



Peramalan Tren Produk Pakaian Menggunakan Metode SARIMA pada Data Google Trends

Muhammad Hilmi Syarif¹, Nana Mulyana Maghfur², Budi Arif Dermawan³

^{1,2,3}Universitas Singaperbangsa Karawang

Abstract

Received: 13 Februari 2023
Revised: 23 Februari 2023
Accepted: 2 Maret 2023

The development of the digital era is very rapid at this time there are so many technologies that are present in human life, one of which is e-commerce. Today's society is very dependent on what is called e-commerce, there are lots of transactions made every year from 2017 there are around 8,657 USD until 2021 which continues to increase to 53,808 USD. This is more or less influenced by advertisements scattered in cyberspace. In 2020, the 1st rank of the highest number of transactions is held by clothing products, so sellers from e-commerce need to be assisted with a recommendation system that displays the best time to display their advertisements. The process uses the time-series forecasting method using the SARIMA algorithm. This algorithm takes the trend of the season from the previous data, so it is very suitable for the cases mentioned earlier. The data retrieval processes up to deployment to the web system uses the CRISP-DM research method. The results of this study have a very good evaluation value, namely with an R2 of 0.36993 so that the model is good to use and in its implementation is a web system that takes the category and how long it will be predicted.

Keywords: Prediction, Fashion, Trend, Google Trends, SARIMA

(* Corresponding Author: hilmi.syarif18228@student.unsika.ac.id

How to Cite: Syarif, M., Maghfur, N., & Dermawan, B. (2023). Forecasting Fashion Product Trends Using the SARIMA Method on Google Trends Data. *Jurnal Ilmiah Wahana Pendidikan*, 9(8), 350-359. <https://doi.org/10.5281/zenodo.7886616>

PENDAHULUAN

Perkembangan era digital sekarang ini terdapat banyak sekali teknologi yang sangat membantu dan bermanfaat bagi manusia. Mulai dari bidang pendidikan, kesehatan, transportasi, ekonomi, dan lain sebagainya (H. Budiman, 2017). Salah satunya adalah di bidang ekonomi, seiring berkembangnya teknologi proses transaksi jual beli saat ini menjadi berbentuk digital. Terdapat aplikasi, sistem, dan platform di bidang ekonomi yang banyak berkembang, salah satunya merupakan e-commerce. E-commerce menjadi salah satu platform yang banyak digunakan saat ini, menurut data statistik yang dibuat oleh Badan Pusat Statistik pada tahun 2020 terdapat sekitar 93% masyarakat Indonesia menggunakan e-commerce sebagai tempat jual beli mereka (A. L. Kusumatriana et al, 2020).

Beberapa tahun belakangan ini e-commerce menjadi primadona bagi penjual dan pelanggan dalam bertransaksi. Ada banyak sekali e-commerce yang ada sekarang ini, seperti Tokopedia, Shopee, Bukalapak, Lazada, dan lain sebagainya. Dari tahun ke tahunnya pendapatan yang dihasilkan oleh e-commerce begitu meningkat, mulai dari tahun 2017 terdapat sekitar 8,657 juta USD sampai tahun 2021 meningkat sampai 53,808 juta USD (Statista, 2015). Dan menurut data statistik yang dilakukan oleh Katadata Insight Center (KIC) produk pakaian



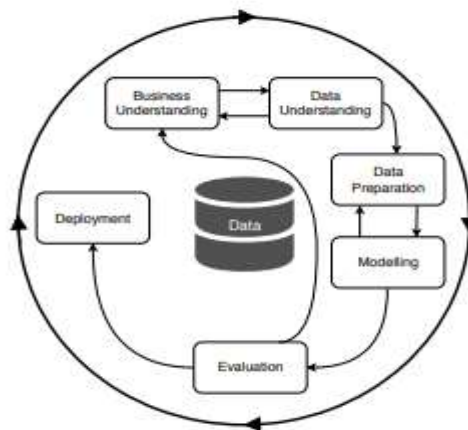
menjadi peringkat 1 dalam hal proporsi jumlah transaksi di *e-commerce* (A. Lidwina, 2021). Hal ini dipengaruhi oleh beberapa faktor diantaranya adalah dengan banyaknya iklan yang beredar di dunia maya.

