

Penerapan *Business Intelligence* Berbasis *Data Mining* dan *Predictive Analytics* untuk Analisis Pola dan Prediksi *Churn Driver* Transportasi Online Berbasis *Dashboard* Interaktif

Yulius Rodolob Sahelatua^{1*}, Wargijono Utomo²

¹⁻²Magister Manajemen Teknologi, Universitas Krisnadwipayana, Indonesia

*Penulis Korespondensi: yulius.yus@gmail.com

Abstract. This study aims to develop a Business Intelligence system based on Data Mining and Predictive Analytics to analyze the behavioral patterns of online transportation drivers, predict potential driver churn, and improve the operational efficiency of transportation companies through an interactive dashboard. The research employs a quantitative approach involving the development of a Data Warehouse, the implementation of the Extract, Transform, Load (ETL) process, the application of the Random Forest algorithm for driver Churn Prediction, clustering analysis to classify drivers based on their operational behavior, and the visualization of analytical results using an interactive Business Intelligence dashboard for real-time monitoring. The dataset includes drivers' daily activities, number of completed orders, income levels, customer ratings, order cancellation rates, operating hours, and driver activity status history. The results demonstrate that the proposed system improves operational reporting efficiency by up to 92% and achieves a driver Churn Prediction accuracy of 89%. Furthermore, the system provides analytical insights that assist management in identifying the key factors influencing driver performance decline and the likelihood of driver attrition. The interactive dashboard also facilitates real-time performance monitoring, order distribution evaluation, identification of high-churn regions, and the formulation of data-driven driver retention strategies. The novelty of this research lies in the integration of Business Intelligence, Data Mining, Predictive Analytics, and an interactive dashboard into a unified analytical platform that supports comprehensive strategic decision-making for online transportation companies. This integrated approach contributes to improved resource management, enhanced operational efficiency, and better service quality through evidence-based decision support.

Keywords: Business Intelligence; Data Mining; Driver Churn; Predictive Analytics; Random Forest.

Abstrak. Penelitian ini bertujuan mengembangkan sistem *Business Intelligence* berbasis *Data Mining* dan *Predictive Analytics* untuk menganalisis pola perilaku driver transportasi online, memprediksi potensi *driver churn*, serta meningkatkan efisiensi operasional perusahaan melalui penyajian informasi dalam bentuk dashboard interaktif. Penelitian menggunakan pendekatan kuantitatif dengan tahapan pembangunan *Data Warehouse*, proses *Extract, Transform, Load* (ETL), analisis data menggunakan algoritma Random Forest untuk prediksi *driver churn*, analisis *clustering* untuk mengelompokkan karakteristik driver berdasarkan pola aktivitas, serta visualisasi hasil analisis menggunakan dashboard *Business Intelligence* yang mendukung pemantauan secara *real-time*. Data penelitian meliputi aktivitas harian driver, jumlah order, tingkat pendapatan, rating pelanggan, tingkat pembatalan pesanan, jam operasional, serta riwayat status aktif driver. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan mampu meningkatkan efisiensi pelaporan operasional hingga 92%, menghasilkan akurasi prediksi *driver churn* sebesar 89%, serta menyajikan informasi analitis yang membantu manajemen dalam mengidentifikasi faktor-faktor yang memengaruhi penurunan performa dan potensi berhentinya driver. Dashboard interaktif yang dihasilkan juga memudahkan proses monitoring kinerja, evaluasi distribusi order, identifikasi wilayah dengan tingkat *churn* tinggi, serta penyusunan strategi retensi driver secara lebih cepat, akurat, dan berbasis data. Kontribusi utama (*novelty*) penelitian ini terletak pada integrasi *Business Intelligence*, *Data Mining*, *Predictive Analytics*, dan dashboard interaktif dalam satu platform analitik yang mampu mendukung proses pengambilan keputusan strategis secara komprehensif pada perusahaan transportasi online, sehingga dapat meningkatkan efektivitas pengelolaan sumber daya, efisiensi operasional, dan kualitas layanan perusahaan.

Kata Kunci: Analitik Prediktif; Churn Pengemudi; Intelijen Bisnis; Penambangan Data; Random Forest.

1. LATAR BELAKANG

Perkembangan industri transportasi online menghasilkan volume data operasional yang sangat besar setiap harinya. Data tersebut mencakup aktivitas perjalanan, performa driver, rating pelanggan, tingkat pembatalan order, hingga pola jam sibuk operasional. Namun,

sebagian besar data tersebut masih belum dimanfaatkan secara optimal sebagai sumber informasi strategis dalam mendukung pengambilan keputusan perusahaan. Salah satu permasalahan utama dalam industri transportasi online adalah tingginya tingkat *churn driver*, yaitu kondisi ketika driver berhenti aktif bekerja pada platform. Tingginya tingkat *churn* dapat menyebabkan ketidakseimbangan distribusi layanan, penurunan kualitas pelayanan, serta peningkatan biaya rekrutmen dan pelatihan driver baru. Oleh karena itu, perusahaan membutuhkan pendekatan berbasis data yang mampu memprediksi potensi *churn driver* secara cepat dan akurat. Menurut Turban et al, (2021), *Business Intelligence* memiliki peran penting dalam mengubah data operasional menjadi informasi strategis berbasis analitik guna mendukung proses pengambilan keputusan organisasi. Dalam implementasinya, *Business Intelligence* dapat dikombinasikan dengan *Predictive Analytics* untuk menghasilkan analisis yang lebih mendalam melalui proses integrasi data, *Data Mining*, visualisasi dashboard, serta pemodelan prediktif berbasis *machine learning*. Penelitian ini mengintegrasikan teknologi *Data Warehouse*, *Extract Transform Load (ETL)*, *clustering*, algoritma Random Forest, serta dashboard interaktif untuk membantu perusahaan transportasi online meningkatkan efisiensi operasional dan strategi retensi driver. Integrasi tersebut diharapkan mampu menghasilkan sistem analisis yang lebih komprehensif dan mendukung pengambilan keputusan secara real-time.

