

FLASH ESTIMATES OF HOUSEHOLD FINAL CONSUMPTION EXPENDITURE IN EAST JAVA

Rizky Zulkarnain*¹

¹Direktorat Neraca Produksi, Badan Pusat Statistik, Indonesia

ABSTRACT

The purpose of this study is to construct a flash estimates model for household final consumption expenditure in East Java. There were 153 indicators that were used, covering Big Data and non Big Data indicators. Indicators from Big Data were the Google Trends indicators. The Google Trends categories were selected based on the highest correlation coefficient. While the non Big Data indicators were the total deposit, consumption credit, consumer lending rate, term deposit rate (1, 3, and 6 months), withdrawal (outflow), and Consumer Price Index (CPI). This study examines several models: ARIMA, ADL, Elastic Net, Support Vector Machine, Random Forest and Ensemble. The models were examined using various scenarios of out-of-sample and estimation periods (t+10 days, t+20 days and t+30 days). The prediction performance of models were evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The evaluation results showed that the Elastic Net and the Ensemble were the best model for any scenarios. Both models had good performance since t+10 days.

*Korespondensi:
Rizky Zulkarnain
E-mail:
zulqarnaen@bps.go.id

Keywords: Big Data, ensemble, Google Trends, machine learning, nowcasting

ABSTRAK

Penelitian ini bertujuan untuk membentuk model estimasi cepat konsumsi rumah tangga di Jawa Timur. Terdapat 153 indikator yang digunakan, yang terdiri atas indikator Big Data dan indikator selain Big Data. Indikator Big Data yang digunakan adalah data Google Trends, yang kategorinya diseleksi berdasarkan koefisien korelasi yang paling kuat. Sedangkan, indikator selain Big Data yang digunakan adalah simpanan masyarakat, pinjaman konsumsi, suku bunga pinjaman konsumsi, suku bunga simpanan berjangka (1, 3, dan 6 bulan), aliran uang keluar, dan Indeks Harga Konsumen (IHK). Penelitian ini menguji beberapa model: ARIMA, ADL, Elastic Net, Support Vector Machine, Random Forest dan teknik Ensemble. Keenam model tersebut diuji pada berbagai skenario out-of-sample dan periode estimasi (H+10, H+20 dan H+30). Kinerja prediksi dari berbagai model dievaluasi menggunakan Root Mean Squared Error (RMSE) dan Mean Absolute Percentage Error (MAPE). Hasil evaluasi menunjukkan bahwa model Elastic Net dan Ensemble menghasilkan kinerja terbaik pada berbagai skenario. Hasil prediksi kedua model tersebut sudah cukup baik pada H+10.

Kata Kunci: Big Data, ensemble, Google Trends, machine learning, nowcasting

JEL : C15; E27; R11

Pendahuluan

Salah satu prasyarat dalam pembuatan kebijakan ekonomi yang baik adalah tersedianya informasi yang tepat waktu tentang kondisi perekonomian terkini, khususnya ketika aktivitas ekonomi berubah dengan cepat (Woloszko, 2020). Indikator yang digunakan secara luas

untuk menilai kinerja perekonomian regional adalah Produk Domestik Regional Bruto (PDRB). PDRB level provinsi dihasilkan oleh Badan Pusat Statistik setiap triwulan pada 35 hari setelah triwulan referensi berakhir. Hal ini tidak terlepas dari ketersediaan data dan indikator yang digunakan dalam penyusunan PDRB, baik yang bersumber dari survei, data administratif maupun laporan keuangan. Kondisi tersebut menimbulkan *lag* dari kebijakan yang diambil.

PDRB di Indonesia disusun dari dua sisi, yaitu sisi produksi dan sisi pengeluaran. PDRB dari sisi produksi disajikan menurut 17 kategori lapangan usaha. Sedangkan, PDRB dari sisi pengeluaran disajikan menurut komponen konsumsi rumah tangga, konsumsi Lembaga Non-Profit yang Melayani Rumah Tangga (LNPRM), konsumsi pemerintah, Pembentukan Modal Tetap Bruto (PMTB), perubahan inventori, dan ekspor neto. Pengeluaran Konsumsi Rumah Tangga (PKRT) merupakan komponen yang sangat dominan dalam PDRB sisi pengeluaran. Di Provinsi Jawa Timur, PKRT memiliki kontribusi sekitar 59% terhadap PDRB Jawa Timur (BPS Jawa Timur, 2022). Dengan demikian, pertumbuhan ekonomi Jawa Timur sangat dipengaruhi oleh konsumsi masyarakat dalam bentuk PKRT.

Kebutuhan estimasi cepat (*flash estimates/nowcasting*) terhadap data PDB dan pertumbuhan ekonomi berkembang pesat dalam beberapa tahun terakhir. Diantara negara-negara Eropa, United Kingdom (UK), Belgia, Lituania, dan Spanyol telah mempublikasikan estimasi cepat PDB pada 30 hari setelah triwulan berakhir, dari sebelumnya 45 hari (Eurostat, 2016). Hal ini didukung oleh perkembangan *Big Data* yang mampu menyediakan data secara masif dan *real time*. Di Indonesia, studi tentang *nowcasting* PDB dan pertumbuhan ekonomi telah dilakukan oleh Luciani dkk. (2015), Tarsidin dkk. (2018), Utari & Ilma (2018), Damuri dkk. (2021) dan Muchisha dkk. (2021). Teknik-teknik yang digunakan juga beragam, seperti *Bridge Equation*, *Mixed Data Sampling*, *Dynamic Factor Model*, dan teknik *Machine Learning*. Namun, studi *nowcasting* untuk provinsi Jawa Timur belum banyak dilakukan, khususnya untuk komponen konsumsi rumah tangga.

Penelitian ini bertujuan untuk membentuk model estimasi cepat (*flash estimates*) konsumsi rumah tangga di Provinsi Jawa Timur menggunakan berbagai indikator, baik dari *Big Data* maupun indikator-indikator konvensional. Penelitian ini membandingkan berbagai model, baik model-model ekonometrik maupun teknik *machine learning*. Disamping itu, penelitian ini juga mengembangkan sebuah model *Ensemble* yang menggabungkan beberapa model-model *machine learning*. Model-model tersebut diuji pada berbagai skenario partisi *out-of-sample* dan periode estimasi. Skenario partisi *out-of-sample* yang dilakukan terdiri atas partisi bertahap dan partisi tak bertahap. Sedangkan, skenario periode estimasi terdiri atas 10 hari setelah triwulan referensi berakhir (H+10), 20 hari setelah triwulan referensi berakhir (H+20), dan 30 hari setelah triwulan referensi berakhir (H+30). Setiap periode memiliki kondisi ketersediaan indikator yang berbeda-beda, sehingga perlu dilakukan estimasi pada bulan yang datanya belum tersedia.

Telaah Literatur

Estimasi cepat (*flash estimates*) PDB adalah pendugaan PDB triwulan terkini ketika sekumpulan data yang paling sesuai untuk menghitung PDB tersebut belum tersedia (Eurostat, 2016). Beberapa istilah lain yang sering digunakan untuk menggambarkan estimasi cepat dengan kondisi data belum lengkap adalah '*advance estimate*' (BEA, 2022) dan '*nowcasting*' (Banbura dkk., 2013). Prinsip dasar dari estimasi cepat, *advance estimate*, ataupun *nowcasting* adalah pemanfaatan informasi tambahan yang ketersediannya lebih cepat atau frekuensinya lebih tinggi daripada variabel referensi yang akan diestimasi. Tujuan dari estimasi cepat, *advance estimate*, ataupun *nowcasting* adalah untuk memberikan gambaran pendahuluan se-

belum angka resmi tersedia. Estimasi cepat, *advance estimate*, ataupun *nowcasting* berfokus pada sesuatu yang sedang berlangsung atau baru saja terjadi, berbeda dengan peramalan (*forecasting*) yang ditujukan untuk mengantisipasi sesuatu yang belum terjadi (Ceron dkk., 2017).

Indikator penunjang dalam estimasi cepat PDB dapat berupa '*soft indicators*' maupun '*hard indicators*'. *Soft indicators* mencerminkan ekspektasi pasar, seperti data hasil survei dan informasi keuangan. Sedangkan, *hard indicators* merupakan ukuran aktivitas ekonomi sebenarnya, seperti data produksi industri dan penjualan eceran (Banbura & Runstler, 2011). *Soft indicators* biasanya tersedia lebih cepat dibandingkan dengan *hard indicators*. Namun, *hard indicators* mencerminkan ukuran PDB yang lebih tepat dibandingkan dengan *soft indicators*. Beberapa tahun terakhir, indikator-indikator *Big Data* juga digunakan dalam estimasi cepat aktivitas ekonomi, seperti *Google Mobility Index* (Sampi & Jooste, 2020), *Facebook Movement Range Data* (Damuri dkk., 2021), dan *Google Trends* (Woloszko, 2020). Kelebihan indikator-indikator *Big Data* adalah ketersediaannya yang *real time*.

Estimasi cepat PDB juga telah dilakukan untuk komponen yang lebih rinci, misalnya konsumsi rumah tangga. Tarsidin dkk. (2018) menggunakan beberapa indikator untuk estimasi cepat konsumsi rumah tangga di Indonesia, yaitu penjualan kendaraan bermotor, indeks penjualan eceran, indeks keyakinan konsumen, indeks produksi industri, Indeks Harga Saham Gabungan (IHSG), total simpanan, kredit, suku bunga kredit konsumsi, jumlah uang beredar (M1 dan M2), dan nilai tukar rupiah (NEER). Toth & Hajdu (2020) menggunakan data *Google Trends* untuk estimasi cepat konsumsi rumah tangga di Hungaria. Gil dkk. (2017) menggunakan indikator tradisional dan non-tradisional dalam estimasi cepat konsumsi rumah tangga di Spanyol. Indikator tradisional yang digunakan misalnya indeks perdagangan eceran, indeks aktivitas jasa, PMI jasa, dan indeks keyakinan konsumen. Sedangkan, indikator non-tradisional yang digunakan misalnya indeks ketidakpastian kebijakan ekonomi (EPU), volatilitas IBEX-35, data kartu kredit dan *Google Trends*. Aastveit dkk. (2020) menggunakan data transaksi kartu debit untuk estimasi cepat konsumsi rumah tangga di Norwegia.

Castle dkk. (2017) menjelaskan bahwa terdapat lima pendekatan utama dalam *nowcasting* variabel ekonomi makro. Pendekatan pertama adalah '*in-filling*', yang dilakukan dengan cara mengestimasi data kosong secara langsung menggunakan metode *Exponentially-Weighted Moving Average* (EWMA) atau *Autoregressive-Integrated Moving Average* (ARIMA). Pendekatan kedua adalah *Mixed-Data Sampling* (MIDAS), yang menghubungkan antara variabel referensi yang frekuensinya lebih rendah dengan indikator yang frekuensinya lebih tinggi menggunakan *parsimonious restricted lag polynomial*. Pendekatan ketiga adalah model faktor, yang terdiri atas model faktor statis dan model faktor dinamis. Model faktor memiliki keunggulan, yaitu dapat digunakan meskipun jumlah indikator sangat banyak. Model faktor memanfaatkan analisis faktor untuk mereduksi dimensi dari indikator tanpa kehilangan banyak informasi. Pendekatan keempat adalah model faktor *with ragged edges*. Pendekatan ini merupakan pengembangan dari pendekatan ketiga, yang memiliki kemampuan untuk mengatasi data kosong pada indikator bulanan. Pendekatan terakhir adalah *Bridge Equations*. Pendekatan ini memiliki dua tahap, yaitu pertama mengestimasi data kosong pada indikator bulanan dan mengagregasinya menjadi data triwulanan, lalu tahap kedua memasukkan indikator agregat triwulanan sebagai prediktor di dalam model untuk mengestimasi variabel referensi triwulanan. Pendekatan ini lebih mudah diinterpretasikan dibandingkan dengan MIDAS dan model faktor. Beberapa tahun terakhir, beberapa peneliti juga mengeksplorasi penggunaan teknik-teknik *machine learning* untuk *nowcasting* PDB dan pertumbuhan ekonomi. Buell dkk. (2021) menggunakan teknik-teknik *machine learning* untuk memprediksi pertumbuhan

ekonomi di Afrika Sub-Sahara. Richardson dkk. (2018) menggunakan algoritma *machine learning* untuk *nowcasting* PDB di New Zealand. Muchisha dkk. (2021) menggunakan model-model *machine learning* untuk *nowcasting* pertumbuhan ekonomi Indonesia.