Pada penelitian sebelumnya membahas tentang prediksi atribut produk yang penting dari *customer review* dan *google trends* (H. Yakubu and C. K. Kwong, 2021). Penelitian tersebut menggunakan algoritma *fuzzy time series* dalam pembuatan modelnya, dan berfokus pada tanggapan para pelanggan dan *google trends* dapat mempengaruhi relevansi rekomendasi produk kepada *customer*. Lalu ada beberapa penelitian menggunakan data dari *google trends* menganalisis ketertarikan masyarakat terhadap anak anjing saat pandemic covid-19 (C. Sietou, 2021).

Beberapa penelitian pada artikel-artikel sebelumnya, belum ada yang berfokus mengenai data tren pada produk pakaian. Sehingga penelitian ini merujuk pada penelitian yang sudah ada dengan menggunakan data tren pada produk pakaian dari *google trends* dan diuji cobakan untuk memberikan prediksi sehingga para penjual dapat mengetahui waktu mana saja yang di rekomendasikan untuk memasang iklan mereka pada internet menggunakan algoritma SARIMA.

METODE PENELITIAN

Penelitian ini menggunakan metode penelitian CRISP-DM (*Cross Industry Standard Process for Data Mining*), yaitu salah satu metode penelitian data mining yang memiliki 6 tahapan yaitu, *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (F. Martinez-Plumed et al, 2021).



Gambar 1. Proses metode CRISP-DM

Tahapan-tahapan metode CRISP-DM dapat dijabarkan sebagai berikut:

Business Understanding

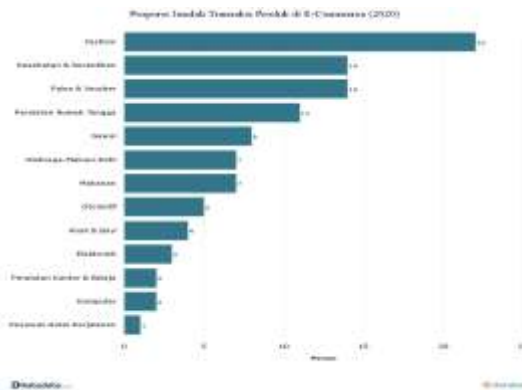
Tahapan ini dilakukan untuk memahami permasalahan apa yang akan diangkat untuk penelitian. Pada penelitian ini permasalahan yang akan diangkat

yaitu sulitnya para penjual/*seller* dalam menentukan waktu mana yang tepat dalam memasang iklan untuk produk mereka. Dengan permasalahan tersebut tujuan dari penelitian ini yaitu membuat sistem rekomendasi untuk penjual dalam mengiklankan produknya dengan tepat di waktu tertentu. Dan batasan masalahnya adalah sebagai berikut:

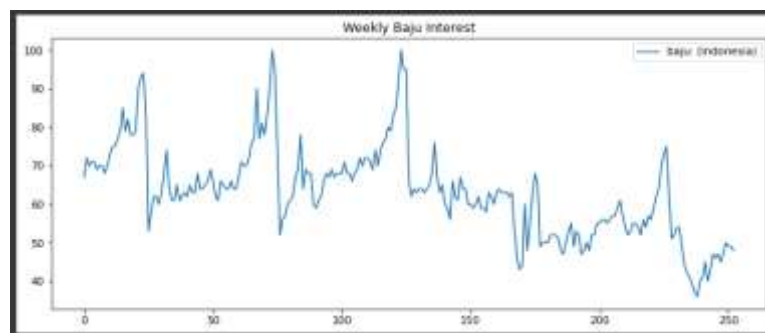
- Data yang digunakan dalam penelitian ini adalah data *time-series* dengan rentang waktu dari tahun 2017 sampai 2021.
- Kategori yang digunakan pada penelitian ini adalah kategori produk *fashion*.
- Hasil dari penelitian ini berupa sistem berbasis *web* yang dapat memberikan rekomendasi waktu terbaik untuk penjual.

