruktur; 0=tidak memadai, 1=cukup memadai, 2=sangat memadai

1

Keberadaan kota metropolitan; 0=tidak ada, 1=ada

0

Cadangan Minyak Bumi; 0=sedikit, 1=banyak

0

Cadangan Gas Alam; 0=sedikit, 1=banyak

0

Cadangan Batu Bara; 0=sedikit, 1=banyak

0

Lengkapi data berikut:

Masukkan Belanja Pegawai

19285486099020

Masukkan Belanja Barang dan Jasa

13258660858311

Masukkan Belanja Modal

8492556614925

Masukkan Belanja Lainnya Selain Transfer Gabungan Pusat & Daerah

3428386300303

Masukkan Usia Harapan Hidup

69

Masukkan Upah Minimum Regional Tahun Sebelumnya

2916810

Masukkan Produk Domestik Regional Bruto Tahun Sebelumnya

164162980000000

Masukkan Jumlah Penduduk

5388100

Masukkan Kepadatan Penduduk per Kilometer Persegi

91

Masukkan Persentase Penyelesaian Tindak Pidana (Crime Clearance Rate) merupakan persentase tindak pidana yang berhasil diselesaikan dibandingkan jumlah tindak pidana yang dilaporkan

37.90

Masukkan Risiko Penduduk Terkena Tindak Pidana (Per 100.000 Penduduk)

149

Prediksi Jumlah Investasi Agregat

Deskripsi *Mock-up* Aplikasi Simulasi Prediksi Investasi Provinsi:

1. Fitur:

- Aplikasi menyajikan *interface* untuk mensimulasikan prediksi investasi di setiap Provinsi. Dalam *interface* tersebut, *user* dapat memilih salah satu provinsi (Provinsi Aceh dalam contoh) yang akan disimulasikan. Ketika Provinsi Aceh dipilih, maka data terbaru (Tahun 2020) untuk setiap variabel prediktor di Provinsi Aceh akan muncul.
- Selanjutnya, *user* dapat melakukan simulasi dengan mencoba mengubah nilai input satu atau beberapa variabel prediktor. Setelah mengubah nilai input, *user* dapat mengklik tombol "Prediksi Jumlah Investasi Agregat". Tombol tersebut akan mengeksekusi model *Extra Trees Regressor* yang terpilih sebagai model terbaik dalam penelitian.
- Setelah mengeksekusi model *Extra Trees Regressor*, aplikasi akan menampilkan prediksi nilai investasi pada Provinsi Aceh dengan nilai-nilai variabel prediktor yang diinput.

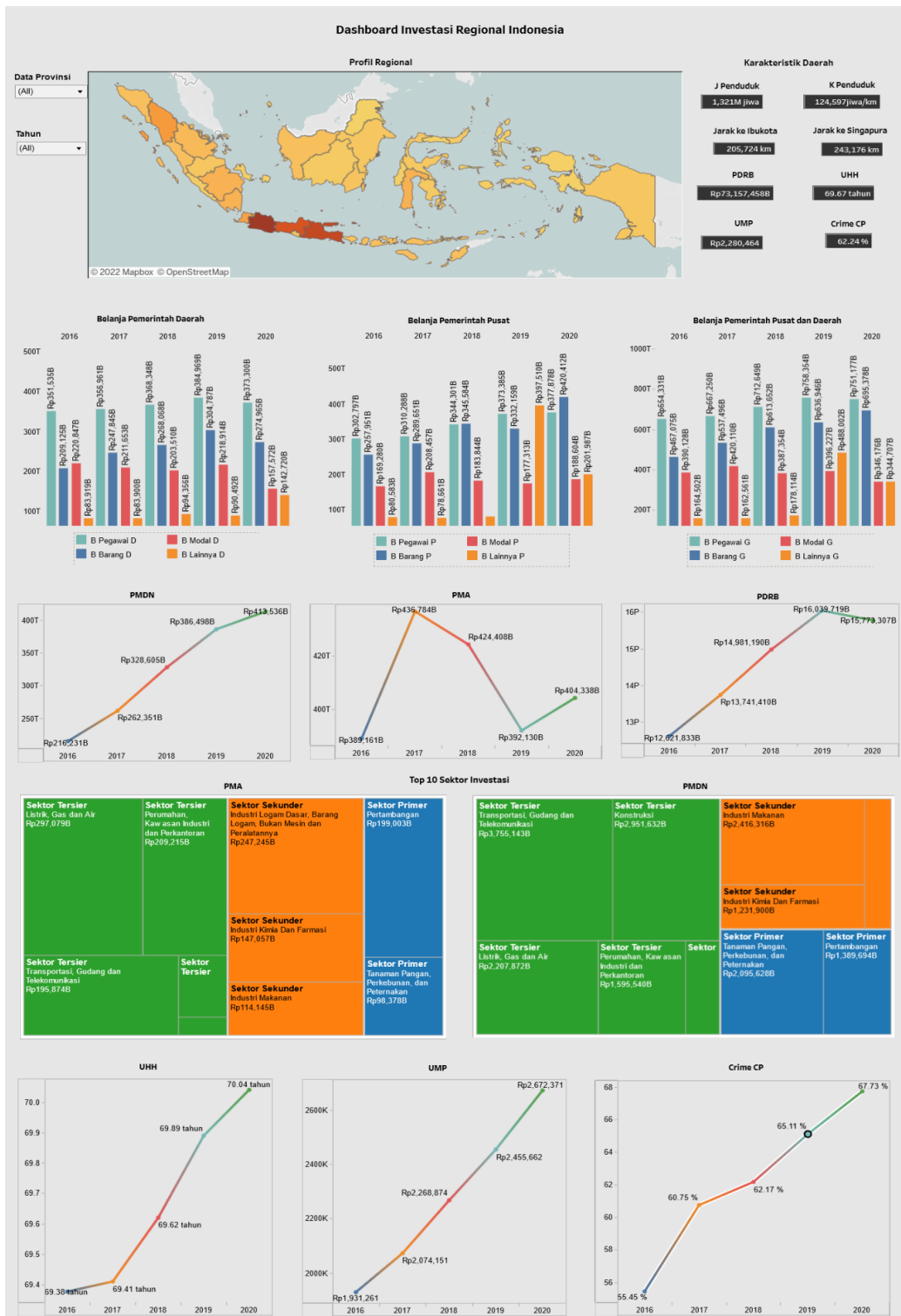
2. Potensi Pemanfaatan :

