

# ANALISIS DAN PERBANDINGAN MODEL KLASIFIKASI PADA PERUSAHAAN RIDERSREPUBLIC MENGGUNAKAN METODE DCOVA

Gabriel Buffon Harahap

*Sistem Informasi, Universitas Multimedia Nusantara  
Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten – 15811 Indonesia  
gabriel.buffon@student.umn.ac.id*

“Retail Business/ RidersRepublic”

**Abstrak**— *Penjualan sepeda dan barang terkait dengan dunia sepeda adalah bagian penting dari bisnis ritel. Tujuan dari penelitian ini adalah menganalisis variabel-variabel yang mempengaruhi penjualan dari sepeda dan barang-barang terkait, serta memprediksi kinerja penjualan dengan menggunakan teknik Random Forest dan Gradient Boosting. Variabel seperti Country, Year, dan Cost adalah beberapa contoh variabel yang dapat mempengaruhi penjualan. Kami menemukan hubungan antara variabel-variabel ini dan penjualan dengan menganalisis data dan melakukan analisis statistik. Selanjutnya, kami membuat prediksi penjualan yang akurat menggunakan model Random Forest dan Gradient Boosting. Hasil penelitian menunjukkan bahwa kedua strategi tersebut dapat memberikan estimasi penjualan yang akurat. Penelitian ini membantu perusahaan ritel meningkatkan kinerja dan mengoptimalkan strategi penjualan mereka.*

**Kata kunci**— **Sepeda, Bisnis, Random Forest, Gradient Boosting**

## I. PENDAHULUAN

Pada era globalisasi yang semakin berkembang serta semakin maju, banyak kota-kota besar di dunia mengalami pergerakan ekonomi yang signifikan termasuk Indonesia. Pergerakan ini memberikan perubahan sehingga membuat perubahan gaya beli dari masyarakat yang membuat perusahaan-perusahaan menjadi berkembang. Dalam industri ritel, bisnis penjualan sepeda dan aksesorisnya telah menjadi salah satu sektor yang semakin berkembang. Bisnis ini menawarkan peluang yang menarik bagi para pengusaha yang tertarik dengan industri transportasi dan gaya hidup aktif. Ritel sepeda memiliki sejumlah keunggulan menarik bagi para pelaku industri. Sepeda telah menjadi alat transportasi yang populer untuk memenuhi kebutuhan mobilitas yang sehat dan ramah lingkungan. Di saat masyarakat peduli terhadap lingkungan dan pola hidup sehat, permintaan akan sepeda terus meningkat, baik untuk keperluan sehari-hari maupun untuk kegiatan rekreasi. Ini merupakan peluang bisnis yang menjanjikan bagi perusahaan maupun penjual pribadi diluar sana yang menjual sepeda.

Selain itu, perkembangan teknologi dan inovasi di industri sepeda juga membuat permintaan aksesoris dan upgrade semakin meningkat. Pelanggan saat ini tidak hanya mencari sepeda berkualitas, tetapi juga aksesoris yang dapat meningkatkan pengalaman berkendara mereka, seperti ban, jok, inflatables, dan lainnya. Ini menciptakan peluang bagi pengecer untuk menawarkan berbagai

macam aksesoris untuk memenuhi kebutuhan dan preferensi pelanggan mereka. Sebagai bisnis ritel, faktor kunci keberhasilan meliputi pemilihan produk yang tepat, layanan pelanggan yang berkualitas, dan upaya pemasaran yang efektif. Lokasi toko fisik yang strategis di area dengan lalu lintas tinggi dan aksesibilitas yang baik dapat membantu meningkatkan visibilitas dan daya tarik bisnis. Memilih produk yang tepat juga penting, menawarkan sepeda dan aksesoris berkualitas dari merek terkenal dan mengikuti tren terkini di industri sepeda.

