

Perancangan Model Peramalan Jangka Pendek Harga Komoditas Pertanian di Indonesia Menggunakan Machine Learning

Muhammad Ammar Erdianto

Fakultas Teknologi Industri, Teknik dan Manajemen Industri, Institut Teknologi Bandung, Bandung, Indonesia

Email: erdiantoammar@gmail.com

Email Penulis Korespondensi: erdiantoammar@gmail.com

Abstrak—Fluktuasi harga komoditas pertanian seringkali menyebabkan kerugian terutama pada masyarakat golongan menengah ke bawah. Prediksi harga di masa mendatang merupakan salah satu *input* penting bagi perumusan kebijakan setiap entitas terkait komoditas pertanian. Penelitian ini bertujuan untuk mengembangkan model peramalan jangka pendek pada harga sepuluh komoditas dengan metode berbasis *machine learning*. Variabel prediktor yang digunakan adalah indeks pencarian kata kunci dari Google Trends yang merepresentasikan ketertarikan masyarakat terhadap komoditas terkait. Penentuan kata kunci dilakukan menggunakan metode *web scraping* pada berbagai kanal berita di Indonesia. Sebelum memasuki tahap pemodelan, seluruh variabel terlebih dahulu melalui pemrosesan awal dan seleksi *lag* optimal. Alternatif model yang digunakan adalah regresi linier, *random forest*, *gradient boosting*, dan *regression tree* yang merupakan beberapa model berbasis *machine learning*. Implementasi model yang diukur menggunakan nilai MAPE memiliki performa akurat dengan tujuh dari sepuluh komoditas memiliki nilai di bawah 10%. Apabila dibandingkan dengan penelitian sebelumnya, model usulan memiliki performa yang lebih baik dengan nilai MAPE yang lebih rendah. Model usulan diharapkan dapat menjadi masukan baru bagi pengambilan kebijakan oleh pemerintah, pelaku bisnis, maupun konsumen terkait komoditas pertanian sehingga dapat membantu mengurangi fluktuasi harga.

Kata Kunci: Google Trends; Gradient Boosting; Komoditas Pertanian; Machine Learning; Random Forest; Regresi Linier;

Abstract—Agricultural commodity prices fluctuations often cause losses, especially for middle to lower-class people. Predicting future prices is essential for formulating policies for each entity related to agricultural commodities. This study aims to develop a short-term forecasting model for the prices of ten commodities using a machine learning-based method. The predictor variable used was the keyword search index from Google Trends, which represents the public interest in the related commodity. The determination of keywords was done using the web scraping method on various news channels in Indonesia. Before entering the modeling stage, all variables undergone pre-processing and optimal lag selection. The alternative models used were linear regression, random forest, gradient boosting, and regression trees, which were several models based on machine learning. Model implementation measured using the MAPE score has accurate performance with seven out of ten commodities having a value below 10%. Compared with previous studies, the proposed model performs better with lower MAPE values. The proposed model is expected to be a new input for policy-making by the government, business players, and consumers related to agricultural commodities to help reduce price fluctuations.

Keywords: Food Commodities; Google Trends; Gradient Boosting; Linear Regression; Machine Learning; Random Forest;

1. PENDAHULUAN

Komoditas pertanian merupakan salah satu sumber kebutuhan primer yang selalu melekat dalam setiap individu manusia. Perannya sebagai bahan makanan membuat komoditas pertanian menjadi kebutuhan hidup utama bagi seluruh masyarakat tak terkecuali masyarakat Indonesia. Secara umum, fluktuasi harga dapat diartikan sebagai perubahan harga komoditas di pasar sebagai hasil dari perubahan situasi pasar, pola musiman, dan perubahan kondisi permintaan dan persediaan [1]. Fluktuasi harga menyebabkan kerugian terutama pada masyarakat golongan menengah ke bawah yang tidak dapat membeli kebutuhan pada saat harga tinggi sehingga menyebabkan pola konsumsi mereka tidak teratur.

Melakukan peramalan harga komoditas pertanian secara akurat merupakan hal penting untuk meminimalkan risiko pasar, meningkatkan pendapatan dari sektor pertanian, dan membantu meningkatkan efektivitas kebijakan makroekonomi pemerintah suatu negara [2]. Di sisi lain, ketidakstabilan harga dan prediksi harga yang tidak akurat dapat memicu krisis baik pada konsumen, entitas bisnis pertanian, hingga ekonomi nasional [3]. Hal ini menunjukkan pentingnya melakukan peramalan harga komoditas pertanian. Meskipun demikian, rendahnya akurasi menjadi kendala utama yang selalu dihadapi seperti pada penelitian [4] yang melakukan *forecasting* pada komoditas daging sapi, daging ayam, bawang, dan cabai di Indonesia dengan nilai MAPE untuk bawang sebesar 32,79% dan cabai sebesar 13,62%. Contoh lain adalah penelitian dari [5] mengenai *forecasting* pada harga komoditas cabai merah di Indonesia menggunakan metode *time-series* seperti *moving-average*, *exponential smoothing*, dan ARIMA dengan nilai MAPE terendah hasil pemodelan sebesar 14%. Rendahnya akurasi peramalan pada penelitian sebelumnya menunjukkan pentingnya pengembangan model peramalan.

Peramalan atau *forecasting* merupakan salah satu metode untuk melakukan perkiraan dan prediksi terhadap suatu variabel di masa mendatang berdasarkan pola yang terjadi di masa lalu [6]. Terdapat tiga komponen penting dari konsep *forecasting* yaitu horizon waktu, jenis data, dan metode yang digunakan. Horizon waktu dapat diklasifikasikan menjadi tiga yaitu jangka pendek, jangka menengah, dan jangka panjang [6]. Dalam konteks harga komoditas pertanian, menurut [7], peramalan secara jangka pendek semakin banyak digunakan untuk penelitian sejak abad 21. Penggunaan jangka pendek didasarkan pada seringkali terdapat faktor yang terjadi secara tiba-tiba dan menyebabkan fluktuasi secara jangka pendek yang dapat menyebabkan dampak negatif dengan efek domino baik dari segi produksi, konsumsi, maupun stabilitas sosial [7]. Disrupsi yang dapat memengaruhi harga komoditas pertanian seringkali terjadi dalam rentang waktu yang pendek [8]. Sebagai contoh yang pernah terjadi, terjadi peningkatan harga bawang putih secara signifikan pada

rentang waktu 6 Mei 2019 hingga 10 Mei 2019 karena permasalahan impor dan lonjakan permintaan. Secara lebih spesifik, diperlukan pendefinisian rentang waktu jangka pendek yang digunakan. Alternatif jangka pendek yang sering digunakan adalah antara satu minggu hingga satu triwulan [7]. Pemodelan kemudian dilakukan untuk membandingkan jangka waktu yang digunakan dan diperoleh bahwa pemodelan dengan jangka waktu satu minggu menghasilkan akurasi lebih baik dibandingkan dengan jangka waktu satu bulan maupun satu triwulan.