Sebagai bahan pertimbangan Pemerintah selaku pengambilan kebijakan, khususnya dalam memperkirakan dampak perubahan kebijakan terhadap nilai investasi yang akan didapatkan di suatu provinsi dengan melakukan simulasi. Contoh-contoh kasus pemanfaatan tersebut antara lain:

Contoh Kasus	Pemanfaatan aplikasi
Pengambil kebijakan ingin menambah Belanja Modal sebesar Rp. 1 Triliun di Provinsi Aceh untuk meningkatkan daya tarik investasi.	Dalam aplikasi simulasi, pengambil kebijakan dapat menginput tambahan Rp. 1 Triliun pada variabel Belanja Modal Konsolidasian, dan melihat potensi manfaatnya berdasarkan nilai investasi yang diprediksi di Provinsi Aceh.
Pengambil kebijakan ingin meningkatkan kapasitas infrastruktur Pelabuhan di Provinsi Maluku Utara hingga setara dengan Pelabuhan di Provinsi DKI Jakarta untuk meningkatkan daya tarik investasi.	Dalam aplikasi simulasi, pengambil kebijakan dapat mengubah variabel Kualitas Infrastruktur Pelabuhan menjadi 0.756 (Setara Kualitas Provinsi DKI Jakarta), dan melihat potensi manfaatnya berdasarkan nilai investasi yang diprediksi di Provinsi Maluku Utara.

Contoh Kasus	Pemanfaatan aplikasi
<p>Pengambil kebijakan ingin menggalakan penegakan hukum di Provinsi Jambi dengan menargetkan POLDA setempat untuk meningkatkan <i>Crime Clearance Rate</i> (Persentase Laporan Tindak Pidana yang berhasil diselesaikan) hingga menjadi 90%.</p>	<p>Meskipun secara umum tidak berkaitan secara langsung dengan investasi, namun pengambil kebijakan dapat mengubah variabel <i>Crime Clearance Rate</i> menjadi 90%, dan melihat bahwa kebijakan tersebut juga berpotensi peningkatan investasi di Provinsi Jambi.</p>
<p>Pengambil kebijakan ingin melihat apakah bauran kebijakan pembangunan infrastruktur hingga Rp. 2 Triliun dapat mengkompensasi dampak negatif rencana kenaikan Upah Minimum Provinsi sebesar Rp. 5 Juta, terhadap daya tarik investasi di Provinsi Kepulauan Riau.</p>	<p>Dalam aplikasi simulasi, pengambil kebijakan dapat menginput tambahan RP. 2 Triliun untuk variabel Belanja Modal dan mengubah input Upah Minimum Provinsi menjadi Rp. 5 Juta di Provinsi Kepulauan Riau. Selanjutnya nilai hasil prediksi investasi dapat dibandingkan dengan nilai investasi sebelumnya untuk melihat apakah bauran tersebut berdampak positif atau negatif.</p>

Lampiran II - *Prototype* Dashboard Investasi Regional (Tableau)



Deskripsi Dashboard:

1. Fitur:

- *Dashboard* memiliki tampilan yang menarik, mudah dipahami dan dapat memberikan informasi secara efektif dan efisien;
- *Dashboard* dapat diakses menggunakan laptop, pc atau *smartphone*;
- *Dashboard* menunjukkan gambaran secara umum faktor-faktor yang mempengaruhi investasi regional di tingkat provinsi;
- *Dashboard* Investasi Regional Indonesia menampilkan data profil regional dalam bentuk peta, *bar chart*, *line chart* dan *treemap* yang interaktif.
- Data yang ditampilkan dapat disesuaikan sesuai dengan provinsi dan tahun yang dipilih untuk menampilkan data secara keseluruhan atau sebagian.

