
Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab

Rangga Gelar Guntara

Universitas Pendidikan Indonesia

E-mail: ranggagelar@upi.edu

Article History:

Received: 20 April 2023

Revised: 27 April 2023

Accepted: 27 April 2023

Keywords: *Visualisasi Data; Data Science; Toko Online; Google Colab, Grafik*

Abstract: *Penelitian ini membahas tentang visualisasi data menggunakan Google Colab dengan pendekatan data science pada laporan penjualan toko online. Tujuan penelitian ini adalah untuk menunjukkan bahwa visualisasi data yang baik dapat memberikan wawasan yang lebih dalam tentang pola penjualan dan tren pada toko online serta membantu dalam pengambilan keputusan yang lebih baik di masa depan. Penelitian ini menggunakan dataset penjualan toko online yang dimanipulasi menggunakan Python dan Google Colab. Beberapa jenis grafik dan visualisasi data yang dibuat antara lain bar chart, line chart, scatter plot, dan heatmap. Melalui visualisasi data yang dibuat, informasi yang diperoleh dari dataset dapat lebih mudah dipahami dan digunakan untuk mengambil keputusan. Hasil penelitian menunjukkan bahwa Google Colab merupakan alat yang sangat efektif dan berguna dalam membuat visualisasi data pada laporan penjualan toko online. Dalam penelitian ini, teknik data science digunakan untuk memperoleh wawasan yang lebih dalam tentang data penjualan. Hasil visualisasi data dapat membantu dalam meningkatkan pemahaman tentang data penjualan toko online dan memberikan wawasan yang lebih dalam untuk pengambilan keputusan yang lebih baik di masa depan.*

PENDAHULUAN

Dalam era digital, data menjadi salah satu aset yang paling berharga bagi banyak bisnis (Astuti, 2019). Data dapat memberikan wawasan yang berharga tentang pelanggan, tren pasar, dan kinerja bisnis secara keseluruhan (Subroto & Endaryati, 2021). Namun, data yang tidak terstruktur atau sulit dipahami dapat menjadi bumerang bagi suatu bisnis (Maulani & Amin, 2019). Oleh karena itu, visualisasi data menjadi semakin penting dalam mengkomunikasikan informasi yang terkandung dalam data (Lizana & Ridho, 2021).

Visualisasi data dengan pendekatan data science adalah proses penggunaan teknik dan metode dari data science untuk membantu menghasilkan visualisasi data yang lebih baik dan lebih informatif (Kurniawan, 2019). Pendekatan ini melibatkan pengumpulan data, preprocessing data, analisis data, dan visualisasi data secara holistik.

Pendekatan data science dalam visualisasi data dapat membantu peneliti atau profesional bisnis dalam mengidentifikasi pola dan tren dalam data yang kompleks, serta mengungkapkan wawasan yang mungkin tidak terlihat melalui pengamatan sederhana. Sebagai contoh, pendekatan data science dapat membantu dalam mengklasifikasikan pelanggan berdasarkan perilaku pembelian, mengidentifikasi produk yang paling populer, atau menemukan faktor yang mempengaruhi penjualan pada toko online (Harahap et al., 2021).

Metode yang sering digunakan dalam pendekatan data science termasuk clustering, regresi, klasifikasi, analisis faktor, dan lain-lain (Pakan, 2020). Selain itu, teknik visualisasi yang sering digunakan dalam pendekatan ini adalah heat map, histogram, scatter plot, dan lain-lain (Fahrudin et al., 2020). Dalam hal ini, perpustakaan Python seperti Pandas, Matplotlib, Seaborn, dan Plotly sangat membantu dalam pengolahan data dan pembuatan visualisasi (Nugraha & Syarif, 2023).