Berdasarkan variabel input, model peramalan dapat diklasifikasikan menjadi dua kategori yaitu metode *time-series* dan metode kausal. Metode *time-series* mencoba membentuk model menggunakan data historis untuk melakukan ekstrapolasi ke masa depan dengan asumsi bahwa pola data selalu berulang [9]. Metode *time-series* memiliki kelebihan yaitu mudah untuk dikembangkan dan diimplementasikan karena hanya memerlukan data historis [10]. Penggunaan data historis juga memungkinkan untuk dapat mengakomodasi adanya pola musiman atau tren yang terjadi pada komoditas pertanian. Di sisi lain, metode kausal berusaha membentuk model menggunakan variabel lain sebagai penjelas dan lebih sering digunakan apabila variabel penjelas tersebut telah diketahui [9]. Kelebihan dari metode kausal adalah kemampuannya untuk memprediksi pola fluktuasi yang tidak umum atau tidak hanya bersifat tren dan musiman sehingga memungkinkan tetap akurat meskipun data bersifat fluktuatif [11]. Selain itu, metode kausal juga memungkinkan untuk digunakan sebagai metode *time-series* melalui penggunaan variabel historis dan variabel waktu. Oleh karena itu, untuk meningkatkan akurasi model, pada penelitian ini digunakan metode kausal.

Pada peramalan jangka pendek dengan metode kausal, diperlukan juga data dengan frekuensi tinggi seperti data harian untuk mengakomodasi peramalan secara *real-time* [12]. Variabel yang memiliki data dengan frekuensi tinggi ini nantinya akan dijadikan variabel prediktor untuk memprediksi variabel target yang dalam penelitian ini adalah harga komoditas pertanian. Berdasarkan penelitian-penelitian sebelumnya, terdapat beberapa data yang dapat digunakan serta bersifat *real-time* dan memiliki frekuensi tinggi. Data pertama yang dapat digunakan adalah jumlah *tweet* pada media sosial Twitter sebagaimana yang telah digunakan oleh [13] dan terbukti efektif untuk melakukan peramalan terhadap harga komoditas daging sapi, daging ayam, bawang putih, dan cabai di Indonesia. Data selanjutnya yang dapat digunakan adalah indeks jumlah pencarian kata dari Google yang dipublikasikan pada Google Trends dalam satuan waktu bulanan, mingguan, hingga harian. Indeks jumlah pencarian kata merepresentasikan seberapa sering sebuah kata (atau kumpulan kata) dicari relatif terhadap total volume pencarian dalam satu waktu dan telah dinormalisasikan ke dalam rentang 0-100 [14]. Indeks pencarian kata dari Google Trends juga telah digunakan oleh [15] untuk membangun model pada penjualan kendaraan bermotor di Chile.

Data yang dipilih untuk digunakan pada penelitian ini adalah indeks pencarian dari Google Trends karena tidak berbayar, mudah dan cepat untuk diperoleh, serta mengakomodasi pengguna yang lebih besar. Dari Google Trends, akan diidentifikasi kata atau kelompok kata yang memiliki korelasi dengan harga pangan setiap komoditas. Diperlukan metodologi untuk mengidentifikasi dan menentukan kata yang sesuai untuk digunakan dalam pemodelan. Kategori yang terdapat pada Google Trends dapat menjadi titik awal proses identifikasi [16]. Selanjutnya digunakan metode *backward induction* dengan cara melakukan ekstraksi kata yang terkait dengan suatu situs web yang paling dicari oleh orang apabila bermaksud mencari informasi tertentu. Sebagai contoh, apabila seseorang bermaksud mencari harga pangan maka orang tersebut akan cenderung mencari informasi dari situs berita harian. Dari artikel pada situs berita tersebut nantinya akan dilakukan ekstraksi pada kata kuncinya. Kata kunci inilah yang akan dijadikan patokan untuk dimasukkan dalam Google Trends untuk memperoleh data indeks pencarian untuk digunakan dalam proses pemodelan.

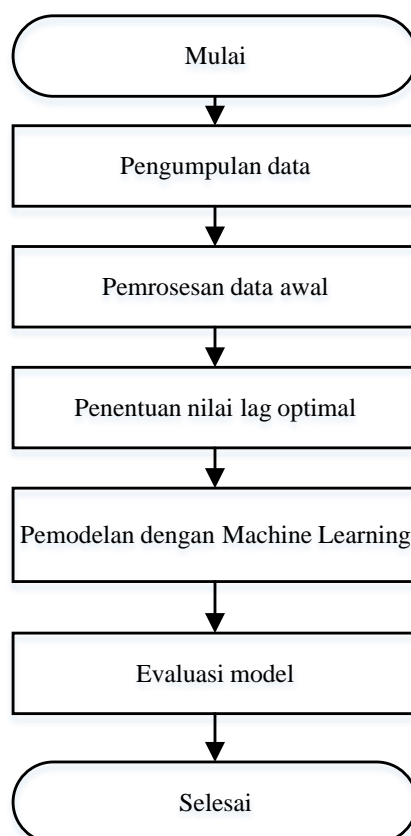
Komponen terakhir dari konsep peramalan adalah pemilihan metode peramalan. Pada penjelasan sebelumnya telah disinggung bahwa kebutuhan metode peramalan adalah memiliki akurasi tinggi dan waktu komputasi yang relatif cepat. Dengan semakin berkembangnya teknologi, semakin besar utilisasi metode berbasis *machine learning* untuk menjawab permasalahan tersebut. *Machine learning* bertujuan untuk membangun sistem komputer yang dapat melakukan perbaikan secara otomatis melalui pengalaman [17] serta telah banyak digunakan dalam pengembangan perangkat lunak, pengenalan suara, pemrosesan bahasa natural, pengontrolan robot, dan masih banyak lagi. Oleh karena itu, pada penelitian ini, pendekatan yang dilakukan untuk melakukan *nowcasting* adalah pendekatan menggunakan prinsip *machine learning* sesuai dengan penelitian sebelumnya [18]. Pemilihan referensi dilakukan berdasarkan semakin luas, beragam, dan berkualitasnya penggunaan *machine learning* sehingga membantu memudahkan metode pengolahan data.