Ritel adalah bisnis yang melibatkan penjualan barang atau jasa langsung ke konsumen akhir untuk penggunaan pribadi. Perdagangan eceran tumbuh secara signifikan di Indonesia sejak tahun 1980-an, seiring dengan pertumbuhan ekonomi negara. Menurut Asosiasi Pengusaha Ritel Indonesia (APRINDO), sektor ritel di Indonesia terus berkembang pesat. Dalam laporan Global Retail Development Index (GRDI) 2019 yang diterbitkan oleh firma konsultan A.T Kearney, Indonesia masih masuk dalam 10 besar negara dengan industri ritel yang berkembang pesat. Bahkan, Indonesia berada di peringkat 5 dari 30 negara berkembang di dunia dengan skor 55,9 dari skala 100 dari 100. Hal ini menunjukkan bahwa pasar ritel di Indonesia memiliki banyak potensi untuk terus tumbuh dan berkembang. Pertumbuhan ekonomi yang kuat, populasi yang besar, gaya hidup konsumen yang berubah, serta kemajuan teknologi dan akses internet yang lebih luas menjadi faktor yang mendukung pertumbuhan sektor ritel di China, Indonesia. Dalam konteks ini, peritel harus mampu mengikuti tren dan inovasi terkini, menghadirkan pengalaman berbelanja yang menyenangkan, serta menjaga kualitas produk dan layanannya agar tetap kompetitif di lanskap persaingan yang semakin ketat ini [1].

RidersRepublic menjadi bisnis ritel yang bersaing dengan market-market lain dengan membuka toko di kota-kota lain di negara-negara besar di dunia. Permasalahan yang dihadapi oleh RidersRepublic merupakan mengidentifikasi produk-produk yang mempunyai nilai jual yang tinggi, tren penjualan antar negara, dan faktor-faktor yang menyebabkan peningkatan atau penurunan penjualan. RidersRepublic menggunakan SAS Viya sebagai solusi untuk memberikan analisis yang mudah dan akurat.

SAS Viya adalah platform analitik yang sangat berguna untuk menganalisis penjualan sepeda di bisnis ritel seperti RidersRepublic. Dengan menggunakan SAS Viya, RidersRepublic dapat mengakses berbagai alat analisis dan teknik pemodelan yang kuat untuk mengekstrak wawasan berharga dari data penjualan mereka. Salah satu kegunaan utama SAS Viya dalam analisis penjualan sepeda adalah kemampuannya untuk mensegmentasi pelanggan [2]. Menggunakan algoritme pemodelan yang canggih, RidersRepublic dapat mengidentifikasi kelompok pelanggan yang memiliki karakteristik serupa, seperti preferensi produk, perilaku pembelian, dan tingkat pengeluaran. Hal ini memungkinkan RidersRepublic untuk mengembangkan strategi pemasaran yang lebih efektif, menargetkan promosi dan penawaran khusus kepada pelanggan yang tepat, serta meningkatkan retensi pelanggan. Selain itu, SAS Viya juga memungkinkan RidersRepublic untuk melakukan analisis tren penjualan secara mendalam. Mereka dapat melihat tren penjualan sepeda dari waktu ke waktu, mengidentifikasi tren musiman, dan memahami faktor-faktor yang mempengaruhi permintaan pelanggan. Informasi ini dapat digunakan untuk mengoptimalkan inventaris, menjalankan kampanye penjualan yang lebih efektif, dan meningkatkan keuntungan bisnis. Oleh karena itu, menggunakan SAS Viya dalam analisis penjualan sepeda membantu RidersRepublic membuat keputusan akurat berdasarkan data dan mendapatkan keunggulan kompetitif di pasar ritel.

## II. KAJIAN LITERATUR

### A. *Random Forest*

Random Forest adalah algoritma yang digunakan dalam analitik prediktif dalam pembelajaran mesin dan statistik. Algoritma ini menggabungkan beberapa pohon prediksi yang saling bergantung dalam sebuah hutan. Setiap pohon di hutan menggunakan nilai vektor acak yang diambil secara independen dan memiliki distribusi yang sama. Kekuatan utama hutan acak adalah kemampuannya menangani hubungan non-linier dalam data, yang seringkali sulit diatasi dengan regresi linier. Faktanya, Random Forest dapat membuat prediksi yang lebih akurat karena dapat mengidentifikasi pola dan interaksi kompleks antar variabel dalam data. Selain itu, algoritma ini juga dapat mengatasi masalah overfitting yaitu kecenderungan model untuk terlalu banyak mempelajari data training yang kemudian tidak dapat digeneralisasikan ke data baru. Untuk alasan ini, hutan acak telah menjadi pilihan populer dalam berbagai aplikasi analitik prediktif, seperti pengenalan pola, analisis keuangan, dan pengambilan keputusan [3].