Di Indonesia, studi-studi tentang *nowcasting* PDB dan pertumbuhan ekonomi umumnya dilakukan untuk total PDB ataupun total pertumbuhan ekonomi. Luciani dkk. (2015) melakukan *nowcasting* pertumbuhan ekonomi Indonesia menggunakan *Dynamic Factor Model* (DFM). Damuri dkk. (2021) melakukan *nowcasting* pertumbuhan ekonomi provinsi-provinsi di Pulau Jawa dan Bali menggunakan *Mixed-Data Sampling* (MIDAS). Utari & Ilma (2018) membandingkan kinerja MIDAS dan *Distributed Lag Model* dalam memprediksi pertumbuhan ekonomi Indonesia, dengan menggunakan indikator ekspor sektor pertanian. Muchisha dkk. (2021) melakukan *nowcasting* pertumbuhan ekonomi Indonesia menggunakan teknik-teknik *machine learning*. Penelitian tentang *nowcasting* PDB Indonesia pada komponen yang lebih rinci masih minim dilakukan. Salah satu dari sedikit penelitian yang pernah dilakukan adalah oleh Tarsidin dkk. (2018), yang melakukan *nowcasting* komponen Pengeluaran Konsumsi Rumah Tangga (PKRT) dan Pembentukan Modal Tetap Bruto (PMTB) di Indonesia.

Metode Penelitian

Data dan Indikator

Data dalam penelitian ini utamanya diperoleh dari Bank Indonesia, Badan Pusat Statistik, dan *Google Trends* (lihat Tabel 1). Data yang diperoleh dari Bank Indonesia mencakup posisi simpanan masyarakat, posisi pinjaman/kredit konsumsi, suku bunga pinjaman konsumsi, suku bunga simpanan berjangka 1 bulan, suku bunga simpanan berjangka 3 bulan, suku bunga simpanan berjangka 6 bulan, dan aliran uang keluar (*outflow*). Sedangkan, data yang diperoleh dari Badan Pusat Statistik mencakup Pengeluaran Konsumsi Rumah Tangga (PKRT) dan Indeks Harga Konsumen (IHK). Penelitian ini juga menambahkan indikator *Big Data* berupa *Google Trends Index* (GTI). Indikator-indikator yang digunakan dalam penelitian ini mengacu pada penelitian-penelitian sebelumnya (Tarsidin dkk., 2018; Woo & Owen, 2018; Varlamova & Larionova, 2015; Bok dkk., 2018).

Tabel 1: Sumber Data

Label	Deskripsi	Satuan	Frekuensi	Lag	Sumber
CONS	Pengeluaran Konsumsi Rumah Tangga (PKRT) (2010=100)	Milyar Rupiah	Triwulanan	H + 35	Badan Pusat Statistik
DEPO	Posisi simpanan masyarakat	Milyar Rupiah	Bulanan	H + 51	Bank Indonesia
LOAN.C	Posisi pinjaman/kredit konsumsi	Milyar Rupiah	Bulanan	H + 51	Bank Indonesia
r.CL	Suku bunga pinjaman konsumsi	%	Bulanan	H + 51	Bank Indonesia
r.D1m	Suku bunga simpanan berjangka 1 bulan	%	Bulanan	H + 51	Bank Indonesia
r.D3m	Suku bunga simpanan berjangka 3 bulan	%	Bulanan	H + 51	Bank Indonesia
r.D6m	Suku bunga simpanan berjangka 6 bulan	%	Bulanan	H + 51	Bank Indonesia
M.OUT	Aliran uang keluar (<i>outflow</i>)	Milyar Rupiah	Bulanan	H + 15	Bank Indonesia
CPI	Indeks Harga Konsumen (IHK) (2018=100)	Indeks	Bulanan	H + 1	Badan Pusat Statistik
GTI	<i>Google Trends Index</i>	Indeks	<i>Real Time</i>	<i>Real Time</i>	https://trends.google.com/trends/?geo=ID

Seluruh data dan indikator selain data suku bunga tersedia untuk level provinsi Jawa Timur. Sedangkan, data suku bunga diasumsikan sama dengan suku bunga di tingkat nasional. Periode yang dicakup dalam penelitian ini adalah tahun 2011-2021 karena data aliran uang keluar tersedia sejak tahun 2011. Data posisi simpanan masyarakat, posisi pinjaman/kredit konsumsi, suku bunga pinjaman konsumsi, suku bunga simpanan berjangka 1 bulan, suku bunga simpanan berjangka 3 bulan, suku bunga simpanan berjangka 6 bulan, aliran uang keluar, dan IHK tersedia bulanan. Sedangkan data PKRT tersedia triwulanan. Frekuensi seluruh data dalam penelitian ini disamakan dengan frekuensi data PKRT karena PKRT adalah variabel referensi yang akan diprediksi. Nilai triwulanan untuk simpanan masyarakat, kredit konsumsi, suku bunga pinjaman konsumsi, suku bunga simpanan berjangka 1 bulan, suku bunga simpanan berjangka 3 bulan, dan suku bunga simpanan berjangka 6 bulan direpresentasikan oleh nilai indikator-indikator tersebut pada bulan terakhir setiap triwulan. Nilai triwulanan untuk aliran uang keluar merupakan jumlah seluruh bulan pada triwulan tersebut. Sedangkan, IHK triwulanan merupakan rata-rata IHK bulanan pada triwulan tersebut.

Google Trends Index (GTI) merupakan ukuran popularitas relatif dari suatu *query* pencarian di suatu lokasi dan waktu tertentu. Indeks dalam GTI merupakan perbandingan popularitas relatif di suatu lokasi atau waktu tertentu terhadap popularitas relatif maksimum. Dengan demikian, angka GTI sebesar 70.21 menunjukkan bahwa popularitas relatif di lokasi atau periode tersebut adalah sebesar 70.21% dari popularitas relatif maksimum. Data *Google Trends* diperoleh dari aktivitas pencarian di Google, dengan mengeliminasi pencarian berulang dari orang yang sama dalam waktu singkat. Selanjutnya, *query* yang volume pencariannya rendah ditampilkan sebagai 0. Data *Google Trends* disajikan menurut waktu dan wilayah. Terdapat pilihan wilayah, periode, kategori dan jenis pencarian di dalam *Google Trends*. Wilayah dapat dipilih hingga level provinsi, sedangkan periodenya mencakup 1 jam yang lalu hingga 2004-sekarang. Kategori data *Google Trends* diklasifikasikan menjadi 1132 kategori, sedangkan jenis pencariannya dibagi menjadi *Web Search*, *Image Search*, *News Search*, *Youtube Search*, dan *Google Shopping*. Penelitian ini melakukan *scraping* data *Google Trends* untuk level provinsi Jawa Timur dan periode 2004-sekarang. Seluruh kategori dan jenis pencarian dimasukkan dalam proses *scraping* data. Agregasi ke data triwulanan dilakukan menggunakan rata-rata sederhana.

Model ARIMA

Model *Autoregressive Integrated Moving Average* (ARIMA) merupakan suatu metode peramalan yang diperoleh melalui gabungan antara *Autoregressive* (AR) dan *Moving Average* (MA), yang dapat diterapkan pada data yang tidak stasioner pada level. Jika data tidak stasioner pada level, maka perlu dilakukan proses *differencing* sebanyak d sampai data menjadi stasioner. Dengan demikian, model ARIMA memerlukan 3 orde, yaitu orde AR (p), orde MA (q) dan orde *differencing* (d), yang disingkat menjadi ARIMA(p,d,q). ARIMA memanfaatkan perilaku historis dari data dalam melakukan prediksi jangka pendek ke depan. Bentuk umum model ARIMA(p,d,q) adalah sebagai berikut (Cryer & Chan, 2008):

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) e_t \quad (1)$$

dimana Y_t adalah variabel yang diamati, B adalah *backshift operator*, dan e_t adalah *noise*.

Penentuan orde ARIMA biasanya dilakukan dengan melihat pola plot *autocorrelation function* (ACF), *partial autocorrelation function* (PACF), dan *extended autocorrelation function* (EACF). Namun, penelitian ini tidak menggunakan teknik tersebut. Penelitian ini memanfaatkan fungsi 'auto.arima' pada *package* R yang secara otomatis mencari orde optimal dengan nilai AIC terkecil. Metode estimasi parameter yang digunakan adalah *maximum likelihood*.

Model ADL

Model *Autoregressive Distributed Lag* (ADL) merupakan suatu teknik peramalan yang menambahkan variabel bebas beserta *lag*-nya sebagai prediktor. Selain itu, model ADL juga memasukkan *lag* dari variabel tak bebas sebagai prediktor. Penambahan variabel *lag* dimaksudkan untuk menangkap sifat dinamis dari sebagian besar proses ekonomi karena dampak dari suatu kebijakan dapat memerlukan waktu untuk dirasakan. Penelitian ini menggunakan model ADL(1,1) yang dapat dituliskan sebagai berikut (Eurostat, 2016):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + e_t \quad (2)$$

dimana Y_t adalah variabel tak bebas, X_t adalah variabel bebas, dan e_t adalah *noise*.

Model ADL memerlukan data yang stasioner untuk menghindari regresi *spurious*. Jika data tidak stasioner maka perlu dilakukan proses *differencing*. Penelitian ini menggunakan *package* 'dLagM' pada *software* R untuk menghasilkan model ADL. Variabel bebas pada model ADL dalam penelitian ini merupakan kombinasi dari berbagai indikator GTI dan non-GTI yang dibentuk menggunakan teknik *Principal Component Analysis* (PCA).

Elastic Net

Elastic Net (Enet) merupakan metode regularisasi yang mengkombinasikan penalti L_1 dan L_2 pada LASSO dan regresi ridge. *Elastic Net* dapat digunakan pada kondisi jumlah parameter lebih besar daripada jumlah amatan ($p > n$). Koefisien *Elastic Net* diestimasi dengan cara meminimumkan fungsi berikut ini (Zou & Hastie, 2005):

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p [(1 - \alpha) \beta_j^2 + \alpha |\beta_j|] \quad (3)$$

dimana y_i adalah variabel tak bebas, x_{ij} adalah variabel bebas ke- j , β_j adalah koefisien untuk variabel bebas ke- j , p adalah banyaknya variabel bebas, dan n adalah banyaknya amatan. Jika $\alpha=0$ maka *Elastic Net* sama dengan regresi ridge. Sedangkan, jika $\alpha=1$ maka *Elastic Net* sama dengan LASSO. Penelitian ini menggunakan *package* 'glmnet' pada *software* R untuk menghasilkan model *Elastic Net*.

Support Vector Machine

Support Vector Machine (SVM) adalah metode nonparametrik yang dapat digunakan pada pemodelan klasifikasi maupun regresi (Muchisha dkk., 2021). Prinsip dari SVM adalah mencari *hyperplane* yang memaksimumkan margin (James dkk., 2013). *Hyperplane* adalah subruang datar berdimensi ($p-1$). SVM mencari fungsi linear dari bentuk berikut (Richardson dkk., 2018):

$$f(x) = \langle w, x \rangle + b \quad (4)$$

dimana w adalah vektor bobot, b adalah bias, dan x adalah vektor fitur sembari memastikan bahwa fungsi pada persamaan (4) sedatar mungkin, yang berarti bahwa mencari nilai w minimum. Salah satu cara untuk mendapatkan nilai w minimum adalah dengan cara meminimumkan $\|w\|^2$ pada fungsi optimasi berikut ini:

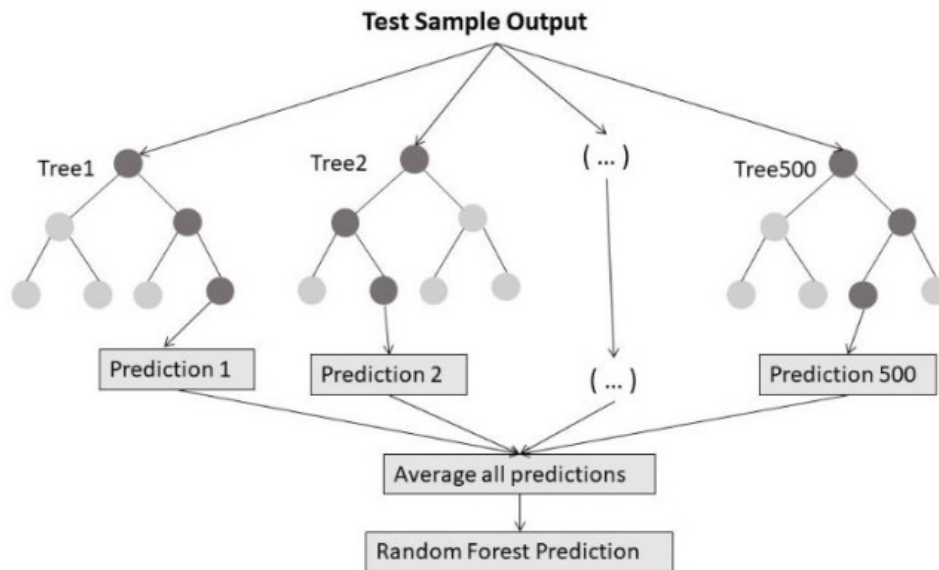
$$0.5 \|w\|^2 + C \sum_{i=1}^t |y_i - f(x_i)|_\epsilon \quad (5)$$

dimana $C > 0$ adalah parameter regularisasi. Komponen pertama pada fungsi galat adalah komponen penalti yang meningkat seiring dengan model yang semakin kompleks. Komponen

kedua adalah fungsi kerugian ϵ -insensitive yang memberikan penalti pada galat yang lebih besar dari ϵ , sehingga memungkinkan fleksibilitas pada model. Penelitian ini menggunakan package 'e1071' pada software R untuk menghasilkan model SVM.