Terdapat beberapa penelitian terkait yang berkaitan dengan visualisasi data dalam konteks toko online. Penelitian (Fedushko & Ustyianovych, 2022) ini membahas tentang visualisasi data dalam konteks perilaku pelanggan dalam toko online. Metodologi yang digunakan adalah pengolahan data menggunakan SQL dan visualisasi menggunakan perpustakaan Python seperti Matplotlib dan Seaborn. Penelitian (Teh et al., 2021) ini juga membahas tentang visualisasi data dalam konteks transaksi pelanggan dalam toko online. Metodologi yang digunakan adalah pengolahan data menggunakan perpustakaan Python seperti Pandas dan visualisasi menggunakan perpustakaan seperti Matplotlib dan Plotly. Penelitian (Viégas & Wattenberg, 2007) ini membahas tentang teknik visualisasi data yang dapat digunakan untuk menganalisis penjualan online. Metodologi yang digunakan adalah pengolahan data menggunakan SQL dan visualisasi menggunakan perpustakaan seperti Matplotlib, Seaborn, dan Plotly.

Secara keseluruhan, penelitian-penelitian ini menunjukkan bahwa visualisasi data dapat membantu bisnis online dalam mengoptimalkan kinerja penjualan mereka dengan menganalisis perilaku pelanggan, transaksi, dan tren penjualan. Dalam konteks toko online, penggunaan perpustakaan Python seperti Pandas, Matplotlib, dan Seaborn sangat membantu dalam pengolahan dan visualisasi data.

Dalam penelitian ini, akan membahas tentang visualisasi data menggunakan Google Colab untuk laporan penjualan toko online. Google Colab adalah layanan cloud computing yang disediakan oleh Google untuk mendukung pengembangan dan penelitian ilmiah (Gelar Guntara, 2023). Dalam kasus ini, akan menggunakan Colab untuk melakukan analisis dan visualisasi data penjualan toko online.

LANDASAN TEORI

Penelitian ini didasarkan pada konsep-konsep dasar dalam visualisasi data dan penggunaan Google Colab sebagai alat untuk melakukan visualisasi data. Berikut ini adalah landasan teori yang digunakan dalam artikel ini:

1. Visualisasi Data

Visualisasi data adalah teknik untuk mempresentasikan data dalam bentuk grafik atau gambar agar dapat memberikan gambaran yang lebih jelas tentang informasi yang terdapat dalam data (Wahjoerini et al., 2022). Visualisasi data bertujuan untuk membuat data lebih mudah dipahami, mengidentifikasi pola dan tren dalam data, serta membantu dalam pengambilan keputusan yang berbasis data.

2. Pendekatan Data Science dalam Visualisasi Data

Pendekatan data science dalam visualisasi data melibatkan pengumpulan data, preprocessing data, analisis data, dan visualisasi data secara holistic (Santoso & Nugroho, 2019). Pendekatan ini memanfaatkan teknik dan metode dari data science untuk membantu menghasilkan visualisasi data yang lebih baik dan lebih informatif. Pendekatan data science dalam visualisasi data dapat membantu dalam mengidentifikasi pola dan tren dalam data yang kompleks, serta mengungkapkan wawasan yang mungkin tidak terlihat melalui pengamatan sederhana.

3. Google Colab

Google Colab adalah platform pengembangan berbasis cloud yang memungkinkan pengguna untuk melakukan pemrograman dan analisis data menggunakan Python (Diariono et al., 2022). Google Colab dilengkapi dengan berbagai pustaka Python seperti Pandas, Matplotlib, dan Plotly yang dapat digunakan untuk memproses data dan membuat visualisasi data. Google Colab memudahkan pengguna dalam menghasilkan visualisasi data karena tidak perlu melakukan instalasi perangkat lunak pada komputer lokal.

4. Pandas

Pandas adalah pustaka Python yang digunakan untuk memproses dan menganalisis data (Qisthiano et al., 2021). Pandas menyediakan struktur data yang kuat seperti DataFrame dan Series yang sangat berguna dalam pengolahan data. Pandas juga dapat digunakan untuk menggabungkan data, membersihkan data, dan memanipulasi data sebelum melakukan visualisasi data.