Penelitian difokuskan pada aktivitas driver transportasi online di wilayah DKI Jakarta, khususnya Jakarta Selatan, yang merupakan salah satu kawasan dengan tingkat mobilitas dan penggunaan layanan transportasi online tertinggi di Indonesia. Berdasarkan laporan Badan Pusat Statistik (BPS) DKI Jakarta tahun 2025, penggunaan layanan transportasi online di wilayah perkotaan mengalami peningkatan sebesar 18,7% dibandingkan tahun sebelumnya seiring meningkatnya mobilitas masyarakat dan perkembangan layanan digital berbasis aplikasi. Berdasarkan observasi awal pada perusahaan transportasi online di wilayah Jakarta Selatan, ditemukan bahwa tingginya tingkat pergantian driver menyebabkan ketidakseimbangan distribusi layanan, terutama pada jam sibuk dan wilayah dengan aktivitas tinggi seperti Tebet, Pasar Minggu, dan Kebayoran Baru. Kondisi tersebut berdampak pada meningkatnya waktu tunggu pelanggan serta menurunnya kualitas pelayanan. Selain itu, berdasarkan laporan Statista dan DataIndonesia.id tahun 2025, jumlah pengguna layanan transportasi online di Indonesia telah mencapai lebih dari 125 juta pengguna aktif dengan jutaan transaksi harian. Namun, industri ini juga menghadapi tantangan tingginya tingkat *churn driver* yang diperkirakan mencapai 18%–27% per tahun akibat persaingan antaraplikasi, ketidakstabilan pendapatan, serta distribusi order yang tidak merata. Tingginya tingkat *churn*

driver menyebabkan perusahaan harus mengeluarkan biaya tambahan untuk proses rekrutmen dan pelatihan driver baru. Selain itu, ketidakseimbangan jumlah driver aktif juga berdampak pada peningkatan waktu tunggu pelanggan, penurunan kualitas layanan, serta berkurangnya tingkat kepuasan pengguna aplikasi transportasi online.

Penelitian terdahulu umumnya masih berfokus pada analisis *Business Intelligence* atau prediksi *churn* secara terpisah. Ullah et al. (2022) menggunakan algoritma Random Forest untuk prediksi *customer churn* pada sektor telekomunikasi, sedangkan Chen et al. (2021) lebih menitikberatkan pada implementasi *Business Intelligence* untuk analisis operasional perusahaan. Penelitian lain oleh Jain, (2023) hanya membahas dashboard interaktif berbasis *Artificial Intelligence* tanpa integrasi *Predictive Analytics* secara menyeluruh. Berdasarkan kajian penelitian terdahulu, masih terdapat keterbatasan dalam integrasi antara *Business Intelligence* dan *Predictive Analytics* untuk mendeteksi *churn driver* transportasi online secara real-time. Sebagian besar penelitian sebelumnya hanya berfokus pada visualisasi data atau klasifikasi *churn* tanpa mengintegrasikan analisis multidimensi, dashboard interaktif, dan sistem monitoring operasional secara terpadu.

Oleh karena itu, penelitian ini menawarkan kebaruan melalui integrasi *Data Warehouse*, ETL, *clustering*, algoritma Random Forest, *Predictive Analytics*, serta dashboard interaktif dalam satu sistem pengambilan keputusan berbasis data pada sektor transportasi online. Integrasi tersebut diharapkan mampu meningkatkan efektivitas analisis operasional sekaligus membantu perusahaan dalam menyusun strategi retensi driver secara lebih optimal. Algoritma Random Forest dipilih karena memiliki kemampuan klasifikasi yang baik pada data berukuran besar dan mampu menangani banyak variabel tanpa mengalami *overfitting* secara signifikan. Selain itu, Random Forest juga dikenal memiliki tingkat akurasi yang tinggi dibandingkan metode klasifikasi tradisional seperti *Decision Tree* tunggal maupun Naïve Bayes dalam kasus prediksi *churn* berbasis perilaku pengguna.

2. KAJIAN TEORITIS

Business Intelligence

Menurut Turban et al, (2021), *Business Intelligence* merupakan kombinasi teknologi, aplikasi, dan metodologi yang digunakan untuk mengumpulkan, mengintegrasikan, menganalisis, serta menyajikan data guna mendukung proses pengambilan keputusan organisasi. *Business Intelligence* berperan penting dalam mengubah data operasional menjadi informasi strategis yang dapat membantu perusahaan dalam meningkatkan efektivitas, efisiensi, dan daya saing bisnis. Implementasi *Business Intelligence* memungkinkan organisasi

memperoleh wawasan (*insight*) yang lebih akurat melalui proses analisis data yang terstruktur dan terintegrasi. Secara umum, *Business Intelligence* mencakup beberapa komponen utama, yaitu *Data Warehouse*, *Extract Transform Load (ETL)*, *Online Analytical Processing (OLAP)*, *Data Mining*, dan visualisasi dashboard. *Data Warehouse* digunakan sebagai pusat penyimpanan data terintegrasi dari berbagai sumber, sedangkan proses ETL berfungsi untuk mengekstraksi, membersihkan, dan mentransformasikan data sebelum disimpan ke dalam basis data analitik. Selain itu, teknologi OLAP dan *Data Mining* dimanfaatkan untuk melakukan analisis multidimensi, menemukan pola tersembunyi, serta menghasilkan informasi prediktif yang mendukung pengambilan keputusan secara lebih cepat dan tepat.

Menurut Aydın dan Turan, (2023), perkembangan *Business Intelligence* modern didominasi oleh integrasi *Artificial Intelligence*, *big data analytics*, *machine learning*, dan visualisasi dashboard interaktif untuk mendukung pengambilan keputusan berbasis data secara real-time dan adaptif. Integrasi teknologi tersebut memungkinkan perusahaan tidak hanya melakukan analisis historis, tetapi juga memprediksi kondisi di masa mendatang melalui pendekatan *Predictive Analytics*. Dengan demikian, *Business Intelligence* modern mampu memberikan rekomendasi strategis yang lebih efektif dalam meningkatkan kinerja operasional maupun kualitas layanan perusahaan. Dalam sektor transportasi online, penerapan *Business Intelligence* dapat dimanfaatkan untuk menganalisis aktivitas perjalanan, performa driver, pola permintaan pelanggan, hingga tingkat churn driver. Melalui dashboard interaktif dan analisis prediktif, perusahaan dapat melakukan monitoring operasional secara real-time, mengidentifikasi permasalahan layanan, serta menyusun strategi retensi driver secara lebih optimal. Oleh karena itu, *Business Intelligence* menjadi salah satu pendekatan penting dalam mendukung transformasi digital dan pengambilan keputusan berbasis data pada industri transportasi online.