Random Forest

Model *Random Forest* (RF) dibangun berdasarkan sejumlah pohon regresi yang saling bebas. Pohon-pohon regresi tersebut dibentuk dari sampel-sampel acak yang diambil dengan metode *bootstrap*. Pada setiap sampel acak, hanya sebagian prediktor yang diambil ($m < p$). Ilustrasi dari RF ditampilkan dalam Gambar 1.



Sumber: Muchisha dkk. (2021)

Gambar 1: Ilustrasi *Random Forest*

Algoritma model RF untuk data numerik adalah sebagai berikut (Hastie dkk., 2017):

1. Partisi dataset L menjadi M daerah $R_1, R_2, \dots, R_M$.
2. Model respon adalah nilai konstan c_m dalam setiap daerah:

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (6)$$

Nilai prediksi dari c_m diperoleh dari rata-rata di setiap daerah R_m :

$$\hat{c}_m = \text{ave}(y_i | x_i \in R_m) \quad (7)$$

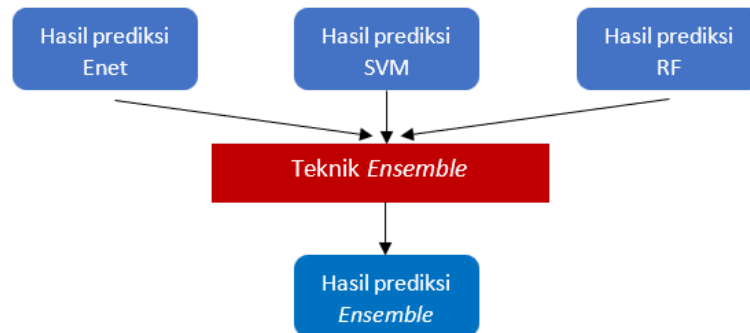
Partisi biner terbaik dalam pohon regresi diperoleh dengan mencoba beberapa *threshold* berbeda dan memilih *threshold* yang memiliki jumlah kuadrat terkecil. Misalnya, untuk daerah R_a dan R_b , kita mencari variabel pemisah j dan titik pemisah s yang memenuhi:

$$\min_{j,s} \left[\min_{c_a} \sum_{x_i \in R_a(j,s)} (y_i - c_a)^2 + \min_{c_b} \sum_{x_i \in R_b(j,s)} (y_i - c_b)^2 \right] \quad (8)$$

Setelah pemisah terbaik ditemukan, dataset dipartisi menjadi dua subset. Proses berlanjut sampai setiap *node* mencapai batas optimal yang ditentukan dan menjadi *node* terakhir. Penelitian ini menggunakan package 'randomForest' pada software R untuk menghasilkan model *Random Forest*.

Teknik Ensemble

Penelitian ini menggabungkan hasil prediksi dari *Elastic Net* (Enet), *Support Vector Machine* (SVM), dan *Random Forest* (RF) untuk menghasilkan suatu prediksi baru. Teknik ini disebut sebagai teknik *Ensemble*. Idenya adalah hasil prediksi dari Enet, SVM dan RF dijadikan sebagai prediktor dalam model prediksi baru. Model prediksi baru tersebut dibangun menggunakan algoritma *Elastic Net*. Ilustrasi dari algoritma ini diberikan pada Gambar 2 sebagai berikut:



Gambar 2: Ilustrasi Teknik Ensemble

Tahapan Analisis

Penelitian ini menggunakan beberapa tahapan dalam menghasilkan model estimasi cepat (*flash estimate*) Pengeluaran Konsumsi Rumah Tangga (PKRT) di Jawa Timur. Tahapan-tahapan tersebut adalah sebagai berikut:

1. Melakukan *scraping* data *Google Trends*, dengan mengatur cakupan wilayah dan periode data. Wilayah yang dicakup hanya provinsi Jawa Timur, dengan cara menentukan kode untuk provinsi Jawa Timur, yaitu 'ID-JI'. Periode yang dicakup adalah 'all', sehingga mencakup tahun 2004-sekarang. Sedangkan, untuk kategori dan jenis pencarian diambil semua saat proses *scraping*. Proses *scraping* data dilakukan menggunakan package 'gtrendsR' di software R.
2. Melakukan *cleaning* data *Google Trends*, misalnya dengan mengganti kode "<1" menjadi 0.
3. Mengagregasi indeks *Google Trends* bulanan menjadi data triwulanan.
4. Menyeleksi kategori *Google Trends* yang korelasinya kuat dengan PKRT di Jawa Timur. Penelitian ini hanya mengambil kategori yang koefisien korelasinya lebih dari 0.90.
5. Memprediksi indikator bulanan non-GTI yang masih kosong, dengan menggunakan ARIMA. Penelitian ini menggunakan berbagai skenario ketersediaan indikator non-GTI, yaitu pada 10 hari setelah triwulan referensi berakhir (H+10), pada 20 hari setelah triwulan referensi berakhir (H+20), dan pada 30 hari setelah triwulan referensi berakhir (H+30). Pada H+10, indikator simpanan masyarakat (DEPO), kredit konsumsi (LOAN.C), suku bunga pinjaman konsumsi (r.CL), dan suku bunga simpanan berjangka 1 hingga 6 bulan (r.D1m, r.D3m, r.D6m) baru tersedia 1 bulan. Sedangkan, indikator aliran uang keluar (M.OUT) baru tersedia 2 bulan. Pada H+20, indikator M.OUT sudah tersedia 3 bulan, namun indikator DEPO, LOAN.C, r.CL, r.D1m, r.D3m, dan r.D6m masih tersedia 1 bulan. Pada H+30, indikator DEPO, LOAN.C, r.CL, r.D1m, r.D3m, dan r.D6m tersedia 2 bulan. Berbagai kondisi ketersediaan indikator pada H+10, H+20 dan H+30 disajikan dalam Tabel 2.

Tabel 2: Ketersediaan Indikator pada H+10, H+20 dan H+30

Indikator	Lag	H + 10			H + 20			H + 30		
		Qt 2021			Qt 2021			Qt 2021		
		Bulan 1	Bulan 2	Bulan 3	Bulan 1	Bulan 2	Bulan 3	Bulan 1	Bulan 2	Bulan 3
DEPO	H + 51	v	x	x	v	x	x	v	v	x
LOAN.C	H + 51	v	x	x	v	x	x	v	v	x
r.CL	H + 51	v	x	x	v	x	x	v	v	x
r.D1m	H + 51	v	x	x	v	x	x	v	v	x
r.D3m	H + 51	v	x	x	v	x	x	v	v	x
r.D6m	H + 51	v	x	x	v	x	x	v	v	x
M.OUT	H + 15	v	v	x	v	v	v	v	v	v
CPI	H + 1	v	v	v	v	v	v	v	v	v
GTI	Real Time	v	v	v	v	v	v	v	v	v

v : Tersedia

x : Belum tersedia

- Mengagregasi indikator bulanan non-GTI (yang telah memprediksi bulan kosong) menjadi data triwulanan.
- Menggabungkan data PKRT triwulanan, indikator non-GTI triwulanan, dan indikator GTI triwulanan yang korelasinya kuat dengan PKRT.
- Melakukan transformasi untuk variabel PKRT, posisi simpanan masyarakat, posisi kredit konsumsi, aliran uang keluar, dan IHK ke dalam bentuk logaritma natural.
- Memberikan fitur tambahan untuk model *machine learning*, yaitu variabel *lag* dari PKRT, *lag* dari indikator GTI, dan *lag* dari indikator non-GTI. Proses ini mengadopsi *feature engineering* yang dilakukan oleh Muchisha dkk. (2021) seperti Tabel 3.

Tabel 3: Fitur Tambahan

	y					x ¹					...	x ¹⁸				
t	t - 1	t - 2	t - 3	t - 4	t - 8	t - 1	t - 2	t - 3	t - 4	t - 8	...	t - 1	t - 2	t - 3	t - 4	t - 8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	...	(91)	(92)	(93)	(94)	(95)
1																
2																
...																
n																

Sumber: Muchisha dkk. (2021)

- Mempartisi data menjadi 2 bagian, yaitu data *in-sample* dan data *out-of-sample*. Penelitian ini melakukan dua skenario estimasi, yaitu estimasi bertahap dan estimasi tak bertahap. Pada estimasi bertahap, data *in-sample* diperbarui secara bertahap sebelum melakukan prediksi pada triwulan berikutnya (lihat Gambar 3). Sedangkan, pada estimasi tak bertahap, data *in-sample* yang sama (Q1 2011-Q4 2020) digunakan dalam memprediksi seluruh titik pada periode *out-of-sample* (lihat Gambar 4).
- Melatih model berdasarkan data *in-sample*. Model yang digunakan adalah ARIMA, ADL, *Elastic Net*, *Support Vector Machine*, *Random Forest*, dan *Ensemble*.
- Melakukan prediksi pada periode *out-of-sample* berdasarkan model-model yang telah dilatih pada langkah 11.
- Mengevaluasi kinerja hasil prediksi dari berbagai model menggunakan kriteria

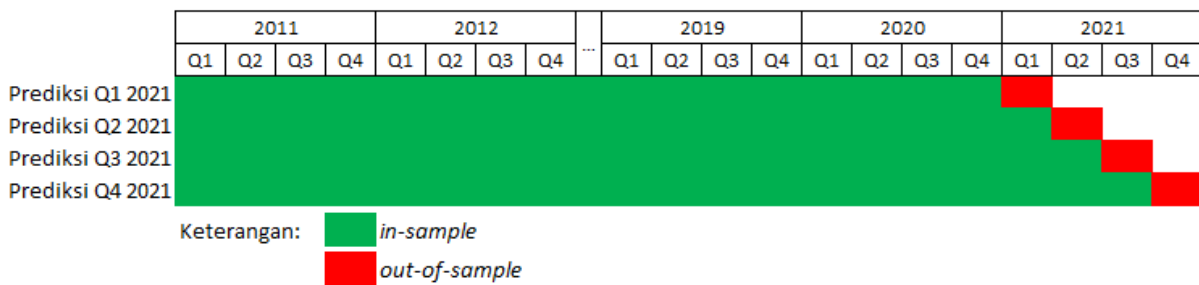
Root Mean Squared Error (RMSE) dan *Mean Absolute Percentage Error (MAPE)*. Formula dari RMSE dan MAPE adalah sebagai berikut (Montgomery dkk., 2015):

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (C_t - \hat{C}_t)^2}{T}} \quad (9)$$

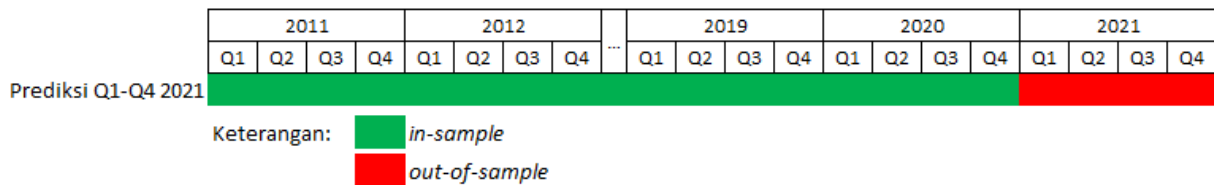
$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{C_t - \hat{C}_t}{C_t} \right| \times 100\%$$

dimana C_t adalah nilai aktual PKRT pada periode ke- t , $\hat{C}_t$ adalah hasil prediksi PKRT pada periode ke- t , dan T adalah banyaknya periode.

14. Memilih model terbaik untuk memprediksi logaritma natural dari PKRT.
15. Mentransformasi kembali hasil prediksi logaritma natural dari PKRT ke bentuk aslinya.