2. Potensi Pemanfaatan :

a. Pertimbangan dalam pengambilan kebijakan, antara lain dengan:

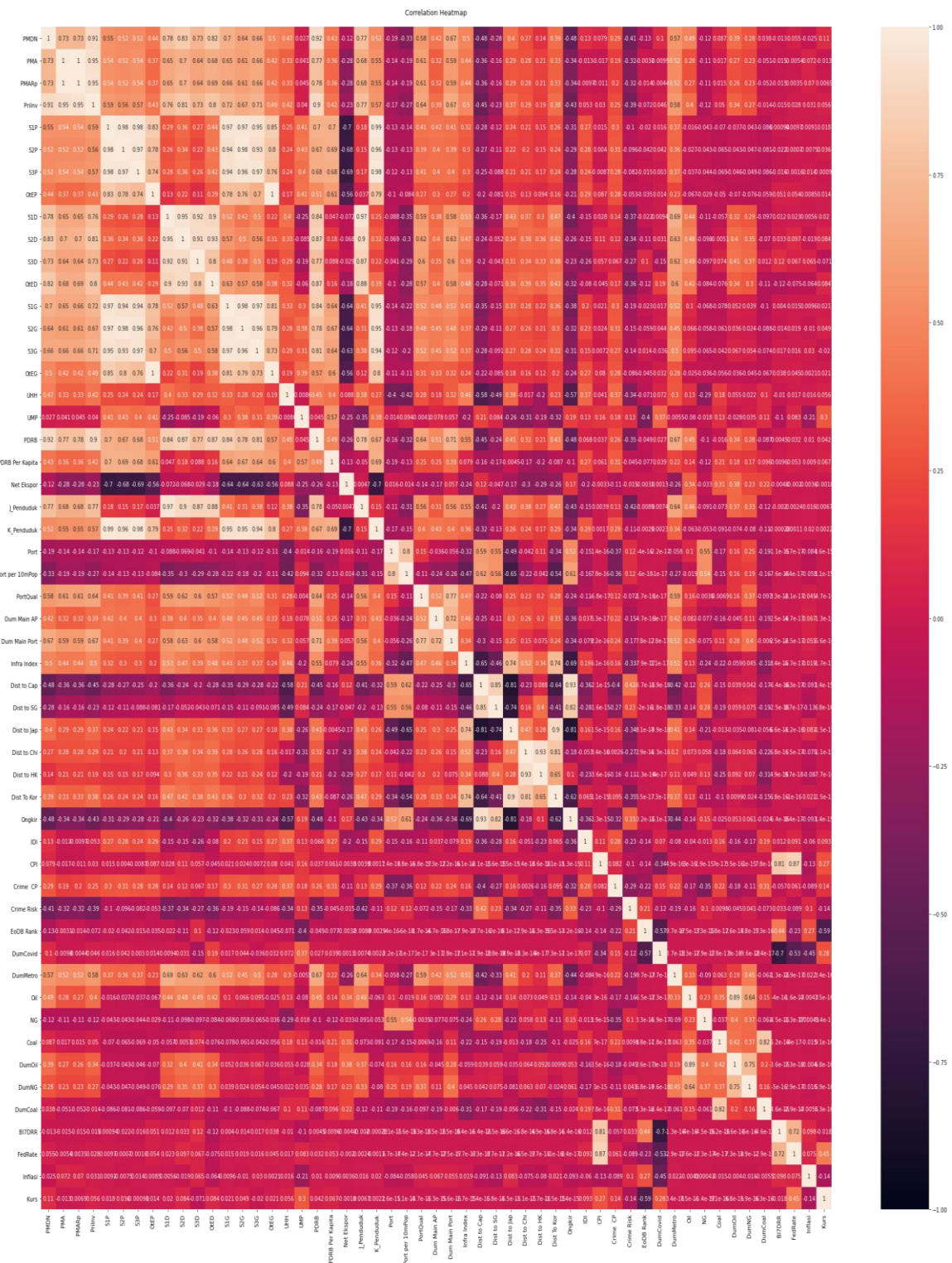
- Mengetahui gambaran secara umum profil investasi setiap regional di Indonesia
- Mengetahui variabel atau faktor yang mengalami peningkatan atau penurunan baik secara signifikan maupun tidak signifikan secara cepat dan dapat segera ditindak lanjuti.
- Mengetahui tren belanja pegawai, barang modal dan variabel lainnya yang setiap tahunnya mengalami peningkatan atau penurunan (fluktuasi),
- Melakukan analisis atau langkah lebih lanjut terhadap profil investasi setiap daerah/regional untuk menjadi bahan evaluasi peningkatan faktor-faktor penunjang investasi di daerah tersebut.
- Mengetahui sektor-sektor investasi yang masih berpotensi untuk lebih ditingkatkan dalam upaya pengembangan investasi yang selaras dengan keunggulan setiap daerah.
- serta manfaat lainnya yang dapat diperoleh oleh pengambil kebijakan.