Data Understanding

Untuk menyelesaikan masalah yang sudah disebutkan sebelumnya, kami mengambil data yang berasal dari *Google Trends* yang berupa data *time-series*. Kategori untuk pengambilan data ini adalah kategori *fashion*, karena penjualan dengan kategori *fashion* dari tahun ke tahun terus meningkat, dan pada dua tahun belakangan ini sesuai dengan yang ditunjukkan pada gambar 2 produk *fashion* menjadi produk dengan transaksi tertinggi di *e-commerce*. Visualisasi bentuk data yang merupakan *time series* tingkat ketertarikan orang - orang terhadap produk *fashion* dapat dilihat pada gambar 3 untuk contoh dari kategori.



Gambar 2. Visualisasi Statistik Data Penjualan Di *E-Commerce* [4].

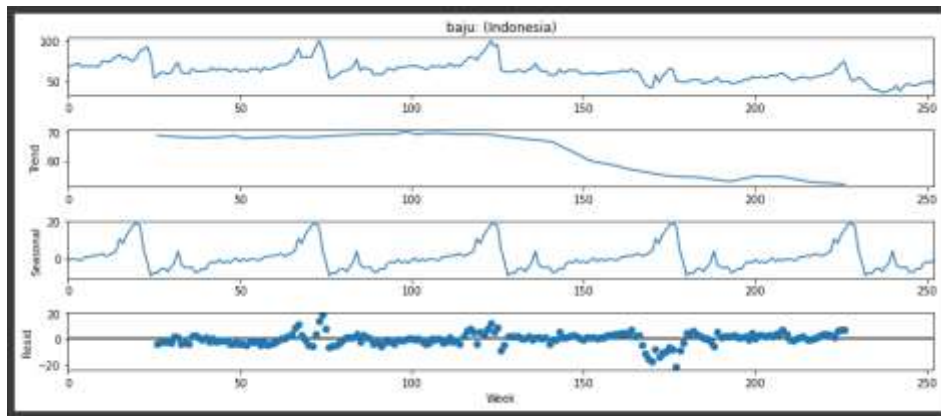


Gambar 3. Visualisasi Data *Time-Series* Produk Kategori Baju

Data Preparation

Seasonality dan Trend

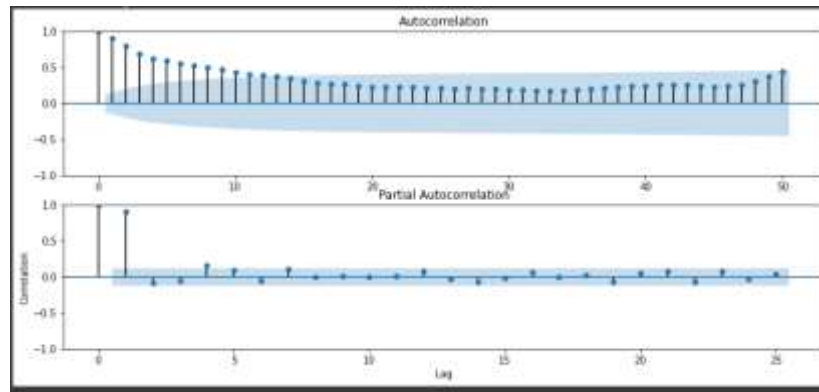
Kami telah melakukan analisis pada data yang telah diambil untuk menentukan penanganan apa saja yang perlu pada data serta menjadi landasan pemilihan algoritma yang tepat. Dari grafik *trend* sebelumnya sekilas dapat dilihat bahwa data selama kurang lebih 4,5 tahun menunjukkan adanya *seasonality* tiap tahunnya dimana ketertarikan orang terhadap produk fashion biasanya meningkat sekitar bulan Mei - Juni. Hal ini dapat lebih jelas terlihat setelah kami melakukan teknik *seasonal decomposition* sehingga pola *seasonal* dan *trend* pada data dapat lebih jelas terlihat seperti ditunjukkan gambar 4 untuk contoh dari kategori.