Gambar 3: Ilustrasi Estimasi Bertahap



Gambar 4: Ilustrasi Estimasi Tak Bertahap

Hasil dan Pembahasan

Ringkasan statistik dari Pengeluaran Konsumsi Rumah Tangga (PKRT) dan indikator non-GTI (selain *Google Trends Index*) disajikan dalam Tabel 4. Rata-rata PKRT di Jawa Timur adalah sekitar 200 triliun rupiah per triwulan, sedangkan aliran uang keluar (*outflow*) adalah sekitar 17 triliun rupiah per triwulan. Rata-rata posisi simpanan masyarakat adalah sekitar 444 triliun rupiah, sedangkan rata-rata posisi kredit konsumsi adalah sekitar 122 triliun rupiah. Suku bunga pinjaman konsumsi, suku bunga simpanan berjangka 1 bulan, suku bunga simpanan berjangka 3 bulan, dan suku bunga simpanan berjangka 6 bulan memiliki rata-rata berturut-turut sebesar 12.82%, 6.26%, 6.49% dan 6.85%. Sedangkan, Indeks Harga Konsumen (IHK) memiliki rata-rata sebesar 91.37. IHK dalam penelitian ini telah di *re-referencing* ke 2018=100.

Simpanan masyarakat dan kredit konsumsi memiliki korelasi yang sangat erat dengan PKRT. Koefisien korelasi yang dihasilkan oleh kedua indikator tersebut masing-masing sebesar 0.9866 dan 0.9927 (Tabel 5). Korelasi positif yang sangat erat juga ditunjukkan oleh Indeks Harga Konsumen (IHK). Korelasi IHK dengan PKRT adalah sebesar 0.9887. Sedangkan, suku bunga kredit konsumsi memiliki korelasi negatif yang kuat dengan PKRT, yaitu sebesar -0.8808. Suku bunga simpanan berjangka 1 hingga 6 bulan memiliki korelasi yang tidak begitu kuat dengan PKRT, dengan nilai koefisien korelasi di bawah 0.50. Di sisi lain, aliran uang keluar

(*outflow*) memiliki korelasi positif yang cukup erat dengan PKRT, yaitu sebesar 0.6703.

Tabel 4: Analisis Deskriptif PKRT dan Indikator Non-GTI

Label	Deskripsi	Satuan	Rataan	Simp. Baku	Min.	Maks.
CONS	Pengeluaran Konsumsi Rumah Tangga (PKRT)	Milyar Rupiah	210,498	27,819	160,970	256,487
DEPO	Posisi simpanan masyarakat	Milyar Rupiah	444,150	135,541	219,026	688,046
LOAN.C	Posisi pinjaman/kredit konsumsi	Milyar Rupiah	122,079	36,795	53,130	169,057
r.CL	Suku bunga pinjaman konsumsi	%	12.82	1.20	10.53	14.83
r.D1m	Suku bunga simpanan berjangka 1 bulan	%	6.26	1.34	2.96	8.58
r.D3m	Suku bunga simpanan berjangka 3 bulan	%	6.49	1.38	3.19	9.37
r.D6m	Suku bunga simpanan berjangka 6 bulan	%	6.85	1.36	3.52	9.33
M.OUT	Aliran uang keluar (<i>outflow</i>)	Milyar Rupiah	17,494	8,471	3,602	37,903
CPI	Indeks Harga Konsumen (IHK)	Indeks	91.37	11.26	71.21	106.64

Tabel 5 juga menyajikan koefisien korelasi antara PKRT dengan beberapa indikator dari *Google Trends Index* (GTI). Indikator GTI yang disajikan dalam Tabel 5 hanya mencakup 20 indikator yang korelasinya paling kuat. Sedangkan, indikator lengkap yang korelasinya kuat disajikan pada Lampiran 1. PKRT memiliki korelasi yang erat dengan beberapa kategori pencarian di *Google Trends*, seperti yang terkait dengan rumah ('Building Materials and Supplies', 'Bathroom', 'Plumbing', 'Home Improvement', 'Doors and Windows', 'Flooring'), pembiayaan ('Credit and Lending', 'Home Financing'), penjualan produk ('Telemarketing', 'E-Commerce Services'), produk kosmetik dan kebersihan ('Perfumes and Fragrances', 'Laundry', 'Hygiene and Toiletries').

Secara keseluruhan terdapat 153 indikator yang digunakan untuk memprediksi Pengeluaran Konsumsi Rumah Tangga (PKRT), yang terdiri atas 8 indikator non-GTI (lihat Tabel 4) dan 145 indikator GTI (lihat Lampiran 1). Penelitian ini juga menambahkan variable *lag* dari PKRT dan *lag* dari masing-masing indikator sebagai prediktor dalam pemodelan *machine learning* (*Elastic Net*, *Support Vector Machine*, *Random Forest*, *Ensemble*), sehingga terdapat 769 prediktor untuk pemodelan *machine learning*. Model ARIMA hanya menggunakan variabel PKRT, sedangkan model ADL menggunakan variabel PKRT dan 153 indikator GTI dan non-GTI. Pada model ADL, indikator GTI dan non-GTI direduksi terlebih dahulu menggunakan *Principal Component Analysis* (PCA) sebelum model dijalankan.

Seluruh model dilatih pada periode *in-sample*, lalu digunakan untuk memprediksi PKRT pada periode *out-of-sample* (Q1 2021-Q4 2021). Seperti yang telah dijelaskan sebelumnya, penelitian ini melakukan dua skenario estimasi, yaitu estimasi bertahap dan estimasi tak bertahap. Pada masing-masing skenario tersebut, disimulasikan berbagai kondisi indikator pada H+10, H+20 dan H+30. Hasil prediksi berbagai model berdasarkan skenario estimasi bertahap disajikan dalam Gambar 5 hingga Gambar 7, sedangkan hasil prediksi berbagai model berdasarkan skenario estimasi tak bertahap disajikan dalam Gambar 8 hingga Gambar 10.

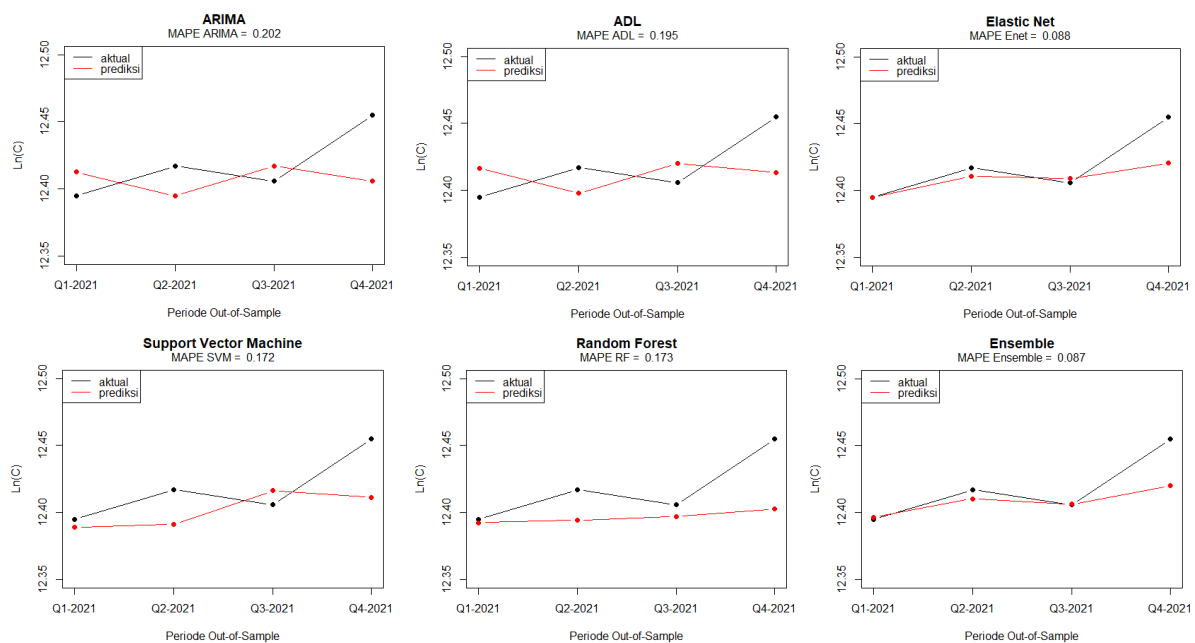
Tabel 5: Korelasi Berbagai Indikator terhadap Pengeluaran Konsumsi Rumah Tangga (PKRT)

No	Label	Indikator	Koefisien Korelasi	No	Label	Indikator	Koefisien Korelasi
1	DEPO	Posisi simpanan masyarakat	0.9866	15	GTI.LY	Indeks <i>Google Trends</i> untuk kategori 'Laundry'	0.9496
2	LOAN.C	Posisi pinjaman/kredit konsumsi	0.9927	16	GTI.P&F	Indeks <i>Google Trends</i> untuk kategori 'Perfumes and Fragrances'	0.9494
3	r.CL	Suku bunga pinjaman konsumsi	-0.8808	17	GTI.FLR	Indeks <i>Google Trends</i> untuk kategori 'Foreign Languages Resources'	0.9491
4	r.D1m	Suku bunga simpanan berjangka 1 bulan	-0.4371	18	GTI.P&P	Indeks <i>Google Trends</i> untuk kategori 'Plastics and Polymers'	0.9488
5	r.D3m	Suku bunga simpanan berjangka 3 bulan	-0.4471	19	GTI.HI	Indeks <i>Google Trends</i> untuk kategori 'Home Improvement'	0.9488
6	r.D6m	Suku bunga simpanan berjangka 6 bulan	-0.4016	20	GTI.C&A	Indeks <i>Google Trends</i> untuk kategori 'Coatings and Adhesives'	0.9487
7	M.OUT	Aliran uang keluar (<i>outflow</i>)	0.6703	21	GTI.ES	Indeks <i>Google Trends</i> untuk kategori 'E-Commerce Services'	0.9470
8	CPI	Indeks Harga Konsumen (IHK)	0.9887	22	GTI.SW	Indeks <i>Google Trends</i> untuk kategori 'Sleepwear'	0.9461
9	GTI.BR	Indeks <i>Google Trends</i> untuk kategori 'Bathroom'	0.9582	23	GTI.D&W	Indeks <i>Google Trends</i> untuk kategori 'Doors and Windows'	0.9459
10	GTI.HF	Indeks <i>Google Trends</i> untuk kategori 'Home Financing'	0.9577	24	GTI.CI	Indeks <i>Google Trends</i> untuk kategori 'Chemical Industry'	0.9459
11	GTI.C&T	Indeks <i>Google Trends</i> untuk kategori 'Coffee and Tea'	0.9543	25	GTI.C&L	Indeks <i>Google Trends</i> untuk kategori 'Credit and Lending'	0.9456
12	GTI.TM	Indeks <i>Google Trends</i> untuk kategori 'Tele-marketing'	0.9528	26	GTI.H&T	Indeks <i>Google Trends</i> untuk kategori 'Hygiene and Toiletries'	0.9450
13	GTI.M&S	Indeks <i>Google Trends</i> untuk kategori 'Building Materials and Supplies'	0.9526	27	GTI.C&P	Indeks <i>Google Trends</i> untuk kategori 'Construction and Power Tools'	0.9441
14	GTI.PB	Indeks <i>Google Trends</i> untuk kategori 'Plumbing'	0.9525	28	GTI.FL	Indeks <i>Google Trends</i> untuk kategori 'Flooring'	0.9441

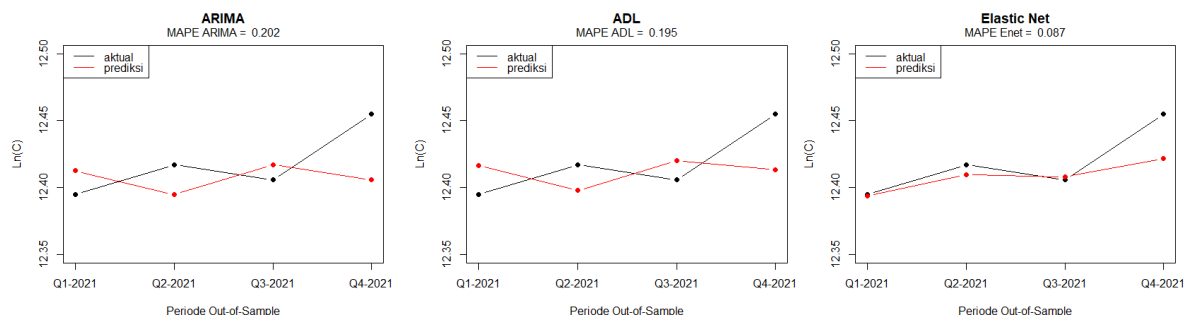
Pada berbagai skenario, model *Ensemble* memberikan hasil prediksi yang paling dekat dengan nilai sebenarnya. Hasil prediksi *Ensemble* sangat baik untuk periode Q1 2021-Q3 2021, namun belum cukup baik untuk Q4 2021. Hal yang menarik adalah pola pergerakan hasil prediksi *Ensemble* sejalan dengan pola pergerakan data aktual. Model *Elastic Net* juga memberikan hasil prediksi yang baik. Sedangkan, keempat model lainnya (ARIMA, ADL, *Support Vector Machine*, dan *Random Forest*) memberikan hasil prediksi yang belum memuaskan.

