



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

Muhamad Ameer Noor¹, Rivai Geraldin Batubara², Kartiko Cokro Sewoyo³,
Bana Ali Fikri⁴, Reghina Ardalova⁵

^{1, 2, 3, 4, 5}Direktorat Jenderal Perbendaharaan, Kementerian Keuangan, Jakarta

Alamat Korespondensi: ameer.noor@kemenkeu.go.id

Abstract

In line with the pro-investment policy of the Indonesian government, investment is one of determining keys to then economic growth. This paper aims to predict investment and identify investment determinants to provide input for the formulation of effective pro-investment policies. This study uses two models, namely machine learning and Ordinary Least Square (OLS) regression, using data from 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 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, Crime, 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 memengaruhi investasi dalam rangka memberikan masukan untuk formulasi kebijakan pro-investasi yang lebih efektif. Penelitian ini 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. 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, kriminalitas, dan sumber daya alam memiliki nilai *feature importance* yang tinggi dalam memprediksi investasi. Hasil model regresi *cross-section*



mengonfirmasi pengaruh signifikan dari belanja modal, kualitas pelabuhan, jumlah penduduk, dan sumber daya alam (SDA) minyak terhadap investasi di Indonesia.

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

JEL Classification: E22, C53, H50

PENDAHULUAN

Investasi memiliki peranan yang penting dalam perekonomian suatu negara, sebagaimana diteliti dalam berbagai literatur, diantaranya Anderson (1990), Blomström et al. (1996), Chow (1993), Gupta (2021), Shabbir et al. (2021), dan Suhendra & Anwar (2014).

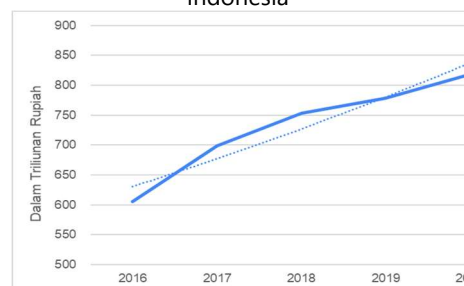
Pentingnya investasi bagi pertumbuhan ekonomi disadari oleh pemerintah Indonesia. Hal tersebut terbukti dengan berbagai kebijakan yang didorong untuk mempermudah investasi. Dari sisi kemudahan berbisnis (*Ease of Doing Business/EoDB*), Indonesia terus berbenah sebagaimana tercermin dalam peningkatan 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.

Gambar 1 menyajikan tren penanaman modal di Indonesia berdasarkan data dari Badan Koordinasi Penanaman Modal (2022). Nilai

penanaman modal yang merupakan gabungan dari Penanaman Modal Asing (PMA) dan Penanaman Modal Dalam Negeri (PMDN) terus meningkat dari Rp605 triliun di tahun 2016 menjadi Rp817 triliun di tahun 2020. Apabila dihitung dengan menggunakan metode *Compounded Annual Growth Rate* (CAGR), rata-rata kenaikan penanaman modal periode tahun 2016-2020 adalah 6,20%. Tren positif tersebut mengindikasikan bahwa kebijakan pro-investasi pemerintah cukup berhasil dalam meningkatkan daya tarik investasi.

Di sisi lain, sebagaimana dilansir oleh berbagai media seperti Kompas (2022) dan Bisnis (2022), pemerintah juga memberikan target investasi yang sangat tinggi, yakni sebesar Rp1.200 triliun di tahun 2022 dan Rp1.900 triliun di tahun 2023. Nilai tersebut mencerminkan kenaikan target sebesar

Gambar 1. Tren Penanaman Modal Di Indonesia



Sumber: BKPM (data diolah, 2022)

30% di tahun 2022 dari target sebesar Rp900 triliun di tahun 2021, serta kenaikan target sekitar 60% di tahun 2023. Dikaitkan dengan tren kenaikan investasi dalam CAGR sebesar 6,20%, kenaikan target yang relatif sangat tinggi tersebut mengimplikasikan bahwa pemerintah Indonesia harus memformulasikan berbagai kebijakan pro-investasi yang lebih efektif dalam meningkatkan daya tarik investasi.

Tantangan lainnya dalam kebijakan pro-investasi ke Indonesia, dapat dilihat dari investasi yang masih cenderung menumpuk di beberapa provinsi. Hal tersebut dapat dilihat pada Gambar 2, yang menunjukkan bahwa provinsi-provinsi dengan warna biru tua memiliki nilai penanaman modal di atas rata-rata nasional.