Pada penelitian ini, akan dilakukan perancangan model untuk memprediksi harga komoditas pertanian dengan metode kausal berbasis *machine learning*. Model yang dirancang akan terdiri dari langkah-langkah pemecahan masalah mulai dari pengumpulan data, penentuan kata kunci, pemodelan, hingga evaluasi model. Secara umum, terdapat dua data yang digunakan pada penelitian ini. Data pertama yaitu indeks pencarian kata dari Google Trends yang akan digunakan sebagai variabel prediktor pada saat proses pemodelan secara kausal sedangkan variabel kedua yang nantinya digunakan sebagai variabel target adalah harga komoditas pertanian yang diperoleh dari situs Pusat Informasi Harga Pangan Strategis (PIHPS). Penentuan kata kunci terlebih dahulu dilakukan untuk dijadikan input dalam pengumpulan data indeks pencarian kata dari Google Trends sehingga proses penentuan kata kunci ini juga dilakukan sebelum dilakukan pemodelan. Meskipun dapat diasumsikan bahwa pemerintah telah mempunyai metode tersendiri untuk melakukan peramalan harga, hasil perancangan model pada penelitian ini tetap dapat digunakan untuk memberikan usulan baru terkait metode yang digunakan. Model dari hasil perancangan diharapkan dapat membantu memberikan tambahan masukan untuk mengatasi fluktuasi harga dari sisi perkiraan harga komoditas pertanian.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Bagian ini memuat urutan metodologi penelitian yang terdiri atas beberapa tahapan. Metodologi penelitian yang pertama adalah pengumpulan data yang terdiri atas pengumpulan data harga komoditas pertanian dan pengumpulan data indeks pencarian kata kunci. Tahapan kedua adalah pemrosesan data awal yang terdiri atas agregasi data dan penggabungan data menjadi *dataset* per komoditas. Tahapan ketiga adalah penentuan nilai *lag* optimal pada setiap komoditas menggunakan metode *Recursive Feature Elimination* (RFE). Tahapan selanjutnya adalah pemodelan pada setiap komoditas dengan empat alternatif model yaitu regresi linier, *random forest*, *gradient boosting*, dan *regression tree*. Tahapan terakhir adalah evaluasi model dengan ukuran performansi *Mean Absolute Percentage Error* (MAPE) untuk menilai performa setiap model dan memilih model terbaik pada setiap komoditas. Bagan alir tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

2.2 Pengumpulan Data

Proses pengumpulan data dibagi menjadi dua tahap yang dapat dilakukan secara paralel. Tahap pertama yaitu pengumpulan data untuk variabel harga pangan nasional setiap komoditas yang diperoleh dari Pusat Informasi Harga Pangan Strategis (PIHPS). Pengambilan data dilakukan pada sepuluh komoditas yaitu bawang merah, bawang putih, cabai merah, cabai rawit, beras, telur ayam, daging ayam, daging sapi, minyak goreng, dan gula pasir. Data yang diperoleh adalah data yang bersifat harian. Tahap kedua merupakan pengumpulan data untuk indeks pencarian kata yang diawali dengan penentuan kata kunci. Metode yang digunakan dalam penentuan kata kunci yaitu *backward induction method* dengan cara melakukan ekstraksi kata yang terkait dengan suatu situs web yang terkait dengan topik dari kata kunci [16].

Dalam menentukan kata kunci, terdapat beberapa tahapan yang dilakukan yaitu penentuan kata kunci awal, *web-scraping* pada Google Berita, dan ekstraksi kata kunci menggunakan *text mining*. Penjelasan dari setiap tahapan adalah sebagai berikut:

a. Penentuan kata kunci awal

Kata kunci awal diperoleh dari studi literatur dan judgment dari penulis. Formula penentuan kata kunci yang digunakan adalah sebagai berikut:

1. Kata kunci yang merupakan kata komoditas itu sendiri (misal: daging ayam, bawang merah, bawang putih, dan seterusnya).
2. Kata kunci harga dari setiap komoditas (misal: harga daging ayam, harga bawang merah, harga bawang putih, dan seterusnya).
3. Kata kunci tambahan seperti harga pangan, harga bahan pangan, sembako, impor komoditas, dan makanan pokok.

b. *Web-scraping* pada Google Berita

Pada metode *backward induction* yang digunakan dalam menentukan kata kunci, salah satu bagian yang paling krusial adalah ekstraksi kata kunci dari suatu situs web yang dikunjungi oleh pengguna untuk suatu topik tertentu. Metode yang digunakan untuk memperoleh data dari Google Berita adalah *web-scraping*. *Web scraping* merupakan teknik yang digunakan untuk memperoleh informasi dari situs web secara otomatis [19]. Tujuan dari *web scraping* adalah melakukan pencarian terhadap informasi, ekstraksi, atau agregasi ke dalam format yang lain. Terdapat beberapa keuntungan dari *web scraping* yaitu mencapai otomatisasi untuk menghemat waktu, memperoleh wawasan bisnis, serta mengakomodasi untuk memperoleh banyak data [20]. Dalam konteks penelitian ini, diperoleh beberapa keuntungan yaitu otomatisasi dalam pengambilan data dengan jumlah besar dalam waktu singkat dan membantu dalam memilih halaman berita yang nantinya akan diekstraksi.

Dari kata kunci awal, dilakukan *web-scraping* pada Google News menggunakan API (*Application Programming Interface*) yang disediakan oleh pihak ketiga yaitu Data for SEO. Tujuan dari *web-scraping* ini adalah menemukan berita-berita yang muncul apabila dilakukan pencarian menggunakan kata kunci awal. Dari berita-berita ini, nantinya akan dilakukan ekstraksi untuk menemukan kata kunci akhir yang merupakan turunan dari kata kunci awal. Output dari Data for SEO adalah kode JSON (*JavaScript Object Notation*) yang berisi judul dan isi berita untuk setiap kata kunci awal. Kode ini kemudian dikonversi menjadi tabel Microsoft Excel menggunakan Python. Untuk setiap kata kunci awal, diperoleh 25 berita yang nantinya akan diekstraksi kata kuncinya.

c. Ekstraksi kata kunci

Dari tabel yang berisi berita, dilakukan ekstraksi kata kunci menggunakan metode RAKE (*Rapid Automatic Keyword Extraction*). Prinsip metode ini adalah melakukan kuantifikasi terhadap tingkat kedekatan suatu kata dengan kata yang lain [21]. Adanya prinsip ini menjadikan hasil ekstraksi dapat berupa kelompok kata yang memang saling berkaitan sehingga kata yang dihasilkan menjadi lebih relevan. Hasil kuantifikasi juga dapat digunakan untuk menilai tingkat kedekatan antar kata yang dapat digunakan sebagai bahan evaluasi terhadap hasil ekstraksi. Algoritma dari metode ini juga diawali dengan pengelompokan kata berdasarkan jenisnya yang akan menghindarkan hasil ekstraksi dari kata hubung yang sering muncul seperti “yang”, “di”, “pada”, dan lain sebagainya karena pengguna dapat memilih batasan dari jenis kata yang akan diekstraksi. Sebagai contoh, pada penelitian ini digunakan formula kata benda (*nouns*) dan kata sifat (*adjective*). Ekstraksi dilakukan menggunakan perangkat lunak Rstudio.