b. Meningkatkan daya tarik Investor dalam melakukan penanaman modal

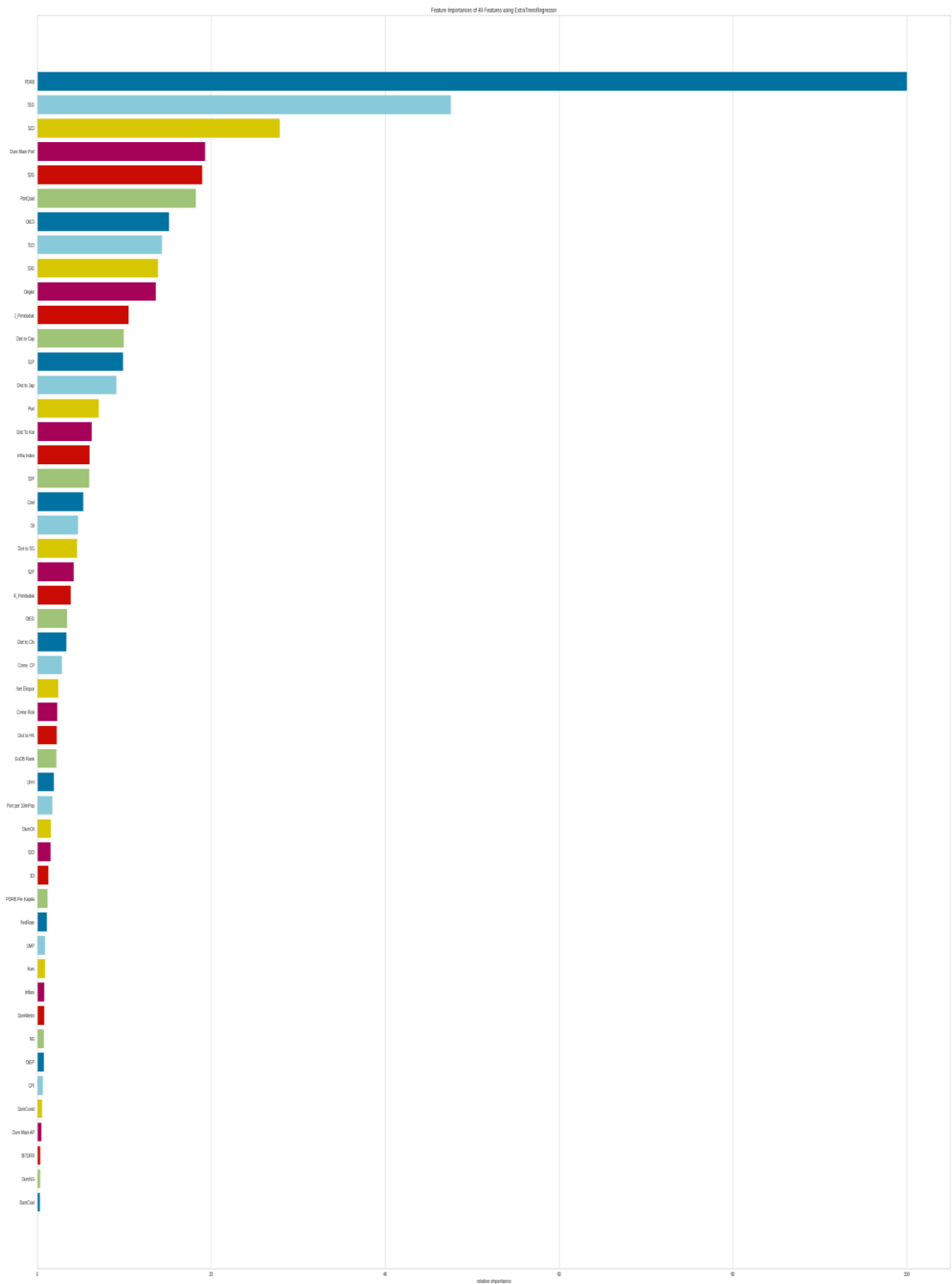
- Investor dapat mengetahui wilayah di Indonesia yang potensial untuk penanaman modal atau investasi dilihat dari sisi belanja pemerintah daerah, letak geografis yang strategis dan sektor-sektor investasi yang menjadi unggulan di wilayah tersebut.
- Investor dapat mengetahui karakteristik setiap daerah di Indonesia berdasarkan jumlah penduduk, kepadatan penduduk, usia harapan hidup, upah minimum pekerja, persentase tindak pidana yang berhasil diselesaikan, dan variabel lainnya.

3. Link Dashboard Tableau : <https://s.id/DashboardDDAC>

Lampiran III - *Heatmap* Koefisien Korelasi



Lampiran IV - Feature Importance Preliminary Model Machine Learning (Extra Trees Regressor)