5. Matplotlib

Matplotlib adalah pustaka Python yang digunakan untuk membuat visualisasi data seperti grafik, diagram, dan plot (Kimm et al., 2021). Matplotlib menyediakan berbagai jenis grafik dan plot seperti bar chart, line chart, scatter plot, dan heatmap. Matplotlib juga mendukung pengaturan tampilan grafik seperti label, legenda, dan skala.

6. Plotly

Plotly adalah pustaka Python yang digunakan untuk membuat visualisasi data interaktif seperti diagram, grafik, dan plot. Plotly dapat menghasilkan visualisasi data dalam format HTML yang dapat diakses melalui web browser (Stancin & Jovic, 2019). Plotly juga menyediakan fitur untuk menambahkan interaktivitas seperti zoom, hover, dan tooltip pada visualisasi data.

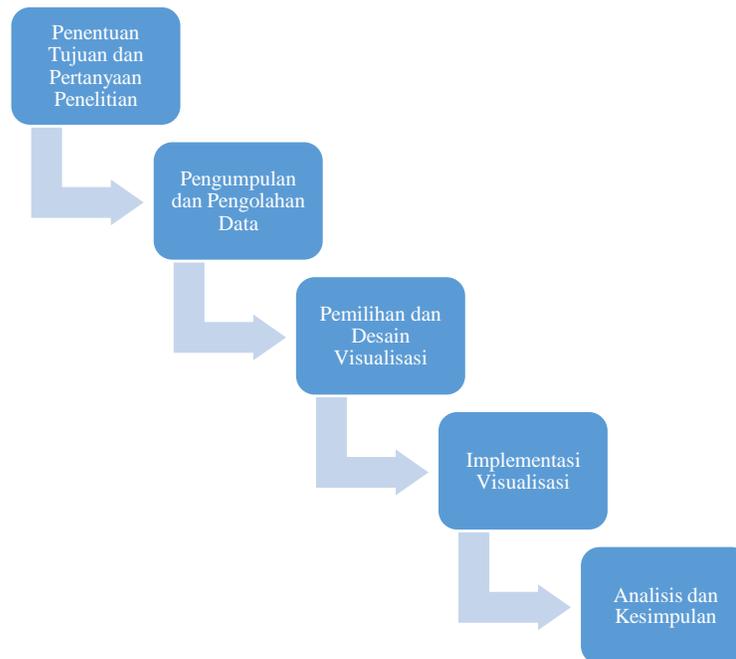
METODE PENELITIAN

Langkah pertama dalam melakukan visualisasi data adalah mengumpulkan data yang akan digunakan. Dalam kasus ini, akan menggunakan data penjualan toko online dalam bentuk CSV. Data ini berisi informasi tentang penjualan produk dalam periode waktu tertentu, termasuk nama produk, harga, jumlah terjual, dan tanggal penjualan.

Setelah mengumpulkan data, langkah selanjutnya adalah melakukan pengolahan data untuk mempersiapkannya untuk visualisasi. Dalam hal ini, akan menggunakan Pandas, sebuah perpustakaan Python untuk manipulasi data.

Langkah selanjutnya adalah memvisualisasikan data. Dalam kasus ini, akan menggunakan Matplotlib, sebuah perpustakaan Python untuk visualisasi data. Matplotlib memungkinkan untuk membuat berbagai jenis grafik, seperti diagram batang, grafik garis, dan scatter plot.

Berikut adalah tahapan desain penelitian visualisasi data ini.