Kata kunci yang telah diekstraksi kemudian dimasukkan ke dalam Google Trends untuk memperoleh indeks pencarian kata. Indeks pencarian kata nantinya akan dijadikan variabel prediktor sedangkan data harga pangan untuk setiap komoditas akan dijadikan sebagai variabel target.

2.3 Pemrosesan Data Awal

Terdapat dua tahapan pemrosesan data awal yaitu agregasi data dan pembuatan *dataset* pada setiap komoditas. Data harga komoditas pertanian yang diperoleh dari situs Pusat Informasi Harga Pangan Strategis (PIHPS) memiliki frekuensi harian. Dengan adanya keterbatasan dari Google Trends, maka data indeks pencarian kata kunci hanya tersedia dalam frekuensi mingguan. Oleh karena itu, diperlukan agregasi pada data harga komoditas pertanian sehingga memiliki frekuensi yang sama dengan data indeks pencarian kata kunci yaitu mingguan. Agregasi dilakukan dengan cara menghitung nilai rata-rata data dalam jangka waktu satu minggu. Data yang telah dilakukan pemrosesan awal kemudian dimasukkan ke dalam *dataset* komoditas yang berisi data harga pangan dan indeks pencarian kata kunci.

2.4 Penentuan Nilai Lag Optimal

Terdapat kemungkinan *lag* antara data indeks pencarian dan data harga. *Lag* yang dimaksud dalam penelitian ini merupakan selisih antara periode indeks pencarian kata dengan periode harga komoditas pertanian. Metode yang dilakukan pada penentuan nilai *lag* optimal ini juga menggunakan prinsip *machine learning* yaitu memisahkan data menjadi data *training* dan *testing* dengan proporsi 80%-20%. Prinsip *machine learning* dipilih karena mengakomodasi jenis model yang beragam dan waktu komputasi yang relatif singkat [18]. Pemodelan kemudian dilakukan pada setiap jenis model untuk nilai *lag* antara nol hingga lima minggu. Pemilihan nilai ini didasarkan pada penelitian [22] yang menyebutkan bahwa rentang waktu prediksi harga adalah di bawah satu bulan. Minggu ke-5 kemudian tetap dicantumkan karena struktur bulan pada kalender masehi yang lebih dari empat minggu namun kurang dari lima minggu. Untuk setiap *lag*, dilakukan seleksi variabel menggunakan metode recursive feature elimination (RFE). Metode RFE dipilih karena efisien serta komprehensif dengan adanya tahapan pemodelan dengan semua variabel lalu menghilangkan variabel yang tidak diperlukan [23]. Proses ini dilakukan untuk melakukan mitigasi terhadap adanya variabel prediktor yang korelasinya dengan variabel target dipengaruhi oleh nilai *lag*. Dari setiap *lag*, akan diperoleh performa model yang direpresentasikan dari nilai RMSE *training* dan *testing* serta *Rsquared training*. Nilai *lag* optimal dipilih berdasarkan nilai RMSE *training* yang terkecil dengan adanya batasan bahwa nilai RMSE *testing* harus lebih besar daripada nilai RMSE *training*. Hal ini karena pada konsep machine learning, model dilatih menggunakan data *training* yang didalamnya terdapat nilai variabel target sehingga apabila nilai RMSE *training* lebih besar, maka terdapat kemungkinan model bias karena jumlah data *testing* yang terlalu kecil [24].

2.5 Pemodelan dengan Machine Learning

Sebelum memulai pemodelan prediksi harga komoditas pertanian, dilakukan terlebih dahulu pemisahan *dataset* ke dalam *dataset training* dan *testing*. Pemisahan dilakukan menggunakan aturan 80%-20% sesuai dengan dasar prinsip Pareto

serta menghindari efek *over-training* karena data *training* yang terlalu banyak. Pemisahan dilakukan secara random dengan membentuk variabel indeks *training* sebagai pemisah antara data *training* dan *testing*. Pemisahan ini dilakukan untuk setiap jenis model pada setiap komoditas serta dilakukan pada *dataset* yang telah disesuaikan dengan nilai *lag* optimalnya.

Pemodelan prediksi harga komoditas pertanian dilakukan untuk setiap jenis model pada setiap komoditas. Pemodelan ini dilakukan menggunakan *dataset* yang telah disesuaikan dengan *lag* serta menggunakan pemisahan data *training* dan *testing* dengan rasio 80%-20%. Proses pemodelan dilakukan menggunakan perangkat lunak Rstudio serta bahasa pemrograman R. Dari tahapan ini, akan diperoleh empat model untuk setiap komoditas, yaitu model dari metode regresi linier, *random forest*, *gradient boosting*, dan *regression tree*. Setiap model akan dilengkapi dengan nilai performanya yang diukur dari nilai RMSE *training* dan *testing* dan *Rsquared training* dan *testing*.

2.6 Evaluasi Model

Nilai MAPE kemudian dihitung untuk setiap model yang diperoleh dari setiap jenis model menggunakan persamaan (1).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (1)$$

Perhitungan nilai MAPE hanya dilakukan pada saat pemodelan prediksi harga komoditas pertanian karena beberapa alasan berikut:

- Nilai MAPE memiliki standar interpretasi yang tidak dimiliki oleh RMSE seperti yang terdapat pada **Tabel 1**. Interpretasi ini berguna untuk menilai kualitas dari performa model dan diperlukan untuk hasil model akhir. Pada penentuan nilai *lag* optimal, hanya diperlukan nilai tingkat kesalahan relatif antar *lag* yang telah diakomodasi oleh RMSE [25].

Tabel 1. Interpretasi nilai MAPE

Nilai MAPE	Interpretasi
Kurang dari 10%	Akurasi peramalan tinggi (<i>highly accurate forecasting</i>)
10-20%	Akurasi peramalan baik (<i>good forecasting</i>)
20-50%	Akurasi peramalan dapat diterima (<i>reasonable forecasting</i>)
Lebih dari 50%	Akurasi peramalan rendah (<i>inaccurate forecasting</i>)

- Nilai MAPE selalu berbanding lurus dengan nilai RMSE. Selain itu, pada perangkat lunak Rstudio, parameter pemodelan yang digunakan adalah RMSE sehingga untuk mempercepat komputasi, pada penentuan nilai *lag* optimal hanya digunakan RMSE.

Dari tahapan ini akan diperoleh satu model terbaik untuk setiap komoditas.