Gambar 2. Sebaran Penanaman Modal di Indonesia



Sumber: BKPM (data diolah, 2022)

Provinsi-provinsi dengan nilai investasi tinggi tersebut masih cenderung terkonsentrasi di Pulau Jawa. Sedangkan di Provinsi DI Yogyakarta masih relatif rendah. Ketidakmerataan tersebut menjadi tantangan sekaligus potensi dalam meningkatkan nilai penanaman modal di Indonesia, mengingat implikasinya bahwa berbagai provinsi di Indonesia terindikasi masih memiliki banyak

potensi yang belum dimanfaatkan untuk menarik penanam modal.

Sebagaimana dilansir oleh Media Indonesia (2022), Menteri Keuangan saat ini mengharapkan Direktorat Jenderal Perbendaharaan (DJPb) untuk menjalankan reformasi jilid kedua yang berfokus pada pengembangan data analitis dan kemampuan *intellectual sharpness* untuk melakukan analisis dan memberikan rekomendasi kebijakan. Salah satu upaya DJPb untuk mewujudkan harapan tersebut ditindaklanjuti dengan meluncurkan inisiatif baru melalui pelaksanaan tugas sebagai *Regional Chief Economist* (RCE).

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. Hal tersebut diharapkan dapat membantu untuk menjawab tantangan nasional mengenai akselerasi peningkatan penanaman modal dan pemerataan investasi antar provinsi.

Agar DJPb 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 memengaruhi 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.

Belum banyak penelitian terdahulu yang telah dilakukan dalam konteks analisis prediksi investasi. 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, 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 Fathia et al. (2021), Soekro & Widodo (2015), dan Suhendra & Anwar (2014) 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. Di sisi lain, studi determinan investasi regional di Rusia oleh Ledyeva (2009), menawarkan model

yang menarik untuk direplikasi, dengan menghasilkan determinan unik yang dapat memengaruhi 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

Konsep Investasi

Studi oleh Anderson (1990) menunjukkan bahwa alokasi investasi yang efisien dalam meningkatkan *output* ekonomi memiliki dampak yang besar terhadap pertumbuhan ekonomi, sedangkan alokasi yang tidak efisien 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 (1993) menemukan bahwa investasi berdampak positif terhadap ekonomi, meskipun di saat yang bersamaan ditemukan adanya potensi ekonomi yang hilang akibat alokasi investasi yang terlalu terpusat pada masa awal pembangunan Tiongkok.

Dalam konteks PMDN atau *Domestic Direct Investment* (DDI) dan 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 (2021) di Indonesia menunjukkan bahwa peningkatan modal atau investasi secara umum dapat mendorong pertumbuhan ekonomi.

Selain itu, literatur yang ditulis oleh Heijman & Van Ophem (2007), Moran et al. (2005), 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.

Machine Learning sebagai Metode Prediksi Investasi

Literatur mengenai prediksi investasi di Indonesia belum banyak ditemui. 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.

Di sisi lain, studi oleh Abd El-Aal et al. (2021) di Mesir, telah memanfaatkan model *machine learning* untuk prediksi investasi. Penelitian tersebut membandingkan tujuh model prediksi yang meliputi *gradient boosting*, *logistic regression*, *neural network*, *random forest*, *naive bayes*, *k-nearest neighbors*, dan *support vector machine*, hasil perbandingan tersebut menemukan bahwa model *gradient boosting* dapat menghasilkan kinerja terbaik untuk memprediksi nilai investasi, dengan nilai akurasi sebesar 87%.

Athey (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 sumber daya yang tersedia dalam setiap studi kasus. Di sisi lain, hal tersebut juga memiliki dampak negatif, yaitu 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.

Determinan Investasi Nasional

Ledyaeva (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*. Hasil penelitian menemukan bahwa faktor penting dari aliran FDI ke wilayah di Rusia meliputi 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 & Lensink (2001) menguji pengaruh ekonomi makro dari kebijakan fiskal pada sampel 30 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.

Ditinjau dari metode analisis yang digunakan, studi oleh Abd El-Aal et al. (2021) mengombinasikan model *machine learning gradient boosting* dan model regresi *Autoregressive Integrated Moving Average* (ARIMA), untuk