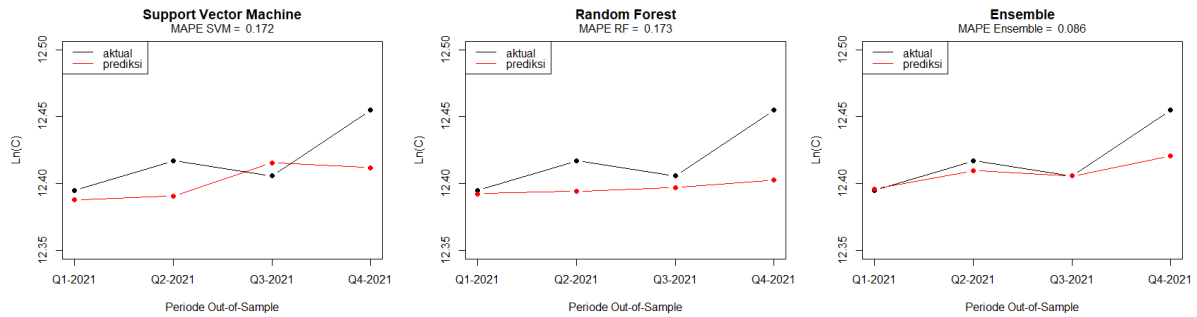
Pada skenario estimasi bertahap, model *Support Vector Machine* dan *Random Forest* sebenarnya cukup baik dalam memprediksi Q1 2021 dan Q3 2021. Namun, keduanya gagal dalam memprediksi Q2 2021 dan Q4 2021. Disamping itu, pola pergerakan hasil prediksi keduanya tidak sejalan dengan pola pergerakan data aktual. Model ARIMA dan ADL menunjukkan pola hasil prediksi yang serupa. Keduanya gagal memprediksi pada hampir seluruh periode *out-of-sample*. Pola prediksi berbagai model pada H+10, H+20 dan H+30 tidak menunjukkan perubahan yang signifikan. Hal tersebut dapat dijelaskan karena pada model ARIMA tidak bergantung pada ketersediaan indikator penunjang, sedangkan pada model lainnya sudah ditunjang oleh 145 indikator GTI yang tersedia *real time*, meskipun 7 indikator non-GTI (DEPO, LOAN.C, r.CL, r.D1m, r.D3m, r.D6m, dan M.OUT) belum tersedia secara lengkap. Perubahan kondisi 7 indikator non-GTI pada H+10, H+20 dan H+30 tidak berdampak signifikan pada model yang dihasilkan.

Pada skenario estimasi tak bertahap, model *Support Vector Machine* sangat baik dalam memprediksi Q1 2021 dan Q3 2021, namun belum cukup baik dalam memprediksi Q2 2021 dan Q4 2021. Model *Random Forest* sangat baik dalam memprediksi Q1 2021, namun gagal memprediksi periode lainnya. Model ADL sangat baik dalam memprediksi Q2 2021, namun gagal memprediksi periode lainnya. Sedangkan, model ARIMA cukup baik dalam memprediksi Q2 2021 dan Q3 2021, namun gagal memprediksi Q1 2021 dan Q4 2021. Prediksi ARIMA pada estimasi tak bertahap cenderung stagnan. Seperti pada estimasi bertahap, pola prediksi berbagai model pada H+10, H+20 dan H+30 tidak menunjukkan perubahan yang signifikan.

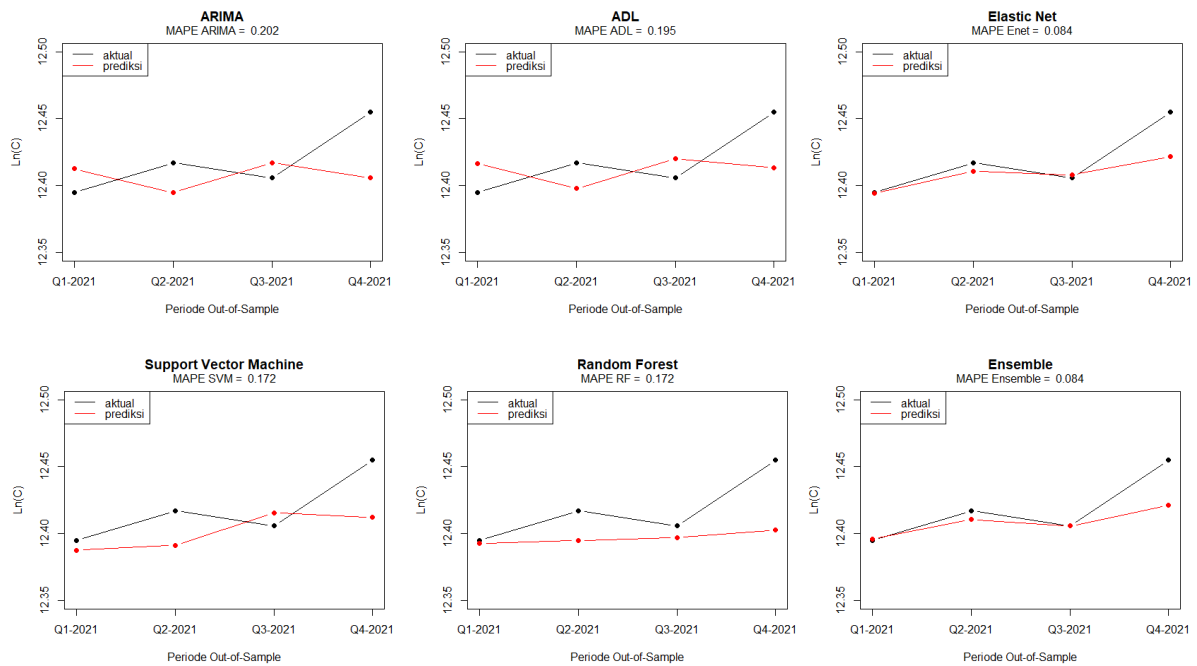


Gambar 5: Hasil Prediksi Berbagai Model (Estimasi Bertahap, Kondisi H+10)

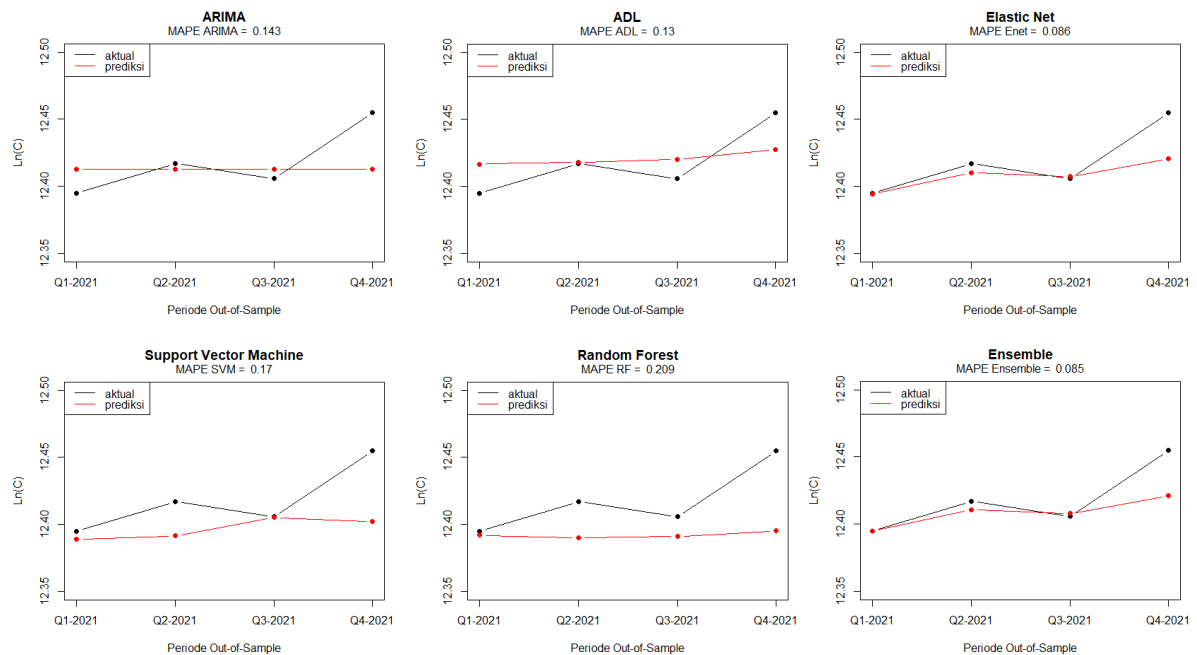




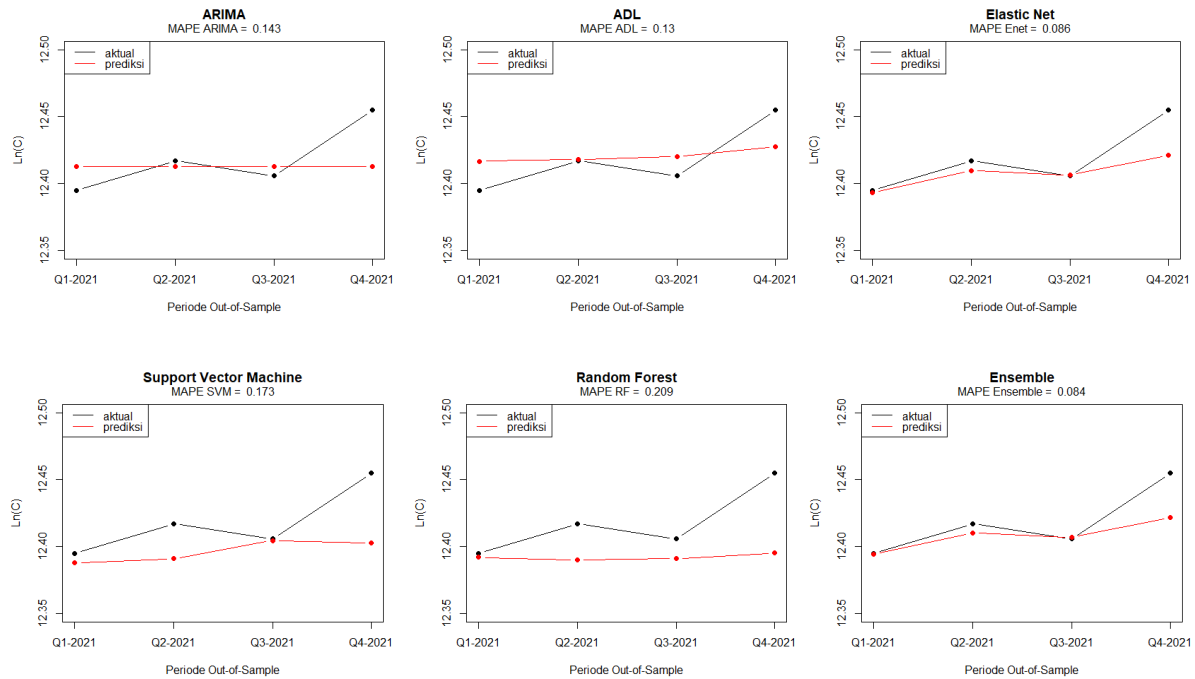
Gambar 6: Hasil Prediksi Berbagai Model (Estimasi Bertahap, Kondisi H+20)



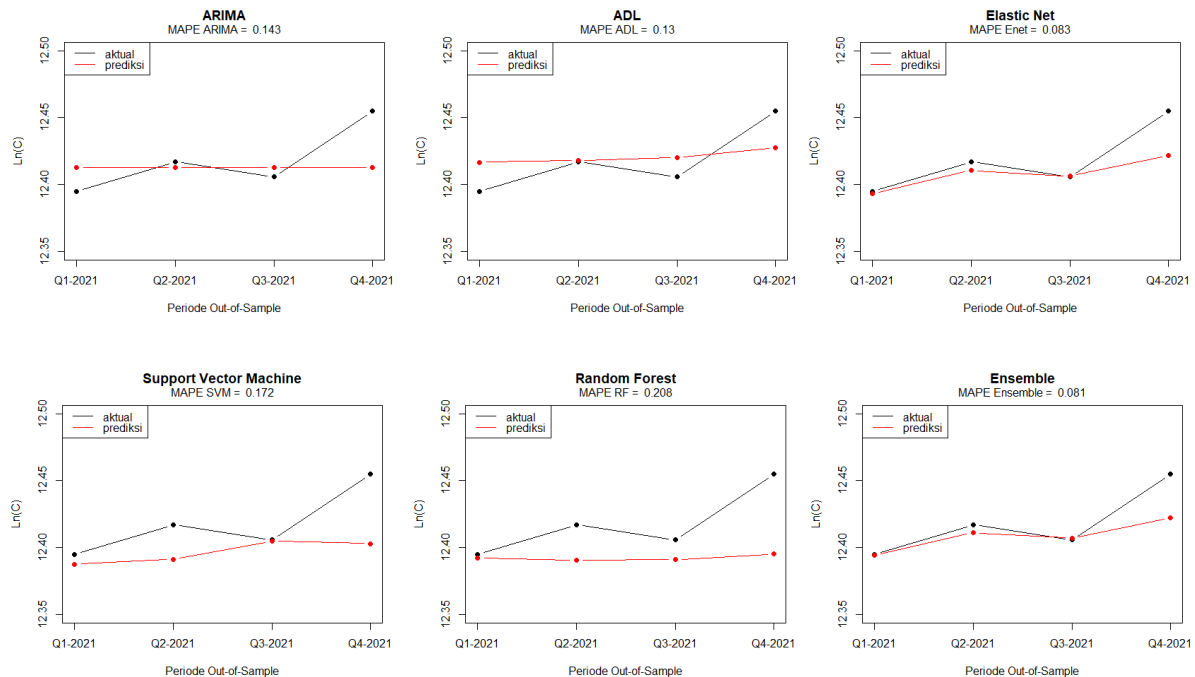
Gambar 7: Hasil Prediksi Berbagai Model (Estimasi Bertahap, Kondisi H+30)



Gambar 8: Hasil Prediksi Berbagai Model (Estimasi Tak Bertahap, Kondisi H+10)



Gambar 9: Hasil Prediksi Berbagai Model (Estimasi Tak Bertahap, Kondisi H+20)



Gambar 10: Hasil Prediksi Berbagai Model (Estimasi Tak Bertahap, Kondisi H+30)

Tabel 6 menyajikan perbandingan kinerja berbagai model pada berbagai skenario, baik menggunakan kriteria *Root Mean Squared Error* (RMSE) maupun *Mean Absolute Percentage Error* (MAPE). Semakin kecil RMSE dan MAPE, semakin baik model yang dihasilkan. Dari tabel tersebut terlihat bahwa *Elastic Net* dan *Ensemble* adalah dua model yang selalu menghasilkan kinerja terbaik pada berbagai skenario. Model *ARIMA* menghasilkan kinerja terburuk pada skenario estimasi bertahap, sedangkan model *Random Forest* menghasilkan kinerja terburuk pada skenario estimasi tak bertahap.

Untuk model *Elastic Net*, *Support Vector Machine*, dan *Ensemble*, kinerja prediksi ketiganya pada skenario estimasi bertahap tidak jauh berbeda dengan skenario estimasi tak bertahap. Sedangkan, model ARIMA, ADL, dan *Random Forest* menghasilkan kinerja yang berbeda signifikan antara skenario estimasi bertahap dan skenario estimasi tak bertahap. Kinerja model ARIMA dan ADL pada estimasi tak bertahap lebih baik dibandingkan dengan estimasi bertahap. Sebaliknya, kinerja model *Random Forest* pada estimasi bertahap lebih baik dibandingkan dengan estimasi tak bertahap.

Pada masing-masing skenario estimasi, kinerja model (khususnya *Elastic Net* dan *Ensemble*) cenderung semakin membaik dari H+10 ke H+30. Hal ini terlihat jelas pada indikator MAPE. Membaiknya kinerja tersebut disebabkan oleh ketersediaan 7 indikator non-GTI (DEPO, LOAN.C, r.CL, r.D1m, r.D3m, r.D6m, dan M.OUT) yang semakin lengkap. Namun, perbaikan kinerja tersebut tidak signifikan. Hal ini menunjukkan bahwa model-model yang dihasilkan sudah cukup stabil karena didukung oleh ketersediaan indikator GTI yang *real time*, *lag* IHK yang minim (H+1), dan estimasi bulan kosong yang cukup baik.

Tabel 6: Perbandingan Akurasi Prediksi PKRT Berbagai Model pada Periode *Out-of-Sample* (Q1 2021-Q4 2021)

Kondisi Indikator	Model	Estimasi Tak Bertahap		Estimasi Bertahap	
		RMSE	MAPE	RMSE	MAPE
H + 10	ARIMA	0.023	0.143	0.029	0.202
	ADL	0.019	0.130	0.026	0.195
	Elastic Net	0.017	0.086	0.017	0.088
	SVM	0.029	0.170	0.026	0.172
	Random Forest	0.034	0.209	0.029	0.173
	Ensemble	0.017	0.085	0.018	0.087
H + 20	ARIMA	0.023	0.143	0.029	0.202
	ADL	0.019	0.130	0.026	0.195
	Elastic Net	0.017	0.086	0.017	0.087
	SVM	0.029	0.173	0.026	0.172
	Random Forest	0.034	0.209	0.029	0.173
	Ensemble	0.017	0.084	0.017	0.086
H + 30	ARIMA	0.023	0.143	0.029	0.202
	ADL	0.019	0.130	0.026	0.195
	Elastic Net	0.017	0.083	0.017	0.084
	SVM	0.029	0.172	0.026	0.172
	Random Forest	0.033	0.208	0.029	0.172
	Ensemble	0.017	0.081	0.017	0.084

Penelitian ini juga menyajikan perbandingan hasil prediksi pertumbuhan PKRT, baik pertumbuhan *year-on-year* (YoY) maupun pertumbuhan *cumulative-to-cumulative* (CtoC). Hasil prediksi pertumbuhan YoY disajikan dalam Tabel 7, sedangkan hasil prediksi pertumbuhan CtoC disajikan dalam Tabel 8. Dari kedua tabel tersebut, terlihat bahwa model *Elastic Net* dan *Ensemble* menunjukkan hasil prediksi pertumbuhan yang lebih baik dibandingkan dengan model-model lainnya.

Pada pertumbuhan YoY (Tabel 7), hanya model *Elastic Net* dan *Ensemble* yang berhasil memprediksi arah pertumbuhan secara benar pada berbagai skenario dan periode *out-of-*

sample. Prediksi pertumbuhan YoY oleh model *Elastic Net* dan *Ensemble* untuk periode Q1 2021-Q3 2021 mendekati pertumbuhan YoY aktual. Namun, prediksi pertumbuhan YoY oleh model *Elastic Net* dan *Ensemble* untuk periode Q4 2021 lebih rendah daripada pertumbuhan YoY aktual. Pada skenario estimasi tak bertahap, model ARIMA dan ADL gagal memprediksi arah pertumbuhan Q1 2021, sedangkan model *Support Vector Machine* dan *Random Forest* gagal memprediksi arah pertumbuhan Q4 2021. Pada skenario estimasi bertahap, model ARIMA gagal memprediksi arah pertumbuhan Q1 2021 dan Q4 2021, sedangkan model ADL gagal memprediksi arah pertumbuhan Q1 2021. Model *Support Vector Machine* dan *Random Forest* gagal memprediksi arah pertumbuhan Q4 2021. Hal ini terjadi pada berbagai kondisi ketersediaan indikator (H+10, H+20 dan H+30).

Tabel 7: Perbandingan Prediksi Pertumbuhan YoY (%) dari Berbagai Model pada Periode Out-of-Sample (Q1 2021-Q4 2021)

Skenario	Model	Estimasi Tak Bertahap				Estimasi Bertahap				
		Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2021	Q2 2021	Q3 2021	Q4 2021	
Pertumbuhan YoY Aktual:		-1.51	5.82	2.27	4.31	-1.51	5.82	2.27	4.31	
Pertumbuhan YoY Prediksi:										
H + 10	ARIMA	0.26	5.34	2.96	0.00		0.26	3.48	3.42	-0.66
	ADL	0.68	5.92	3.76	1.49		0.68	3.79	3.74	0.07
	Elastic Net	-1.56	5.10	2.42	0.80		-1.49	5.13	2.58	0.82
	SVM	-2.09	3.15	2.21	-1.03		-2.09	3.13	3.33	-0.14
	Random Forest	-1.77	3.01	0.76	-1.72		-1.73	3.43	1.39	-0.99
	Ensemble	-1.48	5.16	2.48	0.85		-1.34	5.11	2.30	0.76
H + 20	ARIMA	0.26	5.34	2.96	0.00		0.26	3.48	3.42	-0.66
	ADL	0.68	5.92	3.76	1.49		0.68	3.79	3.74	0.07
	Elastic Net	-1.65	5.03	2.32	0.88		-1.59	5.05	2.48	0.90
	SVM	-2.16	3.08	2.13	-0.96		-2.16	3.06	3.25	-0.07
	Random Forest	-1.77	3.01	0.76	-1.72		-1.73	3.43	1.39	-0.99
	Ensemble	-1.57	5.09	2.38	0.93		-1.39	5.04	2.25	0.84
H + 30	ARIMA	0.26	5.34	2.96	0.00		0.26	3.48	3.42	-0.66
	ADL	0.68	5.92	3.76	1.49		0.68	3.79	3.74	0.07
	Elastic Net	-1.64	5.13	2.32	0.90		-1.57	5.16	2.48	0.93
	SVM	-2.21	3.13	2.16	-0.95		-2.21	3.11	3.28	-0.06
	Random Forest	-1.75	3.04	0.76	-1.72		-1.72	3.48	1.39	-0.99
	Ensemble	-1.55	5.19	2.38	0.95		-1.37	5.14	2.25	0.87

Keterangan:

	Tidak searah dengan pertumbuhan aktual
	Searah dengan pertumbuhan aktual

Pada pertumbuhan CtoC (Tabel 8), seluruh model *machine learning* (*Elastic Net*, *Support Vector Machine*, *Random Forest* dan *Ensemble*) berhasil memprediksi arah pertumbuhan secara benar pada berbagai skenario dan periode *out-of-sample*. Sedangkan, model ARIMA dan ADL gagal memprediksi arah pertumbuhan pada Q1 2021. Model ARIMA dan ADL memprediksi pertumbuhan CtoC pada Q1 2021 adalah positif, sedangkan pertumbuhan aktual adalah negatif. Hal ini terjadi pada berbagai kondisi ketersediaan indikator (H+10, H+20 dan H+30).

Tabel 8: Perbandingan Prediksi Pertumbuhan CtoC (%) dari Berbagai Model pada Periode Out-of-Sample (Q1 2021-Q4 2021)

Skenario	Model	Estimasi Tak Bertahap				Estimasi Bertahap			
		Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2021	Q2 2021	Q3 2021	Q4 2021
Pertumbuhan CtoC Aktual:		-1.51	2.06	2.13	2.69	-1.51	2.06	2.13	2.69
Pertumbuhan CtoC Prediksi:									
H + 10	ARIMA	0.26	2.74	2.81	2.09	0.26	1.83	2.36	1.59
	ADL	0.68	3.24	3.41	2.92	0.68	2.20	2.71	2.04
	Elastic Net	-1.56	1.69	1.93	1.64	-1.49	1.74	2.02	1.71
	SVM	-2.09	0.46	1.05	0.52	-2.09	0.45	1.41	1.02
	Random Forest	-1.77	0.56	0.63	0.03	-1.73	0.79	0.99	0.48
	Ensemble	-1.48	1.76	2.00	1.70	-1.34	1.81	1.97	1.66
H + 20	ARIMA	0.26	2.74	2.81	2.09	0.26	1.83	2.36	1.59
	ADL	0.68	3.24	3.41	2.92	0.68	2.20	2.71	2.04
	Elastic Net	-1.65	1.61	1.84	1.60	-1.59	1.65	1.93	1.67
	SVM	-2.16	0.40	0.97	0.48	-2.16	0.38	1.34	0.98
	Random Forest	-1.77	0.56	0.63	0.03	-1.73	0.79	0.99	0.48
	Ensemble	-1.57	1.68	1.91	1.66	-1.39	1.75	1.91	1.64
H + 30	ARIMA	0.26	2.74	2.81	2.09	0.26	1.83	2.36	1.59
	ADL	0.68	3.24	3.41	2.92	0.68	2.20	2.71	2.04
	Elastic Net	-1.64	1.66	1.88	1.63	-1.57	1.72	1.97	1.71
	SVM	-2.21	0.39	0.98	0.49	-2.21	0.38	1.35	0.99
	Random Forest	-1.75	0.59	0.65	0.04	-1.72	0.82	1.01	0.50
	Ensemble	-1.55	1.73	1.95	1.70	-1.37	1.81	1.95	1.68

Keterangan:

	Tidak searah dengan pertumbuhan aktual
	Searah dengan pertumbuhan aktual

Simpulan

Model *Elastic Net* dan *Ensemble* menghasilkan kinerja prediksi terbaik dibandingkan dengan model-model lainnya. Model *Elastic Net* dan *Ensemble* sangat baik dalam memprediksi level dan pertumbuhan PKRT, khususnya pada periode Q1 2021-Q3 2021. Prediksi dengan model *Elastic Net* dan *Ensemble* dapat dilakukan secara bertahap maupun tak bertahap karena keduanya menghasilkan akurasi yang tidak jauh berbeda. Model *Elastic Net* dan *Ensemble* memberikan keuntungan lain dari sisi statistik karena dapat digunakan pada kondisi jumlah prediktor jauh lebih banyak daripada jumlah amatan. Model-model tersebut juga tidak memerlukan asumsi yang ketat seperti pada model-model ekonometrik.