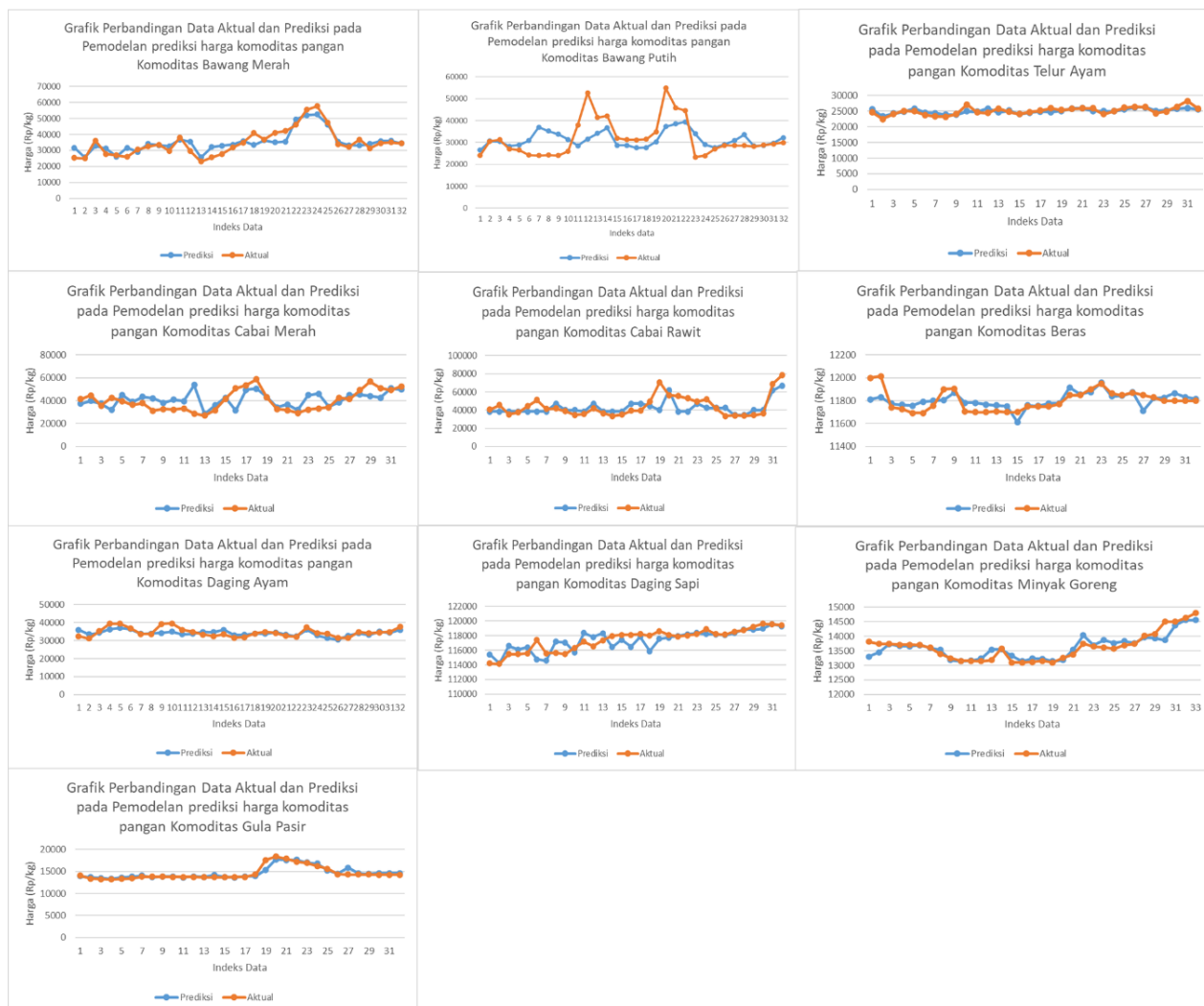
3. HASIL DAN PEMBAHASAN

3.1 Hasil Pemodelan

Melalui metodologi yang dilakukan, diperoleh model terbaik untuk setiap komoditas. Model terbaik diperoleh dengan membandingkan hasil pemodelan dari empat alternatif jenis model pada setiap komoditas. Keempat jenis model tersebut regresi linier, *random forest*, *gradient boosting*, dan *regression tree*. Model terbaik dipilih dari nilai MAPE terkecil dari keempat jenis model. Performa model terbaik pada setiap komoditas dapat dilihat pada Tabel 2 sedangkan perbandingan nilai aktual dan hasil prediksi dapat dilihat pada Gambar 2.

Tabel 2. Hasil pemodelan setiap komoditas

Komoditas	Nilai MAPE <i>test</i> untuk setiap jenis model (%)				Jenis model terbaik	Nilai MAPE terkecil (%)
	Regresi Linier	<i>Random Forest</i>	<i>Gradient Boosting</i>	<i>Regression Tree</i>		
Bawang merah	10,52	9,09	10,95	11,48	<i>Random Forest</i>	9,09
Bawang putih	18,86	16,59	17,47	18,48	<i>Random Forest</i>	16,59
Cabai merah	18,03	17,73	16,63	20,04	<i>Gradient Boosting</i>	16,63
Cabai rawit	18,14	14,05	16,04	13,25	<i>Regression Tree</i>	13,25
Telur ayam	3,65	2,46	3,1	3,34	<i>Random Forest</i>	2,46
Beras	0,44	0,44	0,47	0,47	<i>Regresi Linier</i>	0,44
Daging ayam	5,90	4,68	4,33	4,60	<i>Gradient Boosting</i>	4,33
Daging sapi	1,11	0,87	0,67	0,77	<i>Gradient Boosting</i>	0,67



Gambar 2. Perbandingan nilai aktual dan prediksi seluruh komoditas

Dari hasil pemodelan prediksi harga komoditas pangan, dapat diperoleh informasi bahwa model *random forest* mendominasi model dengan nilai MAPE terkecil dengan terdapat lima dari sepuluh komoditas yang menggunakan model *random forest* sebagai model dengan nilai MAPE terkecil. Selain itu, rentang nilai MAPE relatif jauh dari mulai yang terkecil sebesar 0,44% untuk komoditas beras hingga 16,63% untuk komoditas cabai merah.

3.2 Pembahasan

Dari hasil pemodelan, dilakukan pembahasan pada beberapa aspek yaitu perbandingan dengan hasil penelitian sebelumnya, analisis komoditas dengan nilai MAPE di atas 10%, analisis disparitas nilai MAPE antar komoditas, analisis model terpilih, dan analisis pengaruh inflasi terhadap hasil pemodelan.

