

PENGUNAAN REMOTE SENSING DAN GOOGLE TRENDS UNTUK ESTIMASI PRODUK DOMESTIK BRUTO INDONESIA

Firhand Yusuf Kamal¹, Mutiara Indryan Sari²,
Maulidya Fan Ghul Udzan Utami³, Fitri Kartiasih^{4*}

^{1,2,3,4}Politeknik Statistika STIS, Jakarta, Indonesia

e-mail: fkartiasih@stis.ac.id, 222011513@stis.ac.id

*author corespondence

Abstrak

Pembangunan ekonomi merupakan salah satu topik yang penting untuk dikaji karena memberi gambaran tingkat kesejahteraan suatu negara. Akan tetapi, kebutuhan data yang menggambarkan pembangunan ekonomi, khususnya Produk Domestik Bruto (PDB) belum tersedia secara real-time. Penggunaan big data, seperti Night Time Light (NTL) dan *Google Trends*, untuk mengestimasi PDB menjadi salah satu solusi permasalahan tersebut. Meskipun begitu, kedua jenis big data tersebut masih memiliki kekurangan sebagai proxy untuk variabel PDB. NTL memiliki kekurangan karena data yang dihasilkan tidak mampu membedakan apakah cahaya yang ditangkap merupakan cahaya yang berasal dari listrik atau cahaya temporal, seperti api, pantulan cahaya, dan lain sebagainya. Begitu pula dengan *Google Trends* yang masih memiliki kekurangan dimana kata kunci yang digunakan tidak selalu merepresentasikan pola perilaku masyarakat secara konsisten. Metode penelitian ini bersifat kuantitatif dengan menggunakan data NTL dan *Google Trends* yang dievaluasi dan dibandingkan untuk melihat jenis data mana yang terbaik dalam menghasilkan estimasi PDB. Hasil penelitian menunjukkan bahwa penggunaan NTL, *Google Trends*, dan gabungan keduanya dapat digunakan untuk memprediksi PDB. Hal tersebut dapat terlihat dari hasil pemodelan yang tidak bersifat overfitting dan memiliki nilai MAPE di bawah 10%. Selain itu, penggunaan gabungan kedua data tersebut menjadi pilihan terbaik dalam mengestimasi PDB yang ditunjukkan dengan hasil evaluasi terbaik, yakni nilai RMSE sebesar 15792,73 dan nilai MAPE sebesar 0,52%.

Kata kunci: *google trends; produk domestik bruto; night time light; remote sensing*

Abstract

Economic development is an important topic to study because it provides an overview of the level of welfare of a country. However, the need for data that describes economic development, especially Gross Domestic Product (GDP), is not yet available in real-time. The use of big data, such as Night Time Light (NTL) and Google Trends, to estimate GDP is one solution to this problem. Even so, both types of big data still have shortcomings as proxies for GDP variables. NTL has shortcomings because the data produced is unable to distinguish whether the light captured is light originating from electricity or temporal light, such as fire, reflected light, and so on. Likewise, Google Trends still has shortcomings in that the keywords used do not always consistently represent people's behavior patterns. This research method is quantitative using NTL and Google Trends data which are evaluated and compared to see which type of data is best in producing GDP estimates. The research results show that the use of NTL, Google Trends, and a combination of both can be used to predict GDP. This can be seen from the modeling results which are not overfitting and have a MAPE value below 10%. Apart from that, using a combination of these two data is the best choice in estimating GDP as shown by the best evaluation results, namely an RMSE value of 15792.73 and a MAPE value of 0.52%.

Keywords: *google trends; gross domestic product; nighttime light; remote sensing*

PENDAHULUAN

Sektor ekonomi merupakan salah satu sektor yang berperan penting dalam menunjang seluruh aktivitas yang ada pada suatu negara. Keadaan ekonomi dapat menunjukkan tingkat kesejahteraan dan kemakmuran dari suatu negara. Oleh sebab itu, setiap negara merumuskan sejumlah kebijakan yang dapat mendorong terjadinya pembangunan ekonomi. Pembangunan ekonomi suatu negara yang diukur dengan pertumbuhan ekonomi

mengindikasikan terjadinya peningkatan produksi barang dan jasa dalam perekonomian suatu wilayah pada periode tertentu (Hawari & Kartiasih, 2017; Kartiasih, 2019a). Tingkat produksi ini dihitung berdasarkan total keseluruhan nilai tambah yang dihasilkan oleh sektor-sektor ekonomi yang ada di suatu wilayah atau sering disebut sebagai Produk Domestik Bruto (PDB). Dengan demikian, PDB dapat digunakan sebagai indikator utama untuk mengevaluasi kinerja ekonomi atau dapat juga digunakan sebagai ukuran keberhasilan pemerintah dalam meningkatkan kinerja sektor-sektor ekonomi yang ada. PDB dipublikasikan setiap tiga bulan sekali dan merupakan variabel ekonomi yang penting bagi suatu negara (Adwendi & Kartiasih, 2016; Kartiasih, 2019b).

Pada tahun 2020, kinerja ekonomi secara keseluruhan di seluruh dunia menunjukkan pelemahan. Pandemi Covid-19 menjadi salah satu penyebab utama terjadinya pelemahan kinerja ekonomi di seluruh dunia (Fadillah & Kartiasih, 2023; Umam & Kartiasih, 2023). Tak hanya pandemi Covid-19, beberapa ketidakpastian ekonomi yang berkelanjutan juga berdampak pada perlambatan ekonomi global, seperti eskalasi perang dagang yang terjadi antara Amerika Serikat dan China, ketidakpastian Brexit, aksi protes di Hongkong, dan aksi demonstrasi yang terjadi di Amerika Latin yang berdampak pada pelemahan sentimen sehingga menahan aktivitas bisnis dan investasi (BI, 2019). Sejalan dengan terjadinya ketidakpastian global berkelanjutan ini, perlambatan ekonomi telah menjadi isu yang meluas dan dialami oleh hampir seluruh negara di dunia. Menurut Badan Pusat Statistik (BPS), perekonomian Indonesia di tahun 2020 telah mengalami kontraksi pertumbuhan sebesar 2,07 persen (*c-to-c*) dibandingkan dengan tahun 2019. Di tengah fenomena perlambatan ekonomi global yang sedang terjadi hingga saat ini, penting untuk mengetahui informasi tentang pembangunan ekonomi di Indonesia agar dapat memberikan wawasan atau informasi yang dibutuhkan dalam membuat keputusan, mengelola risiko, serta merumuskan kebijakan yang efektif.

Secara tradisional, data PDB dapat dikumpulkan menggunakan sumber data yang berasal dari hasil survei konvensional, seperti Survei Khusus Konsumsi Rumah Tangga (SKKRT). SKKRT merupakan suatu survei yang bertujuan untuk menghasilkan data terkait konsumsi rumah tangga dan diselenggarakan secara rutin tiap tahun. Pendataan konvensional pada SKKRT diselenggarakan melalui survei rumah tangga di lapangan yang dilakukan setiap tiga bulan sekali. Akan tetapi, pengumpulan data menggunakan metode konvensional ini memiliki cakupan yang terbatas, biaya yang tinggi, sulit, dan waktu yang cukup lama (Putri et al., 2022). Kelemahan ini mengakibatkan keterbatasan dalam merepresentasikan data pertumbuhan ekonomi di Indonesia dari segi cakupan dan waktu. Selain itu, pengumpulan data melalui SKKRT selama masa pandemi Covid-19 juga terhambat.

Secara umum, data PDB membutuhkan waktu 2 bulan, dari proses pengumpulan data hingga data tersebut dapat dirilis secara resmi, seperti PDB Triwulan I 2023 yang membutuhkan waktu kurang lebih 1-2 bulan sejak berakhirnya bulan Maret (Triwulan I). Pengumpulan data menggunakan metode konvensional terkadang juga membutuhkan pertemuan secara langsung (*face-to-face*). Hal ini tentu tidak bisa diimplementasikan secara efektif terlebih ketika masih terjadi pandemi. Oleh karena itu, diperlukan metode pengumpulan data yang bisa mengatasi kelemahan dari pengumpulan data PDB agar para pembuat kebijakan dapat merumuskan kebijakan-kebijakan yang tepat dan relevan dengan kondisi ekonomi yang ada.

Dalam beberapa tahun terakhir, kemajuan teknologi komputer dan munculnya berbagai layanan pengumpulan informasi *online* telah menghadirkan sumber data alternatif baru yang dikenal sebagai *big data* (Heikkinen, 2019). *Big data* menjadi salah satu sumber data tak terstruktur yang paling tepat waktu dan dihasilkan dari ekstensifikasi penggunaan

internet. Salah satu sumber *big data* yang terkenal untuk memprediksi kondisi ekonomi makro dalam jangka pendek adalah *Google Trends*, yakni sumber data yang dapat menyediakan informasi mengenai pencarian istilah tertentu pada pencarian *Google* secara berkala tiap minggu, hari, bahkan *real-time*. Data pencarian *Google* dapat berisi wawasan tentang rencana dan intensi konsumen atau agen lainnya, salah satunya yaitu mengenai pengeluaran konsumen. Seperti contoh, konsumen dapat mencari informasi di mesin pencarian *Google* sebelum membuat keputusan ekonomi terkait pembelian. Dengan demikian, data pencarian *Google* dapat menjadi sumber informasi yang berharga untuk menyediakan variabel ekonomi makro.

Penggunaan variabel ekonomi makro dari *Google Trends* yang sering digunakan salah satunya adalah untuk memprediksi nilai PDB. Seperti halnya pada OECD *Economic Department Working Papers* No. 1634, telah dilakukan prediksi pertumbuhan PDB dengan melakukan pelacakan aktivitas secara *real-time*. Penelitian ini dilaksanakan karena adanya penggunaan indikator berbasis survei untuk peramalan aktivitas ekonomi dan PDB dalam kepentingan pembuatan kebijakan yang kurang bisa diandalkan ketika perubahan aktivitas ekonomi terjadi sangat cepat dan cenderung masif. Oleh karena itu, diperlukan pelacakan aktivitas ekonomi secara *real-time*, salah satu solusinya yaitu dengan menggunakan data dari *Google Trends*. Penelitian ini membahas salah satu indikator berfrekuensi tinggi yang berpengaruh terhadap kegiatan ekonomi berdasarkan *Google Trends*, yang digunakan untuk membuat pelacak mingguan yang dapat memberikan perkiraan pertumbuhan PDB secara *real-time* pada 46 negara G20, OECD, dan beberapa negara mitra.

Hasil penelitian menunjukkan bahwa pelacakan aktivitas masyarakat menggunakan *Google Trends* populer karena kemampuannya untuk merekam aktivitas tersebut secara *real-time* sehingga dapat menyediakan gambaran yang akurat mengenai perubahan pembangunan ekonomi dari waktu ke waktu. Karena kemampuannya tersebut, *Google Trends* dapat digunakan sebagai indikator sentimen publik atau ketertarikan di bidang ekonomi. Sebagai contoh, pencarian yang berhubungan dengan “pekerjaan” atau “pengangguran” dapat merefleksikan perasaan publik atau kekhawatiran tentang pasar tenaga kerja ataupun keseluruhan kondisi perekonomian. Walaupun bukan sebagai pengukuran PDB secara langsung, informasi ini dapat memberikan pemahaman tentang sentimen publik dan persepsi yang mungkin dapat memengaruhi perilaku ekonomi masyarakat.