### B. *Gradient Boosting*

Peningkatan gradien adalah metode pembelajaran mesin yang memungkinkan setiap model dalam satu set

untuk belajar dari kesalahan model sebelumnya dalam potongan data yang masih sulit dipahami. Teknik ini merupakan bagian yang sangat efektif dari pembelajaran holistik dan telah berhasil digunakan dalam banyak aplikasi praktis. Gradient Boosting dapat disesuaikan dengan kebutuhan khusus setiap aplikasi, seperti mempertimbangkan berbagai fungsi kerugian. Dengan pendekatan iteratif, setiap model baru di set berusaha untuk meningkatkan kinerja secara keseluruhan dengan belajar dari dan mengatasi kekurangan dari model sebelumnya. Selain itu, Gradient Boosting juga dapat mengatasi masalah overfitting dan kompleksitas model melalui penggunaan teknik regularisasi. Dalam dunia analitik data dan pembelajaran mesin, Gradient Boosting telah menjadi pilihan populer karena kemampuannya yang kuat dan fleksibel [4].

### C. *EDA (Exploratory Data Analysis)*

Exploratory data analysis (EDA) adalah metode analisis yang digunakan untuk memahami dan meringkas data secara keseluruhan, sehingga lebih mudah dipahami. Pendekatan ini menyediakan berbagai alat dan teknik untuk mendeskripsikan dan memahami kumpulan data dengan lebih baik, menggunakan grafik dan visualisasi data sebagai alat utama tanpa menggunakan model statistik atau teori formulasi semu. EDA memiliki peran penting sebelum melalui tahap modeling dan feature engineering, karena pada tahap ini kita perlu memahami karakteristik data yang ada. Dengan menganalisis distribusi, korelasi, outlier, dan pola yang terlihat dalam data, EDA membantu kita mengidentifikasi potensi masalah, menemukan hubungan antar variabel, dan lebih memahami data yang akan digunakan. Oleh karena itu, EDA merupakan langkah pertama yang penting dalam proses analisis data, memberikan landasan yang kuat untuk tahap selanjutnya dari proses pemodelan dan pengambilan keputusan [5].

### D. *MSE (Mean Squared Error)*

Mean square error (MSE) adalah metrik yang biasa digunakan dalam statistik dan pembelajaran mesin untuk mengukur keakuratan model regresi dalam memprediksi nilai numerik. MSE menghitung rata-rata kuadrat selisih antara nilai prediksi dan aktual. Dengan mengkuadratkan selisihnya, MSE memberikan bobot lebih pada nilai yang lebih jauh dari nilai sebenarnya, sehingga mencerminkan besarnya kesalahan prediksi secara keseluruhan. Semakin kecil nilai MSE, semakin akurat model memprediksi data [6].

Menggunakan MSE memungkinkan untuk mengevaluasi dan membandingkan berbagai model regresi berdasarkan kemampuannya untuk meminimalkan perbedaan antara prediksi dan nilai sebenarnya. Akibatnya, MSE memberikan panduan yang berguna untuk memilih model terbaik untuk memprediksi dan memprediksi data numerik, serta mendukung pengambilan keputusan yang lebih baik di berbagai bidang seperti peramalan, keuangan,

dan masyarakat sains. Berikut adalah contoh rumus dari MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

MSE = Mean Squared Error  
N = Number of Data Points  
Yi = Predicted Values

### III. METODOLOGI

Metode yang digunakan dalam penelitian ini adalah DCOVA (Define, Collect, Organize, Visualize, Analyze) yang dimana DCOVA merupakan kerangka kerja yang digunakan dalam pemecahan masalah untuk mengidentifikasi, mengumpulkan, mengatur, memvisualisasikan, dan menganalisis data. Kerangka kerja ini bertujuan untuk meningkatkan nilai mempelajari statistik dengan memberikan langkah-langkah yang terstruktur dan jelas saat bekerja dengan data [7].

Langkah pertama, Define, melibatkan klarifikasi masalah atau pertanyaan penelitian yang spesifik serta identifikasi data yang diperlukan untuk menjawabnya. Langkah kedua, Collect, berfokus pada pengumpulan data dari berbagai sumber yang relevan seperti survei, eksperimen, atau basis data yang tersedia. Setelah data terkumpul, langkah ketiga, Organize, melibatkan pengaturan dan persiapan data dengan membersihkan, mentransformasi, dan mengatur data sesuai kebutuhan analisis. Langkah keempat, Visualize, menggunakan berbagai teknik visualisasi data untuk mempresentasikan informasi dengan cara yang lebih mudah dipahami dan membantu mengidentifikasi pola atau tren. Terakhir, langkah kelima, Analyze, melibatkan penerapan metode dan teknik statistik untuk memperoleh wawasan dan kesimpulan yang signifikan dari data yang telah disiapkan. Dengan mengikuti kerangka kerja DCOVA, individu dapat secara efektif bekerja dengan data, meningkatkan pemahaman statistik, dan mendapatkan manfaat yang lebih besar dari studi statistik dengan menghubungkan teori dengan keterampilan praktis dalam pemecahan masalah.