Lampiran V - *Metadata / Deskripsi Variabel Dalam Model Machine Learning*

Nama Variabel	Deskripsi	Satuan	Sumber Data	Rentang Waktu
PrInv	Jumlah PMDN dan PMA (Private Investment)	Rupiah	BKPM (Diolah)	2016-2020
51G	Belanja Pegawai Gabungan Pusat & Daerah	Rupiah	DJPb & DJPK	2016-2020
52G	Belanja Barang Gabungan Pusat & Daerah	Rupiah	DJPb & DJPK	2016-2020
53G	Belanja Modal Gabungan Pusat & Daerah	Rupiah	DJPb & DJPK	2016-2020
OtEG	Belanja Lainnya Selain Transfer Gabungan Pusat & Daerah	Rupiah	DJPb & DJPK	2016-2020
UHH	Umur Harapan Hidup / Life Expectancy	Jumlah Tahun	BPS	2016-2020
UMP-1	Upah Minimum Provinsi (UMP) di tahun sebelumnya	Rupiah	BPS	2015-2019
PDRB-1	PDRB atas harga berlaku di tahun sebelumnya	Rupiah	BPS	2015-2019
J_Pen	Jumlah Penduduk Hasil Proyeksi Menurut Provinsi dan Jenis Kelamin (Ribu Jiwa)	Jiwa	BPS	2016-2020
K_Pen	Kepadatan Penduduk dengan rumus jumlah penduduk (jiwa) dibagi dengan luas wilayah (km ²)	jiwa/km ²	BPS	2016-2020
PortQ	Skor Kualitas Infrastruktur Pelabuhan [Adjustment Jabar Banten ikut Jakarta, dan DIY ikut Jateng]	Skor 0-1	Indriastiwi, F. (2017)	data 2017 (diasumsikan konstan)

Nama Variabel	Deskripsi	Satuan	Sumber Data	Rentang Waktu
DMAP	Dummy Provinsi dengan Bandara Utama menurut BPS [Adjustment Pada Bandara Soekarno Hatta yang Dihitung di Banten dan DKI Jakarta]	1 = Provinsi Dengan Bandara Utama 0 = Provinsi Tanpa Bandara Utama	BPS (Diolah)	data 2021 (diasumsikan konstan)
Infralx	Indeks Komposit Pembangunan Infrastruktur, dikonversi dari status kondisi menjadi angka	2 = Memadai 1 = Cukup Memadai 0 = Kurang Memadai	Faradis dan Uswatun (2018)	data 2018 (diasumsikan konstan)
DistCp	Jarak Ibukota Provinsi ke Ibukota Indonesia (Jakarta) dihitung dengan Rumus Vincenty	Kilometer	Simplemaps (Diolah)	permanen
DistSG	Jarak Ibukota Provinsi ke Ibukota Singapura dihitung dengan Rumus Vincenty	Kilometer	Simplemaps (Diolah)	permanen
CrmCP	Persentase Penyelesaian Tindak Pidana (Crime Clearance Rate) merupakan persentase tindak pidana yang berhasil diselesaikan dibandingkan jumlah tindak pidana yang dilaporkan [Backcasting Kaltara 2016-2017]	Persentase	BPS	2016-2020
CRMRI	Risiko Penduduk Terkena Tindak Pidana (Per 100.000 Penduduk) mendeskripsikan rata-rata jumlah penduduk yang terkena tindak pidana per 100.000 penduduk. Adjustment Backcasting Untuk Sulbar di Tahun 2016	Jumlah Penduduk terkena Tindak Pidana per 100.000 Penduduk	BPS	2016-2020

Nama Variabel	Deskripsi	Satuan	Sumber Data	Rentang Waktu
Dmet	Variabel Dummy Provinsi yang memiliki Metropolitan berdasarkan Definisi PP 26/2008 Juncto PP 17/2017 (Jakarta, Surabaya, Bandung, Medan, Palembang, Makassar, Batam, Bandar Lampung)	1 = Provinsi dengan Kota Metropolitan 0 = Provinsi tanpa Kota Metropolitan	BPS, PP 26/2008 & PP 17/2017	data 2008&2017 (diasumsikan konstan)
DOil	Dummy Provinsi Kaya Cadangan Minyak (Proven Reserve) dengan Threshold >100 MMBSTB	1 = Provinsi Kaya 0 = Provinsi Tidak Kaya	Kementerian ESDM (Diolah)	data 2021 (diasumsikan konstan)
DNG	Dummy Provinsi Kaya Cadangan Gas (Proven Reserve) dengan Threshold > 900 Km3	1 = Provinsi Kaya 0 = Provinsi Tidak Kaya	Kementerian ESDM (Diolah)	data 2021 (diasumsikan konstan)
DCoal	Dummy Provinsi Kaya Cadangan Batu Bara (Proven Reserve) dengan Threshold > 3.000 Juta Ton	1 = Provinsi Kaya 0 = Provinsi Tidak Kaya	Kementerian ESDM (Diolah)	data 2021 (diasumsikan konstan)

Lampiran VI - *Metadata / Deskripsi Variabel Dalam Model Regresi OLS*

Nama Variabel	Deskripsi	Satuan	Sumber Data	Rentang Waktu
PrInv	Jumlah PMDN dan PMA (Private Investment)	Rupiah	BKPM (Diolah)	Akumulasi 5 Tahun (2016-2020)
53G	Belanja Modal Gabungan Pusat & Daerah	Rupiah	DJPb & DJPK	Akumulasi 5 Tahun (2016-2020)
UHH	Umur Harapan Hidup / Life Expectancy	Jumlah Tahun	BPS	Rata-rata 5 Tahun (2016-2020)
UMP-1	Upah Minimum Provinsi (UMP) di tahun sebelumnya	Rupiah	BPS	Rata-rata 5 Tahun (2016-2020)
J_Pen	Jumlah Penduduk Hasil Proyeksi Menurut Provinsi dan Jenis Kelamin (Ribu Jiwa)	Jiwa	BPS	Rata-rata 5 Tahun (2016-2020)
PortQ	Skor Kualitas Infrastruktur Pelabuhan [Adjustment Jabar Banten ikut Jakarta, dan DIY ikut Jateng]	Skor 0-1	Indriastiwi, F. (2017)	data 2017 (diasumsikan konstan)
DMAP	Dummy Provinsi dengan Bandara Utama menurut BPS [Adjustment Pada Bandara Soekarno Hatta yang Dihitung di Banten dan DKI Jakarta]	1 = Provinsi Dengan Bandara Utama 0 = Provinsi Tanpa Bandara Utama	BPS (Diolah)	data 2021 (diasumsikan konstan)
Infralx	Indeks Komposit Pembangunan Infrastruktur, dikonversi dari status kondisi menjadi angka	2 = Memadai 1 = Cukup Memadai 0 = Kurang Memadai	Faradis dan Uswatun (2018)	data 2018 (diasumsikan konstan)

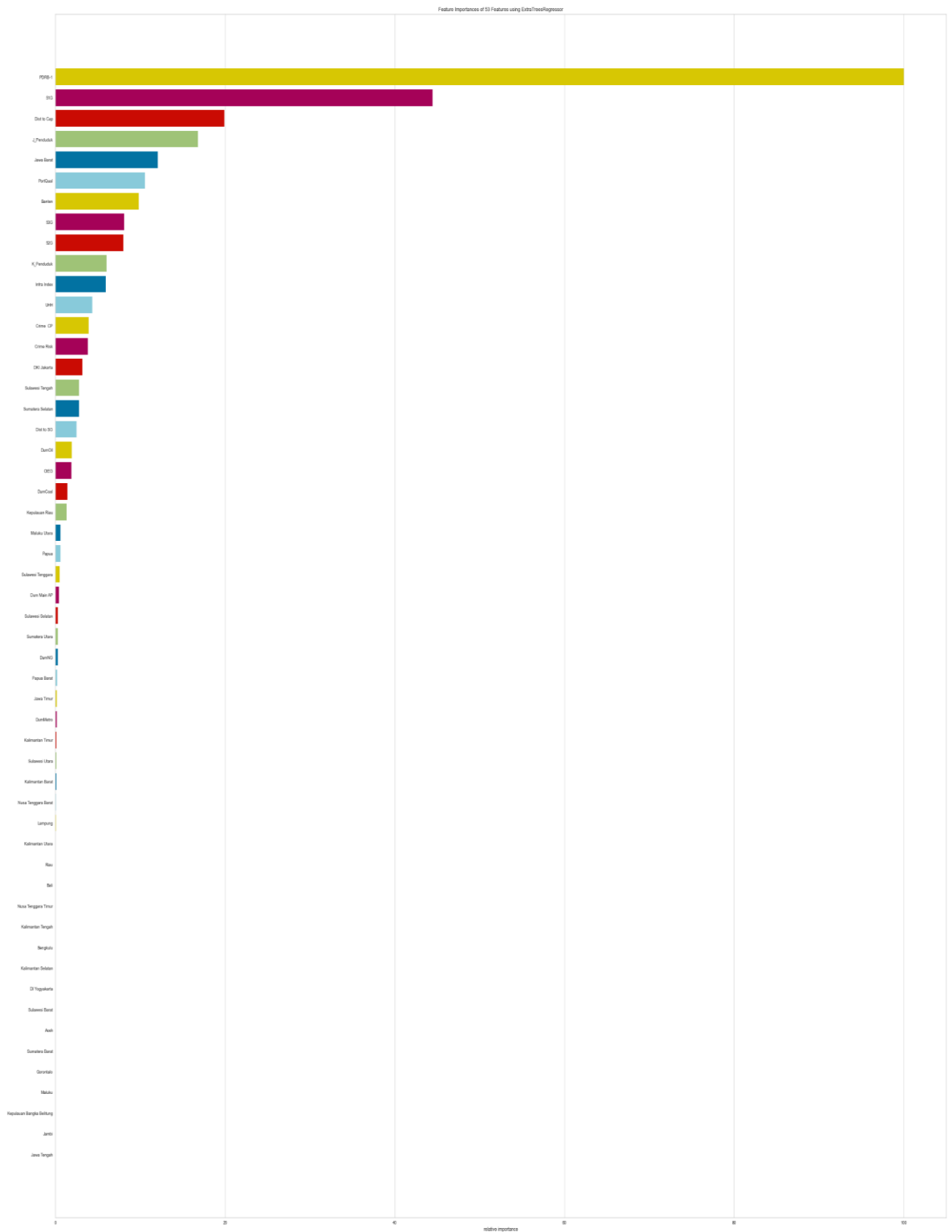
Nama Variabel	Deskripsi	Satuan	Sumber Data	Rentang Waktu
DistSG	Jarak Ibukota Provinsi ke Ibukota Singapura dihitung dengan Rumus Vincenty	Kilometer	Simplemaps (Diolah)	permanen
CrmCP	Persentase Penyelesaian Tindak Pidana (Crime Clearance Rate) merupakan persentase tindak pidana yang berhasil diselesaikan dibandingkan jumlah tindak pidana yang dilaporkan [Backcasting Kaltara 2016-2017]	Persentase	BPS	Rata-rata 5 Tahun (2016-2020)
CRMRI	Risiko Penduduk Terkena Tindak Pidana (Per 100.000 Penduduk) mendeskripsikan rata-rata jumlah penduduk yang terkena tindak pidana per 100.000 penduduk. Adjustment Backcasting Untuk Sulbar di Tahun 2016	Jumlah Penduduk terkena Tindak Pidana per 100.000 Penduduk	BPS	Rata-rata 5 Tahun (2016-2020)
DOil	Dummy Provinsi Kaya Cadangan Minyak (Proven Reserve) dengan Threshold > 100 MMBSTB	1 = Provinsi Kaya 0 = Provinsi Tidak Kaya	Kementerian ESDM (Diolah)	data 2021 (diasumsikan konstan)
DNG	Dummy Provinsi Kaya Cadangan Gas (Proven Reserve) dengan Threshold > 900 Km3	1 = Provinsi Kaya 0 = Provinsi Tidak Kaya	Kementerian ESDM (Diolah)	data 2021 (diasumsikan konstan)
DCoal	Dummy Provinsi Kaya Cadangan Batu Bara (Proven Reserve) dengan Threshold > 3.000 Juta Ton	1 = Provinsi Kaya 0 = Provinsi Tidak Kaya	Kementerian ESDM (Diolah)	data 2021 (diasumsikan konstan)