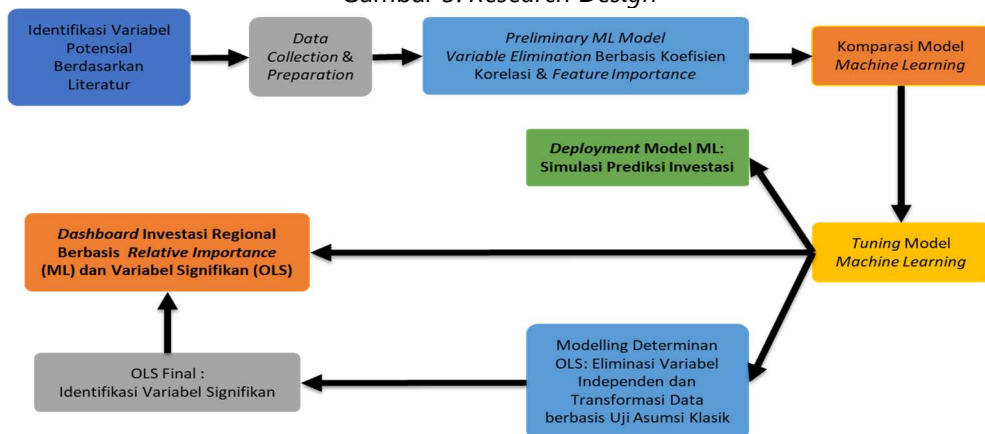
menentukan determinan investasi di Mesir. Hasil penelitian tersebut mengidentifikasi Indeks Pembangunan Manusia (IPM), jumlah populasi, Produk Domestik Bruto (PDB) per kapita, tingkat pinjaman, dan PDB sebagai determinan nilai investasi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain riset *archival* yang menggunakan *dataset* publik yang sudah tersedia. Data yang digunakan merupakan data sekunder yang diperoleh dari studi literatur dan beberapa institusi publik. Analisis data menggunakan teknik korelasi untuk menguji kemungkinan hubungan antar variabel, teknik eksperimental untuk menguji variabel-variabel baru yang berpotensi memiliki hubungan, serta teknik inferensial untuk menguji hubungan antar variabel dan implikasinya.

Secara umum, proses penelitian yang dilakukan dalam tulisan ini dapat

Gambar 3. *Research Design*



Sumber: data diolah (2022)

dijelaskan dalam diagram pada Gambar 3. Pertama, dilakukan identifikasi variabel-variabel yang potensial sebagai determinan investasi, berdasarkan penelitian-penelitian terdahulu. Data atau *proxy* data dari variabel-variabel potensial dicari dalam proses *data collection* dari berbagai sumber baik melalui proses manual maupun *data scrapping*. Selanjutnya, dilakukan analisis sederhana untuk mengeliminasi sebagian variabel independen yang berpotensi tidak signifikan dengan melihat nilai koefisien korelasi dan nilai *feature importance* dalam *preliminary machine learning model*. Setelah variabel-variabel independen final terpilih, dilakukan komparasi 18 model *machine learning*. Model dengan akurasi tertinggi dipilih dan dilakukan *hyperparameter tuning* lebih lanjut untuk meningkatkan akurasi prediksi. Model hasil *tuning* final digunakan sebagai dasar untuk *deployment* aplikasi simulasi dampak perubahan kebijakan terhadap investasi.

Selanjutnya model determinan dengan *Ordinary Least Square* (OLS) dikembangkan untuk melengkapi kesimpulan dari analisis variabel-variabel yang dapat memengaruhi tingkat investasi di suatu provinsi. Hasil analisis OLS dan *machine learning* digabungkan untuk menjadi dasar pengembangan *dashboard* investasi regional yang menyajikan data-data yang paling memengaruhi keputusan investor dalam menanamkan modal.

Analisis Prediksi Investasi dengan Machine Learning

Variabel investasi dapat direpresentasikan dalam dengan PMDN, PMA, atau Penanaman Modal Tetap Bruto (PMTB). Mengingat tujuan

utama dari penelitian ini adalah untuk dasar pertimbangan pemerintah dalam menarik investor melalui instrumen yang dapat diatur pemerintah seperti belanja modal infrastruktur, PMTB tidak digunakan sebagai variabel dependen karena memasukkan belanja modal pemerintah sebagai salah satu komponennya.

Selanjutnya, analisis korelasi dilakukan antara PMDN dan PMA pada periode data observasi tahun 2016-2020, untuk menentukan variabel dependen yang digunakan. Nilai koefisien korelasi antara PMA dan PMDN sebesar 0,73. Menurut Sugiyono (2013), nilai korelasi dalam rentang 0,6 sampai 0,79 menunjukkan korelasi yang kuat, sehingga penelitian ini menggunakan data gabungan PMA dan PMDN (*Prilnv*) sebagai variabel dependen.

Athey (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 sebagaimana dijabarkan dalam bab tinjauan literatur. Penjelasan lebih rinci mengenai 49 variabel independen tersebut disajikan dalam Lampiran I. Untuk menyederhanakan penyajian, 49 variabel independen potensial tersebut dikelompokkan menjadi 12 kelompok besar variabel yang dapat dijabarkan dalam persamaan berikut:

$$Prilnv = \beta XG_1 + \beta XG_2 + \beta XG_3 + \beta XG_4 + \beta XG_5 + \beta XG_6 + \beta XG_7 + \beta XG_8 + \beta XG_9 + \beta XG_{10} + \beta XG_{11} + \beta XG_{12}$$

Penjelasan dari masing-masing variabel tersebut adalah sebagai berikut:

PrInv: *Private investment* atau penanaman modal tetap swasta yang menggabungkan nilai PMA dan PMDN, dalam satuan rupiah.