### IV. HASIL DAN PEMBAHASAN

#### A. Define

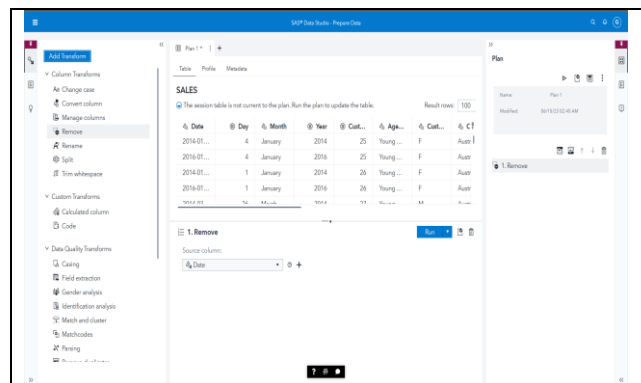
Pada penelitian ini, data yang digunakan merupakan dataset yang diambil dari Kaggle yang berjudul *Bike Sales in Europe* [8]. Data tersebut memiliki jumlah kolom sebanyak 18 buah dan memiliki jumlah baris sebanyak 113,037. Data yang diambil tersebut digunakan dalam melakukan analisis dan perbandingan model klasifikasi pada perusahaan RidersRepublic menggunakan metode DCOVA.

#### B. Collect

Pada perusahaan RidersRepublic digunakan dataset “Bikers Sales in Europe” yang dikumpulkan dari tahun 2011 – 2016. Dataset tersebut berisikan 18 kolom yang diantaranya ada kolom Date, Day, Month, Year, Customer Age, Age Group, Customer Gender, Country, State, Product Category, Sub Category, Product, Order Quantity, Unit Cost, Unit Price, Profit, Cost, Revenue.

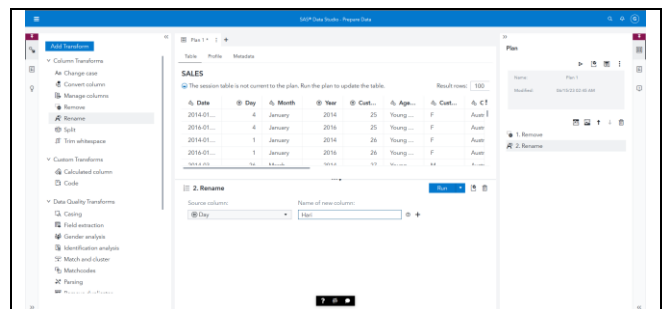
#### C. Organize

Pada tahap ketiga yaitu organize, peneliti melakukan pengolahan data mentah yang dihasilkan dari perusahaan RidersRepublic dan akan dilakukan analisis. Pertama, peneliti menghapus kolom date dalam data dikarenakan kolom tersebut tidak terpakai dalam penelitian ini.



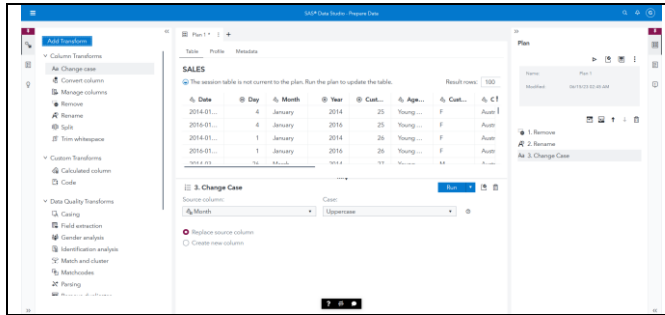
Gambar 4.1 Menghapus Kolom

Kemudian setelah menghapus kolom date dalam data, langkah selanjutnya adalah melakukan rename pada tabel day dan diganti menjadi Hari.



Gambar 4.2 Rename Kolom

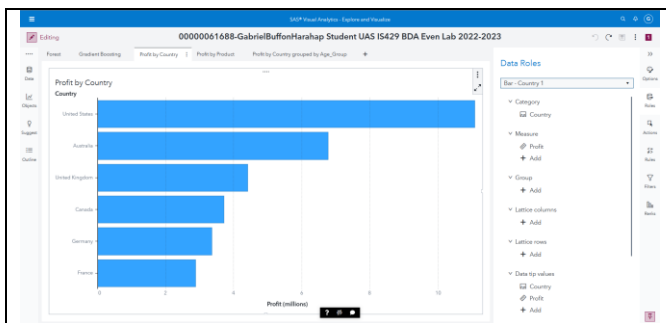
Setelah melakukan rename pada kolom day, Langkah selanjutnya peneliti memberikan Change Case kepada kolom Month yang dimana nantinya huruf dalam kolom Month menjadi huruf kapital.