Penelitian ini menemukan bahwa estimasi cepat (*flash estimates*) PKRT dapat dilakukan sejak 10 hari setelah triwulan referensi berakhir (H+10). Akurasi estimasi pada H+10 tidak jauh berbeda dibandingkan dengan akurasi estimasi pada H+20 dan H+30. Hal ini terjadi karena model yang dibangun didukung oleh indikator *real time* seperti *Google Trends Index* (GTI). Penelitian ini menemukan bahwa beberapa kategori GTI memiliki korelasi yang sangat kuat dengan PKRT. Hal ini tentu menjadi keuntungan untuk pengambilan keputusan yang lebih cepat.

Daftar Pustaka

- Aastveit, K. A., Fastbo, T. M., Granziera, E., Paulsen, K. S., & Torstensen, K. N. (2020). Nowcasting Norwegian Household Consumption with Debit Card Transaction Data. *Norges Bank Research Working Paper No. 17*.
- Banbura, M., & Runstler, G. (2011). A Look into The Factor Model Black Box: Publication Lags and The Role of Hard and Soft Data in Forecasting GDP. *International Journal of Forecasting*, 27(2011), 333-346.
- Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-Casting and The Real-Time Data Flow. *European Central Bank Working Paper Series No. 1564*.
- BEA. (2022). *Technical Note: Gross Domestic Product, First Quarter 2022 (Advance Estimate)*. Bureau of Economic Analysis.
- Bok, B., Caratelli, D., Giannone, D., Sbordon, A., & Tambalotti, A. (2018). Macroeconomic Nowcasting and Forecasting with Big Data. *Annual Review of Economics*, 10, 615-643.
- BPS Jawa Timur. (2022). *Provinsi Jawa Timur dalam Angka, 2022*. Surabaya: BPS Jawa Timur.
- Buell, B., Chen, C., Cherif, R., Seo, H.-J., Tang, J., & Wendt, N. (2021). Impact of COVID-19: Nowcasting and Big Data to Track Economic Activity in Sub-Saharan Africa. *IMF Working Paper, WP/21/124*.
- Castle, J., Hendry, D., & Kitov, O. (2017). *Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview*. Eurostat.
- Ceron, A., Curini, L., & Iacus, S. M. (2017). *Politics and Big Data: Nowcasting and Forecasting Elections with Social Media*. New York: Routledge.
- Cryer, J., & Chan, K-S. (2008). *Time Series Analysis with Applications in R (Second ed.)*. New York: Springer.
- Damuri, Y. R., Tyas, P., Aswicahyono, H., Priyadi, L., Kusumawardhani, S., & Yazid, E. K. (2021). Tracking the Ups and Downs in Indonesia's Economic Activity During COVID-19 Using Mobility Index: Evidence from Provinces in Java and Bali. *ERIA Discussion Paper Series No. 385*.
- Eurostat. (2016). *Overview of GDP Flash Estimation Methods*. Luxembourg: Publications Office of the European Union.
- Gil, M., Perez, J., Sanchez-Fuentes, A., & Urtasun, A. (2017). Nowcasting Private Consumption: Traditional Indicators, Uncertainty Measures, and The Role of Internet Search Query Data. *ISI World Statistics Congress*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second ed.)*. New York: Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer.
- Luciani, M., Pundit, M., Ramayandi, A., & Veronese, G. (2015). Nowcasting Indonesia. *ADB Economics Working Paper Series No. 471*.
- Montgomery, D., Jennings, C., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting (Second ed.)*. New Jersey: John Wiley & Sons, Inc.
- Muchisha, N. D., Tamara, N., Andriansyah, & Soleh, A. M. (2021). Nowcasting Indonesia's GDP Growth Using Machine Learning Algorithms. *Indonesian Journal of Statistics and Its Applications*, 5(2), 355-368.
- Richardson, A., Mulder, T. v., & Vehbi, T. (2018). Nowcasting New Zealand GDP Using Machine Learning Algorithms. *Bank Indonesia International Workshop and Seminar on "Big Data for Central Bank Policies/Building Pathways for Policy Making with Big Data"*.
- Sampi, J., & Jooste, C. (2020). Nowcasting Economic Activity in Times of COVID-19: An Approximation from the Google Community Mobility Report. *World Bank Policy Research Working Paper No. 9247*.

- Tarsidin, Idham, & Rakhman, R. N. (2018). Nowcasting Household Consumption and Investment in Indonesia. *Bulletin of Monetary Economics and Banking*, 20(3), 375-404.
- Toth, I. J., & Hajdu, M. (2020). Google as a Tool for Nowcasting Household Consumption: Estimations on Hungarian Data. *Institute for Economic and Enterprise Research*.
- Utari, D. T., & Ilma, H. (2018). Comparison of Methods for Mixed Data Sampling (MIDAS) Regression Models to Forecast Indonesian GDP Using Agricultural Exports. *AIP Conference Proceeding: The 8th Annual Basic Science International Conference*, (p. 060016).
- Varlamova, J., & Larionova, N. (2015). Macroeconomic and Demographic Determinants of Household Expenditures in OECD Countries. *Procedia Economics and Finance*, 24(2015), 727-733.
- Woloszko, N. (2020). Tracking Activity in Real Time With Google Trends. *OECD Working Paper No. 1634, ECO/WKP(2020)42*.
- Woo, J., & Owen, A. (2018). Forecasting Private Consumption with Google Trends Data. *Journal of Forecasting*, 38(2).
- Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via The Elastic Net. *Journal of The Royal Statistical Society*, 67(2), 301-320.

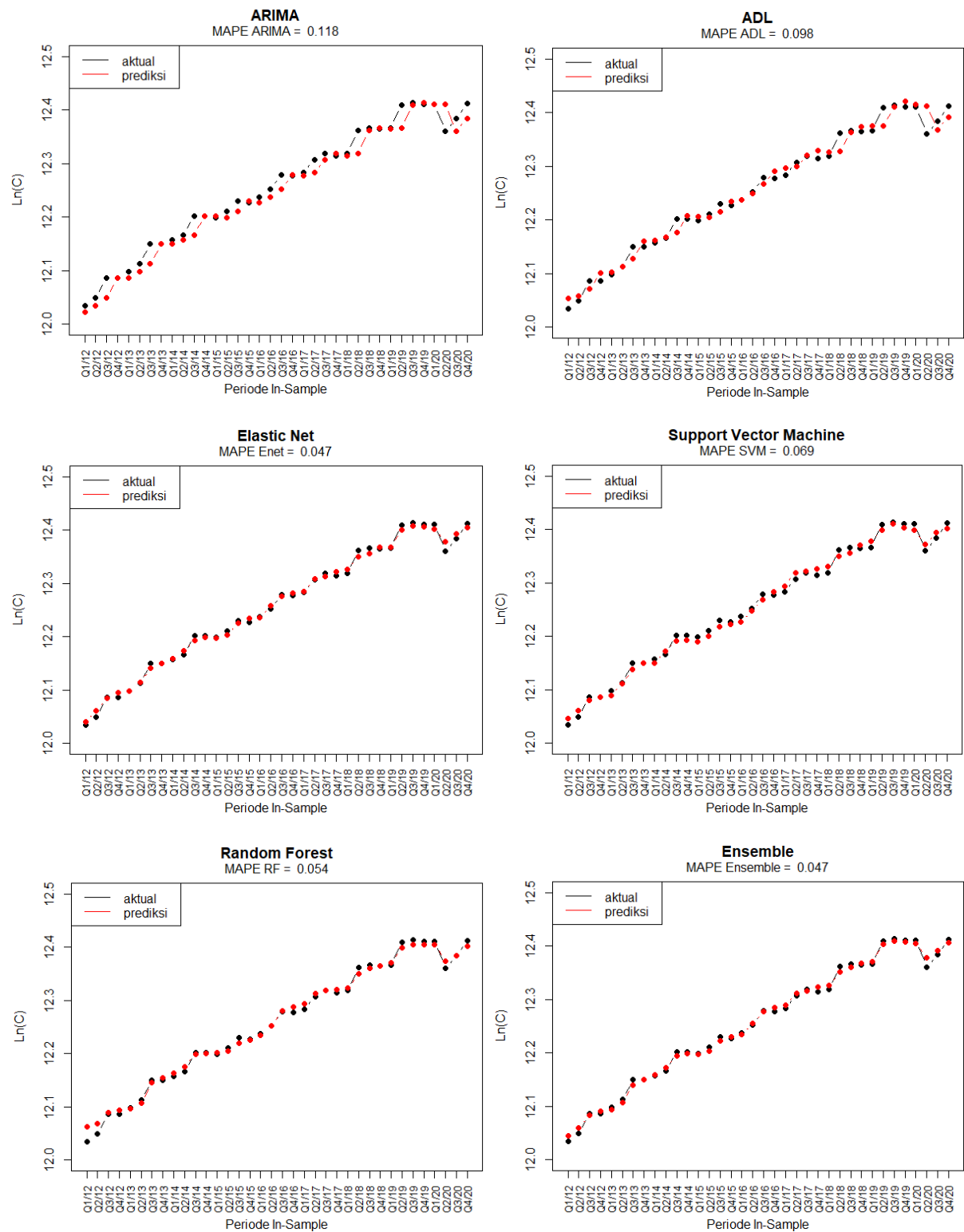
Lampiran

Lampiran 1: Daftar Kategori Data *Google Trends* yang Memiliki Korelasi Kuat dengan Pengeluaran Konsumsi Rumah Tangga ($r > 0.90$)

No	Indeks Data	Kategori GTI	Koefisien Korelasi	No	Indeks Data	Kategori GTI	Koefisien Korelasi
1	85	TV Guides & Reference	0.9180	74	456	GERD & Digestive Disorders	0.9066
2	90	Art & Craft Supplies	0.9197	75	463	Pain Management	0.9346
3	113	Trucks & SUVs	0.9243	76	464	Respiratory Conditions	0.9097
4	134	Hummer	0.9263	77	465	Skin Conditions	0.9364
5	135	Hyundai	0.9066	78	466	Sleep Disorders	0.9055
6	140	Lamborghini	0.9089	79	473	Medical Facilities & Services	0.9417
7	158	Subaru	0.9067	80	475	Hospitals & Treatment Centers	0.9018
8	166	Vehicle Maintenance	0.9115	81	479	Medical Literature & Resources	0.9022
9	168	Auto Exterior	0.9175	82	488	Assisted Living & Long Term Care	0.9135
10	184	Cosmetology & Beauty Professionals	0.9106	83	491	Vitamins & Supplements	0.9198
11	186	Hygiene & Toiletries	0.9450	84	492	Oral & Dental Care	0.9350
12	188	Perfumes & Fragrances	0.9494	85	493	Pediatrics	0.9182
13	189	Skin & Nail Care	0.9275	86	494	Pharmacy	0.9300
14	213	Advertising & Marketing	0.9017	87	495	Drugs & Medications	0.9188
15	217	Telemarketing	0.9528	88	497	Occupational Health & Safety	0.9097
16	225	Food Production	0.9142	89	511	Eyeglasses & Contacts	0.9175
17	229	Business Education	0.9022	90	531	Pet Food & Supplies	0.9124