β : Koefisien determinasi variabel independen.

XG_1 : Variabel kelompok belanja pemerintah konsolidasian yang terdiri dari variabel seluruh belanja pemerintah selain transfer, subsidi, dan pembayaran hutang.

XG_2 : Variabel kelompok ketenagakerjaan yang terdiri dari variabel upah minimum regional 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*/*BI7DRR* dan *fed rate*.

XG_5 : Variabel kelompok infrastruktur yang terdiri dari jumlah pelabuhan, jumlah pelabuhan per 10 juta penduduk, kualitas infrastruktur pelabuhan (Indriastiwi, 2018), *dummy* eksistensi bandara utama, *dummy* eksistensi pelabuhan

utama, dan indeks komposit pembangunan infrastruktur (Faradis & Afifah, 2019).

XG_6 : Variabel kelompok jarak ibukota provinsi ke pusat ekonomi yang terdiri dari variabel jarak ibukota provinsi ke ibukota Indonesia, jarak ibukota provinsi kelima 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).

XG_8 : Variabel *dummy* tahun pandemi Covid-19.

XG_9 : Variabel *dummy* eksistensi kota metropolitan dalam provinsi.

XG_{10} : Variabel kelompok kekayaan sumber daya alam yang terdiri dari cadangan minyak, cadangan gas, cadangan batu bara, *dummy* provinsi kaya minyak, *dummy* provinsi kaya gas, dan *dummy* provinsi kaya batu bara.

XG_{11} : Variabel inflasi.

XG_{12} : 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 $|0,2|$. Pemilihan nilai *threshold* tersebut mengacu pada kategorisasi oleh Sugiyono (2013), yang menjelaskan bahwa nilai korelasi lebih kecil dari 0,19 termasuk ke dalam kategori sangat rendah, sedangkan nilai korelasi dalam rentang 0,2 sampai 0,39 menunjukkan korelasi yang rendah. Koefisien korelasi keseluruhan yang terbentuk disajikan dalam bentuk *Heatmap* dari *library* Seaborn dalam Python yang disajikan dalam Lampiran I. Di samping itu, beberapa variabel independen juga dieliminasi berdasarkan keterkaitan dengan variabel independen lainnya atau normalitas arah hubungan korelasi dengan variabel dependen.

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 I. *Feature importance* turut dipertimbangkan dalam proses eliminasi variabel independen potensial

untuk mencegah tereliminasinya 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:

$$\begin{aligned} \text{PriInv} = & \beta_{51G} + \beta_{52G} + \beta_{53G} + \beta_{0tEG} \\ & + \beta_{UHH} + \beta_{UMP-1} + \beta_{PDRB-1} \\ & + \beta_{L_Pen} + \beta_{K_Pen} + \beta_{PortQ} + \\ & \beta_{DMAP} + \beta_{InfraIx} + \beta_{DistCp} + \\ & \beta_{DistSG} + \beta_{CrmCP} + \beta_{CrmRi} + \\ & \beta_{DMet} + \beta_{DOil} + \beta_{DNG} + \\ & \beta_{Dcoal} \end{aligned}$$

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

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 18 model prediksi *machine learning* yang tersedia dalam PyCaret. Kriteria yang digunakan untuk memilih model terbaik adalah model dengan R^2 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. Selain itu, analisis *feature importance* kembali digunakan untuk melihat variabel-variabel apa saja yang paling dominan dalam memengaruhi 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 lima 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 merata-ratakan nilai data selama lima 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 I.

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 pengolahan data dalam model tersebut disajikan dalam Lampiran II.

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). 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 *dataset* 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'. Dengan nilai *hyperparameter*

n_estimators sebesar 25, maka nilai prediksi *extra trees regressor* yang dihasilkan menggunakan nilai rata-rata prediksi dari 25 *decision trees regressor*. Sampel-sampel dari *decision trees* yang digunakan disajikan dalam Lampiran I.

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 I.

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 (2009) di Rusia dan variabel *PDB* dalam penelitian Fathia et al. (2021). Enam variabel tersebut, baik secara langsung maupun tidak, dapat memengaruhi 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, yang menjelaskan bahwa 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 (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 Badan Koordinasi Penanaman Modal (2022), pada periode tahun 2016-2020, 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* (*CrnCp*) merepresentasikan kinerja penegak hukum, sedangkan variabel persentase penduduk terkena tindak pidana (*CrnRi*) merepresentasikan tingkat keamanan di masyarakat. Signifikansi *relative importance* kedua variabel tersebut mengindikasikan bahwa risiko kelembagaan berupa tingkat keamanan di suatu provinsi juga dapat memengaruhi rasa keamanan penanam modal untuk berinvestasi di provinsi tersebut. Variabel serupa yang merepresentasikan risiko kelembagaan, yakni risiko politik dan risiko legislatif, dalam penelitian Ledyeva (2009) juga terbukti signifikan dalam memengaruhi 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 (2009) di Rusia melalui signifikansi variabel *dummy* Sakhalin (Provinsi Kaya SDA) dan variabel indek 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.