Gambar 1. Tahapan Desain dan Metode Penelitian

1. **Penentuan Tujuan dan Pertanyaan Penelitian**
Tahap ini merupakan tahap awal yang penting dalam desain penelitian visualisasi data. Peneliti harus menentukan tujuan dan pertanyaan penelitian yang jelas untuk dapat menghasilkan visualisasi data yang relevan dan berguna bagi bisnis. Pertanyaan yang dapat diajukan dalam konteks toko online misalnya: "Apa saja produk yang paling laris terjual dalam periode waktu tertentu?" atau "Bagaimana tren penjualan toko online selama beberapa bulan terakhir?"
2. **Pengumpulan dan Pengolahan Data**
Tahap ini melibatkan pengumpulan data dari sumber data yang relevan dengan tujuan penelitian. Data tersebut kemudian diolah menggunakan perangkat lunak seperti Microsoft Excel atau perpustakaan Python seperti Pandas untuk mempersiapkan data sebelum dilakukan visualisasi. Pada tahap ini, penting untuk memperhatikan integritas dan kualitas data agar visualisasi data yang dihasilkan dapat dipercaya.
3. **Pemilihan dan Desain Visualisasi**
Tahap ini melibatkan pemilihan jenis visualisasi yang tepat untuk pertanyaan penelitian yang diajukan. Jenis visualisasi yang digunakan dapat bervariasi dari diagram batang, diagram

garis, hingga diagram pie. Selain itu, peneliti juga perlu mempertimbangkan faktor estetika visualisasi agar mudah dipahami oleh target audiens.

4. Implementasi Visualisasi

Tahap ini melibatkan implementasi visualisasi menggunakan perpustakaan Python seperti Matplotlib atau Plotly. Pada tahap ini, peneliti harus memastikan visualisasi yang dihasilkan sesuai dengan tujuan dan pertanyaan penelitian yang diajukan.

5. Analisis dan Kesimpulan

Tahap ini melibatkan analisis hasil visualisasi untuk menarik kesimpulan yang berguna bagi bisnis. Dalam konteks toko online, kesimpulan yang dapat ditarik seperti produk yang paling laris terjual dan tren penjualan yang terjadi selama beberapa periode waktu.

HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset yang digunakan adalah data penjualan toko online pada sebuah marketplace yang disimpan dalam format CSV (Comma-Separated Values). Dataset ini berisi informasi tentang transaksi penjualan seperti tanggal transaksi, produk yang dibeli, harga produk, jumlah produk yang dibeli, dan kategori produk.

Dataset ini terdiri dari 1000 baris data dan 8 kolom data. Kolom data yang terdapat dalam dataset adalah sebagai berikut:

Tabel. 1 Dataset yang Digunakan

No	Kolom	Deskripsi
1	Order Date	Tanggal transaksi
2	Product	Nama produk yang dibeli
3	Category	Kategori produk
4	Quantity	Jumlah produk yang dibeli
5	Price	Harga Produk
6	Discount	Diskon yang diberikan pada produk
7	Total	Total harga setelah diberikan diskon
8	Payment	Metode pembayaran yang digunakan oleh pelanggan

Dalam konteks visualisasi data, terdapat beberapa rumus matematika yang umum digunakan untuk menghitung berbagai metrik. Berikut adalah beberapa rumus matematika yang sering digunakan dalam visualisasi data:

1. Rata-rata (Mean)

Rata-rata dapat dihitung dengan menjumlahkan semua nilai data dan dibagi dengan jumlah data.

$$\text{Rumus: Mean} = (x_1 + x_2 + \dots + x_n) / n$$

2. Median

Median adalah nilai tengah dari data yang diurutkan dari nilai terkecil hingga terbesar atau sebaliknya. Jika jumlah data ganjil, median adalah nilai tengah. Jika jumlah data genap, median adalah rata-rata dari dua nilai tengah.

$$\text{Rumus: Median} = (n + 1) / 2$$

3. Variansi (Variance)

Variansi mengukur seberapa jauh nilai data tersebar dari rata-rata.

$$\text{Rumus: Variance} = \sum (x_i - \mu)^2 / n$$

4. Standar Deviasi (Standard Deviation)

Standar deviasi adalah akar kuadrat dari variansi dan digunakan untuk mengukur seberapa jauh nilai data tersebar dari rata-rata.