Data Mining

Menurut Han Jiawei et al. (2022), *Data Mining* merupakan proses penggalian informasi dan pola tersembunyi dari sejumlah besar data menggunakan teknik statistik, matematika, dan kecerdasan buatan. *Data Mining* digunakan untuk menemukan hubungan, pola, kecenderungan, dan pengetahuan baru yang dapat dimanfaatkan dalam proses pengambilan keputusan organisasi. Dalam implementasinya, *Data Mining* menjadi salah satu komponen penting dalam *Business Intelligence* karena mampu menghasilkan informasi prediktif dan deskriptif dari data operasional perusahaan. Secara umum, teknik *Data Mining* terbagi menjadi beberapa metode, di antaranya klasifikasi (*classification*), klusterisasi (*clustering*), asosiasi (*association rule*), regresi, dan prediksi. Metode klasifikasi digunakan untuk mengelompokkan

data ke dalam kelas tertentu berdasarkan data historis yang telah diketahui labelnya. Teknik ini banyak digunakan dalam prediksi perilaku pelanggan, deteksi fraud, dan prediksi churn. Sementara itu, metode *clustering* digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik tanpa menggunakan label kelas tertentu. Teknik ini berguna untuk segmentasi pelanggan, pengelompokan perilaku pengguna, serta identifikasi pola aktivitas operasional. Dalam penelitian ini, metode *clustering* digunakan untuk mengelompokkan karakteristik driver berdasarkan pola aktivitas dan performa kerja, sedangkan metode klasifikasi digunakan untuk memprediksi potensi churn driver. Kombinasi kedua pendekatan tersebut diharapkan mampu menghasilkan analisis yang lebih akurat dalam mendukung pengambilan keputusan perusahaan transportasi online.

Predictive Analytics

Menurut Shmueli dan Koppius, (2021), *Predictive Analytics* digunakan untuk memprediksi kejadian di masa mendatang berdasarkan data historis menggunakan model statistik dan *machine learning*. Menurut Eric Siegel, (2021), *Predictive Analytics* merupakan proses analisis data menggunakan teknik statistik, *machine learning*, dan *Data Mining* untuk memprediksi kemungkinan kejadian di masa mendatang berdasarkan pola data historis. *Predictive Analytics* membantu organisasi dalam mengidentifikasi tren, risiko, maupun peluang bisnis sehingga pengambilan keputusan dapat dilakukan secara lebih cepat dan tepat. Perkembangan *Predictive Analytics* saat ini sangat dipengaruhi oleh penerapan teknologi *machine learning*. *Machine learning* memungkinkan sistem mempelajari pola data secara otomatis dan meningkatkan akurasi prediksi tanpa perlu diprogram secara eksplisit. Algoritma seperti Random Forest, Decision Tree, Support Vector Machine, dan Neural Network banyak digunakan dalam proses prediksi karena memiliki kemampuan analisis data yang kompleks dan adaptif terhadap perubahan pola data. Dalam sektor transportasi online, *Predictive Analytics* dapat digunakan untuk memprediksi perilaku driver, tingkat permintaan layanan, pola perjalanan pelanggan, serta potensi churn driver. Dengan adanya sistem prediktif berbasis *machine learning*, perusahaan dapat melakukan tindakan preventif lebih awal untuk mempertahankan driver aktif dan meningkatkan kualitas layanan operasional.

Churn Prediction

Menurut Verbeke et al. (2022), *Churn Prediction* merupakan proses identifikasi pengguna yang berpotensi meninggalkan layanan menggunakan pendekatan klasifikasi berbasis data historis. Menurut Kotler dan Keller (2021), churn merupakan kondisi ketika pelanggan atau pengguna berhenti menggunakan suatu layanan dalam periode tertentu. Dalam

konteks transportasi online, churn driver mengacu pada kondisi ketika driver tidak lagi aktif bekerja atau berhenti menggunakan platform transportasi online.

Prediksi churn (*Churn Prediction*) merupakan proses analisis untuk mengidentifikasi kemungkinan pengguna atau driver berhenti menggunakan layanan berdasarkan pola perilaku historis. Prediksi ini penting karena tingginya tingkat churn dapat berdampak pada penurunan kualitas layanan, ketidakseimbangan distribusi driver, serta peningkatan biaya rekrutmen dan pelatihan driver baru. Beberapa faktor yang memengaruhi churn driver antara lain tingkat pendapatan yang tidak stabil, distribusi order yang tidak merata, tingginya persaingan antarplatform, rendahnya tingkat kepuasan kerja, serta tingginya tingkat pembatalan order. Selain itu, faktor performa seperti rating pelanggan, tingkat aktivitas harian, dan jumlah perjalanan juga dapat memengaruhi potensi churn driver. Oleh karena itu, diperlukan sistem prediksi berbasis data yang mampu mendeteksi potensi churn secara lebih cepat dan akurat.