Selanjutnya, untuk mengidentifikasi apakah pengaruh suatu *feature* atau variabel independen, cenderung positif atau negatif terhadap variabel dependen, dilakukan analisis dengan menggunakan SHAP (*SHapley Additive exPlanations*) *values*. SHAP *values* menginterpretasikan dampak suatu *feature* dengan nilai tertentu terhadap nilai prediksi dibandingkan dengan nilai *baseline* prediksi. Hasil SHAP *values* lebih lengkap disajikan dalam Lampiran I.

Hasil SHAP *values* menunjukkan bahwa *features* yang cenderung memiliki pengaruh positif adalah PDRB (*PDRB-1*), belanja pemerintah (*51G*, *52G*, dan *53G*), jumlah penduduk (*J_Pen*), kepadatan penduduk (*K_Pen*), sumber daya alam (*DOil*, *DCoal*, dan *DNG*), kualitas pelabuhan (*PortQual*), indeks komposit infrastruktur (*InfraIx*), dan keberadaan bandara utama dalam suatu provinsi (*DMAP*). *Features* yang cenderung memiliki pengaruh negatif adalah jarak ke ibukota (*DistCp*) dan risiko kriminalitas (*CriRi*), sedangkan *features* lainnya tidak dapat diidentifikasi dengan pasti apakah pengaruhnya positif atau negatif.

Analisis Determinan Investasi

Model regresi OLS dan uji asumsi klasik untuk memastikan keandalan model dilakukan menggunakan aplikasi Eviews, dengan hasil yang disajikan

pada Lampiran III sampai dengan VI. Berdasarkan hasil regresi OLS untuk analisis determinan, didapati nilai koefisien determinasi (R^2) dari regresi OLS adalah 0,8362 yang berarti bahwa 83,62% variasi dari *PrIInv* dapat dijelaskan oleh seluruh variabel bebas. Sementara itu 16,38% variasi dari *PrIInv* 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.

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 *PrInv*.

Analisis determinan dilakukan dengan melihat signifikansi masing-masing variabel independen terhadap *PrInv*. Hal tersebut dapat dilakukan dengan uji t.

Dalam model determinan tersebut, variabel bebas yang berpengaruh signifikan secara individu terhadap variabel *PrInv* 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 & Lensink (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), yang menemukan

bahwa investasi perbaikan kualitas pelabuhan memengaruhi 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

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.

Dalam konteks prediksi investasi, model *extra trees regressor* tersebut menemukan bahwa faktor penentu penanaman modal meliputi variabel-variabel terkait *market size*, jarak terhadap pusat ekonomi, infrastruktur, kesehatan, kriminalitas, dan kekayaan sumber daya alam. Variabel independen yang berkaitan dengan *market size* juga memiliki *relative importance* yang paling dominan sebagai prediktor model *machine learning*, namun variabel tersebut cenderung bersifat *given* sehingga sulit diintervensi. Khusus untuk variabel-variabel terkait infrastruktur, kualitas pelabuhan (*PortQ*) merupakan variabel yang paling penting dalam meningkatkan daya tarik investasi di suatu provinsi, di antara tiga variabel independen terkait infrastruktur yang digunakan. Sementara itu, untuk variabel-variabel independen yang berkaitan dengan sumber daya alam ditemukan bahwa variabel kekayaan gas tidak memiliki pengaruh signifikan, sehingga mengindikasikan perlunya

evaluasi kebijakan yang berkaitan dengan investasi gas di Indonesia.

Dalam konteks determinan investasi, hasil dari regresi OLS sejalan dengan hasil model prediksi *machine learning*. Variabel-variabel belanja modal pemerintah (53G), kualitas pelabuhan (*PortQ*), jumlah penduduk (*J_Pen*) dan provinsi dengan cadangan minyak besar (*DOil*) memiliki pengaruh yang signifikan terhadap investasi.

Dalam membuat kebijakan untuk meningkatkan investasi di Indonesia, pemerintah, khususnya Kementerian Koordinator Bidang Kemaritiman dan Investasi, Kementerian Koordinator Bidang Perekonomian, Badan Koordinasi Penanaman Modal, Badan Perencanaan Pembangunan Nasional (BAPPENAS), Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR), Kementerian Keuangan, Kementerian Hukum dan Hak Asasi Manusia, Kementerian Kesehatan, serta Pemerintah daerah (Pemda) yang menjadi tujuan investasi, dapat mempertimbangkan untuk membuat kebijakan yang berkaitan dengan variabel-variabel prediktor dengan *feature importance* tinggi dalam model *machine learning* dan variabel-variabel determinan yang signifikan dalam model regresi OLS. 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.