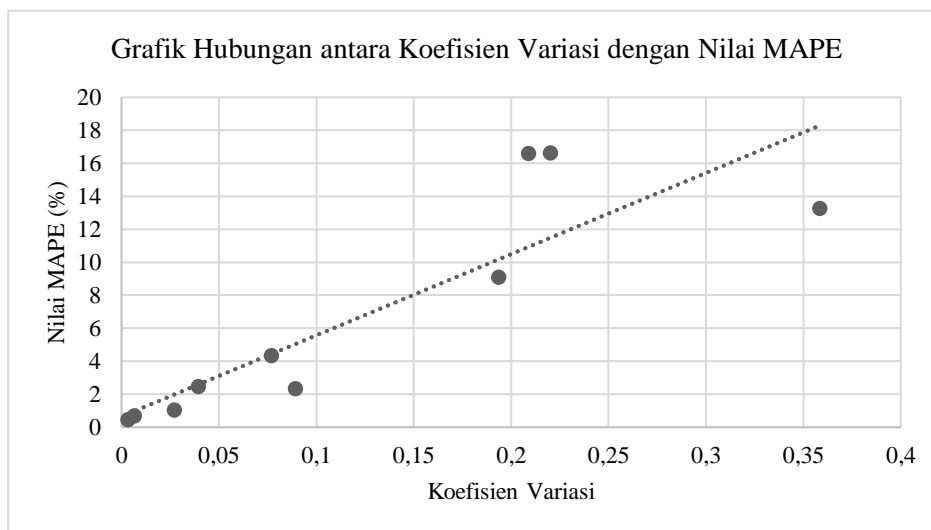
a. Perbandingan dengan penelitian sebelumnya

Penelitian yang akan dibahas sebagai pembanding adalah penelitian dari Kim pada tahun 2017 bekerja sama dengan Badan Perencanaan Pembangunan Nasional (Bappenas) mengenai *nowcasting* pada komoditas daging sapi, daging ayam, bawang, dan cabai di Indonesia menggunakan variabel prediktor berupa data harga yang diperoleh dari media sosial *Twitter* [13]. Penelitian ini menggunakan model yang memberikan bobot terhadap harga dari *Twitter* di masa sekarang dan satu periode ke belakang. Dari hasil pemodelan, diperoleh nilai MAPE untuk daging sapi sebesar 4,86%, daging ayam sebesar 9,41%, bawang sebesar 32,79%, dan cabai sebesar 13,62%.

Dari keempat komoditas tersebut, diperoleh kesimpulan bahwa penelitian ini menghasilkan performa yang lebih baik daripada penelitian [13] untuk tiga dari empat komoditas dengan nilai MAPE untuk daging sapi sebesar 0,67%, daging ayam sebesar 4,33%, bawang merah sebesar 9,09%, dan bawang putih sebesar 16,59%. Sementara itu, hasil prediksi terhadap komoditas cabai merah pada penelitian ini menghasilkan nilai MAPE yang lebih besar dibandingkan penelitian pembanding dengan nilai 16,63% sedangkan komoditas cabai rawit memperoleh nilai MAPE yang lebih kecil yaitu sebesar 13,25%. Performa model dari penelitian [13] lebih buruk karena hanya terdapat satu model untuk digunakan pada proses prediksi seluruh komoditas. Hal ini berbeda dengan pemodelan pada penelitian ini yang berusaha memperoleh model yang berbeda dan sesuai dengan komoditas yang dimodelkan.

b. Analisis disparitas nilai MAPE antar komoditas

Dari hasil pemodelan prediksi harga komoditas pangan, diperoleh bahwa terdapat komoditas dengan nilai MAPE sangat kecil seperti beras dan daging sapi yang memiliki nilai MAPE di bawah 1% sedangkan terdapat komoditas dengan nilai MAPE yang jauh lebih besar yaitu bawang putih dan cabai merah dengan nilai lebih dari 16%. Kondisi ini disebabkan oleh kondisi fluktuasi harga dari masing-masing komoditas. Kecenderungan ini dapat dilihat pada Gambar 3 yang menunjukkan bahwa semakin besar nilai koefisien variasi atau tingkat fluktuasi harga, semakin tinggi juga nilai MAPE hasil prediksi yang dihasilkan.



Gambar 3. Grafik hubungan antara koefisien variasi (tingkat fluktuasi harga) dengan nilai MAPE

Secara teknis, semakin tinggi fluktuasi harga, semakin tinggi pula variansi dari data. Untuk mengakomodasi setiap nilai dari data yang bersifat fluktuatif, diperlukan model yang kompleks sehingga setiap nilai dari variabel target pada data *training* dapat diikuti polanya. Contoh model yang bersifat kompleks ini antara lain yaitu regresi polinomial yang mengakomodasi hubungan yang tidak linear antara variabel prediktor dan variabel target. Risiko penggunaan model seperti ini pada data yang bersifat fluktuatif adalah model terlalu mengikuti setiap data termasuk pada data yang berbeda jauh dari nilai rata-ratanya. Hal ini berpotensi menyebabkan *overfitting* atau nilai *error* pada data *testing* jauh lebih besar dibandingkan dengan pada data *training*. Dengan banyaknya variabel yang tersedia, maka penggunaan model yang kompleks akan semakin meningkatkan risiko terjadinya *overfitting*. Pada penelitian ini, tidak digunakan model yang dikategorikan sebagai model yang kompleks sehingga tidak terjadi *overfitting* pada hasil pemodelan

c. Analisis karakteristik komoditas ditinjau dari nilai MAPE

Kesepuluh komoditas dapat dikelompokkan menjadi dua kelompok berdasarkan nilai MAPE hasil pemodelan prediksi yang diperoleh. Kelompok pertama merupakan komoditas yang memiliki nilai MAPE di bawah 10% yaitu komoditas bawang merah, telur ayam, beras, daging ayam, daging sapi, minyak goreng, dan gula pasir sedangkan kelompok kedua merupakan komoditas yang memiliki nilai MAPE hasil prediksi lebih dari 10% yaitu komoditas cabai merah, cabai rawit, dan bawang putih. Karakteristik dari setiap komoditas akan terlebih dahulu dibahas dari segi pasokan baik dari sisi produksi maupun distribusi. Karakteristik dari sisi permintaan tidak dilakukan karena keterbatasan informasi serta dapat diasumsikan antar komoditas memiliki kesamaan karakteristik permintaan karena kesepuluh komoditas merupakan kebutuhan pokok sehari-hari dan bukan kebutuhan yang bersifat musiman.

Terdapat beberapa karakteristik dari komoditas yang masuk ke dalam kategori dua atau yang memiliki nilai MAPE lebih dari 10%. Cabai merupakan komoditas yang mudah terserang hama serta memerlukan lingkungan yang kering agar dapat bertahan. Hal ini menjadikan produksi cabai tidak stabil karena sangat bergantung pada kondisi lingkungan yang tidak dapat dikendalikan. Di sisi lain, komoditas bawang putih masih sangat bergantung pada impor dari China dengan persentase mencapai 95%. Hal ini disebabkan oleh mahalnya biaya produksi bawang putih. Pada penelitian ini, kontribusi dari fluktuasi harga serta tingginya nilai MAPE hasil prediksi juga terjadi pada periode yang mengalami kendala impor.

Sebaliknya, komoditas yang masuk ke dalam kategori pertama terdiri atas komoditas dengan produksi nasional yang surplus dan tingkat produksi yang cenderung stabil seperti bawang merah, telur ayam, beras, minyak goreng, daging sapi, gula pasir, dan daging ayam. Adanya surplus produksi dengan tingkat produksi yang stabil akan membuat harga di pasaran juga menjadi relatif stabil. Hal ini juga berkontribusi dalam membuat nilai MAPE hasil prediksi menjadi kecil.

d. Analisis model terpilih

Pada penelitian ini, digunakan empat jenis model yaitu regresi linear, *random forest*, *gradient boosting*, dan *regression tree*. *Random forest* dan *gradient boosting* merupakan jenis model pengembangan dari *regression tree*. Cara kerja dari keduanya adalah membangkitkan banyak pohon yang lebih sederhana dari satu pohon *regression tree* [26]. Perbedaannya adalah pada *random forest* membangkitkan pohon dilakukan secara paralel untuk kemudian diambil nilai rata-rata dari hasil prediksi antar pohon sehingga antar pohon bersifat independent. Hal ini berbeda dengan *gradient boosting* yang membangkitkan pohon secara serial sehingga satu pohon dengan yang lainnya dependen.

Dari keempat model, model regresi linear cenderung menghasilkan nilai MAPE terbesar. Hal ini ditunjukkan dengan enam dari sepuluh komoditas memiliki nilai MAPE terbesar pada model regresi linear sementara empat komoditas tersisa memiliki nilai MAPE terbesar pada model *regression tree*. Hal ini terjadi karena kedua jenis model ini merupakan jenis model yang paling sederhana dan lebih diperuntukkan untuk pemodelan yang membutuhkan waktu singkat. Hal ini berbeda dengan model *random forest* dan *gradient boosting* yang lebih diperuntukkan untuk meningkatkan akurasi model sehingga nilai MAPE yang dihasilkan juga cenderung lebih kecil dibandingkan dengan nilai MAPE dari model regresi linear dan *regression tree*. Nilai MAPE yang dihasilkan dari model *random forest* dan *gradient boosting* juga cenderung tidak berbeda jauh karena keduanya merupakan metode *ensemble* yang menggabungkan banyak model untuk memperoleh hasil prediksi yang lebih akurat.

Meskipun regresi linear merupakan salah satu model yang paling sederhana, terdapat satu komoditas yang memiliki nilai MAPE terkecil dari model regresi linear yaitu komoditas beras. Komoditas beras sendiri merupakan komoditas yang menghasilkan nilai MAPE terkecil dibandingkan dengan komoditas yang lainnya. Meskipun model terpilih adalah regresi linear, dapat dilihat bahwa performa antar model pada komoditas beras tidak berbeda secara signifikan atau hampir sama. Hal ini menunjukkan bahwa komoditas beras cukup *robust* pada seluruh jenis model yang digunakan. Performa dari model regresi linear yang lebih baik dapat disebabkan oleh tren harga beras yang sangat stabil dan tidak terdapat lonjakan harga sehingga model yang berbasis pohon seperti *regression tree*, *random forest*, dan *gradient boosting* cenderung tidak memperoleh pembagian keputusan yang secara signifikan lebih baik daripada metode regresi linear.

Selain itu, terdapat satu komoditas dengan nilai MAPE terkecil diperoleh dari model *regression tree* yaitu komoditas cabai rawit. Komoditas cabai rawit merupakan komoditas yang memiliki nilai koefisien variasi tertinggi dibandingkan dengan komoditas lain yaitu sebesar 0,27. Salah satu kelebihan dari model *regression tree* pada komoditas dengan fluktuasi tinggi seperti ini yaitu model *regression tree* membangkitkan satu pohon yang lebih kompleks dibandingkan dengan pohon dari model *random forest* dan *gradient boosting* yang lebih sederhana namun berjumlah banyak. Dengan jumlah variabel prediktor dari komoditas cabai rawit hanya berjumlah delapan, maka pohon sederhana yang dibangkitkan oleh metode *random forest* dan *gradient boosting* dapat menghasilkan performa yang lebih buruk karena mengalami *underfit* atau kemampuan prediksi yang kurang. Kombinasi dari banyak pohon yang tidak memiliki performa yang baik berpotensi membuat hasil prediksi secara keseluruhan juga tidak menghasilkan performa yang lebih baik. Dengan adanya satu pohon yang kompleks dari model *regression tree*, proses prediksi dapat dilakukan secara lebih rinci dan dapat menghasilkan performa yang lebih baik.

Meskipun fluktuasi dari bawang putih dan cabai merah juga cenderung tinggi seperti cabai rawit, namun model dengan MAPE terkecil dari kedua komoditas ini bukan *regression tree*. Model *random forest* dan *gradient boosting* cenderung bernilai lebih baik pada dua komoditas ini dibandingkan dengan *regression tree*. Hal ini disebabkan oleh jumlah variabel prediktor pada kedua komoditas ini lebih besar yaitu 13 variabel untuk komoditas cabai merah dan 26 variabel untuk komoditas bawang putih. Dengan jumlah variabel yang lebih besar, lebih memungkinkan bagi jenis model *random forest* dan *gradient boosting* menghasilkan pohon individu yang lebih baik dan ketika dikombinasikan juga memperoleh hasil yang lebih baik.

4. KESIMPULAN

Berdasarkan penelitian, pengolahan data dimulai dari penentuan komoditas pangan. Hasil penentuan komoditas pangan kemudian menjadi *input* untuk proses pengumpulan data harga dan penentuan kata kunci terkait komoditas yang ditentukan tersebut. Data yang telah dikumpulkan kemudian dijadikan satu *dataset* yang akan menjadi *input* untuk tahap penentuan *lag* optimal. Nilai *lag* optimal digunakan sebagai *input* pada proses pemodelan harga komoditas pangan untuk setiap jenis model. Setiap model yang digunakan pada tahapan pemodelan harga ini akan dievaluasi menggunakan nilai MAPE. Model yang dapat digunakan untuk melakukan prediksi harga komoditas dikategorikan berdasarkan masing-masing komoditas. Hasil prediksi yang diperoleh dapat dikategorikan menjadi dua yaitu akurat (nilai MAPE <10%) dan baik (nilai MAPE antara 10% dan 20%). Secara umum, penggunaan model *random forest* dan *gradient boosting* menghasilkan nilai MAPE yang lebih kecil dibandingkan dengan model regresi linier dan *regression tree*. Model *regression tree* memiliki nilai MAPE kecil pada komoditas cabai rawit karena komoditas cabai rawit memiliki fluktuasi harga tinggi dan jumlah variabel prediktor yang kecil sehingga kurang dapat diakomodasi oleh model *random forest* dan *gradient boosting*. Hal ini berbeda dengan komoditas bawang putih dan cabai merah yang juga memiliki fluktuasi tinggi namun lebih dapat diakomodasi oleh model *random forest* dan *gradient boosting*. Model regresi linier menghasilkan nilai MAPE terkecil pada komoditas beras. Pada penelitian ini, diperoleh informasi bahwa terdapat tujuh komoditas yang memiliki hasil prediksi akurat dengan nilai MAPE di bawah 10% dan tiga komoditas memiliki hasil prediksi *good forecast* dengan nilai MAPE di antara 10% dan 20%.