Gambar 4. Grafik *Seasonal* Dan *Trend* Ketertarikan Orang Dengan Produk Baju

2.3.1 Stationary

Selanjutnya kami perlu mengecek apakah data yang digunakan *stationary* atau tidak. Hal ini penting untuk dilakukan karena sebagian besar metode *forecasting* statistik berlandaskan pada asumsi bahwa data *time series* yang diolah adalah data yang *stationary* (tidak bergantung pada waktu dan *trend*) (G. Reikard, 2021; A. Aggarwal et al., 2020). Untuk mengeceknya dapat dilakukan visualisasi dari *autocorrelation function* (ACF) dan *partial autocorrelation function* (PACF), jika suatu *time series stationary* (K. Tanaka, 2010), maka grafik ACF/PACF akan menunjukkan batasan pada area interval kepercayaan 95%. Lalu dari hasil pengecekan yang dapat dilihat pada gambar 5 menunjukkan bahwa grafik (area biru) tidak melewati interval tingkat kepercayaan 95% (-0,5 s/d 0,5) yang mengindikasikan bahwa data ini merupakan data *time series* yang *stationary*.



Gambar 5. Grafik Interval Tingkat Kepercayaan Pada Produk Baju

Selain menggunakan grafik ACF/PACF, untuk lebih memastikan dilakukan pula tes *Augmented Dickey-Fuller* dimana H_0 dari tes ini yaitu *time series* memiliki unit akar dan berarti datanya tidak *stationary*. Kita dapat memahami hasil tes menggunakan *P-value*, jika *P-value* lebih rendah dari nilai ambang (5% atau 1%), kita bisa menolak H_0 dan menyimpulkan bahwa data *time series* tersebut *stationary*. Berdasarkan hal ini maka hasil tes yang dapat dilihat pada tabel 1 kembali menunjukkan bahwa data yang digunakan merupakan data yang *stationary*, namun berbeda dengan yang ditunjukkan pada tabel 2 dimana *P-value* yang ditunjukkan meski lebih kecil dari nilai ambang 1% namun lebih besar dari nilai ambang 5%, yang mengindikasikan bahwa data celana belum cukup *stationary* dan perlu dilakukan transformasi agar data ini menjadi lebih *stationary* dan tidak dependant kepada nilai data sebelumnya.

Tabel 1. Hasil tes *Augmented Dickey-Fuller* pada produk baju

<i>Deskripsi</i>	<i>Hasil</i>
<i>Test Statistic</i>	-2,628609
<i>P-value</i>	0,087227
<i>Lags Used</i>	4,000000
<i>Number of Observation Used</i>	248,0000
<i>Critical Value (1%)</i>	-3,456996
<i>Critical Value (5%)</i>	-2,873266
<i>Critical Value (10%)</i>	-2,573019

Data Transformation

Transformasi data hanya diperlukan untuk data yang tidak *stationary*, yaitu pada data celana, kami menggunakan metode log scaling untuk mentransformasi data menjadi *stationary*, setelah itu kembali melakukan *Dickey-Fuller test* yang hasilnya ditunjukkan pada tabel 2. Dari hasil tersebut sudah dapat dilihat bahwa data sudah *stationary* karena *P-value* sudah dibawah nilai ambang dan bisa dilakukan pemodelan dengannya.