Dalam konteks yang lebih teknis operasional, model prediksi *machine learning* dalam penelitian ini dapat

dilakukan *deployment* sebagai *tools* simulasi untuk pengambil kebijakan, khususnya dari kementerian/badan yang berfokus pada peningkatan investasi di Indonesia. Simulasi tersebut dapat membantu pengambil kebijakan untuk melihat proyeksi dampak kebijakan yang akan ditempuh, terhadap kenaikan investasi pada 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 I. *Dashboard* tersebut dapat mendukung pemerintah, khususnya Badan Koordinasi Penanam Modal, Unit Vertikal DJPb, dan Dinas Penanaman Modal Daerah 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, Kementerian Keuangan, 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,

seperti Badan Perencanaan dan Pembangunan Daerah dan Dinas Penanaman 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 antara penanam modal asing dan domestik.

Dalam penelitian ini, data yang digunakan untuk pembentukan model *machine learning extra trees regressor* dan analisis regresi OLS belum dilakukan normalisasi atau standarisasi data. Hal tersebut belum dilakukan dengan pertimbangan untuk menyederhanakan proses *modelling* dan *deployment*. Pada penelitian selanjutnya, dapat dilakukan proses normalisasi atau standarisasi data dengan menggunakan metode *simple feature scaling*, *min-max* atau *z-score*, untuk menyamakan satuan data atau skala data sehingga berpotensi menghasilkan model yang lebih akurat.

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.

REFERENSI

- Abd El-Aal, M. F., Algarni, A., Fayomi, A., Abdul Rahman, R., & Alrashidi, K. (2021). Forecasting foreign direct investment inflow to Egypt and determinates: Using machine learning algorithms and ARIMA model. *Journal of Advanced Transportation*, 1-7. <https://doi.org/10.1155/2021/9614101>
- Anderson, D. (1990). Investment and economic growth. *World Development*, 18(8), 1057–1079. [https://doi.org/10.1016/0305-750X\(90\)90088-F](https://doi.org/10.1016/0305-750X(90)90088-F)
- Athey, S. (2018). The impact of machine learning on economics. *The Economics of Artificial Intelligence*, 507–552. <https://doi.org/10.7208/chicago/9780226613475.003.0021>
- Badan Koordinasi Penanaman Modal. (2022). *Data perkembangan realisasi penanaman modal*. https://nswi.bkpm.go.id/data_statistik
- Bisnis. (2022). *Target fantastis! Jokowi*

- patok investasi Rp1.900 triliun pada 2023. <https://ekonomi.bisnis.com/read/20220216/9/1501289/target-fantastis-jokowi-patok-investasi-rp1900-triliun-pada-2023>
- Blomström, M., Lipsey, R. E., & Zejan, M. (1996). Is fixed investment the key to economic growth? *Quarterly Journal of Economics*, 111(1), 269–276. <https://doi.org/10.2307/2946665>
- Budiono, S., & Purba, J. T. (2019). Data panel model solution in forecasting investments through energy electricity and government policy in Indonesia. *Proceedings of the International Conference on Industrial Engineering and Operations Management, January*, 1132–1132.
- Chow, G. C. (1993). Capital formation and economic growth in China. *Quarterly Journal of Economics*, 108(3), 809–842. <https://doi.org/10.2307/2118409>
- Faradis, R., & Afifah, U. N. (2019). Indeks komposit pembangunan infrastruktur provinsi-provinsi di Indonesia. *Jurnal Ekonomi dan Pembangunan Indonesia*, 20(1), 33–55. <https://doi.org/10.21002/jepi.v20i1.1108>
- Fathia, N., Silvia, V., & Gunawan, E. (2021). Analysis of foreign investment determinants in Indonesia. *Economics Development Analysis Journal*, 3, 338–350.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Gupta, K. (2021). The importance of financial liberalisation for economic growth: The case of Indonesia. *Bulletin of Indonesian Economic Studies*, 57(2), 175–201. <https://doi.org/10.1080/00074918.2020.1747596>
- Heijman, W., & Van Ophem, J. (2007). Abatement of commuting's negative externalities by regional investment in houses and buildings. *Regional Externalities*, 245–254. https://doi.org/10.1007/978-3-540-35484-0_12
- Hermes, N. & Lensink, R. (2001). Fiscal policy and private investment in less developed countries. *WIDER Discussion Paper, No. 2001/32, The United Nations University World Institute for Development Economics Research (UNU_WIDER), Helsinki*.
- Indriastiwi, F. (2018). Identifikasi fasilitas 24 pelabuhan di Indonesia menggunakan analisis cluster dan analysis hierarchy process. *Jurnal Penelitian Transportasi Laut*, 19(1), 25–39. <https://doi.org/10.25104/transla.v19i1.322>
- Kompas. (2022). *Target investasi 2022*. <https://www.kompas.id/baca/opini/2022/01/18/target-investasi-2022>
- Ledyaeva, S. (2009). Spatial econometric analysis of foreign direct investment determinants in Russian regions. *World Economy*, 32(4), 643–666. <https://doi.org/10.1111/j.1467->