REFERENCES

- [1] H. Huka, H. Huka, C. Ruoja, dan A. Mchopa, "Price Fluctuation of Agricultural Products and its Impact on Small Scale Farmers Development: Case Analysis from Kilimanjaro Tanzania," *European Journal of Business and Management*, vol. 6, no. 36, hlm. 155–160, Nov 2014.

- [2] Y. Zhang dan S. Na, "A Novel Agricultural Commodity Price Forecasting Model Based on Fuzzy Information Granulation and MEA-SVM Model," 2018, doi: 10.1155/2018/2540681.
- [3] Z. Wang, O. Kwon, dan F. Liu, "Applying Keyword Analysis to Predicting Agriculture Product Price Index: The Case of the Chinese Farming Market," *Asia Pacific Journal of Business Review*, vol. 1, no. 1, hlm. 1–22, Agu 2016, doi: 10.20522/apjbr.2016.1.1.1.
- [4] J. Kim, M. Cha, dan J. G. Lee, "Nowcasting commodity prices using social media," *PeerJ Comput Sci*, vol. 2017, no. 7, 2017, doi: 10.7717/peerj-cs.126.
- [5] K. Sukiyono dan M. Janah, "Forecasting Model Selection of Curly Red Chili Price at Retail Level," *Indonesian Journal of Agricultural Research*, vol. 2, no. 1, hlm. 1–12, Mar 2019, doi: 10.32734/injar.v2i1.859.
- [6] S. Nahmias dan T. Olsen, "Production and operations analysis : strategy, quality, analytics, application," hlm. 820, 2015.
- [7] G. Q. Li, S. W. Xu, dan Z. M. Li, "Short-term price forecasting for agro-products using artificial neural networks," dalam *Agriculture and Agricultural Science Procedia*, 2010, vol. 1, hlm. 278–287. doi: 10.1016/j.aaspro.2010.09.035.
- [8] C. Elleby, I. P. Domínguez, M. Adenauer, dan G. Genovese, "Impacts of the COVID-19 Pandemic on the Global Agricultural Markets," *Environ Resour Econ (Dordr)*, vol. 76, no. 4, hlm. 1067–1079, Agu 2020, doi: 10.1007/s10640-020-00473-6.
- [9] Benkachcha. S, Benhra. J, dan E. Hassani. H, "Causal Method and Time Series Forecasting model based on Artificial Neural Network," *Int J Comput Appl*, vol. 75, no. 7, hlm. 37–42, Agu 2013, doi: 10.5120/13126-0482.
- [10] V. Gosasang, W. Chandraprakaikul, dan S. Kiattisin, "A comparison of traditional and neural networks forecasting techniques for container throughput at bangkok port," *Asian Journal of Shipping and Logistics*, vol. 27, no. 3, hlm. 463–482, 2011, doi: 10.1016/S2092-5212(11)80022-2.
- [11] B. Wang dkk., "Research on Hybrid Model of Garlic Short-term Price Forecasting based on Big Data," *Computers, Materials & Continua*, vol. 57, no. 2, hlm. 283–296, Jan 1970, doi: 10.32604/CMC.2018.03791.
- [12] M. Bañbura, D. Giannone, M. Modugno, dan L. Reichlin, "Now-Casting and the Real-Time Data Flow," *Handbook of Economic Forecasting*, vol. 2, hlm. 195–237, Jan 2013, doi: 10.1016/B978-0-444-53683-9.00004-9.
- [13] J. Kim, M. Cha, dan J. G. Lee, "Nowcasting commodity prices using social media," *PeerJ Comput Sci*, vol. 2017, no. 7, 2017, doi: 10.7717/peerj-cs.126.
- [14] J. M. Barros dkk., "The validity of google trends search volumes for behavioral forecasting of national suicide rates in Ireland," *Int J Environ Res Public Health*, vol. 16, no. 17, Sep 2019, doi: 10.3390/ijerph16173201.
- [15] Y. Carrière-Swallow dan F. Labbé, "Nowcasting with Google Trends in an Emerging Market," *J Forecast*, vol. 32, no. 4, hlm. 289–298, Jul 2013, doi: 10.1002/FOR.1252.
- [16] A. Ross, "Nowcasting with Google Trends: a keyword selection method," *Fraser of Allander Economic Commentary*, 2013.
- [17] H. Jiang, "Machine learning fundamentals : A concise introduction," Cambridge University Press, hlm. 423, 2021.
- [18] N. D. Muchisha, N. Tamara, A. Andriansyah, dan A. M. Soleh, "Nowcasting Indonesia's GDP Growth Using Machine Learning Algorithms," *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 2, hlm. 355–368, Jun 2021, doi: 10.29244/ijsa.v5i2p355-368.
- [19] E. Vargiu dan M. Urru, "Exploiting web scraping in a collaborative filtering- based approach to web advertising," *Artif Intell Res*, vol. 2, no. 1, hlm. 44, Des 2012, doi: 10.5430/air.v2n1p44.
- [20] V. Krotov, L. Johnson, dan L. Silva, "Tutorial: Legality and Ethics of Web Scraping," *Communications of the Association for Information Systems*, vol. 47, no. 1, hlm. 22, Des 2020, doi: 10.17705/1CAIS.04724.
- [21] S. Rose, D. Engel, N. Cramer, dan W. Cowley, "Automatic Keyword Extraction from Individual Documents," *Text Mining: Applications and Theory*, hlm. 1–20, Mar 2010, doi: 10.1002/9780470689646.CH1.
- [22] G. Q. Li, S. W. Xu, dan Z. M. Li, "Short-term price forecasting for agro-products using artificial neural networks," dalam *Agriculture and Agricultural Science Procedia*, 2010, vol. 1, hlm. 278–287. doi: 10.1016/j.aaspro.2010.09.035.
- [23] B. F. Darst, K. C. Malecki, dan C. D. Engelman, "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data," *BMC Genet*, vol. 19, Sep 2018, doi: 10.1186/s12863-018-0633-8.
- [24] B. Neal dkk., "A Modern Take on the Bias-Variance Tradeoff in Neural Networks," Okt 2018, doi: 10.48550/arxiv.1810.08591.
- [25] C. D. (Colin D. Lewis, "Industrial and business forecasting methods : a practical guide to exponential smoothing and curve fitting," hlm. 143, 1982.
- [26] G. James, D. Witten, T. Hastie, dan R. Tibshirani, "An Introduction to Statistical Learning with Applications in R Second Edition," 2021.