Tabel 2. Hasil Tes Statistik *Dickey-Fuller* pada Produk Celana yang Telah di Transformasi

Deskripsi	Hasil
<i>Test Statistic</i>	-6,34535
<i>P-value</i>	2,674409
<i>Lags Used</i>	1,000000
<i>Number of Observation Used</i>	2,4,00000
<i>Critical Value (1%)</i>	-3,457894
<i>Critical Value (5%)</i>	-2,873659
<i>Critical Value (10%)</i>	-2,573229

Modeling

Pemodelan awal dilakukan dengan menggunakan algoritma SARIMA (*Seasonal Autoregressive Integrated Moving Average*). Model ini memprediksi tren yang muncul dalam bentuk seasonal (K. He et al, 2021; A. Kumar Dubey et al, 2021).

Evaluation

Tahapan ini merupakan hasil evaluasi pada model apakah model memiliki tingkat error yang tinggi atau tidak, sehingga dapat ditentukan model yang telah dibuat dapat dilanjutkan ke proses *deployment* atau tidak.

Deployment

Proses ini merupakan tahapan terakhir dari metode penelitian CRISP-DM yang dimana model yang telah dibuat dan dievaluasi sebelumnya dapat di *deploy* ke sistem.

HASIL DAN PEMBAHASAN

Hasil Evaluasi

Pemodelan awal telah kami lakukan pada data kategori “baju” untuk mengukur seberapa baik model yang dihasilkan tanpa melakukan proses - proses tambahan. Hasil evaluasi dari model awal ini dapat dilihat pada tabel 3.

Tabel 3. Hasil Evaluasi Model Sebelum Optimasi

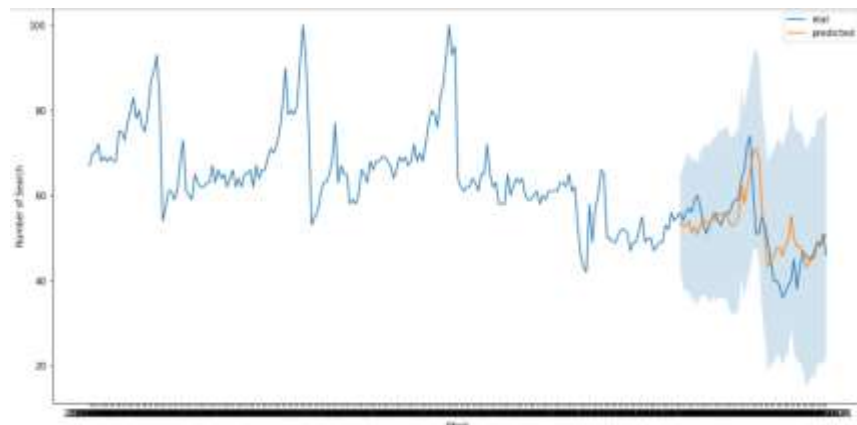
Model Evaluasi	Hasil
R2	-0,240668
MAE	7,235702
MSE	90,26489
RMSE	14,545295

Selanjutnya kami telah melakukan improvisasi performa model dengan tuning *hyperparameter* dari model SARIMAX, juga dengan mengubah *parameter seasonality* yang tadinya kami mengalami mispersepsi dengan diisikan angka 12