Lampiran VII - Komparasi Kinerja Model *Machine Learning* dalam Prediksi Investasi

Kode Model	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
et	Extra Trees Regressor	5.1117E+12	6.8923E+25	7.3741E+12	0.90	0.55	0.52	0.04
catboost	CatBoost Regressor	5.5389E+12	7.8478E+25	8.0337E+12	0.88	0.60	0.65	0.75
rf	Random Forest Regressor	6.0784E+12	9.42E+25	8.8469E+12	0.86	0.59	0.68	0.63
gbr	Gradient Boosting Regressor	6.0246E+12	9.5705E+25	8.6649E+12	0.85	0.61	0.69	0.02
xgboost	Extreme Gradient Boosting	5.9911E+12	1.0618E+26	9.1372E+12	0.84	0.55	0.52	0.15
omp	Orthogonal Matching Pursuit	6.5267E+12	8.7631E+25	9.0304E+12	0.84	0.75	1.00	0.01
lightgbm	Light Gradient Boosting Machine	7.386E+12	1.1112E+26	9.9526E+12	0.82	0.87	0.94	0.11
dt	Decision Tree Regressor	6.9014E+12	1.2034E+26	9.904E+12	0.80	0.68	0.66	0.00
ada	AdaBoost Regressor	8.0185E+12	1.2041E+26	1.0252E+13	0.78	0.92	1.65	0.02

Kode Model	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
knn	K Neighbors Regressor	9.6973E+12	2.0865E+26	1.3643E+13	0.70	0.84	1.19	0.01
huber	Huber Regressor	9.3794E+12	2.5935E+26	1.495E+13	0.63	0.74	0.67	0.01
llar	Lasso Least Angle Regression	7.5885E+12	3.1E+26	1.3797E+13	0.62	0.69	0.69	0.01
lasso	Lasso Regression	8.361E+12	5.1171E+26	1.6476E+13	0.40	0.72	0.72	0.01
br	Bayesian Ridge	1.0285E+13	5.0215E+26	1.7133E+13	0.36	0.82	1.25	0.02
par	Passive Aggressive Regressor	1.5623E+13	1.0485E+27	2.7503E+13	-0.26	1.10	1.07	0.00
lr	Linear Regression	1.2954E+13	1.6175E+27	2.5887E+13	-0.87	0.84	1.25	0.35
ridge	Ridge Regression	1.1657E+13	2.031E+27	2.7873E+13	-1.25	0.69	0.76	0.00
en	Elastic Net	1.4365E+13	2.736E+27	3.1774E+13	-2.04	0.97	1.45	0.00

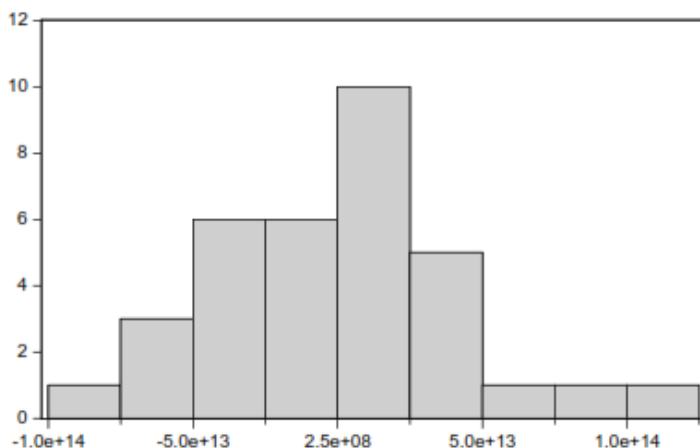
Lampiran VIII - Feature Importance Model Machine Learning Final (Tuned Extra Trees Regressor)



Lampiran IX - Hasil Regresi OLS Untuk Analisis Determinan Investasi

Dependent Variable: PRIINV				
Method: Least Squares				
Date: 03/10/22 Time: 14:58				
Sample: 1 34				
Included observations: 34				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.53E+13	3.47E+14	0.188300	0.8525
_53G	0.501292	0.196676	2.548816	0.0191
UHH	-3.23E+12	5.22E+12	-0.618309	0.5433
UMP_1	47800760	35773466	1.336207	0.1965
J_PEN	5648366.	1732711.	3.259843	0.0039
DISTSG	-8.63E+09	1.38E+10	-0.625971	0.5384
PORTQ	1.44E+14	6.64E+13	2.167849	0.0424
DMAP	-3.17E+13	3.44E+13	-0.921462	0.3678
INFRAIX	3.22E+12	2.24E+13	0.143701	0.8872
CRMCP	1.69E+12	1.29E+12	1.311743	0.2045
CRMRI	-2.74E+11	1.67E+11	-1.640310	0.1166
DOIL	9.46E+13	4.53E+13	2.089578	0.0496
DNG	-4.51E+13	3.97E+13	-1.136928	0.2690
DCOAL	-2.53E+13	4.16E+13	-0.607158	0.5506
R-squared	0.900763	Mean dependent var	1.07E+14	
Adjusted R-squared	0.836259	S.D. dependent var	1.36E+14	
S.E. of regression	5.51E+13	Akaike info criterion	66.41248	
Sum squared resid	6.08E+28	Schwarz criterion	67.04098	
Log likelihood	-1115.012	Hannan-Quinn criter.	66.62682	
F-statistic	13.96445	Durbin-Watson stat	2.490439	
Prob(F-statistic)	0.000000			

Lampiran X - Hasil Uji Normalitas



Series: Residuals	
Sample 1 34	
Observations 34	
Mean	-0.058824
Median	8.80e+12
Maximum	1.02e+14
Minimum	-8.09e+13
Std. Dev.	4.29e+13
Skewness	0.313975
Kurtosis	2.740745
Jarque-Bera	0.653842
Probability	0.721141

Lampiran XI - Hasil Uji Multikolinearitas

Variance Inflation Factors			
Date: 03/10/22 Time: 14:40			
Sample: 1 34			
Included observations: 34			
Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	1.20E+29	1345.260	NA
_53G	0.038682	4.422229	3.013811
DUMCOAL	1.73E+27	2.849324	2.430306
DUMNG	1.57E+27	4.144288	3.169162
DUMOIL	2.05E+27	4.722280	3.750046
CRIME_CP	1.66E+24	75.01229	3.122728
CRIME_RISK	2.78E+22	9.601206	1.687357
DIST_TO_SG	1.90E+20	6.131365	1.782000
DUM_MAIN_AP	1.18E+27	2.333853	1.921996
INFRA_INDEX	5.03E+26	7.938196	2.957367
PORTQUAL	4.41E+27	7.208859	2.778393
J_PENDUDUK	3.00E+12	6.022731	3.996756
UHH	2.73E+25	1481.642	1.995050
UMP_1	1.28E+15	64.76677	2.556563

Lampiran XII - Hasil Uji Heteroskedastisitas

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
F-statistic	1.549951	Prob. F(13,20)	0.1834	
Obs*R-squared	17.06325	Prob. Chi-Square(13)	0.1964	
Scaled explained SS	5.138884	Prob. Chi-Square(13)	0.9720	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 03/10/22 Time: 14:41				
Sample: 1 34				
Included observations: 34				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.39E+28	1.37E+28	1.751208	0.0952
_53G	-1.12E+13	7.74E+12	-1.448109	0.1631
DUMCOAL	-2.31E+27	1.64E+27	-1.407516	0.1746
DUMNG	-1.59E+27	1.56E+27	-1.015472	0.3220
DUMOIL	8.72E+26	1.78E+27	0.489023	0.6301
CRIME_CP	6.56E+25	5.07E+25	1.292556	0.2109
CRIME_RISK	-4.24E+24	6.57E+24	-0.646168	0.5255
DIST_TO_SG	-4.60E+23	5.43E+23	-0.847351	0.4068
DUM_MAIN_AP	-1.75E+27	1.35E+27	-1.288946	0.2121
INFRA_INDEX	-6.91E+26	8.83E+26	-0.782595	0.4430
PORTQUAL	9.00E+27	2.62E+27	3.443290	0.0026
J_PENDUDUK	1.86E+19	6.82E+19	0.272987	0.7877
UHH	-3.58E+26	2.06E+26	-1.739159	0.0974
UMP_1	-3.24E+20	1.41E+21	-0.230195	0.8203
R-squared	0.501860	Mean dependent var	1.79E+27	
Adjusted R-squared	0.178069	S.D. dependent var	2.39E+27	
S.E. of regression	2.17E+27	Akaike info criterion	129.0210	
Sum squared resid	9.43E+55	Schwarz criterion	129.6495	
Log likelihood	-2179.357	Hannan-Quinn criter.	129.2353	
F-statistic	1.549951	Durbin-Watson stat	1.892755	
Prob(F-statistic)	0.183426			