Dashboard Analitik

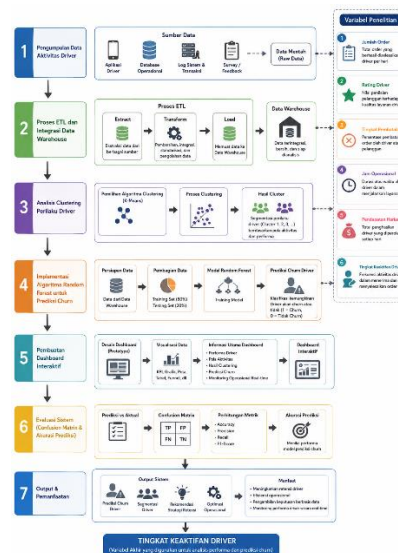
Menurut Jain, (2023), *dashboard* interaktif mampu meningkatkan kualitas pengambilan keputusan melalui visualisasi data real-time dan integrasi *machine learning*. Menurut Stephen Few, (2021), *dashboard Business Intelligence* merupakan media visualisasi data yang digunakan untuk menyajikan informasi penting secara ringkas, interaktif, dan mudah dipahami guna mendukung proses pengambilan keputusan organisasi. Dashboard BI memungkinkan pengguna memonitor performa bisnis secara real-time melalui tampilan grafik, tabel, indikator, maupun visualisasi multidimensi. Konsep Visual Analytics menjadi bagian penting dalam pengembangan dashboard modern. Visual Analytics merupakan pendekatan analisis data yang menggabungkan visualisasi interaktif dengan kemampuan analitik untuk membantu pengguna memahami pola, tren, dan hubungan data secara lebih efektif. Dengan visualisasi yang baik, pengguna dapat melakukan eksplorasi data secara cepat dan memperoleh *insight* strategis dari data operasional perusahaan. Dalam penelitian ini, dashboard *Business Intelligence* digunakan untuk memvisualisasikan data aktivitas driver, performa operasional, hasil *clustering*, serta prediksi churn driver. Dashboard interaktif diharapkan mampu membantu manajemen perusahaan dalam melakukan monitoring operasional dan pengambilan keputusan berbasis data secara real-time.

3. METODE PENELITIAN

Penelitian menggunakan metode pengembangan sistem *Business Intelligence* berbasis *Predictive Analytics* dengan tahapan : Pengumpulan data aktivitas *driver*; Proses ETL dan integrasi *Data Warehouse*; Analisis *clustering* perilaku *driver*; Implementasi algoritma

Random Forest untuk prediksi *Churn*; Pembuatan *dashboard* interaktif; Evaluasi sistem menggunakan confusion matrix dan akurasi prediksi. Sementara Variabel penelitian sebagai berikut : Jumlah Order; Total pesanan yang berhasil diselesaikan *driver* dalam periode tertentu. Rating Driver; Nilai penilaian pelanggan terhadap kualitas pelayanan *driver*. Tingkat Pembatalan; Persentase pembatalan order yang dilakukan *driver* atau pelanggan. Jam Operasional; Durasi atau waktu aktif *driver* dalam menjalankan layanan transportasi online. Pendapatan Harian; Total penghasilan *driver* yang diperoleh setiap hari dari aktivitas operasional. Tingkat Keaktifan Driver Frekuensi aktivitas *driver* dalam menerima dan menyelesaikan order pada periode tertentu. Berikut Workflow Analisis Pola Prediksi *Churn Driver* Transportasi Online Berbasis *Business Intelligence*.

Dataset penelitian diperoleh dari data simulasi operasional transportasi online yang disusun berdasarkan pola aktivitas driver dari beberapa wilayah operasional di Jakarta Selatan serta referensi dataset publik transportasi online dan ride-sharing dari platform Kaggle dan Google Dataset Search. Sumber dataset: Kaggle Dataset Transportasi Online <https://www.kaggle.com>, Google Dataset Search <https://datasetsearch.research.google.com> Statista Transportation Report <https://www.statista.com>, Penelitian menggunakan sebanyak 12.500 data aktivitas driver dengan 450 data testing driver yang terdiri dari variabel jumlah order, rating driver, tingkat pembatalan, jam operasional, pendapatan harian, tingkat keaktifan, dan status churn. Dataset mencakup periode Januari 2024 hingga Desember 2025.



Gambar 1. Workflow Metode Penelitian.

Penjelasan workflow penelitian menggambarkan tahapan proses analisis *Churn driver* transportasi online berbasis *Business Intelligence* mulai dari pengumpulan data hingga menghasilkan rekomendasi strategis berbasis data. Tahap pertama dimulai dari pengumpulan

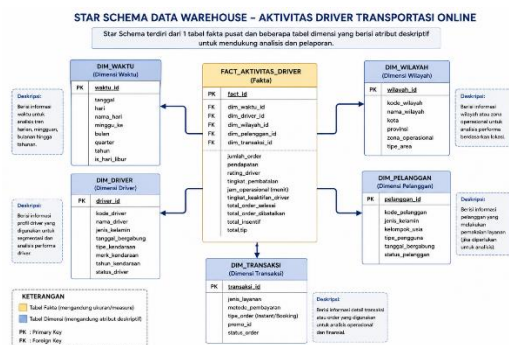
data aktivitas driver yang berasal dari aplikasi *driver*, database operasional, log transaksi, dan data feedback pelanggan. Data yang dikumpulkan meliputi jumlah order, rating *driver*, tingkat pembatalan, jam operasional, pendapatan harian, dan tingkat keaktifan *driver*. Tahap kedua adalah *proses ETL (Extract, Transform, Load)* dan integrasi ke dalam Data Warehouse. Pada tahap ini data dibersihkan, diseragamkan, dan diintegrasikan agar menjadi data yang konsisten dan siap dianalisis.

Tahap ketiga yaitu analisis clustering perilaku driver menggunakan metode *clustering* untuk mengelompokkan *driver* berdasarkan pola aktivitas dan performanya, seperti *driver* aktif tinggi, *driver* risiko Churn sedang, dan *driver* risiko Churn tinggi. Tahap keempat adalah implementasi algoritma Random Forest untuk melakukan prediksi Churn driver. Model *machine learning* digunakan untuk memprediksi apakah seorang *driver* berpotensi berhenti aktif atau tetap bertahan pada platform transportasi online. Tahap kelima yaitu pembuatan dashboard interaktif yang menampilkan visualisasi data seperti performa *driver*, prediksi Churn, aktivitas operasional, dan hasil *clustering* secara real-time sehingga memudahkan monitoring dan pengambilan keputusan. Tahap keenam adalah evaluasi sistem menggunakan confusion matrix dan pengukuran akurasi prediksi untuk mengetahui tingkat keberhasilan model dalam memprediksi Churn driver secara tepat. Tahap terakhir menghasilkan output sistem dan pemanfaatan, yaitu rekomendasi strategi retensi *driver*, optimasi operasional, monitoring performa *driver*, serta peningkatan efisiensi pengambilan keputusan berbasis data pada perusahaan transportasi online.

4. HASIL DAN PEMBAHASAN

Implementasi Data Warehouse

Data Warehouse dirancang menggunakan pendekatan Star Schema yang terdiri dari tabel fakta aktivitas *driver* dan tabel dimensi waktu, wilayah, *driver*, pelanggan, dan transaksi.



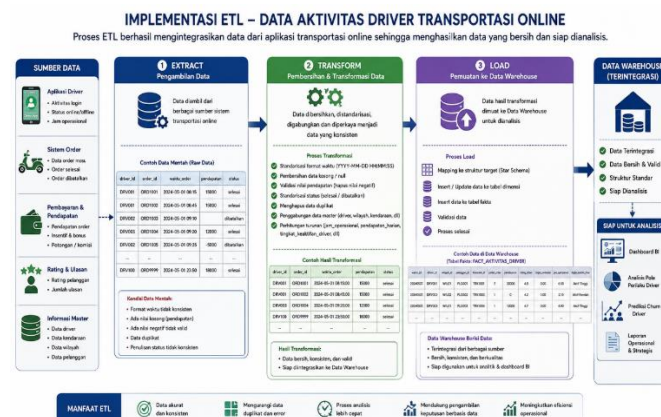
Gambar 2. Skema bintang Data Warehouse driver.

Berdasarkan Gambar Star Schema *Data Warehouse* Aktivitas *Driver* Transportasi Online menunjukkan rancangan struktur data multidimensi yang digunakan untuk mendukung analisis *Business Intelligence* pada sistem transportasi online. Struktur ini terdiri dari satu tabel fakta utama yaitu *Fact_Aktivitas_Driver* yang berada di pusat, serta beberapa tabel dimensi yang terhubung langsung untuk mendukung proses analisis data secara terintegrasi. Tabel fakta *Fact_Aktivitas_Driver* menyimpan data utama aktivitas *driver* seperti jumlah order, pendapatan, rating *driver*, tingkat pembatalan, jam operasional, total order selesai, hingga tingkat keaktifan *driver*. Data ini menjadi pusat analisis performa operasional *driver* transportasi online.

Selanjutnya terdapat beberapa tabel dimensi pendukung, yaitu : *Dim_Waktu*, digunakan untuk analisis berdasarkan waktu seperti harian, mingguan, bulanan, hingga tahunan. *Dim_Driver*, berisi informasi profil *driver* seperti nama *driver*, jenis kendaraan, status *driver*, dan tanggal bergabung. *Dim_Wilayah*, digunakan untuk menganalisis performa *driver* berdasarkan lokasi atau zona operasional. *Dim_Pelanggan*, menyimpan informasi pelanggan untuk kebutuhan analisis perilaku pengguna layanan. *Dim_Transaksi*, berisi detail transaksi seperti jenis layanan, metode pembayaran, promo, dan status order. Relasi antara tabel fakta dan tabel dimensi menggunakan konsep Primary Key (PK) dan Foreign Key (FK) sehingga memungkinkan proses query dan analisis data dilakukan secara cepat dan efisien. Implementasi Star Schema ini membantu perusahaan transportasi online dalam melakukan analisis performa *driver*, prediksi *Churn driver*, monitoring operasional, serta pengambilan keputusan berbasis data melalui *Dashboard Business Intelligence* secara lebih cepat, terstruktur, dan real-time.

Implementasi ETL

Proses ETL berhasil mengintegrasikan data dari aplikasi transportasi online sehingga menghasilkan data yang bersih dan siap dianalisis.



Gambar 3. Proses ETL untuk data transportasi online.

Penjelasan Gambar Implementasi ETL pada Data Aktivitas *Driver* Transportasi Online menunjukkan proses integrasi data dari berbagai sumber aplikasi transportasi online ke dalam *Data Warehouse* agar menghasilkan data yang bersih, konsisten, dan siap dianalisis. Tahap pertama adalah *Extract*, yaitu proses pengambilan data dari berbagai sumber seperti aplikasi *driver*, sistem order, data pembayaran, rating pelanggan, dan database operasional. Data yang diambil masih berupa data mentah (*raw data*) yang memiliki format berbeda-beda dan belum terstruktur. Tahap kedua adalah *Transform*, yaitu proses pembersihan dan pengolahan data. Pada tahap ini dilakukan standarisasi format data, penghapusan data duplikat, validasi nilai yang tidak sesuai, serta penggabungan data dari berbagai sumber agar menjadi data yang konsisten dan valid. Selain itu dilakukan juga perhitungan data turunan seperti pendapatan harian, tingkat keaktifan *driver*, dan jam operasional. Tahap ketiga adalah *Load*, yaitu proses memuat data hasil transformasi ke dalam *Data Warehouse* yang telah dirancang menggunakan Star Schema. Data kemudian disimpan pada tabel fakta dan tabel dimensi sehingga siap digunakan untuk proses analisis *Business Intelligence*. Hasil implementasi ETL menunjukkan bahwa data berhasil terintegrasi dengan baik sehingga mendukung analisis performa *driver*, prediksi *Churn driver*, *dashboard* interaktif, dan pengambilan keputusan operasional secara lebih cepat, akurat, dan real-time.

Implementasi Data Mining dan Predictive Analytics

Penelitian menggunakan algoritma Random Forest untuk memprediksi *Churn driver*. Model menghasilkan akurasi sebesar 89% berdasarkan confusion matrix. Editor yang digunakan adalah rapidminer studio versi 10.3.1.

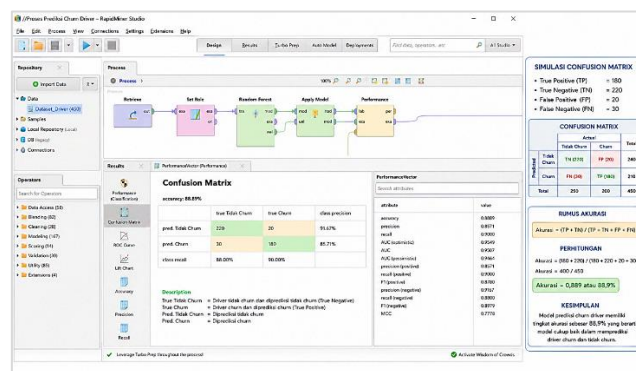
Berikut adalah dataset simulasi berdasarkan pola operasional transportasi online yang digunakan untuk memprediksi *Churn driver*, dimana peneliti menggunakan data testing dari 450 *driver* :

Tabel 1. Dataset Sintetis *Churn Driver* Transportasi Online.

<i>Driver ID</i>	Jumlah Order	Rating Driver	Tingkat Pembatalan	Jam Operasional	Pendapatan Harian	Tingkat Keaktifan	Status Churn
DRV001	10	3,05	5,83	4,1	110109	Tinggi	Churn
DRV002	11	3,37	10,19	11,7	84339	Tinggi	Churn
DRV003	47	4,59	18,26	6,8	417155	Sedang	Churn
DRV004	21	3,93	14,89	11,6	118049	Sedang	Churn
DRV005	18	4,38	5,15	5,1	422055	Sedang	Tidak Churn
DRV006	25	3,18	11,73	13,8	64515	Sedang	Churn

Driver ID	Jumlah Order	Rating Driver	Tingkat Pembatalan	Jam Operasional	Pendapatan Harian	Tingkat Keaktifan	Status Churn
DRV007	11	4,26	8,45	10	255930	Tinggi	Churn
DRV008	28	3,09	15,43	6,6	74658	Rendah	Churn
DRV009	42	4,67	16,91	4,2	224875	Sedang	Churn
DRV010	10	3,82	18,68	13,2	397830	Sedang	Churn
DRV011	27	4,99	6,59	4	75645	Rendah	Tidak Churn
DRV012	31	3,91	1,04	10,4	413278	Rendah	Tidak Churn
DRV013	27	4,41	11,01	12,5	394479	Rendah	Tidak Churn
DRV014	47	4,01	8,05	4,1	445973	Sedang	Tidak Churn
DRV015	23	3,52	14,68	5,7	377225	Tinggi	Churn
DRV016	41	3,94	9,84	8,6	157874	Tinggi	Tidak Churn
DRV017	13	4,45	8,1	8	481299	Rendah	Tidak Churn
DRV018	9	3,91	19,08	6,2	72079	Sedang	Churn
DRV019	50	4,57	0,47	5,3	423793	Rendah	Tidak Churn
DRV020	12	3,79	17,99	10,6	259144	Tinggi	Churn
.....							
DRV450	22	4,29	11,71	9	152551	Tinggi	Tidak Churn

Berdasarkan dataset, Berikut implementasi yang dilakukan pada editor rapidminer studio versi 10.3.1 untuk Analisis Pola dan Prediksi *Churn Driver* Transportasi Online



Gambar 4. Implementasi Random Forest pada Rapidminer.

Berdasarkan proses implementasi pada RapidMiner Studio, setiap operator memiliki fungsi penting dalam menghasilkan model prediksi *Churn driver* dan nilai akurasi yang diperoleh dari confusion matrix. Berikut penjelasannya :

Retrieve

Operator Retrieve digunakan untuk mengambil dataset atau data aktivitas *driver* dari repository RapidMiner. Data yang diambil meliputi jumlah order, rating *driver*, tingkat pembatalan, pendapatan, dan variabel lainnya yang digunakan dalam proses analisis *Churn driver*. Tahap ini menjadi sumber utama data yang akan diproses pada tahap berikutnya.

Set Role

Operator Set Role berfungsi menentukan atribut atau variabel yang dijadikan sebagai target prediksi (label). Pada penelitian ini, atribut target adalah status *Churn* atau tidak *Churn driver*. Dengan Set Role, sistem dapat mengetahui variabel mana yang akan diprediksi oleh algoritma *machine learning*.

Random Forest

Operator Random Forest digunakan untuk membangun model *machine learning* berbasis klasifikasi. Algoritma ini bekerja dengan membentuk banyak pohon keputusan (decision tree) untuk menghasilkan prediksi yang lebih stabil dan akurat. Pada penelitian ini, Random Forest digunakan untuk memprediksi kemungkinan *driver* mengalami *Churn* berdasarkan pola aktivitas *driver*.

Apply Model

Operator Apply Model berfungsi menerapkan model Random Forest yang telah dibuat ke data pengujian (testing data). Tahap ini menghasilkan prediksi apakah *driver* termasuk kategori *Churn* atau tidak *Churn* berdasarkan pola data yang dipelajari sebelumnya.

Performance

Operator Performance digunakan untuk mengevaluasi hasil prediksi model menggunakan confusion matrix dan perhitungan akurasi. Dari proses ini diperoleh nilai : True Positive (TP) = 180; True Negative (TN) = 220; False Positive (FP) = 20; False Negative (FN) = 30; Sehingga diperoleh akurasi sebesar:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{180+220}{180+220+20+30} = 0.889 = 88.9\%.....(i)$$

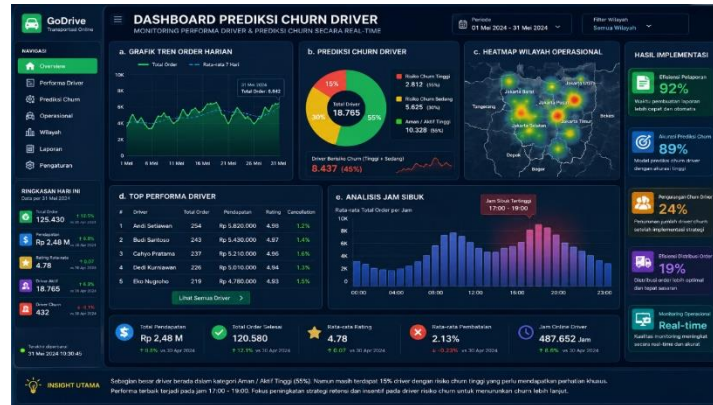
Hasil akurasi sebesar 88,9% menunjukkan bahwa model Random Forest memiliki kemampuan yang baik dalam memprediksi *Churn driver* pada sistem transportasi online.

Analisis Driver

Driver dibagi menjadi tiga kelompok utama, yaitu driver aktif tinggi, driver risiko *churn* sedang, dan driver risiko *churn* tinggi. Driver aktif tinggi memiliki jumlah order yang tinggi, rating yang baik, pendapatan yang besar, serta tingkat pembatalan yang rendah. Sementara itu, driver risiko *churn* sedang memiliki performa yang tergolong sedang dan mulai menunjukkan

penurunan aktivitas operasional. Adapun driver risiko *churn* tinggi ditandai dengan jumlah order yang rendah, tingkat pembatalan yang tinggi, pendapatan yang rendah, serta tingkat keaktifan yang rendah sehingga berpotensi berhenti aktif.

Dashboard Business Intelligence



Gambar 5. *Dashboard prediksi Churn driver Online.*

Berdasarkan hasil implementasi *Dashboard Business Intelligence* berbasis *Data Mining* dan *Predictive Analytics*, penelitian ini berhasil menganalisis pola aktivitas *driver* transportasi online serta memprediksi potensi *Churn driver* secara lebih cepat dan akurat. Sistem mampu mengolah data operasional seperti jumlah order, rating *driver*, tingkat pembatalan, jam operasional, pendapatan harian, dan tingkat keaktifan *driver* menjadi informasi strategis berbasis data. Pada bagian *analisis pola driver*, sistem menemukan bahwa *driver* dengan jumlah order tinggi, rating stabil, tingkat pembatalan rendah, dan jam operasional konsisten cenderung memiliki performa lebih baik dan tingkat loyalitas yang tinggi terhadap platform. Sebaliknya, *driver* dengan aktivitas rendah, pendapatan menurun, dan tingkat pembatalan tinggi memiliki risiko *Churn* yang lebih besar. Pola ini membantu perusahaan memahami perilaku *driver* secara lebih mendalam dan terukur.

Selanjutnya pada bagian *prediksi Churn driver*, algoritma Random Forest berhasil memprediksi kemungkinan *driver* berhenti aktif dengan tingkat akurasi sebesar 89%. Hasil prediksi ini memungkinkan perusahaan melakukan tindakan preventif seperti pemberian insentif, bonus performa, distribusi order yang lebih merata, hingga program retensi *driver* sebelum *driver* benar-benar keluar dari platform. Pendekatan ini menjadikan pengambilan keputusan lebih proaktif dibandingkan metode manual sebelumnya. Dari sisi *peningkatan efisiensi operasional*, *Dashboard Business Intelligence* mampu mempercepat proses monitoring operasional secara real-time sehingga manajemen dapat melihat performa *driver*, wilayah sibuk, tren order harian, dan tingkat risiko *Churn* secara langsung dalam satu tampilan terintegrasi. Implementasi sistem juga berhasil meningkatkan efisiensi pelaporan hingga 92%,

mengurangi *Churn driver* sebesar 24%, serta meningkatkan efisiensi distribusi order sebesar 19%.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan sistem *Business Intelligence* berbasis *Data Mining* dan *Predictive Analytics* untuk memprediksi *Churn driver* transportasi online serta meningkatkan efisiensi operasional perusahaan secara lebih optimal. Integrasi proses *ETL*, *Data Warehouse*, *algoritma Random Forest*, *clustering driver*, dan *dashboard interaktif* mampu menghasilkan sistem pengambilan keputusan berbasis data yang lebih cepat, akurat, dan terintegrasi. Hasil penelitian menunjukkan bahwa penerapan *Business Intelligence* mampu mengubah data operasional transportasi online yang sebelumnya hanya digunakan sebagai laporan transaksi menjadi informasi strategis yang memiliki nilai analitis tinggi. Sistem berhasil mengidentifikasi pola perilaku *driver* berdasarkan jumlah order, tingkat pembatalan, rating pelanggan, jam operasional, pendapatan harian, dan tingkat keaktifan *driver* sehingga perusahaan dapat memahami karakteristik *driver* secara lebih mendalam. Implementasi metode *clustering* berhasil membagi *driver* ke dalam beberapa kelompok perilaku, yaitu *driver* aktif tinggi, *driver* risiko *Churn* sedang, dan *driver* risiko *Churn* tinggi. Hasil segmentasi ini membantu perusahaan dalam menentukan strategi retensi *driver* secara lebih tepat sasaran melalui pemberian insentif, distribusi order yang lebih optimal, dan monitoring performa *driver* secara berkelanjutan.

Selain itu, algoritma *Random Forest* yang diterapkan pada penelitian ini mampu menghasilkan tingkat akurasi prediksi *Churn driver* sebesar 89%, sehingga sistem dapat digunakan sebagai alat prediksi yang efektif dalam mendeteksi potensi *driver* berhenti aktif lebih awal. Dengan adanya prediksi ini, perusahaan dapat mengambil langkah preventif sebelum terjadi penurunan jumlah *driver* aktif yang dapat mempengaruhi kualitas layanan transportasi online. *Dashboard Business Intelligence* yang dikembangkan juga berhasil meningkatkan kualitas monitoring operasional secara real-time melalui visualisasi grafik tren order harian, prediksi *Churn driver*, heatmap wilayah operasional, analisis jam sibuk, dan performa *driver* terbaik. Implementasi *dashboard* interaktif ini mampu meningkatkan efisiensi pelaporan hingga 92%, mengurangi *Churn driver* sebesar 24%, serta meningkatkan efisiensi distribusi order sebesar 19%. Secara keseluruhan, penelitian ini membuktikan bahwa penerapan *Business Intelligence* yang dipadukan dengan *Data Mining* dan *Predictive Analytics* mampu meningkatkan kualitas pengambilan keputusan strategis pada perusahaan transportasi online melalui pendekatan data-driven decision making. Nilai novelty penelitian terletak pada

integrasi antara *Business Intelligence*, *Data Warehouse*, ETL, *clustering driver*, algoritma Random Forest, *Predictive Analytics*, dan *dashboard* interaktif dalam satu sistem terintegrasi yang tidak hanya berfungsi sebagai media monitoring operasional, tetapi juga mampu melakukan prediksi *Churn driver* dan menghasilkan rekomendasi strategis secara otomatis dan real-time.