No	Indeks Data	Kategori GTI	Koefisien Korelasi	No	Indeks Data	Kategori GTI	Koefisien Korelasi
18	242	Human Resources	0.9083	91	532	Veterinarians	0.9133
19	244	Business Services	0.9366	92	535	Cats	0.9060
20	247	E-Commerce Services	0.9470	93	543	Zoos-Aquariums-Pre-serves	0.9016
21	248	Fire & Security Services	0.9279	94	551	Gifts	0.9040
22	252	Outsourcing	0.9243	95	552	Party & Holiday Supplies	0.9332
23	258	Chemicals Industry	0.9459	96	559	Boating	0.9282
24	260	Cleaning Agents	0.9299	97	560	Diving & Underwater Activities	0.9191
25	261	Coatings & Adhesives	0.9487	98	563	Home & Garden	0.9085
26	262	Dyes & Pigments	0.9210	99	564	Bed & Bath	0.9262
27	263	Plastics & Polymers	0.9488	100	565	Bathroom	0.9582
28	264	Construction & Maintenance	0.9358	101	566	Bedroom	0.9129
29	265	Building Materials & Supplies	0.9526	102	569	Gardening & Landscaping	0.9243
30	266	Doors & Windows	0.9459	103	570	Home Appliances	0.9375
31	269	Construction Consulting & Contracting	0.9119	104	571	Major Kitchen Appliances	0.9088
32	277	Water Supply & Treatment	0.9142	105	574	Home Furnishings	0.9284
33	278	Enterprise Technology	0.9401	106	575	Clocks	0.9305
34	285	Grocery & Food Retailers	0.9128	107	576	Lamps & Lighting	0.9064
35	286	Industrial Materials & Equipment	0.9101	108	578	Sofas & Chairs	0.9065
36	287	Fluid Handling	0.9301	109	579	Home Improvement	0.9488
37	288	Generators	0.9001	110	580	Construction & Power Tools	0.9441
38	289	Heavy Machinery	0.9239	111	581	Flooring	0.9441
39	294	Printing & Publishing	0.9305	112	583	Plumbing	0.9525
40	295	Document & Printing Services	0.9199	113	585	Home Storage & Shelving	0.9124
41	296	Journalism & News Industry	0.9282	114	586	Homemaking & Interior Decor	0.9212
42	298	Retail Trade	0.9306	115	587	Kitchen & Dining	0.9314
43	299	Retail Equipment & Technology	0.9069	116	588	Cookware & Diningware	0.9406
44	304	Textiles & Nonwovens	0.9353	117	589	Laundry	0.9496
45	310	Mail & Package Delivery	0.9024	118	593	Yard & Patio	0.9098
46	333	Electronic Accessories	0.9306	119	628	Study Abroad	0.9201
47	341	Optoelectronics & Fiber	0.9157	120	653	State & Local Government	0.9142
48	349	VPN & Remote Access	0.9061	121	661	Family Law	0.9300
49	370	Finance	0.9177	122	667	Air Force	0.9032

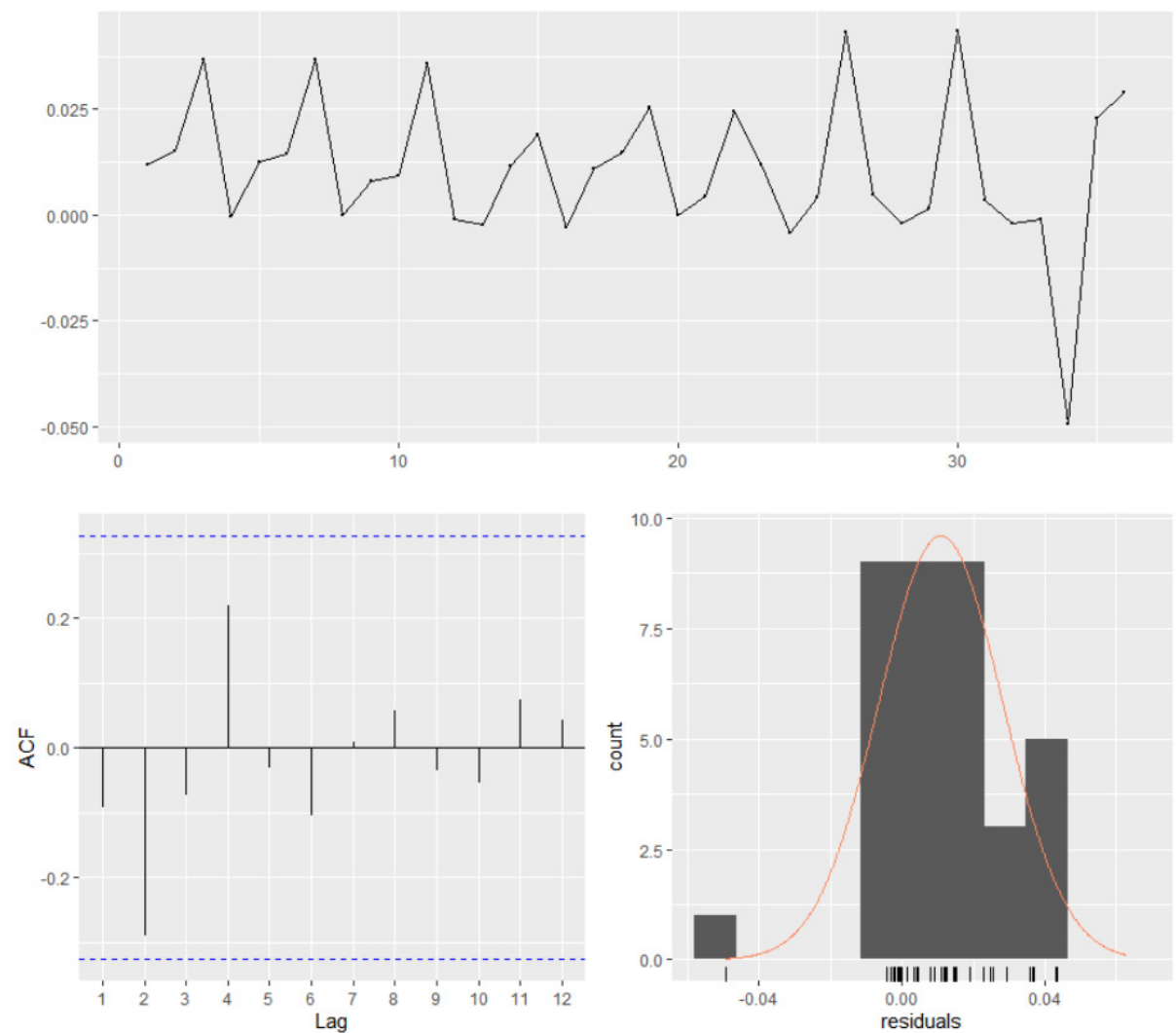
No	Indeks Data	Kategori GTI	Koefisien Korelasi	No	Indeks Data	Kategori GTI	Koefisien Korelasi
50	374	Credit & Lending	0.9456	123	671	Veterans	0.9343
51	377	Credit Cards	0.9165	124	724	Western Europeans	0.9179
52	378	Debt Management	0.9295	125	740	Islam	0.9130
53	379	Home Financing	0.9577	126	751	Charity & Philanthropy	0.9066
54	380	Currencies & Foreign Exchange	0.9065	127	757	Privacy Issues	0.9169
55	383	Insurance	0.9189	128	768	Commercial & Investment Real Estate	0.9092
56	387	Investing	0.9129	129	774	Timeshares & Vacation Properties	0.9050
57	390	Food & Drink	0.9013	130	775	Reference	0.9100
58	393	Liquor	0.9182	131	779	Calculators & Reference Tools	0.9281
59	394	Wine	0.9216	132	783	How-To, DIY & Expert Content	0.9113
60	395	Candy & Sweets	0.9232	133	793	Language Resources	0.9385
61	399	Meat & Seafood	0.9159	134	794	Foreign Language Resources	0.9491
62	401	Vegetarian Cuisine	0.9136	135	816	Statistics	0.9145
63	404	Non-Alcoholic Beverages	0.9335	136	828	Eyewear	0.9210
64	405	Coffee & Tea	0.9543	137	832	Men's Clothing	0.9194
65	406	Restaurants	0.9241	138	834	Sleepwear	0.9461
66	407	Dining Guides	0.9064	139	848	Luxury Goods	0.9284
67	408	Fast Food	0.9130	140	856	Wholesalers & Liquidators	0.9359
68	444	Cleansing & Detoxification	0.9123	141	858	College Sports	0.9126
69	445	Health Conditions	0.9175	142	868	Golf	0.9119
70	447	Allergies	0.9249	143	877	Baseball	0.9407
71	451	Diabetes	0.9016	144	902	Ecotourism	0.9223
72	452	Ear Nose & Throat	0.9031	145	909	Mountain & Ski Resorts	0.9012
73	454	Endocrine Conditions	0.9027				

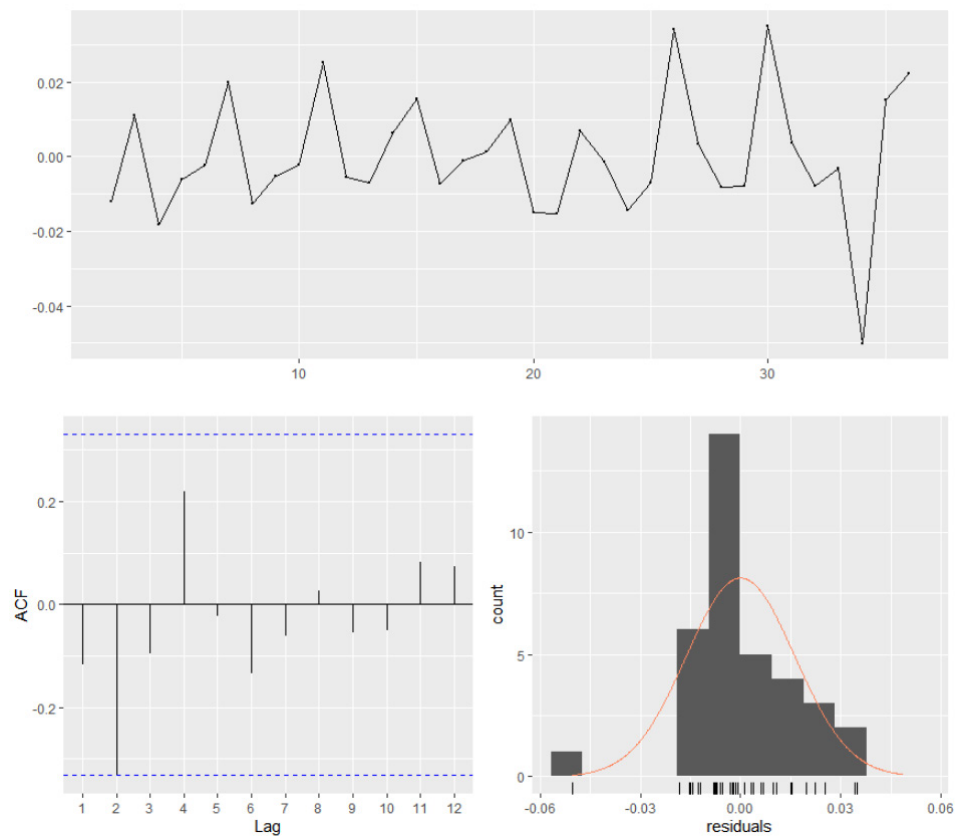
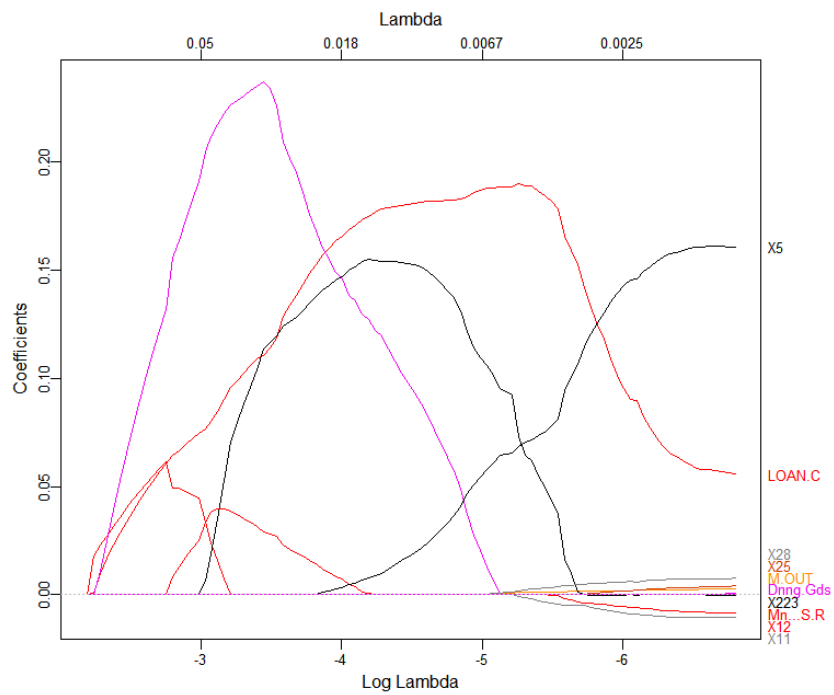
Lampiran 2: Hasil Prediksi Berbagai Model pada Periode *In-Sample*

Lampiran 3: Perbandingan Akurasi Prediksi PKRT dari Berbagai Model pada Periode *In-Sample*

Model	Ukuran Akurasi	
	RMSE	MAPE
ARIMA	0.020	0.118
ADL	0.016	0.098
Elastic Net	0.007	0.047
SVM	0.009	0.069
Random Forest	0.009	0.054
Ensemble	0.007	0.047

Lampiran 4: Diagnostik Residual pada Model ARIMA (Contoh pada Estimasi Bertahap, H+10)



Lampiran 5: Diagnostik Residual pada Model ADL (Contoh pada Estimasi Bertahap, H+10)**Lampiran 6: Penyusutan Koefisien *Elastic Net* (Contoh pada Estimasi Bertahap, H+10)**

Lampiran 7: Variable Importance *Random Forest* (Contoh pada Estimasi Bertahap, H+10)