- 9701.2008.01145.x
- Media Indonesia. (2022). *Regional chief economist dukung reformasi jilid II DJPB*.
<https://mediaindonesia.com/ekonomi/477914/regional-chief-economist-dukung-reformasi-jilid-ii-djpb>
- Moran, T., Graham, E., & Blomström, M. (2005). *Does foreign direct investment promote development?*. Washington DC.
- 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.
<https://doi.org/10.1021/acsomega.0c05846>
- 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.
<https://doi.org/10.14710/jwl.8.3.230-246>
- 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.
<https://doi.org/10.1080/19186444.2020.1858676>
- Soekro, S., & Widodo, T. (2015). Mapping and determinants of intra-ASEAN foreign direct investment (FDI): Indonesia case study. *Working Paper Bank Indonesia*.
- Sugiyono, D. (2013). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*. Bandung: Alfabeta.
- 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), 128–133.
<https://doi.org/10.7603/s40706-014-0028-4>
- 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.
<https://doi.org/10.1177/0974930616667875>
- 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 - Daftar Tautan

1. Tautan *Deployment* Aplikasi Simulasi Prediksi Investasi Provinsi:
<https://bit.ly/SEDAPDeploy>
2. Tautan *Prototype Dashboard* Investasi Regional (Tableau):
<https://s.id/DashboardDDAC>
3. Tautan Metadata / Deskripsi Variabel dalam Model *Preliminary*:
<https://bit.ly/SEDAPDataPreliminary>
4. Tautan *Heatmap* Koefisien Korelasi:
<https://bit.ly/SEDAPHeatmapKorelasi>
5. Tautan *Feature Importance Preliminary Model Machine Learning (Extra Trees Regressor)*:
<https://bit.ly/SEDAPFeaturePreliminary>
6. Tautan Metadata / Deskripsi Variabel dalam Model *Machine Learning*:
<https://bit.ly/SEDAPDataMachineLearning>
7. Tautan Metadata / Deskripsi Variabel dalam Model Regresi OLS:
<https://bit.ly/SEDAPDataOLS>
8. Sampel *Decision Trees Regressor* yang Digunakan dalam Model Final *Extra Trees Regressor*:
<https://bit.ly/SEDAPSampelTrees>
9. Tautan *Feature Importance Model Machine Learning Final (Tuned Extra Trees Regressor)*:
<https://bit.ly/SEDAPFeatureFinal>
10. Tautan *SHAP Values Model Machine Learning Final (Tuned Extra Trees Regressor)*:
<https://bit.ly/SEDAPSHAPFinal>