Meskipun begitu, *Google Trends* merekam aktivitas masyarakat hanya berdasarkan apa yang mereka cari pada mesin pencarian *Google*. Oleh karena itu, jika menggunakan *Google Trends* untuk membuat prediksi, hasilnya hanya akan dipengaruhi oleh sentimen dan tren dalam perilaku masyarakat, tanpa mempertimbangkan faktor-faktor lain yang mungkin penting dan berpengaruh. Keterbatasan tersebut dapat mengakibatkan prediksi yang tidak akurat, seperti yang terjadi pada kasus *Google Trends Flu Fail* di Amerika Serikat pada tahun 2009 dan 2012.

Hal tersebut perlu menjadi pertimbangan dalam menggunakan data *Google Trends* untuk menyusun penelitian pada prediksi PDB. Dalam mengestimasi variabel ekonomi, termasuk penggunaan *Google Trends* untuk memprediksi PDB, hasil yang diperoleh akan cenderung tidak *robust*. Oleh karena itu, dalam proses estimasi, penting untuk menggunakan sumber data lain. Selain *Google Trends*, *big data* yang sering digunakan untuk memprediksi PDB adalah data yang berasal dari *remote sensing*, salah satunya yaitu dengan menggunakan pendekatan variabel *Night Time Light* (NTL). Variabel NTL diperoleh dari hasil *remote sensing* yang menjelaskan intensitas cahaya yang dipancarkan permukaan bumi saat malam hari. Penjelasan mengenai intensitas cahaya di malam hari ini dapat memberikan gambaran

yang cukup mengenai kondisi dan aktivitas ekonomi manusia pada waktu tersebut (Pemayun et al., 2024).

Meskipun begitu, NTL hanya benar-benar merekam bagaimana intensitas cahaya di malam hari pada permukaan bumi, padahal aktivitas ekonomi manusia sangat beragam dan tidak selalu berhubungan dengan tingkat intensitas cahaya di area tersebut. Selain itu, hanya menggunakan NTL sebagai *benchmark* dalam menjelaskan aktivitas ekonomi manusia dapat mengabaikan faktor-faktor lain yang memengaruhi perubahan intensitas NTL di area tersebut. Hal ini disebabkan karena variabel NTL seringkali tidak mampu membedakan antara berbagai sumber cahaya secara jelas. Adanya keterbatasan pemahaman NTL untuk menjelaskan kondisi pada permukaan bumi yang berubah dapat menghalangi analisis lebih mendalam dari NTL dalam menangkap dinamika dan konteks pemandangan saat malam hari dengan akurat (Zhao et al., 2019). Sebagai contoh, data NTL mungkin tidak mampu membedakan antara cahaya yang dihasilkan oleh area industri pada perkotaan, petir, ataupun cahaya yang dihasilkan oleh kebakaran hutan. Oleh karena itu, estimasi PDB menggunakan variabel NTL belum cukup baik karena hasil yang didapatkan tidak disertai dengan konteks yang tepat.

Berdasarkan penjelasan di atas, dapat diketahui bahwa data yang dihasilkan *Google Trends* dan data NTL masing-masing menunjukkan sejumlah kelemahan yang dapat menyebabkan hasil estimasi PDB menjadi kurang akurat. Dengan menggabungkan data *Night Time Light* (NTL) dari citra satelit dan data *Google Trends*, informasi yang diperoleh untuk mengestimasi pertumbuhan ekonomi menjadi lebih komprehensif. Contohnya, data NTL dapat memberikan pola spasial dari aktivitas ekonomi secara umum, sedangkan data *Google Trends* dapat memberikan informasi yang lebih rinci mengenai ketertarikan konsumen dan tren yang dapat menjadi konteks untuk melengkapi gambaran yang diperoleh dari variabel NTL.

Tujuan penelitian ini adalah untuk mendapatkan hasil estimasi model nilai PDB triwulanan di Indonesia. Penelitian ini akan mencoba untuk membuktikan bagaimana kombinasi kedua sumber data, yakni *Google Trends* dan variabel NTL, akan memengaruhi nilai estimasi PDB. Selain itu, penelitian ini bertujuan untuk membandingkan hasil model estimasi PDB ketika model hanya menggunakan data dari *Google Trends* atau nilai NTL, dengan model estimasi yang menggunakan gabungan kedua sumber data tersebut sebagai faktor komplemen.

Produk Domestik Bruto (PDB) adalah salah satu indikator penting dari suatu negara untuk menentukan kondisi perekonomian dari negara tersebut pada periode waktu tertentu. Pada dasarnya, PDB adalah besarnya nilai tambah yang dihasilkan dari seluruh satuan bisnis di negara tertentu, atau total nilai dari produk barang dan jasa akhir yang dihasilkan oleh seluruh satuan ekonomi, termasuk barang dan jasa yang dihasilkan oleh warga negara asing yang berasal dari wilayah teritori negara tersebut (BPS, 2022). Penghitungan PDB terdiri dari dua konsep harga, yaitu PDB Atas Dasar Harga Berlaku (ADHB) dan PDB Atas Dasar Harga Konstan (ADHK). PDB ADHB merujuk pada nilai tambah barang dan jasa yang dihasilkan suatu negara dalam periode waktu tertentu dan dihitung menggunakan harga yang berlaku pada waktu tersebut. Dengan menggunakan harga pada tahun dasar, besar PDB akan menunjukkan pergeseran dan struktur ekonomi (Kusumasari & Kartiasih, 2017).

Sementara itu, PDB ADHK atau PDB riil menjelaskan nilai tambah barang dan jasa yang dihitung dengan harga yang berlaku pada tahun dasar tertentu. Saat ini, tahun dasar yang digunakan BPS dalam menghitung PDB adalah tahun 2010. PDB ADHK dapat digunakan untuk menentukan tingkat pertumbuhan ekonomi dari tahun ke tahun sebagai gambaran capaian pembangunan yang dihasilkan pada periode waktu tertentu. Pertumbuhan ekonomi dihitung menggunakan PDB riil untuk memastikan gambaran pertumbuhan

sebenarnya dalam bentuk jumlah barang dan jasa, tanpa mempertimbangkan adanya perubahan harga. Dalam konteks prediksi PDB, PDB ADHK akan memberikan hasil yang lebih konsisten dan relevan karena menghilangkan adanya pengaruh inflasi dan perubahan harga, sehingga dapat digunakan untuk mengukur pertumbuhan ekonomi sebenarnya dalam bentuk jumlah barang dan jasa (Weley et al., 2019). Dengan menggunakan PDB ADHK, prediksi PDB akan lebih berfokus pada perubahan nyata dari aktivitas ekonomi tanpa pengaruh fluktuasi harga (Ginting & Kartiasih, 2019; Kartiasih, 2019b; Kusumasari & Kartiasih, 2017; Ningsih & Kartiasih, 2019). Oleh karena itu, pada penelitian ini, nilai PDB yang digunakan adalah PDB ADHK menurut pengeluaran dengan tahun dasar 2010. Data yang diperoleh merupakan data PDB triwulanan dengan rentang tahun 2014-2022. Selain itu, sumber data selanjutnya adalah *big data*. Ada dua jenis *big data* yang digunakan dalam penelitian ini, yaitu data dari *Google Trends* dan data dari *remote sensing* berupa *Night-Time Light* (NTL).

Selanjutnya, juga dilakukan pengambilan data dari *Google Trends* dengan menggunakan sejumlah kata kunci pencarian. *Google Trends Index* adalah suatu alat yang dapat digunakan untuk mempelajari tren dan pola dari *query* mesin pencarian *Google* (Vishal, 2019). Data *Google Trends* tersedia dalam bentuk ekstensif yang dipublikasikan *Google* secara berkala mingguan dan harian secara *real-time*. Kemudian, pengguna juga dapat memilih data berdasarkan negara dan dalam beberapa kasus, bahkan dapat hingga level wilayah administratif tertentu.

Google Trends dapat memberikan indeks *time series* dari *query* volume yang dimasukkan oleh pengguna ke dalam *Google* pada area geografis tertentu. *Query* yang dibentuk dapat berdasarkan kata kunci, kategori kata kunci, atau topik. *Query* yang berdasarkan kata kunci bersifat spesifik terhadap bahasa dan rentan akan ambiguitas. Contohnya, *query* "Apple", *Google Trends* menggabungkan pencarian tentang buah apel dan sebuah perusahaan. Sementara itu, *query* berdasarkan kategori dan topik dapat diharmonisasikan antarbahasa dan *query* berdasarkan topik menghasilkan kumpulan data yang dapat dibandingkan antar negara. Selain itu, menggunakan topik dan kategori juga dapat menggunakan model umum yang lebih banyak karena topik dan kategori adalah kata kunci yang disempurnakan dan memberikan representasi ketertarikan atas pencarian sesuatu secara umum dibandingkan dengan istilah tertentu (OECD, 2020).

Beberapa penelitian sebelumnya telah menemukan bahwa *Google Trends* dapat menghasilkan perkiraan yang akurat untuk berbagai variabel ekonomi, termasuk PDB (Heikkinen, 2019). Götz dan Knetsch (2019) menggunakan data *Google Trends* untuk mengestimasi PDB Jerman menggunakan *bridge equation model*. Hasil penelitian tersebut menunjukkan bahwa variabel *Google Trends* dapat memberikan informasi tambahan yang berguna untuk estimasi PDB jangka panjang dan menengah. Ferrara dan Simoni (2019) juga menunjukkan bahwa *Google Trends* dapat memberikan prediksi yang akurat untuk pertumbuhan PDB hingga empat minggu pertama periode prediksi yaitu sebelum data resmi tersedia. Selain itu, Bouayad et al., (2022) menemukan bahwa penambahan data *Google Trends Index* dapat meningkatkan akurasi prediksi pertumbuhan PDB. Dengan demikian, terlihat bahwa *Google Trends Index* merupakan alat yang baik untuk dimanfaatkan dalam estimasi pertumbuhan PDB. Oleh karena itu, penelitian ini juga akan memanfaatkan *Google Trends Index* dalam mencari data rasa ingin tahu, keresahan, dan ketertarikan masyarakat terhadap kata kunci mengenai ekonomi dan perkembangannya pada tahun 2014-2022.

Untuk memperoleh data NTL diperlukan metode penginderaan jarak jauh dan diproses di *Google Earth Engine*. NTL (*Night Time Light*) merujuk pada penggunaan data intensitas cahaya malam untuk mempelajari pola yang dihasilkan aktivitas manusia di malam hari. Data variabel ini dapat diperoleh dari metode pengumpulan data yang disebut *remote*

sensing. NTL (*Night Time Light*) merupakan citra satelit yang menampilkan kekuatan dari adanya cahaya pada malam hari di seluruh dunia. Citra satelit ini dihasilkan oleh sensor cahaya yang sensitif pada gelombang radiasi elektromagnetik yang dihasilkan dari permukaan bumi dan menutupi spektrum *range* cahaya yang dapat dilihat oleh mata manusia serta sinyal *infrared* terdekat (Afrianto & Graha, 2023). Sensor NTL dan produk data terus dikembangkan, sejalan dengan munculnya beragam aplikasi *remote sensing* berbasis citra satelit NTL sejak akhir tahun 1990-an (Latifa et al., 2023).