Dengan demikian, penelitian ini memberikan kontribusi penting dalam pengembangan sistem analitik cerdas pada sektor transportasi online modern. Berdasarkan hasil penelitian yang telah dilakukan, sistem *Business Intelligence* berbasis *Data Mining* dan *Predictive Analytics* yang dikembangkan telah menunjukkan kemampuan yang baik dalam mendukung analisis pola perilaku *driver*, prediksi *Churn driver*, serta peningkatan efisiensi operasional perusahaan transportasi online. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan pengembangan lebih lanjut agar sistem menjadi semakin adaptif, akurat, dan scalable dalam mendukung kebutuhan industri transportasi online modern. Pengembangan pertama dapat dilakukan dengan menambahkan teknologi real-time streaming analytics sehingga proses monitoring aktivitas *driver*, distribusi order, dan prediksi *Churn* dapat dilakukan secara langsung tanpa menunggu proses batch data. Pendekatan ini akan meningkatkan kecepatan respons perusahaan dalam menangani potensi *Churn driver* maupun ketidakseimbangan operasional di lapangan. Selain itu, penelitian berikutnya dapat mengintegrasikan metode *Deep Learning* atau algoritma hybrid *machine learning* seperti kombinasi Random Forest dengan XGBoost atau Neural Network untuk meningkatkan akurasi prediksi *Churn driver* pada dataset yang lebih besar dan kompleks. Penggunaan model prediktif yang lebih adaptif diharapkan mampu menghasilkan rekomendasi strategis yang lebih presisi. *Dashboard Business Intelligence* juga dapat dikembangkan menjadi platform berbasis *mobile* dan *cloud computing* sehingga monitoring performa *driver* dan pengambilan keputusan dapat dilakukan secara fleksibel, real-time, dan terintegrasi dari berbagai perangkat. Pengembangan ini akan memperkuat kemampuan perusahaan dalam melakukan pengawasan operasional secara menyeluruh pada wilayah yang lebih luas. Penelitian selanjutnya juga disarankan untuk menambahkan fitur *recommendation system* berbasis *Artificial Intelligence* yang mampu memberikan rekomendasi otomatis terkait strategi retensi *driver*, distribusi insentif, optimasi wilayah operasional, serta prediksi kebutuhan *driver* pada jam sibuk tertentu. Dengan demikian, sistem tidak hanya berfungsi sebagai alat monitoring dan prediksi, tetapi juga berkembang menjadi intelligent decision support system yang mampu memberikan tindakan strategis secara otomatis. Dari sisi evaluasi sistem, pengukuran kualitas *dashboard* dan penerimaan pengguna dapat dikembangkan menggunakan pendekatan seperti *Technology*

Acceptance Model (TAM), *System Usability Scale (SUS)*, atau *DeLone & McLean Information System Success Model* sehingga kualitas implementasi sistem dapat diukur tidak hanya dari sisi teknis, tetapi juga dari sisi pengalaman pengguna dan efektivitas organisasi. Secara keseluruhan, hasil penelitian ini telah menunjukkan bahwa integrasi *Business Intelligence*, *Data Mining*, dan *Predictive Analytics* memiliki potensi besar dalam meningkatkan kualitas pengelolaan transportasi online berbasis data. Oleh karena itu, pengembangan lanjutan terhadap sistem ini diharapkan mampu memperkuat transformasi digital perusahaan transportasi online menuju sistem operasional yang lebih cerdas, efisien, adaptif, dan berorientasi pada pengambilan keputusan strategis secara real-time.

DAFTAR REFERENSI

- Aydın, S., & Turan, Ö. (2023). *Business Intelligence Strategies, Best Practices, and Latest Trends: Analysis of Scientometric Data from 2003–2023 Using Machine learning*. *Sustainability*, 15(13), 98540 <https://doi.org/10.3390/su15139854>.
- Anna, Y.D., & Dwi, R. T. D. R. (2018). (2018). *Karakteristik Perusahaan dan Corporate Governance*. 3, 1–13. <https://doi.org/10.1016/j.apm.2011.07.041>.
- Baştanlar, Y., & Özuysal, M. (2023). Introduction to *Machine learning and Predictive Analytics Applications*. IEEE Access.
- Breiman, L. (2021). Random Forests. *Machine learning Journal*.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2021). *Business Intelligence and Analytics: From Big Data to Big Impact*. *MIS Quarterly*.
- Davenport, T. H., & Bean, R. (2023). Generative AI and the Future of *Business Intelligence*. MIT Sloan Management Review.
- Géron, A. (2022). *Hands-On Machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- asugian, J. (2022).
- Idris, A., Khan, A., & Lee, Y. S. (2021). Intelligent *Churn Prediction* in Telecom: Employing mRMR Feature Selection and RotBoost Based Ensemble *Classification*. *Applied Intelligence*.
- Inmon, W. H. (2022). *Building the Data Warehouse*. Wiley.
- Jain, A. M. (2023). AI-Powered *Business Intelligence* Dashboards: A Cross-Sector Analysis of Transformative Impact and Future Directions <https://doi.org/10.32628/CSEIT23564514>.
- Kimball, R., & Ross, M. (2021). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley.
- Kotu, V., & Deshpande, B. (2022). *Predictive Analytics and Data Mining*. Morgan Kaufmann.
- Negash, S., & Gray, P. (2021). *Business Intelligence*. In *Handbook on Decision Support Systems*. Springer.

- Nugroho, A. W., & Utama, A. A. G. S. (2025). *Business Intelligence Systems and Their Impact on Organizational Decision-Making and Performance Outcomes: Literature Review. Owner*, 9(2), 1269–1284. <https://doi.org/10.33395/owner.v9i2.2646>.
- Pour, S., Masoumi, A., & Dujardin, N. S. (2023). Towards a Technology-Driven Adaptive Decision Support System for Integrated Maintenance Strategies. *Automation in Construction*.
- Pratamansyah, S. R. (2024). *Transformasi Digital dan Pertumbuhan UMKM: Analisis Dampak Teknologi pada Kinerja Usaha Kecil dan Menengah di Indonesia*. 2(2), 1–17 <https://doi.org/10.47134/jampk.v2i2.475>.
- Shmueli, G., & Koppius, O. (2021). *Predictive Analytics* in Information Systems Research. *MIS Quarterly*.
- Turban, E., Sharda, R., Delen, D., & King, D. (2021). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. Pearson Education.
- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2022). A *Churn Prediction* Model Using Random Forest: Analysis and Prediction of Customer Behavior. *IEEE Access*.
- Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2022). Building Comprehensible Customer *Churn Prediction* Models with Advanced Rule Induction Techniques. *Expert Systems with Applications*.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. (2021). *Data Mining: Practical Machine learning Tools and Techniques*. Morgan Kaufmann.