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

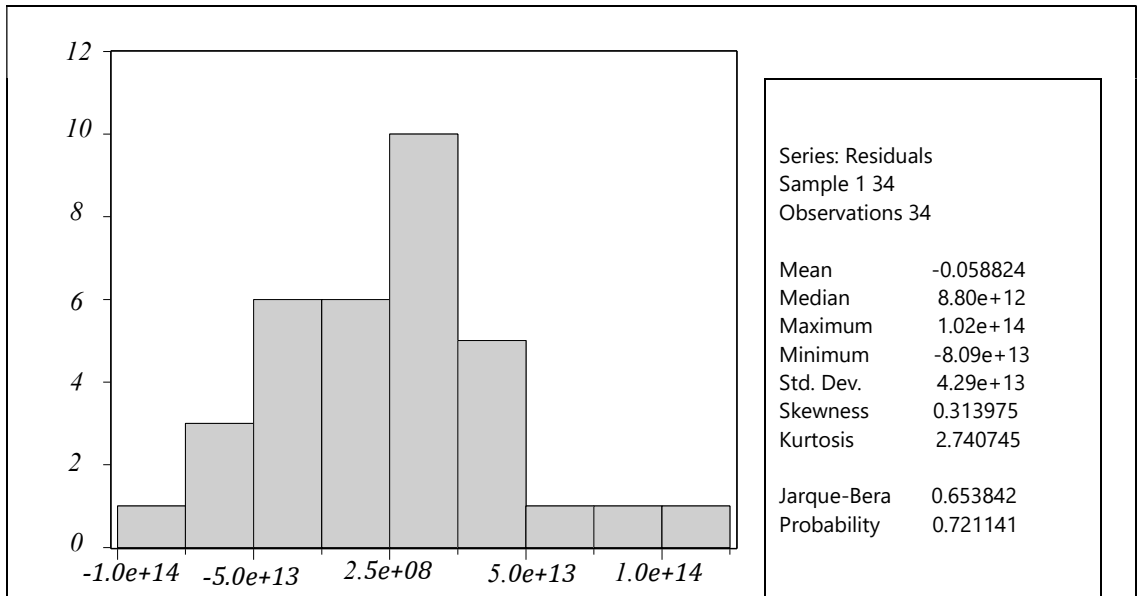
Kode Model	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
et	Extra Trees Regressor	5.1117E+12	6.8923E+25	7.3741E+12	0.90	0.55	0.52
catboost	CatBoost Regressor	5.5389E+12	7.8478E+25	8.0337E+12	0.88	0.60	0.65
rf	Random Forest Regressor	6.0784E+12	9.42E+25	8.8469E+12	0.86	0.59	0.68
gbr	Gradient Boosting Regressor	6.0246E+12	9.5705E+25	8.6649E+12	0.85	0.61	0.69
xgboost	Extreme Gradient Boosting	5.9911E+12	1.0618E+26	9.1372E+12	0.84	0.55	0.52
omp	Orthogonal Matching Pursuit	6.5267E+12	8.7631E+25	9.0304E+12	0.84	0.75	1.00
lightgbm	Light Gradient Boosting Machine	7.386E+12	1.1112E+26	9.9526E+12	0.82	0.87	0.94
dt	Decision Tree Regressor	6.9014E+12	1.2034E+26	9.904E+12	0.80	0.68	0.66
ada	AdaBoost Regressor	8.0185E+12	1.2041E+26	1.0252E+13	0.78	0.92	1.65
knn	K Neighbors Regressor	9.6973E+12	2.0865E+26	1.3643E+13	0.70	0.84	1.19
huber	Huber Regressor	9.3794E+12	2.5935E+26	1.495E+13	0.63	0.74	0.67

Kode Model	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
llar	Lasso Least Angle Regression	7.5885E+12	3.1E+26	1.3797E+13	0.62	0.69	0.69
lasso	Lasso Regression	8.361E+12	5.1171E+26	1.6476E+13	0.40	0.72	0.72
br	Bayesian Ridge	1.0285E+13	5.0215E+26	1.7133E+13	0.36	0.82	1.25
par	Passive Aggressive Regressor	1.5623E+13	1.0485E+27	2.7503E+13	- 0.26	1.10	1.07
lr	Linear Regression	1.2954E+13	1.6175E+27	2.5887E+13	- 0.87	0.84	1.25
ridge	Ridge Regression	1.1657E+13	2.031E+27	2.7873E+13	- 1.25	0.69	0.76
en	Elastic Net	1.4365E+13	2.736E+27	3.1774E+13	- 2.04	0.97	1.45

Lampiran III - Hasil Regresi OLS Untuk Analisis Determinan Investasi

Dependent Variable: PRIINV Method: Least Squares				
Date: 03/10/22 Time: 14:39				
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
DUMCOAL	-2.53E+13	4.16E+13	-0.607158	0.5506
DUMNG	-4.51E+13	3.97E+13	-1.136928	0.269
DUMOIL	9.46E+13	4.53E+13	2.089578	0.0496
CRIME_ CP	1.69E+12	1.29E+12	1.311743	0.2045
CRIME_RISK	-2.74E+11	1.67E+11	-1.64031	0.1166
DIST_TO_SG	-8630000000	13800000000	-0.625971	0.5384
DUM_MAIN_AP	-3.17E+13	3.44E+13	-0.921462	0.3678
INFRA_INDEX	3.22E+12	2.24E+13	0.143701	0.8872
PORTQUAL	1.44E+14	6.64E+13	2.167849	0.0424
J_PENDUDUK	5648366	1732711	3.259843	0.0039
UHH	-3.23E+12	5.22E+12	-0.618309	0.5433
UMP_1	47800760	35773466	1.336207	0.1965
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 IV - Hasil Uji Normalitas



Lampiran V - 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	$2.05E+27$	4.72228	3.750046
CRIME_ CP	$1.66E+24$	75.01229	3.122728
CRIME_RISK	$2.78E+22$	9.601206	1.687357
DIST_TO_SG	$1.9E+20$	6.131365	1.782
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	$3E+12$	6.022731	3.996756
UHH	$2.73E+25$	1481.642	1.99505
UMP_1	$1.28E+15$	64.76677	2.556563

Lampiran VI - Hasil Uji Heteroskedastisitas

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
F-statistic	1.549951	Prob. F(13,20)	0.1834	
Obs*R-squared Scaled explained SS	17.06325	Prob. Chi-Square(13)	Prob.	0.1964
	5.138884	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	0.322
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	-4.6E+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	0.443
PORTQUAL	9E+27	2.62E+27	3.44329	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			