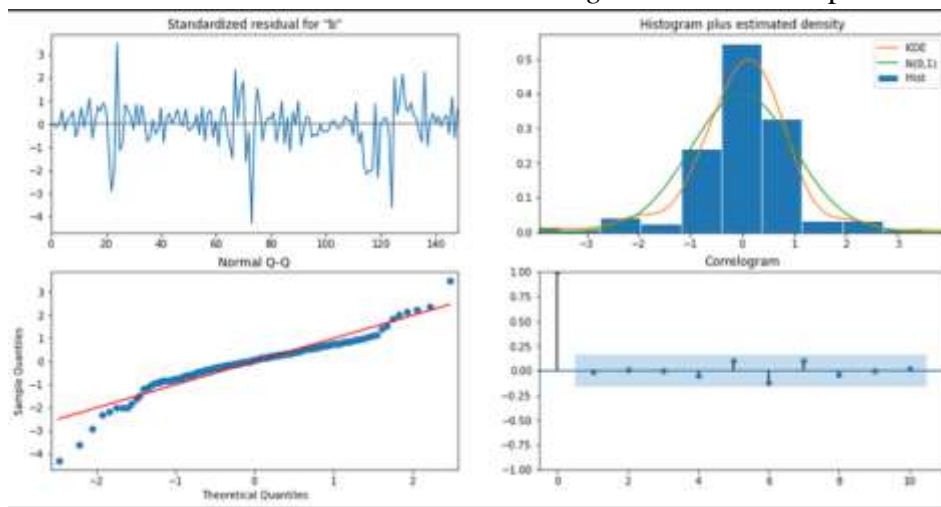
dalam setahun yang mengindikasikan data perbulan padahal data yang kami pakai merupakan data per minggu dimana terdapat 52 minggu dalam satu tahun. Hasil performa model dapat dilihat pada tabel 4, gambar 6 dan 7 yang dapat dikomparasi langsung dengan hasil pemodelan awal sebelumnya.

Tabel 4. Hasil Evaluasi Model Sesudah Optimasi

Model Evaluasi	Hasil
R2	0,3699
	3
MAE	4,8983
	82
MSE	44,664
	985
RMSE	10,091
	467



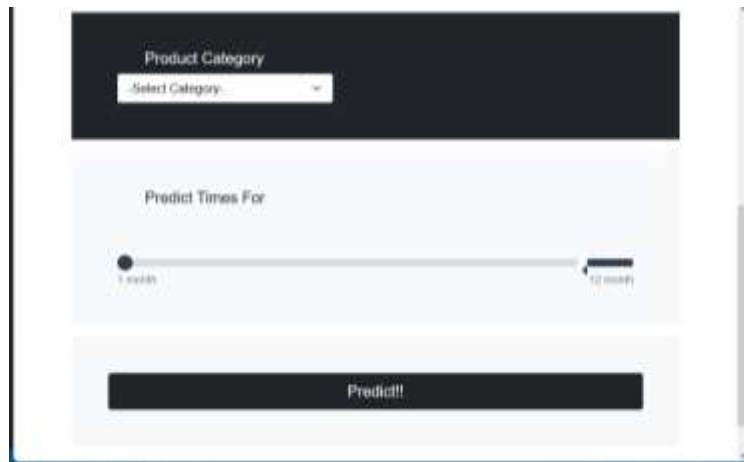
Gambar 6. Grafik Hasil *Forecasting* Model Setelah Optimasi



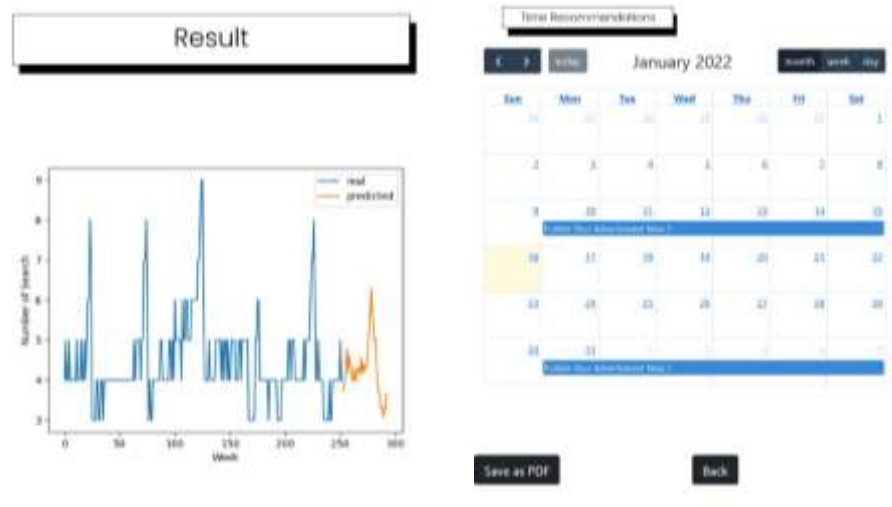
Gambar 7. Diagnosa Model Setelah Optimasi

Hasil Implementasi

Setelah model telah dibuat dan di evaluasi sehingga model menjadi optimal langkah selanjutnya model tersebut diimplementasikan ke dalam sebuah *web* yang hasilnya terdapat pada gambar 8 dan 9.



Gambar 8. Tampilan Awal Web



Gambar 9. Tampilan Web Hasil Dari Prediksi

KESIMPULAN

Dari hasil yang didapatkan, baik dilihat dari nilai akurasi dan error maupun dari grafik prediksi yang didapatkan proses optimasi model awal telah cukup baik meningkatkan performa prediksi model dimana keterlambatan seasonality yang terjadi sebelumnya tidak lagi terjadi dan hanya terjadi galat pada tingkat kecocokan nilai prediksi dan nilai sebelumnya. Berikutnya model ini bisa platform yang lain seperti *mobile Android* maupun *iOS* atau ditingkatkan lebih jauh maupun dikomparasi dengan model lainnya.

REFERENSI

- A. Aggarwal, M. Alshehri, M. Kumar, O. Alfarraj, P. Sharma, and K. R. Pardasani, "Landslide data analysis using various time-series forecasting models," *Computers and Electrical Engineering*, vol. 88, Dec. 2020, doi: 10.1016/j.compeleceng.2020.106858.
- A. Kumar Dubey, A. Kumar, V. García-Díaz, A. Kumar Sharma, and K. Kanhaiya, "Study and analysis of SARIMA and LSTM in forecasting time series data," *Sustainable Energy Technologies and Assessments*, vol. 47, Oct. 2021, doi: 10.1016/j.seta.2021.101474.
- A. Lidwina, "Produk Fesyen Jadi Primadona di E-Commerce | Databoks," 2021. <https://databoks.katadata.co.id/datapublish/2021/06/09/produk-fesyen-jadi-primadona-di-e-commerce>
- A. L. Kusumatriana *et al.*, *Statistik E-Commerce 2020*. 2020.
- C. Sietou, "Societal interest in puppies and the Covid-19 pandemic: A google trends analysis," *Preventive Veterinary Medicine*, vol. 196, Nov. 2021, doi: 10.1016/j.prevetmed.2021.105496.
- F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.
- G. Reikard, "Forecasting paleoclimatic data with time series models," *Results in Geophysical Sciences*, vol. 6, p. 100015, Jun. 2021, doi: 10.1016/j.ringps.2021.100015.
- H. Budiman, "PERAN TEKNOLOGI INFORMASI DAN KOMUNIKASI DALAM PENDIDIKAN," *Al-Tadzkiyyah: Jurnal Pendidikan Islam*, vol. 8, 2017.
- H. Yakubu and C. K. Kwong, "Forecasting the importance of product attributes using online customer reviews and Google Trends," *Technological Forecasting and Social Change*, vol. 171, Oct. 2021, doi: 10.1016/j.techfore.2021.120983.
- K. He, L. Ji, C. W. D. Wu, and K. F. G. Tso, "Using SARIMA–CNN–LSTM approach to forecast daily tourism demand," *Journal of Hospitality and Tourism Management*, vol. 49, pp. 25–33, Dec. 2021, doi: 10.1016/j.jhtm.2021.08.022.
- K. Tanaka, "A sales forecasting model for new-released and nonlinear sales trend products," in *Expert Systems with Applications*, Nov. 2010, vol. 37, no. 11, pp. 7387–7393. doi: 10.1016/j.eswa.2010.04.032.

Statista, “eCommerce - Indonesia | Statista Market Forecast,” 2021.
<https://www.statista.com/outlook/dmo/ecommerce/indonesia#revenue>
(accessed Jan. 15, 2022).