Gambar 4.3 Change Case Kolom

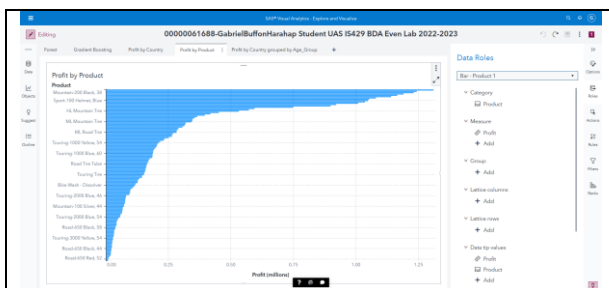
#### D. Visualize

Pada tahap visualize, peneliti menggunakan proses EDA (Exploratory Data Analysis) yang bertujuan untuk memahami isi dari data yang diteliti. Dengan menggunakan EDA, peneliti juga mengharapkan untuk menemukan pola atau korelasi antar variable secara visual. Pada visualisasi pertama, peneliti menggunakan visualisasi bar chart yang dimana peneliti ingin menampilkan nilai profit dari masing-masing negara yang ada dalam data.



Gambar 4.4 Visualisasi Bar Chart

Kemudian pada visualisasi kedua, peneliti menggunakan visualisasi bar chart Kembali namun pada visualisasi ini, peneliti ingin menampilkan profit yang dihasilkan dari masing-masing produk yang ada pada perusahaan RidersRepublic. Negara dengan profit tertinggi adalah negara Amerika Serikat.



Gambar 4.5 Visualisasi Bar Chart

Pada visualisasi EDA yang terakhir peneliti menggunakan bar chart Kembali, namun pada visualisasi ketiga ini peneliti ingin menampilkan profit yang dihasilkan dari masing-masing negara yang diambil dari grup berisi rentang usia orang-

orang yang membeli produk dari RidersRepublic. Rata-rata pembeli dari masing-masing negara adalah orang-orang dewasa yang berumur 35-64 tahun.

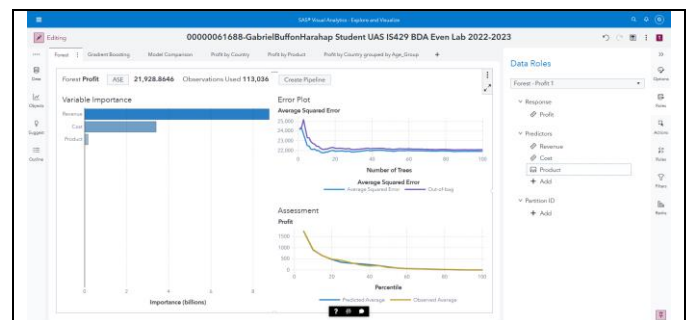


Gambar 4.6 Visualisasi Bar Chart

#### E. Analyze

##### 1. Random Forest

Pada pembuatan model random forest dalam melakukan klasifikasi didapatkan hasil number dari tree dengan jumlah 100 menghasilkan Average Squared Error sebesar 21,928.8646 yang dimana semakin rendah jumlah yang dihasilkan maka semakin bagus.

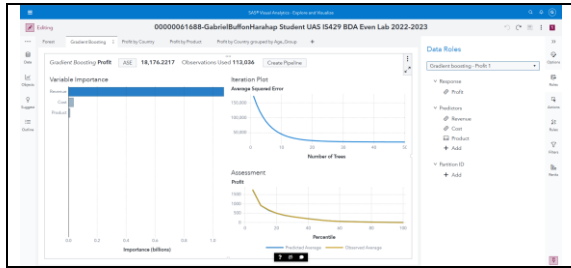


Gambar 4.7 Visualisasi Forest

Model ini dibuat oleh peneliti dengan menggunakan Profit sebagai response. Revenue, Cost dan Product sebagai predictors untuk menampilkan hasil diatas dengan baik dan akurat.