Data NTL dapat digunakan untuk mengawasi aktivitas manusia, termasuk pertumbuhan penduduk perkotaan, pola pemukiman, tingkat penggunaan energi, dan perubahan lingkungan. Dengan menganalisis pola cahaya malam dari data NTL, peneliti dapat memperoleh informasi mengenai distribusi populasi, tingkat industrialisasi, dan aktivitas ekonomi di suatu wilayah. Penggunaan data NTL sangat beragam, termasuk dalam bidang ilmu pengetahuan, geografi, ilmu kelingkungan, dan penginderaan jauh dan perkotaan. Data tersebut dapat digunakan untuk menghasilkan peta cahaya malam global, perubahan pada pola cahaya dari waktu ke waktu, dan mendukung pembuatan kebijakan untuk menciptakan perencanaan perkotaan yang berkelanjutan dan pengelolaan sumber daya energi.

METODE

Penelitian ini menggunakan data sekunder yang berasal dari tiga sumber, yaitu Badan Pusat Statistik (BPS), *Google Earth Engine*, dan *Google Trends*. Data yang digunakan adalah PDB, *Google Trends Index*, dan NTL. Data dari variabel NTL ini merupakan cara yang tepat untuk menjelaskan kondisi dan aktivitas manusia di permukaan bumi, yang mana hal ini dapat menjadi sumber informasi mengenai bagaimana pertumbuhan ekonomi di suatu wilayah. Dengan kata lain, tingkat intensitas cahaya di malam hari dapat menjadi sumber data yang berguna untuk mengawasi krisis kemanusiaan (Li et al., 2014). Oleh karena itu, data NTL adalah pilihan yang tepat untuk mengestimasi PDB melalui intensitas cahaya. Pada *Google Earth Engine* bisa didapatkan nilai-nilai dari variabel NTL di lokasi yang dipilih, yaitu Indonesia. Informasi lebih rinci mengenai seluruh data yang digunakan terdapat pada Tabel 1.

Tabel 1. Ringkasan data dalam penelitian

Data	Definisi	Periode
Produk Domestik Bruto (PDB)	Data PDB ini diperoleh dari website Badan Pusat Statistik (BPS), PDB yang digunakan pada penelitian ini adalah PDB atas harga konstan menurut pengeluaran dengan tahun dasar 2010.	2014-2022
<i>Night-Time Light</i> (NTL)	Data ini diperoleh dari metode <i>remote sensing</i> dengan <i>Google Earth Engine</i> . Pada penelitian ini variabel dari <i>remote sensing</i> yang digunakan adalah NTL atau besarnya intensitas cahaya di permukaan bumi pada malam hari	2014-2022

Google Trends

Pengambilan data dari *Google Trends* 2014-2022 adalah menggambarkan tren yang sedang banyak dicari. Perolehan data ini dilakukan dengan menggunakan beberapa kata kunci pencarian yang berkaitan dengan ekonomi

Untuk menghindari terjadinya model yang *overfitting* terhadap data, penelitian ini membagi data menjadi data latih (*train*) dan data uji (*uji*) untuk melihat konsistensi performa model pada 2 data yang berbeda (Parsa et al., 2020). Data latih yang digunakan adalah data dari tahun 2014 – 2021, sedangkan untuk data uji digunakan data tahun 2022.

Penelitian ini akan menggunakan beberapa skenario dengan tahapan pengolahan yang sama. Untuk membandingkan performa prediksi antara penggunaan variabel *Google Trends*, NTL, dan gabungan keduanya dalam model, penelitian ini akan mencoba tiga skenario sebagai berikut:

- Penggunaan *Google Trends* dalam model,
- Penggunaan NTL dalam model, dan
- Penggunaan *Google Trends* dan NTL dalam model

Skenario ini disusun untuk membandingkan penggunaan data mana yang paling baik digunakan untuk mengestimasi PDB Indonesia 2014-2022. Seleksi variabel adalah suatu cara yang penting untuk mereduksi dimensi data yang digunakan dalam model dan mencegah *overfitting*. Selain itu, seleksi variabel juga dilakukan untuk menghapus variabel redundan dengan memilih variabel prediktor yang paling optimal (Akbar & Kurniawan, 2020). Metode pemilihan variabel telah banyak diterapkan di sejumlah cabang ilmu komputasi antara lain *computer vision*, pengenalan pola, dan *machine learning*. Seleksi variabel memiliki kelebihan sebagai berikut: pertama, pemilihan variabel biasanya memungkinkan seluruh metode diimplementasikan secara komputasional dengan lebih efisien. Kedua, seleksi variabel biasanya mencapai peningkatan akurasi dari metode tersebut (Song et al., 2010).

Salah satu proses pemilihan variabel yang dapat digunakan ke model prediksi adalah metode yang didasarkan pada korelasi *Pearson* dan *Spearman*. Korelasi *Pearson* menghasilkan koefisien korelasi sampel (r) yang mengukur kekuatan dan arah hubungan linier antara pasangan variabel kontinu. Nilai r atau koefisien korelasi pada korelasi *Pearson* dan *Spearman* berada pada rentang $[-1, 1]$. Tanda dari koefisien korelasi mengindikasikan arah hubungan, sedangkan besarnya korelasi (seberapa dekat dengan -1 atau $+1$) mengindikasikan kekuatan hubungan. Nilai absolut dari koefisien *Pearson* dan *Spearman* digunakan untuk menyederhanakan proses pengolahan. Hanya variabel yang memiliki hubungan "sedang", "kuat", dan "sangat kuat" yang digunakan untuk proses pengolahan selanjutnya (Kapetanakis et al., 2017). Interpretasi kekuatan koefisien korelasi dapat dilihat pada tabel berikut.

Tabel 2. Interpretasi kekuatan koefisien korelasi

Koefisien Korelasi	Interpretasi
0,00 - 0,20	Sangat Lemah
0,21 - 0,40	Lemah
0,41 - 0,70	Sedang

0,71 - 0,90	Kuat
0,90 - 0,99	Sangat Kuat
1,00	Sempurna

Sumber: Sujarweni (2014)

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) atau analisis komponen utama adalah teknik untuk mengurangi dimensi dari sekumpulan data, meningkatkan interpretabilitas, serta meminimalkan kehilangan informasi (Jolliffe & Cadima, 2016). Metode ini merupakan metode yang kuat dan serbaguna yang mampu memberikan gambaran data multivariat yang kompleks. PCA dapat digunakan untuk mengungkap hubungan antar variabel dan hubungan antarsampel (misalnya *clustering*), mendeteksi *outlier*, menemukan dan mengukur pola, menghasilkan hipotesis baru serta banyak hal lainnya (Bro & Smilde, 2014).

PCA mereduksi data dengan memproyeksikannya secara geometris ke dimensi yang lebih rendah yang disebut komponen utama (PC), dengan tujuan menemukan ringkasan data terbaik menggunakan PC dalam jumlah terbatas. Komponen utama (PC) pertama dipilih untuk meminimalkan total jarak antara data dan proyeksi mereka ke PC. Dengan meminimalkan jarak ini, varian dari titik yang diproyeksikan juga dimaksimalkan, σ^2 . Komponen utama kedua (dan selanjutnya) dipilih dengan cara yang sama, dengan persyaratan tambahan bahwa mereka tidak berkorelasi dengan semua PC sebelumnya. Misalnya, proyeksi ke PC1 tidak berkorelasi dengan proyeksi ke PC2, dan PC dapat dianggap sebagai ortogonal geometris. Persyaratan tanpa korelasi ini berarti bahwa jumlah maksimum PC yang mungkin adalah jumlah sampel atau jumlah fitur, mana yang lebih kecil (Lever et al., 2017). Untuk lebih jelasnya, berikut merupakan langkah-langkah dalam melakukan reduksi dimensi dengan PCA.

Langkah pertama adalah standardisasi pada data. Tujuan dari langkah ini adalah untuk menormalkan rentang variabel awal kontinu sehingga masing-masing variabel memberikan kontribusi yang sama untuk analisis. Secara lebih spesifik, alasan mengapa standardisasi sebelum PCA penting untuk dilakukan adalah karena PCA sangat sensitif terhadap varians dari variabel awal. Artinya, jika terdapat perbedaan besar antara rentang variabel awal, variabel-variabel dengan rentang yang lebih besar akan mendominasi variabel-variabel dengan rentang yang lebih kecil (misalnya, variabel yang rentangnya antara 0 dan 100 akan mendominasi variabel yang rentangnya antara 0 dan 1), sehingga akan menghasilkan hasil yang bias. Oleh karena itu, dengan mentransformasikan data ke skala yang dapat dibandingkan, masalah ini dapat dicegah. Secara matematis, standardisasi dapat dilakukan dengan mengurangi nilai dengan rata-rata dan membaginya dengan standar deviasi untuk setiap nilai dari setiap variabel.

Langkah selanjutnya, yaitu menghitung matriks kovarians antara variabel-variabel dalam data. Matriks kovarians ini menggambarkan tingkat hubungan linier antara setiap pasangan variabel. Matriks kovarians (C_x) dapat dihitung dengan persamaan sebagai berikut.

$$C_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (1)$$

Setelah mendapatkan matriks kovarians, PCA menghitung vektor-vektor *eigen* dan nilai *eigen* dari matriks kovarians untuk mengidentifikasi komponen utama. Vektor *eigen*

merepresentasikan arah dan ukuran variasi yang terkandung dalam data. Sementara itu, komponen utama PCA adalah kombinasi linear dari variabel awal yang diberikan oleh vektor *eigen* dan diurutkan berdasarkan nilai *eigen* yang terkait.

Dalam penelitian ini, PCA dipilih sebagai salah satu metode karena tujuan utamanya adalah untuk mengompres data dan mendapatkan jumlah variabel yang lebih sedikit. Metode PCA pada penelitian ini hanya digunakan untuk data yang diambil dari *Google Trends* karena data tersebut memiliki jumlah variabel yang sangat banyak, yaitu sebanyak 50 variabel. Dari 50 variabel tersebut, akan dilakukan reduksi dimensi dengan mengelompokkan variabel-variabel menjadi hanya beberapa komponen utama yang signifikan. Dengan hanya memilih sejumlah komponen utama yang signifikan, dimensi data dapat berkurang tanpa kehilangan terlalu banyak informasi dan juga memiliki interpretabilitas yang lebih baik daripada variabel-variabel asli karena mereka mewakili kombinasi linear dari variabel awal yang berkorelasi. Oleh karena itu, PCA dapat digunakan untuk menggambarkan data *Google Trends* yang cukup kompleks ke dalam dimensi ruang yang lebih rendah sehingga mempermudah pemahaman dan analisis data selanjutnya.

Pada penelitian ini, teknik yang digunakan adalah melakukan normalisasi dengan menggunakan *z-score*. Normalisasi *z-score* mengacu pada proses normalisasi setiap nilai dalam data sehingga rata-rata dari semua nilai adalah 0 dan standar deviasi adalah 1.

$$z = \frac{x - \bar{x}}{\sigma_x} \quad (2)$$

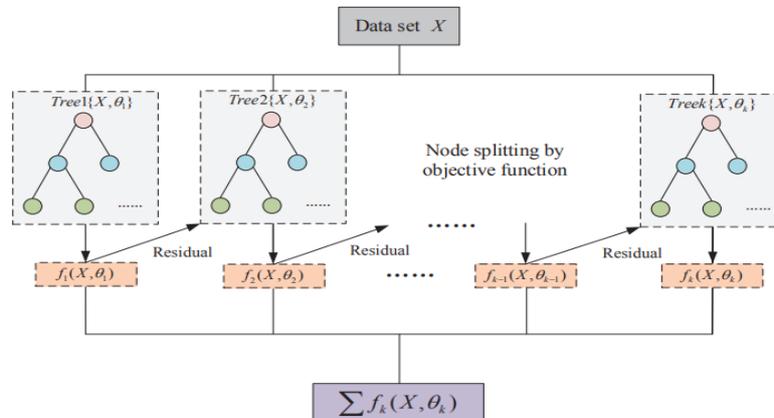
Di mana x = nilai aktual; \bar{x} = rata-rata dari x ; σ_x = standar deviasi x

Beberapa penelitian telah membuktikan bahwa terjadi peningkatan kinerja prediksi setelah dilakukan normalisasi daripada sebelum dilakukan normalisasi pada data. Penelitian yang dilakukan oleh Panigrahi dan Behera (2013) mengenai pengaruh teknik normalisasi pada peramalan *time series* univariat menunjukkan bahwa teknik normalisasi memiliki dampak yang signifikan dalam peningkatan kinerja prediksi.

Cabang ilmu komputer ini menggunakan algoritma dan model statistik untuk melatih komputer dalam mempelajari pola dan memprediksi untuk membuat keputusan. Setelah proses *training*, model dapat digunakan untuk melakukan prediksi atau keputusan pada data baru yang belum pernah digunakan. *Machine learning* merupakan bagian dari *Artificial Intelligence* (AI), dan melibatkan beberapa paradigma pembelajaran, seperti *Supervised Learning* (SL), *Unsupervised Learning* (UL), and *Reinforcement Learning* (RL) (Fauzan et al., 2024; Innayah et al., 2024; Pemayun et al., 2024). Kemudian, untuk melakukan estimasi menggunakan *machine learning*, penelitian ini akan menggunakan metode regresi. Terdapat beberapa metode untuk menghasilkan model regresi pada *machine learning*, antara lain sebagai berikut: *Support Vector Regression* (SVR), *Random Forest Regression*, *XGBoost*, *Ridge Regression* dan *Lasso*, *Decision Tree Regression*, dll. Namun, dalam penelitian ini akan memilih menggunakan metode *machine learning XGBoost* dalam melakukan estimasi PDB.

XGBoost merupakan salah satu model dalam *machine learning* yang dapat difungsikan untuk mengestimasi suatu nilai yang berbasis pohon *regressor* (Jiang et al., 2019). *XGBoost* dapat digunakan sebagai model untuk memprediksi nilai PDB dengan hasil lebih baik dibandingkan model *machine learning* tradisional (Aguilar & Xavier, 2022; Qureshi et al., 2020). *XGBoost* pada awalnya adalah sebuah proyek penelitian yang diinisiasi oleh Tianqi Chen. Sistem ini dapat digunakan sebagai model untuk menghasilkan nilai prediksi PDB yang baik (Qureshi et al., 2020). Sistem ini memberikan banyak dampak pada komunitas *machine learning* dan penelitian *data mining*. Sebagai contoh, dari 29 kompetisi yang

diadakan pada situs *Kaggle.com*, 17 solusi pemenangnya menggunakan *XGBoost* sebagai model utama yang digunakan (T. Chen & Guestrin, 2016). Model ini menjadi digemari oleh para praktisi *data mining* karena kecepatan kalkulasinya dan performa model yang tinggi (J. Chen et al., 2020).



Gambar 1. Flowchart XGBoost

Ide dari *XGBoost* ini adalah perbaikan pada kecepatan dan performa *gradient boosting*. Konsep *gradient boosting* ini memiliki tiga langkah dasar. Pertama, penentuan *loss function* atau fungsi error yang teridentifikasi sesuai dengan kasus yang diberikan. Kedua, pembuatan *weak learner* untuk melakukan prediksi. Dalam *gradient boosting*, pohon regresi dipilih sebagai *weak learner*. Secara khusus, digunakan pohon regresi yang menghasilkan nilai-nilai prediksi. Nilai prediksi ini kemudian akan dibandingkan dengan nilai sebenarnya yang menghasilkan suatu residu. Residu ini kemudian akan dijadikan sebagai dasar pembentukan *weak learner* berikutnya. Pendekatan ini memungkinkan menghasilkan prediksi yang lebih akurat. Pohon-pohon ini dibuat secara *greedy* dan seringkali ada batasan-batasan tertentu yang diberlakukan untuk memastikan bahwa *weak learner* dibuat dengan metode *greedy*.

Ketiga, pembuatan model aditif untuk menjumlahkan seluruh prediksi-prediksi dari kumpulan *weak learners* yang dibuat dengan tujuan mengurangi fungsi kerugian. Proses penambahan pohon-pohon ini dilakukan satu per satu. Keluaran yang dihasilkan pada pohon kemudian ditambahkan ke keluaran urutan pohon yang sudah ada untuk meningkatkan keluaran akhir dari model. Proses ini berhenti ketika nilai yang dioptimalkan dengan baik untuk fungsi kerugian tercapai. Perbedaan yang mencolok dari ide *gradient boost* dan *XGBoost* adalah adanya penambahan *regularization term* dan penggunaan paralel *processing* sehingga dapat meningkatkan performa dan kecepatan penghitungan (Subian et al., 2024). Algoritma ini tidak hanya dirancang untuk mengumpulkan beberapa *weak learner* menjadi satu *strong learner*, tetapi juga untuk meningkatkan kecepatan komputasi dan kinerja model (Jange, 2022).

Pada penelitian ini digunakan *package XGBoost* pada pemrograman *python* untuk melakukan pemodelan *XGBoost*. Adapun dalam menjalankan *XGBoost* dibutuhkan beberapa parameter yang tepat agar mendapatkan hasil yang terbaik (Pandika Pinata et al., 2020). Untuk itu, dilakukan *tuning* parameter dengan menggunakan *5-fold cross-validation* sebagai berikut:

Tabel 3. Cross Validation

Parameter	Keterangan	Rentang Parameter
<i>max_depth</i>	parameter untuk menentukan kedalaman pohon yang dibangun	1,2,3,4,5,6,7,8,9

<i>n_estimators</i>	banyak pohon yang dibangun	700,800,900,1000
<i>learning_rate</i>	parameter untuk menentukan kecepatan pembelajaran model	0,0001;0,001;0,01;0,1;0,15

Sebelum melakukan evaluasi pada model, model tersebut harus dipastikan tidak *overfit*. Hal tersebut dilakukan dengan metode *learning curve*.

Learning curve atau kurva pembelajaran adalah kurva yang menggambarkan perubahan tingkat kesalahan pada model dengan jumlah data yang dimasukkan ke dalam model (Abidin et al., 2021). Untuk melihat apakah suatu model *overfit* atau tidak, kita akan melihat apakah kurva pembelajaran untuk fase latih (*training*) yang menggunakan data latih dan fase validasi (*validation*) dengan menggunakan data validasi cenderung mendekat atau tidak. Jika trennya cenderung mendekat, model tersebut tidak *overfit* dan sebaliknya (Zhang et al., 2018). Setelah memastikan model tidak *overfit*, Dalam mengevaluasi hasil estimasi dari model yang digunakan, digunakan metode evaluasi, yaitu *learning curve* (Jange, 2022; Pandika Pinata et al., 2020) (Sitepu et al., 2023; Widayati, 2009).

Seperti namanya, *Root Mean Square Error* (RMSE) adalah nilai akar kuadrat dari *Mean Square Error* (MSE). Nilai RMSE menghasilkan skor dalam satuan yang sama dengan variabel yang diobservasi sehingga nilai RMSE dengan mudah mewakili kesalahan tipikal atau "standar" untuk kesalahan yang berdistribusi normal. Penghitungan RMSE dapat dilakukan dengan formula berikut (Hodson, 2022).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

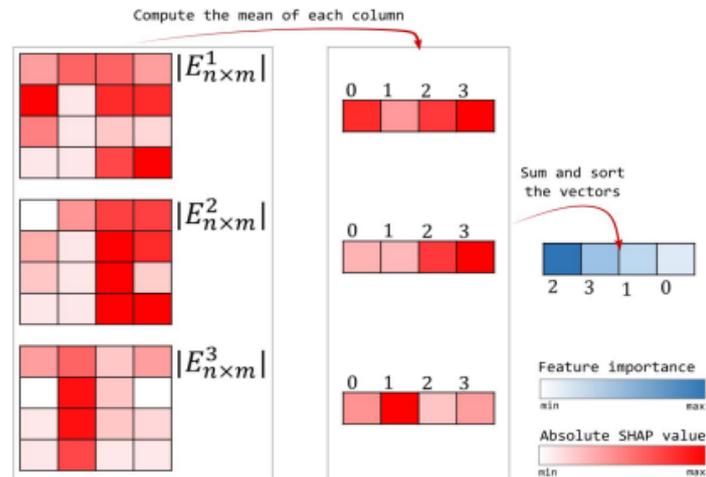
Mean Absolute Percentage Error (MAPE) merupakan salah satu ukuran evaluasi yang digunakan untuk mengukur tingkat kesalahan antara nilai aktual dan nilai prediksi dalam suatu model. MAPE menghitung rata-rata dari selisih persentase antara nilai aktual dan nilai prediksi dalam bentuk persentase absolut. Untuk menghitung nilai MAPE, dapat menggunakan rumus sebagai berikut.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (4)$$

Nilai MAPE yang lebih kecil menunjukkan tingkat kesalahan prediksi yang lebih rendah atau tingkat akurasi yang lebih tinggi. Penggunaan MAPE direkomendasikan dalam tugas-tugas yang lebih mementingkan kepekaan terhadap variasi relatif daripada variasi absolut. Namun, MAPE juga memiliki sejumlah kelemahan, yang paling kritis adalah pembatasan dalam penggunaannya untuk data yang benar-benar positif menurut definisi dan bias terhadap prediksi rendah, yang membuatnya tidak cocok untuk model prediksi di mana kesalahan besar diharapkan (Chicco et al., 2021).

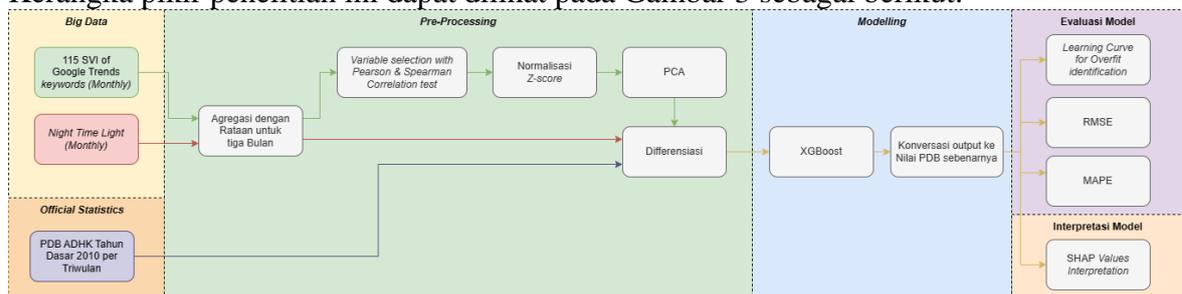
Shapley Additive Explanations atau nilai SHAP adalah sebuah pendekatan untuk menginterpretasikan model, dimana setiap fitur akan dihitung kontribusinya terhadap *output* model. Dengan kata lain, SHAP akan menghitung nilai *shapley*, yaitu menghitung kontribusi setiap subset fitur terhadap prediksi model dalam suatu *dataset* dengan *m* fitur didalamnya. Cara kerja pendekatan ini adalah dengan sedikit mengubah input dan mengukur perubahan

dalam prediksi model. Jika prediksi model tidak mengalami perubahan yang signifikan ketika nilai input untuk suatu fitur diubah sedikit, maka fitur tersebut mungkin bukan prediktor yang penting untuk observasi tertentu (Marcilio & Eler, 2020). Jadi, jika secara kumulatif nilai *Shapley* cenderung sedikit dibandingkan model lainnya, fitur tersebut tidak berpengaruh besar pada model.



Gambar 2. Proses Penghitungan Nilai *Shapley*

Kerangka pikir penelitian ini dapat dilihat pada Gambar 3 sebagai berikut.

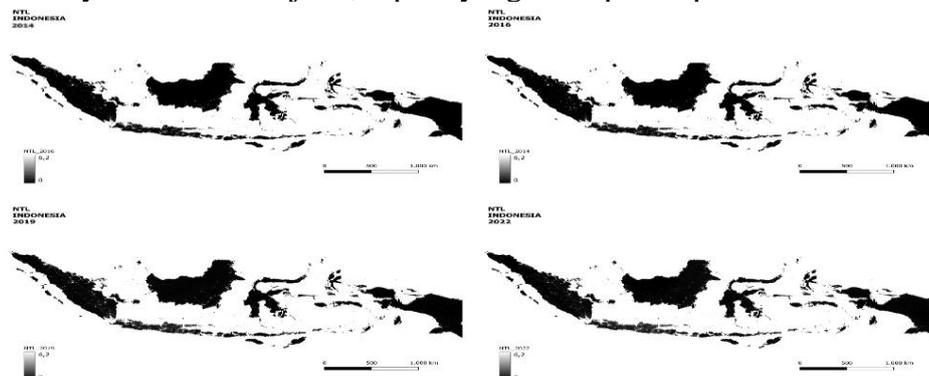


Gambar 3. Kerangka Pikir Penelitian

HASIL DAN PEMBAHASAN

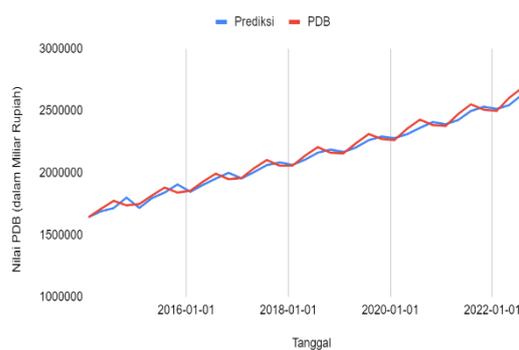
a. Hasil estimasi PDB dengan data *Night-Time Light* (NTL)

Data cahaya langit malam dikumpulkan dari NOAA VIIRS dengan media *Earth Engine*. Data yang dikumpulkan adalah data median NTL bulanan tiap tahunnya dari 2014 hingga 2022. Untuk melihat visualisasinya, dilakukan normalisasi terlebih dahulu agar perubahan cahaya terlihat lebih jelas, seperti yang ditampilkan pada Gambar 4.

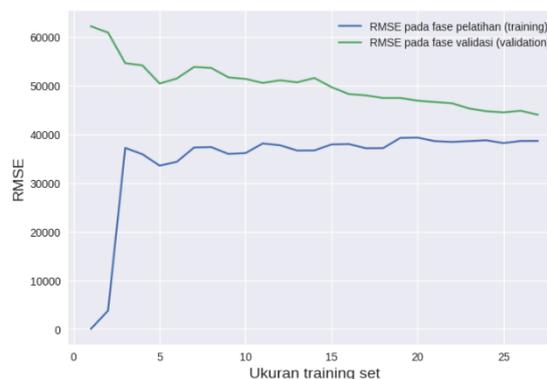


Gambar 4. Visualisasi NTL 2014-2022

Selanjutnya, data NTL diagregasi dengan rata-rata per 3 bulan. Lalu dilakukan uji korelasi dengan hasil uji 0,758 dan $p\text{-value } 8 \times 10^{-8}$ yang mana nilai tersebut kurang dari 0,05 sehingga menunjukkan bahwa terdapat korelasi yang kuat dan signifikan antara variabel NTL dan PDB. Dengan demikian, NTL dapat digunakan sebagai variabel prediktor dalam penelitian ini. NTL yang telah dikumpulkan melalui satelit NOAA-VIIRS secara bulanan akan dilakukan agregasi dengan rata-rata per 3 bulannya. Setelah itu, data dipisahkan menjadi dua bagian, yaitu data *training* (2014–2022) dan data *testing* (2022). Selanjutnya, data akan dimasukkan ke dalam model dan dilakukan *tuning* parameter dengan menggunakan *cross-validation 5-fold*.



Gambar 5. Nilai PDB dan Hasil Estimasi PDB dengan NTL



Gambar 6. *Learning Curve* Model Estimasi PDB dengan Variabel NTL

Dari Gambar 5, dapat dilihat bahwa nilai PDB dan nilai PDB hasil estimasi dengan NTL memiliki tren yang sama meskipun nilainya cenderung berbeda. Hal ini menunjukkan bahwa estimasi dengan menggunakan data NTL dapat dilakukan untuk melihat tren PDB secara triwulanan. Hal ini juga menunjukkan bahwa NTL dapat menjadi *proxy* variabel yang cukup baik untuk estimasi PDB pada skala triwulanan. Untuk mengetahui performa model dan melihat apakah model *overfitting* atau tidak, dilakukan evaluasi dengan mengukur nilai kesalahan terhadap ukuran data dengan pendekatan *learning curve*. Seperti yang tertera pada Gambar 6, *gap* antara kurva pembelajaran (biru) dengan kurva validasi cenderung menyempit seiring bertambahnya ukuran data yang dimasukkan ke dalam model. Hal ini menunjukkan bahwa mesin dapat belajar dengan baik dan tidak terjadi *overfitting* pada model *XGBoost* dengan variabel NTL. Selanjutnya, untuk mengukur performa model dalam melakukan estimasi, dilakukan pengukuran dengan metrik RMSE dan MAPE.

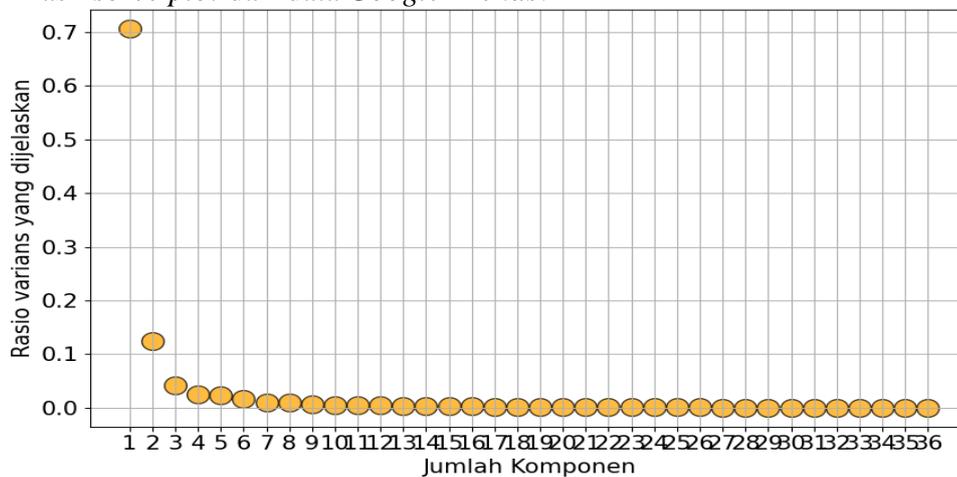
Tabel 4. Evaluasi Model Estimasi PDB dengan Variabel NTL

Metrik	Hasil Uji	Hasil Keseluruhan
RMSE	38396,1	37967,1
MAPE(%)	1,2	1,56

Dari Tabel 4, diperoleh bahwa RMSE pada data *testing* nilai RMSE sebesar 38396,1 dan MAPE sebesar 1,2%. Sementara itu, jika model diuji dengan keseluruhan data, diperoleh RMSE sebesar 37967,1 dan MAPE sebesar 1,56%. Ukuran ini menunjukkan model dengan data NTL cukup baik untuk digunakan sebagai variabel prediktor PDB.

b. Hasil estimasi PDB dengan data Google Trends

Estimasi PDB dilakukan dengan menggunakan variabel *Google Trends* yang berasal dari *API Gtrends*. Data dikumpulkan dari 115 *keywords* dengan asumsi *keywords* tersebut berpengaruh pada PDB dan menunjukkan pola ekonomi masyarakat (OECD, 2020). Selanjutnya, karena jumlah variabel atau *keywords* yang digunakan melebihi jumlah amatan (jumlah *keywords* (50) > jumlah amatan (36)) dan untuk mengurangi tingkat multikolinearitas pada masing-masing *keywords*, dilakukan pengelompokkan dengan PCA. Untuk mendapatkan berapa jumlah komponen utama yang signifikan, salah satu cara yang dapat digunakan adalah dengan melihat titik pada *scree plot* ketika kurva mulai melandai. Titik potongan atau "*elbow*" pada kurva *scree plot* merupakan titik yang menggambarkan perubahan tajam dalam nilai *eigenvalue* antara satu komponen utama dan komponen berikutnya. Pada titik ini, penambahan komponen utama berikutnya tidak memberikan peningkatan signifikan dalam varians yang dijelaskan. Jumlah komponen utama di atas titik *elbow* biasanya dipilih sebagai jumlah komponen yang akan dipertahankan. Berikut disajikan hasil *scree plot* dari data *Google Trends*.



Gambar 7. Scree Plot

Berdasarkan kurva yang tertera pada Gambar 7, dapat diketahui bahwa titik *elbow* berada di kisaran angka 3 atau 4. Namun, interpretasi dengan *scree plot* saja tidak cukup untuk menentukan berapa jumlah komponen utama yang dapat digunakan karena *scree plot* melibatkan penilaian visual yang cukup subjektif. Maka dari itu, dibutuhkan kriteria pemilihan jumlah komponen utama yang lebih objektif, yaitu dengan *eigenvalue* dan keragaman kumulatif. Komponen utama yang dapat digunakan adalah komponen utama yang memiliki *eigenvalue* ≥ 1 dan keragaman kumulatif minimal 80%. Berikut disajikan tabel statistik komponen utama yang signifikan atau telah memenuhi kriteria yang ditentukan.

Tabel 5. Statistik Komponen Utama

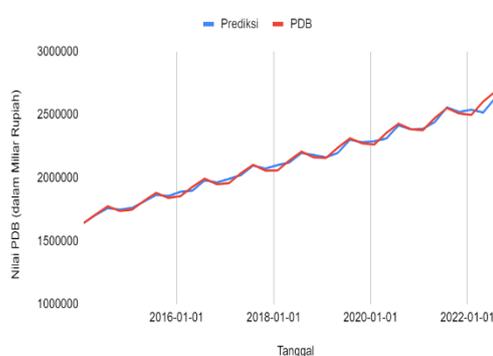
	PC1	PC2	PC3	PC4
<i>Eigenvalue</i>	36	6	2	1
Varians Kumulatif	0,70702548	0,83166016	0,87305316	0,89749456

Berdasarkan Tabel 5, varians atau keragaman kumulatif dengan 4 komponen utama dapat menjelaskan sebesar 89,75% dari keseluruhan data yang berjumlah 50 variabel. Nilai

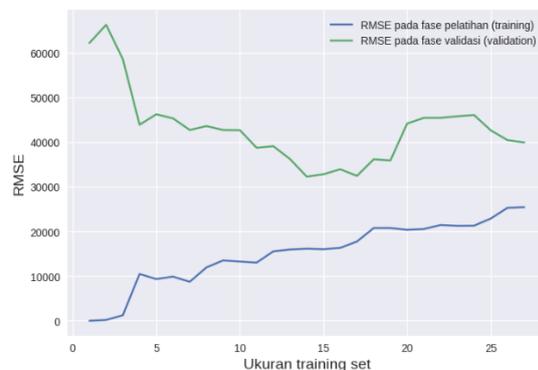
persentase varians kumulatif tersebut sudah mendekati 100%, hal ini menunjukkan bahwa 4 komponen utama yang dipertahankan sudah cukup untuk menjelaskan sebagian besar variasi dalam data.

Komponen utama pertama (PC1) berisi variabel Kebisingan, Pengangguran, Kesejahteraan dan Pengangguran, Minat, Makanan dan Minuman, Hari Ulang Tahun, Energi dan Utilitas, Kesehatan, Farmasi, Produk dan Layanan Hewani, Agen Perjalanan dan Layanan, Resume dan Portofolio, Kebangkrutan, Perbaikan Rumah, Konstruksi dan Pemeliharaan, Perencanaan Keuangan, Manajemen Data, Teknologi Perusahaan, Alat Pengembangan, Mencetak dan Menerbitkan, Acara Perusahaan, Distribusi dan Logistik, Konsultasi, Operasi Bisnis, Pertanian dan Kehutanan, Bisnis dan Industri, Transportasi dan Logistik, serta Manufaktur. Dari variabel-variabel yang menyusun komponen utama 1 ini, maka komponen utama 1 dapat dikategorikan sebagai variabel konsumsi, pekerjaan dan manajemen bisnis. Komponen utama 1 ini mampu menjelaskan 70,7% keragaman dari seluruh keragaman data *Google Trends*.

Komponen utama kedua (PC2) mencakup variabel Pinjaman Pelajar, Kredit dan Pinjaman, Meminjamkan, Hipotek, Rumah dan Taman, Pekerjaan Pengembang, serta variabel Apartemen dan Penyewaan Hunian sehingga komponen utama 2 ini dapat dikategorikan sebagai variabel Kredit. Keragaman data yang dapat digambarkan oleh komponen utama 2 ini sudah mencapai 83,17% dari seluruh keragaman data. Sementara itu, secara kumulatif, komponen utama 3 (PC3) mampu menjelaskan 87,3% variasi data dari *Google Trends*. Variabel-variabel yang termasuk ke dalam komponen utama 3 ini yaitu variabel Krisis Ekonomi, Resesi, Pengecer Bahan Makanan dan Makanan, Peralatan Rumah Tangga, Truk dan SUV, Dokter Hewan, Lantai, Pelapis dan Perekat, Produksi Makanan, Pewarna dan Pigmen, serta Pengiriman dan Pengangkutan. Maka dari itu, komponen utama 3 dapat dikategorikan sebagai Krisis dan Kebutuhan Sekunder.



Gambar 8. Nilai PDB dan Hasil Estimasi PDB dengan *Google Trends*



Gambar 9. *Learning Curve* Model Estimasi PDB dengan Variabel *Google Trends*

Komponen utama 4 (PC4) mencakup variabel Olahraga, Pekerjaan Sementara, Investasi, dan Berinvestasi sehingga komponen utama 4 dapat dikategorikan sebagai Investasi dan Aktivitas Ekonomi Tambahan. Komponen utama 4 memiliki varians kumulatif sebesar 0,8975. Artinya, dengan 4 komponen utama, mampu menjelaskan 89,75% keragaman data *Google Trends* dengan baik. Dengan demikian, jumlah variabel dari *Google Trends* yang semula berjumlah 50 *keywords* dapat direduksi dimensinya menjadi 4 komponen utama atau 4 kategori saja. Dari 4 komponen utama tersebut juga telah mampu menjelaskan keragaman data sebesar 89,75% dari total keseluruhan data dan meminimalkan

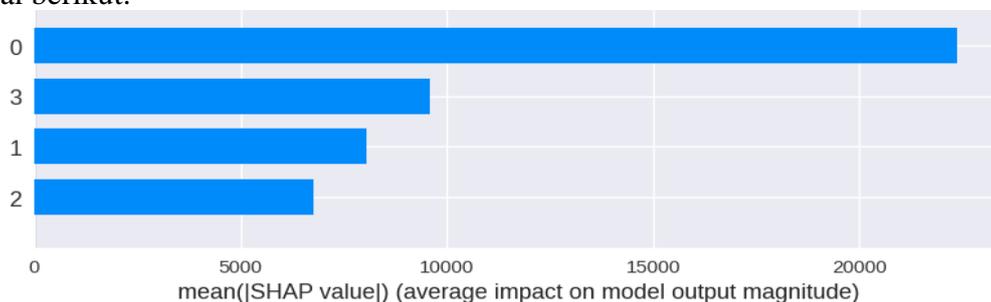
potensi kehilangan informasi. Dari hasil PCA tersebut, dilakukan estimasi pada model menggunakan data hasil PCA *Google Trends*.

Berdasarkan Gambar 8, ditemukan bahwa nilai PDB dan nilai PDB hasil estimasi dengan *Google Trends* cenderung sama, meskipun pada tahun 2022 nilai estimasi cenderung meleset dari nilai sebenarnya. Hal ini menunjukkan bahwa estimasi dengan menggunakan *Google Trends* dengan *keywords* tertentu dapat dilakukan dengan *keywords* yang tepat. Hal ini juga menunjukkan *Google Trends* dapat menjadi variabel prediktor yang baik untuk PDB pada skala triwulanan. Untuk mengetahui performa model dan melihat apakah model *overfitting* atau tidak, dilakukan evaluasi dengan mengukur nilai kesalahan terhadap ukuran data dengan pendekatan *learning curve*. Berdasarkan *learning curve* yang tertera pada Gambar 9, *gap* antara kurva pembelajaran (biru) dengan kurva validasi cenderung menyempit seiring bertambahnya ukuran data yang dimasukkan ke dalam model. Hal ini menunjukkan bahwa mesin dapat belajar dengan baik dan tidak terjadi *overfitting* pada model *XGBoost* dengan variabel *Google Trends*. Lalu, untuk mengukur performa model dalam melakukan estimasi, dilakukan pengukuran dengan metrik RMSE dan MAPE dengan hasil sebagai berikut:

Tabel 6. Evaluasi Model Estimasi PDB dengan Variabel *Google Trends*

Metrik	Hasil Uji	Hasil Keseluruhan
RMSE	61352,98	26639,702
MAPE(%)	2,02	0,92

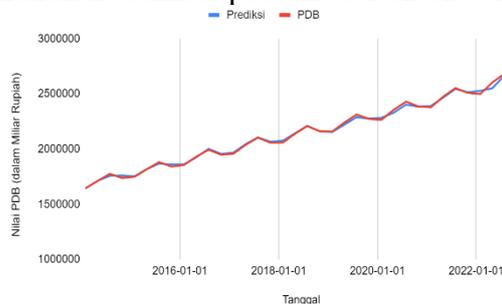
Berdasarkan Tabel 6, Nilai RMSE pada data *testing* nilai RMSE sebesar 61352,98 dan MAPE sebesar 2,02. Hasil ini menunjukkan tingkat kesalahan yang lebih tinggi jika menggunakan NTL sebagai variabel prediktor untuk PDB. Sementara itu, jika model diuji dengan keseluruhan data, diperoleh nilai RMSE sebesar 26639,702 dan MAPE sebesar 0,92. Angka ini menunjukkan tingkat kesalahan yang lebih kecil. Oleh karena itu, ukuran ini menunjukkan model *Google Trends* mampu bekerja lebih baik dibandingkan dengan NTL sebagai variabel prediktor. Hal ini juga sesuai dengan interpretasi pada visualisasi pada Gambar 8 dimana hasil visualisasi menunjukkan estimasi PDB dengan *Google Trends* cenderung sama. Dari hasil tersebut, model akan diinterpretasikan dengan plot nilai SHAP sebagai berikut:



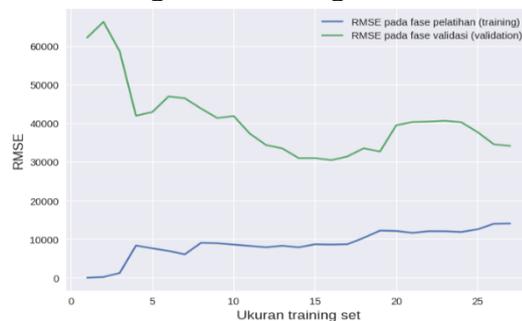
Gambar 10. Nilai SHAP dari Model *XGBoost* dengan Variabel *Google Trends*

Berdasarkan Gambar 10, diperoleh bahwa komponen 0 atau kelompok kata dengan topik konsumsi, pekerjaan dan manajemen bisnis menjadi variabel paling berpengaruh pada model dengan nilai rata-rata SHAP tertinggi dibandingkan variabel lainnya. Sementara itu, komponen 2 atau kelompok kata dengan topik krisis dan kebutuhan sekunder menjadi variabel yang paling tidak berpengaruh dengan nilai SHAP paling rendah. Data *Google*

Trends yang telah dikelompokkan dengan PCA dan data NTL akan dimasukkan ke dalam model *XGBoost*. Model *XGBoost* di-*tuning* dengan *cross validation 5 fold*. Parameter terbaik dari model *XGBoost* kemudian dipilih untuk digunakan dalam estimasi. Setelah di-*tuning* model *XGBoost* terbaik adalah model dengan parameter *max_depth* sebesar 1, *n_estimators* sebanyak 800 pohon, dan *learning_rate* sebesar 0,01. Dari model tersebut, dilakukan estimasi pada model untuk semua *dataset* dengan hasil sebagai berikut.



Gambar 11. Nilai PDB dan Hasil Estimasi PDB dengan NTL dan *Google Trends*



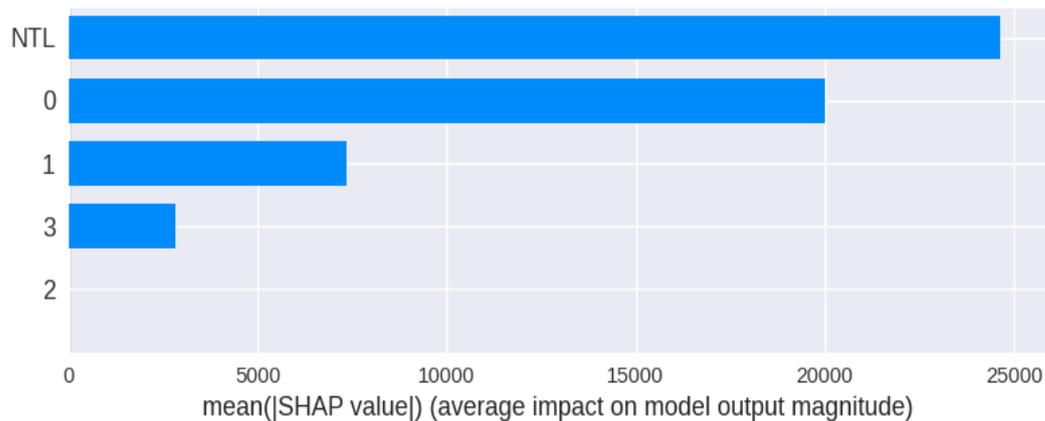
Gambar 12. *Learning Curves* Model *XGBoost* dengan Variabel *Google Trends* & NTL

Pada Gambar 11, ditunjukkan bahwa nilai PDB dan nilai PDB hasil estimasi dengan NTL dan *Google Trends* cenderung sama. Hal ini menunjukkan bahwa estimasi dengan menggunakan data NTL dan *Google Trends* dengan *keywords* tertentu dapat dilakukan. Hal ini juga menunjukkan bahwa NTL dan *Google Trends* dapat menjadi *proxy* variabel yang baik untuk PDB pada skala triwulanan. Untuk mengetahui performa model dan melihat apakah model *overfitting* atau tidak, dilakukan evaluasi dengan mengukur nilai kesalahan terhadap ukuran data dengan pendekatan *learning curve*. Dari *learning curve* yang tertera pada Gambar 12, ditunjukkan bahwa *gap* antara kurva pembelajaran (biru) dengan kurva validasi cenderung menyempit seiring bertambahnya ukuran data yang dimasukkan ke dalam model. Hal ini menunjukkan bahwa mesin dapat belajar dengan baik dan tidak terjadi *overfitting* pada model *XGBoost* dengan variabel NTL dan *Google Trends*. Selanjutnya, untuk mengukur performa model dalam melakukan estimasi, dilakukan pengukuran dengan metrik RMSE dan MAPE.

Tabel 7. Evaluasi Model Estimasi PDB dengan Variabel NTL dan *Google Trends*

Metrik	Hasil Uji	Hasil Keseluruhan
RMSE	71499,03	15793,72
MAPE(%)	2,1	0,523

Nilai RMSE pada data *testing* sebesar 71499,03 dan MAPE sebesar 2,1% (lihat Tabel 7). Hasil ini cenderung lebih buruk dibandingkan dengan hanya menggunakan NTL saja atau *Google Trends* saja sebagai variabel prediktor. Akan tetapi, jika model diuji dengan keseluruhan data, diperoleh nilai RMSE sebesar 15793,72 dan MAPE sebesar 0,523%. Hal ini menunjukkan tingkat kesalahan yang lebih baik dari penggunaan NTL saja dan *Google Trends* saja sebagai variabel prediktor. Maka dari itu, ukuran ini menunjukkan bahwa model dengan data NTL dan *Google Trends* mampu bekerja dengan lebih baik dibandingkan dengan memasukkan salah satu dari kedua data tersebut. Hal ini sesuai dengan interpretasi pada visualisasi pada Gambar 11. Dari hasil tersebut, model akan diinterpretasikan dengan plot nilai SHAP sebagai berikut:



Gambar 13. Nilai SHAP dari Model *XGBoost* dengan variabel *Google Trends* dan NTL

Dari Gambar 13, ditemukan bahwa variabel NTL menjadi variabel paling berpengaruh pada model dengan nilai rata-rata SHAP tertinggi dibandingkan variabel lainnya. Selanjutnya, komponen 0 atau kelompok kata dengan topik konsumsi, pekerjaan dan manajemen bisnis menjadi komponen dari *Google Trends* paling berpengaruh pada model dibandingkan komponen lainnya. Sementara itu, komponen 2 atau kelompok kata dengan topik krisis dan kebutuhan sekunder menjadi variabel yang paling tidak berpengaruh dengan nilai SHAP mendekati nol.

Untuk membandingkan dan mengevaluasi kinerja prediksi ketiga model, yaitu model dengan NTL, *Google Trends*, serta kombinasi antara NTL dan *Google Trends*, berikut disajikan tabel perbandingan ketiga model berdasarkan metrik evaluasinya, yaitu RMSE dan MAPE.

Tabel 8. Perbandingan evaluasi ketiga model

Model	RMSE	MAPE (%)
NTL	37967,1	1,56
Google Trends	26639,7	0,92
NTL dan Google Trends	15793,7	0,52

Berdasarkan Tabel 8 dapat diketahui bahwa model dengan NTL dan *Google Trends* memiliki MAPE yang paling rendah diantara kedua model lainnya, yaitu sebesar 0,52%. Selain itu, nilai RMSE model dengan variabel NTL dan *Google Trends* juga yang paling rendah, yaitu sebesar 15793,7. Dengan demikian, dapat dikatakan bahwa model dengan *Google Trends* dan NTL memiliki kinerja prediksi yang paling baik dibandingkan dengan model yang menggunakan NTL saja maupun model yang menggunakan *Google Trends* saja. Hasil penelitian ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa data dari *Google Trends* dan NTL dapat menjadi *proxy variable* untuk memprediksi PDB, dengan data NTL yang mampu menggambarkan aktivitas manusia serta pertumbuhan ekonomi di suatu wilayah, sementara *Google Trends* menunjukkan pola ekonomi masyarakat melalui *keywords* yang dicari (OECD, 2020; Li & Li, 2014).

KESIMPULAN

Berdasarkan hasil dan pembahasan diatas dapat disimpulkan bahwa estimasi dengan menggunakan *Night Time Light* (NTL), *Google Trends*, dan penggunaan keduanya dapat digunakan sebagai *proxy variable* untuk PDB karena model yang terbentuk tidak mengalami *overfitting* dan memiliki hasil evaluasi uji dengan $MAPE < 10\%$. Penggunaan data NTL dan *Google Trends* secara bersamaan sebagai variabel *proxy* dalam memprediksi PDB menjadi skenario terbaik untuk memprediksi PDB karena menghasilkan model dengan hasil evaluasi keseluruhan terbaik dengan RMSE 15793,72 dan MAPE sebesar 0,52%.

DAFTAR PUSTAKA

- Abidin, M. I., Notodiputro, K. A., & Sartono, B. (2021). Improving Classification Model Performances using an Active Learning Method to Detect Hate Speech in Twitter: Peningkatan Kinerja Model Klasifikasi dengan Pembelajaran Aktif dalam Mendeteksi Ujaran Kebencian di Twitter. *Indonesian Journal of Statistics and Its Applications*, 5(1), 26–38. <https://doi.org/10.29244/ijsa.v5i1p26-38>
- Adwendi, S. J., & Kartiasih, F. (2016). Penggunaan Error Correction Mechanism dalam Analisis Pengaruh Investasi Langsung Luar Negeri Terhadap Pertumbuhan Ekonomi Indonesia. *Statistika: Journal of Theoretical Statistics and Its Applications*, 16(1), 17–27. <https://doi.org/10.29313/jstat.v16i1.1767>
- Afrianto, F., & Graha, D. (2023). Morfologi Kota Malang: Sebuah Tinjauan dari Nighttime Light Satellite Imagery. 3, 68–76. <https://doi.org/10.36456/jpb.v3i2.7002>
- Aguilar, M., & Xavier, K. (2022). A comparison between machine learning and classic algorithms for GDP forecast [Master thesis, Universitat Politècnica de Catalunya]. <https://upcommons.upc.edu/handle/2117/375350>
- Akbar, I. A., & Kurniawan, R. (2020). Pemodelan Nowcasting Tingkat Pengangguran Terbuka Menggunakan Data Google Trends Dengan Metode Antlion Optimization-Support Vector Regression. *Seminar Nasional Official Statistics, 2020(1)*, 60-770. <https://doi.org/10.34123/semnasoffstat.v2020i1.504>.
- Arora, V. S., McKee, M., & Stuckler, D. (2019). Google Trends: Opportunities and limitations in health and health policy research. *Health Policy*, 123(3), 338-341. <https://doi.org/10.1016/j.healthpol.2019.01.001>.
- Badan Pusat Statistik. (2022). Produk Domestik Bruto Indonesia 2018-2022. Jakarta: BPS RI.
- Bank Indonesia. (2019). Perkembangan Ekonomi Global. Perkembangan Ekonomi Keuangan dan Kerja sama Internasional, Jakarta, Bank Indonesia.
- Bantis, E., Clements, M. P., & Urquhart, A. (2022). Forecasting GDP growth rates in the United States and Brazil using Google Trends. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2022.10.003>.
- Bouayad, I., Zahir, J., & Ez-Zetouni, A. (2022). Nowcasting and Forecasting Morocco GDP growth using Google Trends data. *IFAC-PapersOnLine*, 55(10), 3280–3285. <https://doi.org/10.1016/j.ifacol.2022.10.129>
- Bro, R., & Smilde, A. K. (2014). Principal Component Analysis. *Analytical methods*, 6(9), 2812-2831.
- Brown, R. H., Vitullo, S. R., Corliss, G. F., Adya, M., Kaefer, P. E., & Povinelli, R. J. (2015). Detrending Daily Natural Gas Consumption Series to Improve Short-Term Forecasts. *2015 IEEE Power & Energy Society General Meeting*. doi:10.1109/pesgm.2015.7286138
- Chen, J., Zhao, F., Sun, Y., & Yin, Y. (2020). Improved XGBoost model based on genetic algorithm. *International Journal of Computer Applications in Technology*, 62(3),

240. <https://doi.org/10.1504/IJCAT.2020.106571>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Coefficient of Determination R-Squared is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>
- Choi, H. & Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record*, 88, 2-9. DOI: 10.1111/j.1475-4932.2012.00809.x.
- Das, S., Seferbekov, S., Datta, A., Islam, M. S., & Amin, M. R. (2021). Towards Solving the DeepFake Problem: An Analysis on Improving DeepFake Detection Using Dynamic Face Augmentation. 3776–3785. https://openaccess.thecvf.com/content/ICCV2021W/RPRMI/html/Das_Towards_Solving_the_DeepFake_Problem_An_Analysis_on_Improving_DeepFake_ICCV_W_2021_paper.html
- Fadillah, K., & Kartiasih, F. (2023). The effect of COVID-19 and population mobility on the underemployment rate in Indonesia. *Jurnal Kependudukan Indonesia*, 18(2), 217–236. <https://doi.org/10.55981/jki.2023.2037>
- Fauzan, F. D., Rayhan, D. A., Putri, H. M., & Kartiasih, F. (2024). Peramalan Konsentrasi PM2.5 Menggunakan Model ARCH/GARCH dan Long Short-Term Memory (Studi Kasus Kota Jakarta Pusat). *INFOMATEK: Jurnal Informatika, Manajemen Dan Teknologi*, 26(1), 27–44. <https://doi.org/10.23969/infomatek.v26i1.12603>
- Ferrara, L., & Simoni, A. (2023). When are Google Data Useful to Nowcast GDP? An Approach via Preselection and Shrinkage. *Journal of Business and Economic Statistics*, 41(4), 1188–1202. <https://doi.org/10.1080/07350015.2022.2116025>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Ginting, C. P., & Kartiasih, F. (2019). Analisis Ekspor Kopi Indonesia ke Negara-Negara ASEAN. *Jurnal Ilmiah Ekonomi Dan Bisnis*, 16(2), 143–157. <https://doi.org/10.31849/jieb.v16i2.2922>
- Götz, T. B., & Knetsch, T. A. (2019). Google data in bridge equation models for German GDP. *International Journal of Forecasting*, 35(1), 45–66. <https://doi.org/10.1016/j.ijforecast.2018.08.001>
- Guo, R., Zhao, Z., Wang, T., Liu, G., Zhao, J., & Gao, D. (2020). Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. *Applied Sciences*, 10, 6593. <https://doi.org/10.3390/app10186593>
- Hawari, R., & Kartiasih, F. (2017). Kajian Aktivitas Ekonomi Luar Negeri Indonesia Terhadap Pertumbuhan Ekonomi Indonesia Periode 1998-2014. *Media Statistika*, 9(2), 119. <https://doi.org/10.14710/medstat.9.2.119-132>
- Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian di Kabupaten Bojonegoro. *JIEET (Journal of Information Engineering and Educational Technology)*, 5(2), 49-54. <https://doi.org/10.26740/jieet.v5n2.p49-54>
- Heikkinen, J. (2019). Nowcasting GDP Growth Using Google. *Master's Thesis*.
- Hindle, B. R., Keogh, J. W. L., & Lorimer, A. V. (2020). Validation of Spatiotemporal and Kinematic Measures in Functional Exercises Using a Minimal Modeling Inertial Sensor Methodology. *Sensors*, 20(16), Article 16. <https://doi.org/10.3390/s20164586>

- Hodson, T. O. (2022). Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Innayah, A. N., Sulistiana, D. I., Febrian, M. Y., & Kartiasih, F. (2024). Utilizing The Sarima Model and Support Vector Regression to Forecast Monthly Rainfall in Bandung City. *JITTER (Jurnal Ilmiah Teknologi Informasi Terapan)*, 10(2), 88–104. <https://doi.org/10.33197/jitter.vol10.iss2.2024.1663>
- Jange, B. (2022). Prediksi Harga Saham Bank BCA Menggunakan XGBoost. *ARBITRASE: Journal of Economics and Accounting*, 3(2), Article 2. <https://doi.org/10.47065/arbitrase.v3i2.495>
- Jiang, Y., Tong, G., Yin, H., & Xiong, N. (2019). A Pedestrian Detection Method Based on Genetic Algorithm for Optimize XGBoost Training Parameters. *IEEE Access*, 7, 118310–118321. <https://doi.org/10.1109/ACCESS.2019.2936454>
- Jolliffe, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). <https://doi.org/10.1098/rsta.2015.0202>
- Jovic, A., Brkic, K., & Bogunovic, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200–1205. doi:10.1109/mipro.2015.7160458.
- Kapetanakis, D.S., Mangina, E., & Finn, D. P. (2017). Input variable selection for thermal load predictive models of commercial buildings. *Energy and Buildings*, 137, 13–26. doi:10.1016/j.enbuild.2016.12.016.
- Kartiasih, F. (2019). Dampak Infrastruktur Transportasi Terhadap Pertumbuhan Ekonomi Di Indonesia Menggunakan Regresi Data Panel. *Jurnal Ilmiah Ekonomi Dan Bisnis*, 16(1), 67–77. <https://doi.org/10.31849/jieb.v16i1.2306>
- Kartiasih, F. (2019). Transformasi Struktural dan Ketimpangan Antardaerah di Provinsi Kalimantan Timur. *INOVASI: Jurnal Ekonomi, Keuangan Dan Manajemen*, 15(1), 105–113. <https://doi.org/https://doi.org/10.30872/jinv.v15i1.5201>
- Kusumasari, A., & Kartiasih, F. (2017). Aglomerasi Industri dan Pengaruhnya Terhadap Pertumbuhan Ekonomi Jawa Barat 2010-2014. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 9(2), 28–41. <https://doi.org/https://doi.org/10.34123/jurnalasks.v9i2.143>
- Latifa, A., Primadani, A. D. P., Fitriyyah, N. R., & Kartiasih, F. (2023). Mapping and Estimating the Impact of Drought on Food Crop Farmers Using Remote Sensing in East Nusa Tenggara Province. *TheJournalish: Social and Government*, 4(5), 309–335.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014b). The Parable of Google Flu: Traps in *Big data* Analysis. *Science*, 343(6176), 1203–1205. <https://doi.org/10.1126/science.1248506>
- Lever, J., Krzywinski, M., & Altman, N. (2017). Points of Significance: Principal Component Analysis. *Nature Methods*, 14(7), 641–642. doi:10.1038/nmeth.4346.
- Li, M., Wang, F., Jia, X., Li, W., Li, T., & Rui, G. (2021). Multi-source data fusion for economic data analysis. *Neural Computing and Applications*, 33, 4729–4739.
- Li, X., Ge, L., & Chen, X. (2014). Quantifying Contribution of Land Use Types to Nighttime Light Using an Unmixing Model. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1667–1671. <https://doi.org/10.1109/LGRS.2014.2304496>
- Marcilio, W. E., & Eler, D. M. (2020). From explanations to feature selection: Assessing

- SHAP values as feature selection mechanism. *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 340–347. <https://doi.org/10.1109/SIBGRAPI51738.2020.00053>
- Ningsih, Y. P., & Kartiasih, F. (2019). Dampak Guncangan Pertumbuhan Ekonomi Mitra Dagang Utama terhadap Indikator Makroekonomi Indonesia. *Jurnal Ilmiah Ekonomi Dan Bisnis*, 16(1), 78–92. <https://doi.org/https://doi.org/10.31849/jieb.v16i1.2307>
- OECD. (2020). Tracking GDP growth in real time. <https://www.oecd.org/economy/weekly-tracker-of-gdp-growth>
- Pandika Pinata, N. N., Sukarsa, I. M., & Dwi Rusjayanthi, N. K. (2020). Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 188. <https://doi.org/10.24843/JIM.2020.v08.i03.p04>
- Panigrahi, S., & Behera, H. S. (2013). Effect of Normalization Techniques on Univariate Time Series Forecasting Using Evolutionary Higher Order Neural Network. *International Journal of Engineering and Advanced Technology*, 3(2), 280–285.
- Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. (Kouros). (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136, 105405. <https://doi.org/10.1016/j.aap.2019.105405>
- Pemayun, A. A. G. R. B. D., Azizi, M. Z., Dauly, N. A., Apriliani, N. H., & Kartiasih, F. (2024). Estimation of Java GRDP in Regency/City Level: Satellite Imagery and Machine Learning Approaches. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, X(2), 379–386. <http://dx.doi.org/10.33330/jurteks.v10i2.2993>
- Putri, S. R., Wijayanto, A. W., & Sakti, A. D. (2022). Developing Relative Spatial Poverty Index Using Integrated Remote Sensing and Geospatial Big Data Approach: A Case Study of East Java, Indonesia. *ISPRS International Journal of Geo-Information*, 11(5). <https://doi.org/10.3390/ijgi11050275>
- Qureshi, S., Chu, B. M., & Demers, F. S. (2020). Forecasting Canadian GDP Growth Using XGBoost (No. 20-14). Carleton University, Department of Economics.
- Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., Chen, L., & Wu, J. (2014). Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data. *Remote Sensing*, 6(2), Article 2. <https://doi.org/10.3390/rs6021705>
- Sitepu, A. A., Tantular, B., Darmawan, G., Pontoh, R. S., & Faidah, D. Y. (2023). Pemodelan Produk Domestik Bruto (Pdb) Dengan Pendekatan Vector Error Correction Model (VECM). *PRIMER: Jurnal Ilmiah Multidisiplin*, 1(2), Article 2. <https://doi.org/10.55681/primer.v1i2.50>
- Song, F., Guo, Z., & Mei, D. (2010). Feature Selection Using Principal Component Analysis. *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*. 27–30. doi:10.1109/icsem.2010.14
- Song, F., Guo, Z., & Mei, D. (2010). Feature Selection Using Principal Component Analysis. *Proceedings - 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, ICSEM 2010, 1*, 27–30. <https://doi.org/10.1109/ICSEM.2010.14>
- Subian, A. R., Mulkan, D. A., Ahmady, H. H., & Kartiasih, F. (2024). Comparison Methods of Machine Learning and Deep Learning to Forecast The GDP of Indonesia. *SISTEMASI: Jurnal Sistem Informasi*, 13(1), 149–166.

- <https://doi.org/10.32520/stmsi.v13i1.3445>
- Tinungki, G. M., & Sunusi, N. (2019). Penerapan Sparse Principal Component Analysis dalam Menghasilkan Matriks Loading yang Sparse. *Jurnal Matematika, Statistika Dan Komputasi*, 15(2), 44-53. <https://doi.org/10.20956/jmsk.v15i2.5713>.
- Umam, K., & Kartiasih, F. (2023). Resiliensi Wilayah terhadap Shock Pandemi COVID-19 : Studi Kasus Kabupaten/Kota di Pulau Jawa. *Jurnal Ekonomi Indonesia*, 12(3), 283–304. <https://doi.org/10.52813/jei.v12i3.415>
- Weley, I. R., Kumenaung, A. G., & Sumual, J. I. (2019). Analisis Pengaruh Inflasi dan Produk Domestik Regional Bruto (PDRB) Terhadap Pendapatan Asli Daerah di Kota Manado. *Jurnal Pembangunan Ekonomi dan Keuangan Daerah*, 19(3). <https://doi.org/10.35794/jpekd.16457.19.3.2017>
- Widayati, C. S. W. (2009). Komparasi Beberapa Metode Estimasi Kesalahan Pengukuran. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 13(2), Article 2. <https://doi.org/10.21831/pep.v13i2.1409>
- Zhang, C., Vinyals, O., Munos, R., & Bengio, S. (2018). A Study on Overfitting in Deep Reinforcement Learning (arXiv:1804.06893). arXiv. <http://arxiv.org/abs/1804.06893>.