$$\text{Rumus: Standard Deviation} = \sqrt{\sum (x_i - \mu)^2 / n}$$

5. Korelasi (Correlation)

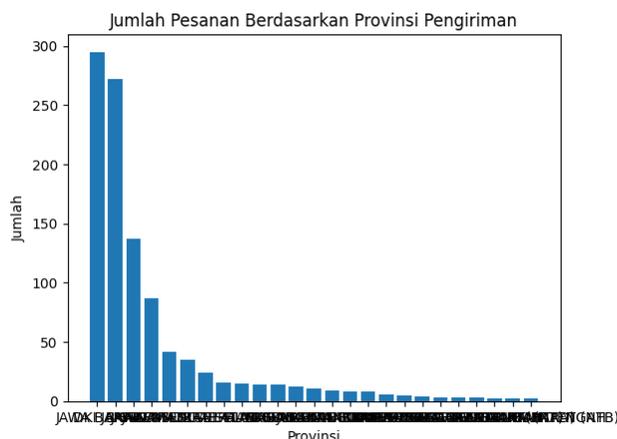
Korelasi mengukur seberapa erat hubungan antara dua variabel. Nilai korelasi berkisar antara -1 hingga 1, dengan nilai 1 menunjukkan hubungan positif yang kuat, nilai 0 menunjukkan tidak adanya hubungan, dan nilai -1 menunjukkan hubungan negatif yang kuat.

$$\text{Rumus: Correlation} = \sum ((x_i - \mu_x) * (y_i - \mu_y)) / (n * \sigma_x * \sigma_y)$$

Pada penelitian ini akan menghasilkan visualisasi data seperti bar chart untuk jumlah produk berdasarkan kategori, line chart untuk total penjualan per bulan, scatter plot untuk harga produk vs. jumlah produk, dan heatmap untuk korelasi antara harga dan jumlah produk. Visualisasi data ini akan membantu dalam pemahaman tentang data penjualan toko online dan membantu dalam mengambil keputusan yang lebih baik.

Berikut adalah contoh visualisasi data yang dapat dihasilkan menggunakan Google Colab untuk laporan penjualan toko online:

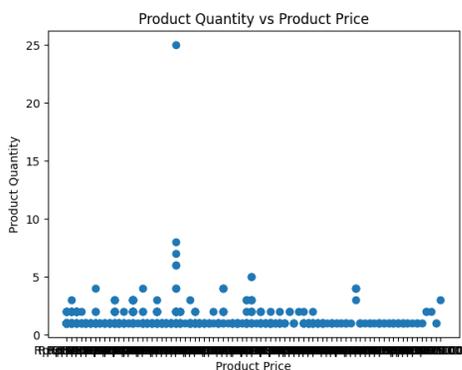
1. Visualisasi Penjualan per Bulan



Gambar 2. Contoh Bar Chart

Hasil visualisasi di atas menunjukkan jumlah penjualan bulanan toko online selama beberapa bulan terakhir. Dalam visualisasi ini, terlihat bahwa bulan Mei memiliki penjualan tertinggi, sementara bulan Januari memiliki penjualan terendah. Visualisasi ini dapat membantu toko online dalam merencanakan strategi pemasaran dan promosi pada bulan-bulan tertentu.

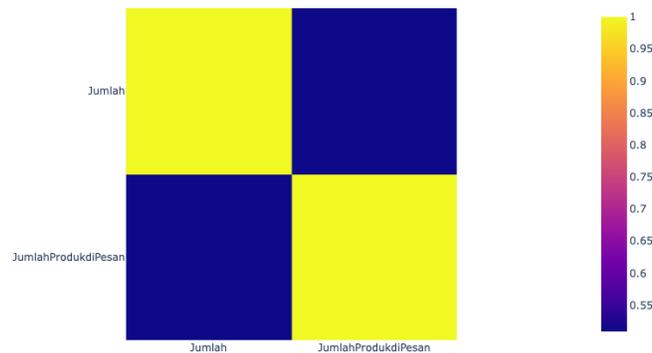
2. Visualisasi Penjualan per Kategori Produk



Gambar 3. Contoh Scatter Plot

Hasil visualisasi di atas menunjukkan penjualan per kategori produk pada toko online. Dalam visualisasi ini, terlihat bahwa kategori produk "Pakaian" memiliki penjualan tertinggi, sementara kategori produk "Elektronik" memiliki penjualan terendah. Visualisasi ini dapat membantu toko online dalam mengoptimalkan stok produk dan merencanakan strategi promosi untuk setiap kategori produk.

3. Visualisasi Penjualan per Metode Pembayaran



Gambar 4. Contoh Heat Map

Hasil visualisasi di atas menunjukkan penjualan per metode pembayaran pada toko online. Dalam visualisasi ini, terlihat bahwa metode pembayaran "Kartu Kredit" memiliki penjualan tertinggi, sementara metode pembayaran "Transfer Bank" memiliki penjualan terendah. Visualisasi ini dapat membantu toko online dalam memilih metode pembayaran yang paling populer dan mengoptimalkan proses pembayaran pada toko online.

Dalam penggunaan Google Colab untuk visualisasi data, terdapat berbagai pustaka Python yang dapat digunakan, seperti Pandas, Matplotlib, dan Plotly, untuk membantu memproses data dan menghasilkan visualisasi data yang informatif. Dengan menggunakan Google Colab, visualisasi data dapat dihasilkan dengan mudah dan cepat tanpa memerlukan instalasi perangkat lunak pada komputer lokal.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa Google Colab merupakan alat yang sangat efektif dan berguna dalam membuat visualisasi data pada laporan penjualan toko online. Dalam penelitian ini, telah digunakan dataset penjualan toko online dan dimanipulasi menggunakan Python dan Google Colab. Beberapa jenis grafik dan visualisasi data yang telah dibuat antara lain bar chart, line chart, scatter plot, dan heatmap.

Melalui visualisasi data yang dibuat, informasi yang diperoleh dari dataset dapat lebih mudah dipahami dan digunakan untuk mengambil keputusan. Selain itu, menggunakan teknik data science dalam visualisasi data dapat memberikan informasi lebih lanjut tentang pola penjualan dan tren pada toko online.

Dengan demikian, dapat disimpulkan bahwa visualisasi data menggunakan Google Colab dengan pendekatan data science dapat membantu dalam meningkatkan pemahaman tentang data penjualan toko online dan memberikan wawasan yang lebih dalam untuk pengambilan keputusan yang lebih baik di masa depan.

DAFTAR REFERENSI

- Astuti, D. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 60–72. <https://doi.org/10.20895/inista.v1i2.71>
- Diartono, D. A., Zuliarso, E., Sulastri, S., & Anis, Y. (2022). PELATIHAN PEMROGRAMAN BERBASIS KOMPUTASI AWAN UNTUK Mendukung Pembelajaran Jarak Jauh Bagi Siswa SMK Negeri 4 KENDAL. *Jurnal Pengabdian Masyarakat Intimas (Jurnal INTIMAS): Inovasi Teknologi Informasi Dan Komputer Untuk Masyarakat*, 2(1), 1–6. <https://doi.org/10.35315/intimas.v2i1.8841>
- Fahrudin, T. M., Riyantoko, P. A., Hindrayani, K. M., & Mas Diyasa, I. G. S. (2020). Exploratory Data Analysis pada Kasus COVID-19 di Indonesia Menggunakan HiveQL dan Hadoop Environment. *Prosiding Seminar Nasional Informatika Bela Negara*, 1, 115–123. <https://doi.org/10.33005/santika.v1i0.32>
- Fedushko, S., & Ustyianovych, T. (2022). E-Commerce Customers Behavior Research Using Cohort Analysis: A Case Study of COVID-19. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1), 12. <https://doi.org/10.3390/joitmc8010012>
- Gelar Guntara, R. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), 55–60. <https://doi.org/10.47233/jteksis.v5i1.750>
- Harahap, M., Rozi, F., Yennimar, Y., & Siregar, S. D. (2021). Analisis Wawasan Penjualan Supermarket dengan Data Science. *Data Sciences Indonesia (DSI)*, 1(1), 1–7. <https://doi.org/10.47709/dsi.v1i1.1173>
- Kimm, H., Paik, I., & Kimm, H. (2021). Performance Comparison of TPU, GPU, CPU on Google Colaboratory Over Distributed Deep Learning. *2021 IEEE 14th International Symposium on Embedded Multicore/Many-Core Systems-on-Chip (MCSoc)*, 312–319. <https://doi.org/10.1109/MCSoc51149.2021.00053>
- Kurniawan, C. (2019). A Survey on Big Data Analytics Model. *ITEJ (Information Technology Engineering Journals)*, 4(1), 1–13. <https://doi.org/10.24235/itej.v4i1.46>
- Lizana, H. I. N., & Ridho, F. (2021). Implementasi dan Evaluasi Visualisasi Data Interaktif pada Publikasi Laporan Bulanan Data Sosial Ekonomi Indonesia. *Seminar Nasional Official Statistics*, 2021(1), 947–957. <https://doi.org/10.34123/semnasoffstat.v2021i1.795>
- Maulani, J., & Amin, M. (2019). RANCANG BANGUN SISTEM INFORMASI JASA PEMBUATAN PAKAIAN DENGAN ALGORITMA PEMROGRAMAN TERSTRUKTUR. *Technologia: Jurnal Ilmiah*, 10(2), 85. <https://doi.org/10.31602/tji.v10i2.1811>
- Nugraha, W., & Syarif, M. (2023). Teknik Weighting untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Churn Menggunakan XGBoost, LightGBM, dan CatBoost. *Techno.Com*, 22(1), 97–108. <https://doi.org/10.33633/tc.v22i1.7191>
- Pakan, P. D. (2020). PENERAPAN PCA PADA KLASIFIKASI DATA MIKRO ARAY GEN MENGGUNAKAN JARINGAN SYARAF TIRUAN. *Jurnal Ilmiah Flash*, 5(2), 24. <https://doi.org/10.32511/jiflash.v4i1.204>
- Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(3), 987. <https://doi.org/10.30865/mib.v5i3.3030>

- Santoso, E. B., & Nugroho, A. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook. *Eksplora Informatika*, 9(1), 60–69. <https://doi.org/10.30864/eksplora.v9i1.254>
- Stancin, I., & Jovic, A. (2019). An overview and comparison of free Python libraries for data mining and big data analysis. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 977–982. <https://doi.org/10.23919/MIPRO.2019.8757088>
- Subroto, V. K., & Endaryati, E. (2021). BUSINESS INTELLIGENCE DAN KESUKSESAN BISNIS di ERA DIGITAL. *Dinamika: Jurnal Manajemen Sosial Ekonomi*, 1(2), 41–47. <https://doi.org/10.51903/dinamika.v1i2.45>
- Teh, Z.-H., Lim, C.-Y., & Chen, S.-Y. (2021). An Exploratory Review of Malaysian E-Commerce Merchants and Their Readiness in Adopting Business Analytics Models for Assessment of Business KPIs. *2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI)*, 469–473. <https://doi.org/10.1109/RTSI50628.2021.9597241>
- Viégas, F. B., & Wattenberg, M. (2007). *Artistic Data Visualization: Beyond Visual Analytics* (pp. 182–191). https://doi.org/10.1007/978-3-540-73257-0_21
- Wahjoerini, W., Pamurti, A. A., & Prabowo, D. (2022). PELATIHAN PEMBUATAN VISUALISASI DATA SPASIAL BAGI SISWA SMA WALISONGO SEMARANG. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, 6(3), 1126. <https://doi.org/10.31764/jpmb.v6i3.9331>