##### 2. Gradient Boosting

Pada pembuatan model gradient boosting dalam melakukan klasifikasi didapatkan hasil number dari tree dengan jumlah 50 menghasilkan Average Squared Error sebesar 18,155.3149 yang dimana dengan hasil tersebut dapat dikatakan bahwa model tersebut semakin akurat dengan analisis yang dilakukan.

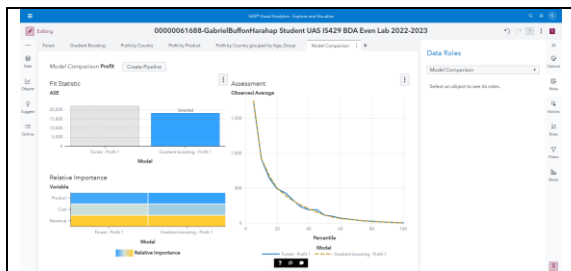


Gambar 4.8 Visualisasi Gradient Boosting

Model ini dibuat oleh peneliti dengan menggunakan Profit sebagai response. Revenue, Cost dan Product sebagai predictors untuk menampilkan hasil diatas dengan baik dan akurat.

### 3. Model Comparison

Pada pembuatan model ini, peneliti mengambil 2 model yang sudah dibuat diawal dan dibandingkan menggunakan objek model comparison dalam SAS Drive.



Gambar 4.9 Visualisasi Model Comparison

Dalam model ini, peneliti membandingkan 2 model algoritma yaitu Forest dan Gradient Boosting. Pada model Forest dihasilkan Average Squared Error sebesar 21% sedangkan model dari Gradient Boosting menghasilkan Average Squared Error sebesar 18%. Dapat dilihat bahwa Gradient Boosting memiliki model yang lebih akurat dibandingkan dengan model Random Forest.

## V. KESIMPULAN

Kesimpulan yang dapat ditarik dari penelitian ini adalah peneliti menemukan model terbaik dari hasil analisis yang dilakukan. Setelah dibandingkan kedua

model yang sudah dicoba Random Forest dan Gradient Boosting, didapatkan bahwa Gradient Boosting menjadi model terbaik yang dimana Average Squared Error yang dihasilkan lebih rendah dibanding dengan model Random Forest. Model Gradient Boost sendiri menghasilkan ASE sebesar 18% dan model Random Forest menghasilkan ASE sebesar 21%.

Kemudian dengan didapatkannya model terbaik, maka ini menjadi jawaban dari permasalahan yang dihadapi oleh RidersRepublic yang dimana perusahaan menjadi mengetahui bahwa produk yang memiliki nilai jual tinggi berada di negara Amerika Serikat. Banyak factor yang juga menjadi pengaruh dari peningkatan dan penurunan dari harga jual, seperti stok yang terbatas, harga yang mungkin kurang pas, dan masih banyak lagi.

Sekian yang dapat disimpulkan terkait dengan analisis yang dilakukan oleh peneliti, diharapkan penelitian ini dapat berguna dan membantu penelitian lainnya di kemudian hari.

## REFERENSI

- [1] Soliha, E. (2008). Analisis industri ritel di Indonesia. ANALISIS INDUSTRI RITEL DI INDONESIA. <https://media.neliti.com/media/publications/24251-ID-analisis-industri-ritel-di-indonesia.pdf>
- [2] Dixon, M. (2021). What is SAS Viya and should you invest in it?. Selerity. <https://seleritysas.com/blog/2019/06/19/what-is-sas-viya-and-should-you-invest-in-it/>
- [3] Schonlau, M., & Zou, R. Y. (2020). The Random Forest Algorithm for Statistical Learning. The random forest algorithm for statistical learning. <https://journals.sagepub.com/doi/pdf/10.1177/1536867X20909688>
- [4] Korstanje, J. (1970). Gradient boosting with XGBoost and lightgbm. SpringerLink. [https://link.springer.com/chapter/10.1007/978-1-4842-7150-6\\_15](https://link.springer.com/chapter/10.1007/978-1-4842-7150-6_15)
- [5] ws, Y. (2021). Mengenal Exploratory Data Analysis. MENGENAL EXPLORATORY DATA ANALYSIS – eksplorasidata.mipa.ugm.ac.id. <https://eksplorasidata.mipa.ugm.ac.id/2021/08/16/mengenal-exploratory-data-analysis/>
- [6] Nuha, H. (2023). Mean squared error (MSE) Dan Penggunaannya. SSRN. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4420880](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4420880)
- [7] Levine, D. M., & Stephan, D. M. (2011). Teaching Introductory Business Statistics Using the DCOVA Framework. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4609.2011.00316.xss>
- [8] <https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe>