

Pemanfaatan *Business Intelligence* berbasis OLAP dan Analisis Prediktif (*Random Forest Classifier* dan ARIMA) pada Perusahaan *Fintech P2P Lending*

Application of OLAP Based Business Intelligence and Predictive Analysis (Random Forest Classifier and ARIMA) in a Fintech P2P Lending Company

¹Kristiana Olivia, ²Frederik Samuel Papilaya*

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
^{1,2}Jl. Dr. O. Notohamidjojo, Blotongan, Salatiga, Jawa Tengah, Indonesia

*e-mail: krislv253@gmail.com

(received: 4 August 2025, revised: 12 August 2025, accepted: 13 August 2025)

Abstrak

Meningkatnya angka kredit macet di industri *financial technology* P2P *lending* menjadi tantangan serius, termasuk bagi PT. XYZ yang fokus pada pendanaan sektor UMKM. Penelitian ini menerapkan *Business Intelligence* dengan metode OLAP dan analisis prediktif pada PT. XYZ selama 2 tahun. Data yang digunakan merupakan data *record* sejak Januari 2023 hingga Desember 2024, mencakup informasi 276.320 baris peminjaman dan 267.542 baris pembayaran. Pendekatan OLAP dimanfaatkan untuk mengeksplorasi transaksi pinjaman, performa pembayaran, serta peforma keterlambatan berdasarkan berbagai dimensi, seperti akun, waktu, metode pembayaran, dan bank yang digunakan. Hasil eksplorasi ini divisualisasikan dalam tiga *dashboard* interaktif sesuai fokus analisis. Analisis prediktif dilakukan dengan menggunakan model *Random Forest Classifier* yang mengidentifikasi *loan_amount* sebagai variabel paling berpengaruh terhadap potensi keterlambatan, meskipun performa model masih terbatas dengan akurasi 61,47%, *precision* 38,33%, dan *recall* 36,89%. Prediksi tren keterlambatan menggunakan model ARIMA (2,1,1) menunjukkan peningkatan jumlah keterlambatan hingga Maret 2025, kemudian menurun di bulan-bulan berikutnya, dengan evaluasi MAPE 39,74%, sMAPE 33,32%, dan RMSE 1086,77 yang mengindikasikan akurasi prediksi yang cukup memadai namun dengan tingkat kesalahan relatif tinggi. Penelitian ini menyarankan penerapan insentif bagi OBA dengan performa pembayaran baik dan pembatasan plafon pinjaman berbasis skor risiko otomatis. Pengembangan lebih lanjut direkomendasikan dengan memperkaya data menggunakan variabel tambahan dan meningkatkan model prediksi melalui teknik *oversampling* seperti SMOTE serta *hyperparameter tuning* untuk memaksimalkan akurasi.

Kata kunci: *financial technology*, kredit macet, *business intelligence*, analisis prediktif

Abstract

The increasing number of non-performing loans in the financial technology P2P lending industry poses a serious challenge, including for PT. XYZ, which focuses on funding the MSME sector. This study applies Business Intelligence using OLAP methods and predictive analysis at PT. XYZ over a two year period. The data used consists of records from January 2023 to December 2024, covering 276,320 loan transactions and 267,542 payment records. The OLAP approach is utilized to explore loan transactions, payment performance, and delay performance based on various dimensions such as account, time, payment method, and bank used. The exploratory results are visualized in three interactive dashboards according to the focus of the analysis. Predictive analysis is conducted using a Random Forest Classifier model, which identifies loan_amount as the most influential variable affecting the likelihood of delay, although the model's performance remains limited with an accuracy of 61.47%, precision of 38.33%, and recall of 36.89%. The ARIMA (2,1,1) model is used to forecast delay trends, showing an increase in delays until March 2025 followed by a decline in subsequent months. Evaluation metrics of MAPE 39.74%, sMAPE 33.32%, and RMSE 1086.77 indicate a reasonably adequate prediction accuracy but with a relatively high error rate. This study recommends implementing incentives for OBAs with good payment performance and limiting loan

<http://sistemasi.ftik.unisi.ac.id>

ceilings based on automatic risk scores. Further development is suggested by enriching the data with additional variables and improving predictive models through oversampling techniques such as SMOTE and hyperparameter tuning to maximize accuracy.

Keywords: financial technology, non-performing loan, business intelligence, predictive analysis

1 Pendahuluan

Peningkatan digitalisasi di sektor keuangan membuat *fintech peer-to-peer (P2P) lending* sebagai media pinjaman bagi masyarakat terus meningkat sejak awal 2024. Pada Februari 2024, total pembiayaan yang tersalurkan tercatat sebesar Rp61,09 triliun, mengalami pertumbuhan sebesar 21,98% dibandingkan dengan tahun sebelumnya. Namun, di balik pertumbuhan tersebut, kredit macet masih menjadi tantangan utama bagi *fintech peer-to-peer (P2P) lending*. Otoritas Jasa Keuangan (OJK) mencatat tingkat wanprestasi di atas 90 hari (TWP90) pada Februari 2024 mencapai Rp1,79 triliun, sedikit naik dari Rp1,78 triliun di bulan sebelumnya. Secara keseluruhan, TWP90 berada di angka 2,95%, meningkat sekitar 27% dibandingkan Februari 2023 yang tercatat Rp1,40 triliun [1]. Maka dari itu perlu adanya pembenahan dari dalam Perusahaan *fintech peer-to-peer (P2P) lending* untuk mencegah peningkatan kredit macet lebih lanjut.

PT. XYZ merupakan salah satu perusahaan *fintech peer-to-peer (P2P) lending* di Indonesia yang memfokuskan pendanaannya untuk sektor Usaha Mikro, Kecil, dan Menengah (UMKM). PT. XYZ menyediakan solusi pendanaan bagi UMKM dengan pendekatan produktif, yaitu tidak memberikan dana tunai secara langsung, melainkan menyediakan barang yang dibutuhkan terlebih dahulu. Namun, dalam proses bisnisnya, PT. XYZ menghadapi tantangan dalam pengolahan data, yaitu dalam mengambil keputusan di 2 tahun belakang ini. Untuk mengatasi kredit macet dengan lebih melirik pada performa keterlambatan OBA dan agar dapat bersaing di industri P2P *lending*, PT. XYZ membutuhkan penerapan *Business Intelligence (BI)* untuk pengolahan data yang lebih efektif dan menghasilkan laporan interaktif guna mendukung pengambilan solusi dan keputusan strategis. Selain itu, metode analisis prediktif seperti *Random Forest Classifier* dan ARIMA diperlukan untuk mengantisipasi dan memitigasi risiko kredit macet, sehingga mendukung pertumbuhan dan keberlanjutan bisnis secara berkelanjutan.

Business Intelligence (BI) memiliki istilah yang luas yang meliputi berbagai alat, infrastruktur, aplikasi dan hal yang memfasilitasi akses dan penganalisisan informasi, serta proses mengumpulkan, mengolah, dan menganalisis data untuk menghasilkan informasi yang guna meningkatkan serta mengoptimalkan proses pengambilan keputusan dan performa bisnis [2]. Selain itu BI juga menggabungkannya dengan *knowledge management* dipadukan dengan alat analisis untuk menyediakan informasi internal dan kompetitif yang kompleks kepada perencana dan pengambil keputusan di perusahaan. Sistem BI menyajikan informasi yang dapat ditindaklanjuti dengan waktu, lokasi, dan format yang tepat, untuk mendukung proses pengambilan keputusan. Tujuannya adalah memperbaiki ketepatan waktu dan mutu informasi yang digunakan dalam pengambilan keputusan, sekaligus mendukung efisiensi dalam pelaksanaan tugas-tugas manajerial [3]. Salah satu teknik analisis yang efektif digunakan dalam *Business Intelligence (BI)* adalah *Online Analytical Processing (OLAP)*. Metode ini memungkinkan pengguna untuk menganalisis data dari berbagai dimensi dan tingkat hierarki secara interaktif, sehingga memungkinkan analisis multidimensi yang memberikan pemahaman yang lebih mendalam [4].

Random Forest Classifier adalah metode klasifikasi yang membangun sejumlah pohon keputusan (*decision tree*) berdasarkan data latih yang dipilih secara acak, baik dari sisi observasi maupun variabel, untuk meningkatkan akurasi model. Tujuan utama dari pendekatan ini adalah untuk meningkatkan akurasi model prediksi. Metode ini menggabungkan menggabungkan teknik *bootstrap aggregating (bagging)* dan *random feature selection*, di mana setiap pohon dibentuk dari sampel acak dengan pengembalian (*bootstrap*) tanpa proses pemangkasan (*pruning*). Hasil klasifikasi akhir ditentukan melalui *majority voting* dari seluruh pohon yang terbentuk [5].

ARIMA merupakan metode pendekatan yang umum digunakan dalam analisis deret waktu (*time series*), terutama untuk keperluan peramalan dan pengendalian. Metode ini terdiri dari dua komponen utama, yaitu *autoregressive (AR)* dan *moving average (MA)*, serta satu komponen tambahan berupa pembedaan (*differencing*) yang digunakan untuk membuat data menjadi stasioner. Model ini

diimplementasikan dalam notasi ARIMA (p, d, q), di mana p menunjukkan orde dari komponen AR, d merupakan jumlah diferensiasi yang dilakukan, dan q adalah orde dari komponen MA [6].

Peneliti ingin memadukan antara kekuatan *Business Intelligence* dengan metode OLAP dan kemampuan prediksi dalam memecahkan permasalahan PT. XYZ, sehingga tidak hanya mampu meningkatkan efisiensi analisis data, namun juga memfasilitasi perencanaan strategis jangka panjang yang didukung oleh data yang terukur dan prediktif. Penelitian ini juga bertujuan untuk memberikan pedoman kepada PT. XYZ dalam menerapkan BI, sehingga dapat mengidentifikasi pola dan tren dalam data pinjaman, dan data pembayaran yang belum diolah secara lebih efektif serta memberikan landasan yang kokoh untuk penelitian lebih lanjut mengenai kolaborasi antara BI dengan metode prediksi.

2 Tinjauan Literatur

Seiring meningkatnya kebutuhan bisnis dalam memahami pola-pola tersembunyi dalam data operasional, banyak penelitian terkini yang mengkaji peran *Business Intelligence* (BI), khususnya melalui metode OLAP (*Online Analytical Processing*), dalam membantu pengambilan keputusan strategis. OLAP memungkinkan eksplorasi data secara multidimensi, seperti berdasarkan waktu, lokasi, jenis produk, atau sumber transaksi, sehingga sangat ideal dalam menganalisis data keuangan dan operasional yang kompleks. Penelitian yang dilakukan oleh Sulistyoningih menyoroti efektivitas OLAP dalam mengembangkan *dashboard* pemasaran berbasis *marketing mix* pada sektor retail, yang mampu menyajikan ringkasan performa produk dan promosi berdasarkan lokasi dan waktu. Hasilnya, perusahaan dapat melihat tren penjualan berdasarkan waktu, produk yang menghasilkan pendapatan tertinggi dan perubahan performa akibat pandemi [7]. Penelitian tersebut berhasil memperlihatkan bagaimana OLAP memberikan pemahaman visual dan deskriptif yang kuat, namun belum dilengkapi dengan model analisis prediktif untuk mengantisipasi perubahan pola di masa depan.

Dalam konteks serupa, Sipayung menggunakan OLAP untuk menganalisis preferensi pembayaran pelanggan di sebuah *coffee shop* bernama Cafe Le Kahve, yang menghasilkan visualisasi trend transaksi berdasarkan waktu dan metode pembayaran. Meski demikian, studi ini tetap berfokus pada visualisasi dan eksplorasi data historis, tanpa melibatkan model analitik lanjutan untuk memahami faktor penyebab atau memprediksi perilaku pelanggan [8]. Hal ini menunjukkan bahwa sebagian besar implementasi OLAP di tingkat operasional masih bersifat deskriptif dan kurang dilengkapi dengan komponen prediktif untuk mendukung keputusan berbasis proyeksi.

Anis mengembangkan sistem keputusan berbasis OLAP untuk PT. Tirtamarta yang mana sistem ini membantu dalam mengambil keputusan keuangan. Sistem ini telah diuji dan mendapatkan tanggapan positif dari user yang menandakan bahwa *Business Intelligence* dengan menggunakan OLAP mampu memberikan solusi dalam mendukung perencanaan strategi keuangan yang lebih baik, bahkan diproyeksikan dapat digunakan lintas divisi untuk efisiensi maksimal [9].

Pada penelitian dengan metode *Random Forest Classifier* oleh Alvanof, Bustami, dan Dinata didapatkan bahwa metode ini menunjukkan efektivitas tinggi dalam mengidentifikasi serta mengklasifikasikan file *ransomware*, dengan tingkat akurasi sebesar 98,79%. Dengan pelatihan pada 966 sampel dan pengujian terhadap 414 sampel *ransomware*, pendekatan ini terbukti efektif dalam mendeteksi *ransomware* untuk tujuan keamanan siber [10]. Namun, pengujian yang dilakukan dalam penelitian tersebut masih terbatas pada tahap pemodelan, tanpa disertai implementasi visualisasi atau *dashboard* interaktif yang dapat mendukung analisis secara lebih komprehensif oleh pengguna non-teknis maupun pemangku kebijakan.

Fitri membandingkan metode *Random Forest Classifier*, regresi linier, dan *Gradient Boosted Trees Regression Method* untuk memprediksi harga rumah. Hasil dari penelitian ini adalah akurasi tertinggi diperoleh dari model *Random Forest Classifier* dengan nilai 81,5% [11]. Sama seperti Fitri, pada penelitian Rofi, Setiawan dan Riana membandingkan metode *Random Forest Classifier* dengan metode yang lain, yaitu *K-Nearest Neighbor*. Didapatkan bahwa *Random Forest Classifier* memiliki tingkat akurasi yang lebih tinggi dengan nilai 99,05% dibandingkan dengan *K-Nearest Neighbor* yang bernilai 98,10% [12]. Penelitian Zhang juga menunjukkan bahwa algoritma *Random Forest* terbukti sangat efektif dalam memprediksi risiko kredit dan default pada platform P2P *lending*, dengan performa superior dibandingkan metode seperti regresi logistik dan *decision tree* [13]. Meskipun penelitian-penelitian ini menunjukkan performa yang baik dalam hal akurasi, pendekatannya masih

terbatas pada aspek komparatif model prediksi saja, tanpa mengembangkan integrasi lebih lanjut seperti visualisasi interaktif atau sistem pendukung keputusan yang aplikatif.

Penelitian yang dilakukan oleh Putri dan Junaedi menggunakan ARIMA mengatasi permasalahan perencanaan dan pengadaan bahan baku yang masih manual, sehingga dibutuhkan peramalan atau prediksi penjualan kedepan untuk mengetahui kebutuhan permintaan. Didapatkan nilai MAPE 14,81% dan peramalan 7 periode kedepan [14]. Patokan penelitian hanya berfokus dalam prediksi masa depan, belum didukung dengan elemen analisis lain yang divisualisasikan untuk mendukung proses bisnis perusahaan.

Sedangkan Fahmuddin, Ruliana dan Mustika mencoba untuk perbandingan metode peramalan antara ARIMA dan *Single Exponential Smoothing* untuk meramalkan nilai Kakao Indonesia. Hasil dari penelitian tersebut ARIMA lebih baik dengan nilai MAPE 10,38060% yang lebih kecil dari *Single Exponential Smoothing* yang bernilai 10,92874% [15]. Dalam konteks perbandingan metode peramalan, Pratama dan Firdaus melakukan perbandingan antara ARIMA dengan Fuzzy Time Series Model Chen untuk meramal curah hujan dan didapatkan bahwa metode ARIMA lebih dapat diandalkan. Hal ini dibuktikan dengan ARIMA(4,1,0) yang lebih mendekati pola data aktual daripada FTS Model Chen [16]. Namun penelitian-penelitian ini hanya fokus pada perbandingan metode peramalan saja yang membuktikan bahwa ARIMA lebih baik dibandingkan dengan metode lain dalam studi kasusnya.

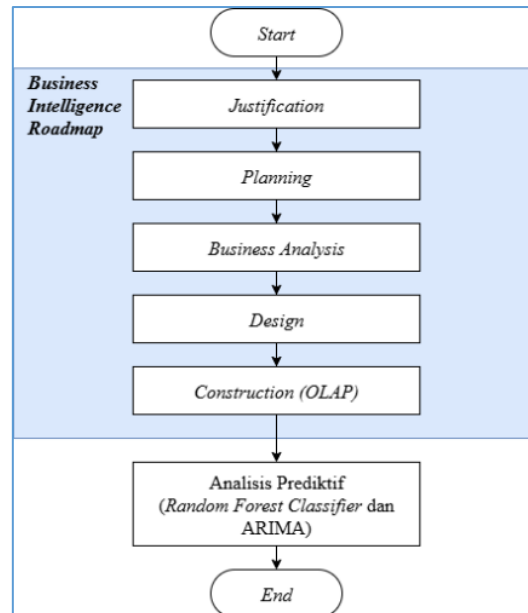
Hanifah telah mencoba menerapkan *Business Intelligence* (BI) dan model regresi linear untuk menganalisis dan memprediksi penjualan serta breakage di PT. XYZ yang merupakan sebuah perusahaan ritel minuman. Dengan *dashboard* yang interaktif dan prediksi yang akurat, manajemen bisa mengambil keputusan lebih cepat dan tepat, seperti menentukan produk yang harus difokuskan atau vendor mana yang performanya menurun. Penelitian ini memberi gambaran nyata bagaimana integrasi BI dan statistik dari regresi linear bisa mendongkrak efisiensi kerja di perusahaan retail. Lebih dari sekadar tampilan data, BI mampu menjadi alat strategis untuk merespons dinamika pasar [17]. Meskipun berhasil mengintegrasikan BI dan model prediktif dengan regresi linear, penelitian tersebut belum menerapkan metode lain, sehingga diperlukan eksplorasi metode prediktif lain yang dapat mendukung perusahaan.

Dengan demikian, gap dalam penelitian sebelumnya cukup jelas bahwa *Business Intelligence* dengan pendekatan OLAP dan analisis prediktif masih sering digunakan secara terpisah. Penelitian berbasis OLAP umumnya berakhir pada visualisasi dan deskripsi tren, tanpa tindak lanjut ke arah prediksi. Sebaliknya, analisis prediktif seperti *Random Forest Classifier* dan ARIMA digunakan secara langsung pada data mentah atau agregat tanpa proses eksplorasi mendalam. Padahal, integrasi keduanya akan memperkaya analisis karena memungkinkan peneliti tidak hanya melihat tren masa lalu, tapi juga memahami penyebab dan memperkirakan kejadian masa depan secara lebih tajam. Mengacu pada penelitian Hanifah, penyempurnaan penelitian perlu dilakukan dengan mengintegrasikan OLAP dan analisis prediktif dalam satu sistem BI yang utuh.

3 Metode Penelitian

Dalam penelitian ini, langkah-langkahnya mencakup beberapa tahap, dimulai dengan studi lapangan dan studi literatur untuk mengumpulkan data dan informasi yang relevan dengan topik penelitian. Selanjutnya metode roadmap dari *Business Intelligence* yang merupakan metode untuk mendesain *Business Intelligence*, yang melibatkan serangkaian langkah penting yaitu *justification*, *planning*, *business analysis*, *design*, dan *construction* [18]. Setelah itu melakukan analisis prediksi dengan menggunakan model *Random Forest Classifier* dan ARIMA.

Penelitian ini memanfaatkan dua jenis sumber data, yakni primer dan sekunder. Data primer dikumpulkan secara langsung dari sumber utama, berupa informasi terkait pinjaman dan pembayaran yang diakses langsung dari pihak perusahaan. Sementara itu, data sekunder diperoleh melalui kajian literatur dan hasil analisis data yang mendukung topik penelitian. Alur tahapan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1 Diagram alur metode penelitian

Tahapan penelitian dijelaskan berikut ini:

- a. *Justification*. Mendefinisikan masalah atau peluang bisnis dari objek penelitian. Hal ini termasuk mengidentifikasi mengenai mengapa suatu proyek BI harus dilaksanakan, dengan mempertimbangkan manfaat yang akan diperoleh serta biaya yang diperlukan. *Justification* bertujuan untuk memastikan bahwa setiap inisiatif BI yang diambil akan memberikan nilai tambah yang signifikan bagi perusahaan [18]. Peneliti mengidentifikasi masalah perusahaan yang akan diteliti dengan BI dan analisis prediktif.
- b. *Planning*. Membangun dan mengevaluasi infrastruktur yang mencakup komponen teknis dan nonteknis yang melibatkan perangkat yang diperlukan, standar, metodologi, dan prosedur yang mendukung kelancaran BI. Selain itu, perencanaan proyek BI harus dilakukan secara rinci dan hati-hati karena proyek BI dapat dipengaruhi oleh perubahan yang terjadi selama pelaksanaannya. Pemantauan dan pelaporan kemajuan proyek sangat penting untuk memastikan kesuksesan implementasi BI kedepannya [18]. Peneliti melakukan perencanaan perancangan proyek BI, serta merencanakan perangkat atau *tools* yang dibutuhkan.
- c. *Business Analysis*. Melakukan analisis menyeluruh terhadap business process serta aliran data untuk memperoleh wawasan yang mendalam tentang sumber data dan kebutuhan informasi perancangan Business Intelligence perusahaan [19]. Peneliti memahami dan menganalisis proses bisnis perusahaan, agar memahami kebutuhannya.
- d. *Design*. Perancangan skema *database* dengan merancang data mart yang selaras dengan kebutuhan informasi. Perancangan data menghasilkan tabel *fact_loan*, *fact_payment*, *dim_account*, *dim_method*, *dim_bank* dan *dim_date* yang berfungsi untuk pembuatan *dashboard* dan prediksi dengan linear regression model [18].
- e. *Construction*. Peneliti membuat data mart pada *database SQL Server*, dan *Extract, Transform, Load (ETL)* menggunakan python. Selain itu, dilakukan *data mining* untuk dapat menggali informasi tersembunyi yang bernilai dari data. Terakhir, pengembangan repositori metadata bisnis yaitu dengan pembuatan *dashboard*. Pembuatan *dashboard* menggunakan Power BI. Ketiga langkah ini dapat memberikan wawasan yang lebih dalam dan bermanfaat bagi pengambilan keputusan perusahaan [18].
- f. Analisis Prediktif dengan Metode *Random Forest Classifier*. *Random Forest Classifier* adalah metode prediktif berbasis klasifikasi yang membangun sejumlah pohon keputusan dari data latih dan fitur acak untuk meningkatkan akurasi. Dengan teknik *bagging* dan *random feature selection*, model ini meminimalkan *overfitting* dan menghasilkan prediksi berdasarkan suara mayoritas (*majority voting*) [5]. Pemodelan ini dilakukan dengan menggunakan python.

- g. Analisis Prediktif dengan Metode *Autoregressive Integrated Moving Average*. ARIMA merupakan model analisis deret waktu yang menggabungkan komponen *autoregressive* (AR) dan *moving average* (MA), serta membutuhkan data yang bersifat stasioner. Proses analisis dengan ARIMA mencakup empat tahap utama, yaitu identifikasi model, estimasi parameter, pemeriksaan diagnostik, dan peramalan [6]. Peneliti melakukan pembuatan model ARIMA dengan python.

4 Hasil dan Pembahasan

a. Justification

Peneliti mengidentifikasi bahwa PT XYZ, perusahaan *fintech* P2P *lending* yang memfokuskan pendanaan pada sektor UMKM, menghadapi tantangan dalam pengolahan data dan pengambilan keputusan selama dua tahun terakhir, khususnya terkait pemantauan kinerja keterlambatan OBA untuk mengantisipasi risiko kredit macet. Untuk meningkatkan daya saing di industri P2P *lending*, diperlukan penerapan *Business Intelligence* (BI) dengan metode OLAP guna menghasilkan laporan interaktif dan analisis multidimensi yang mendukung pengambilan keputusan strategis. Selain itu, metode analisis prediktif seperti *Random Forest Classifier* dan ARIMA diharapkan mampu mengidentifikasi pola, memprediksi potensi keterlambatan, serta memitigasi risiko kredit macet secara lebih efektif.

b. Planning

Dataset yang dianalisis terdiri dari dua file CSV berisi data *loan* dan *payment* milik PT XYZ, dengan cakupan waktu dari Januari 2023 hingga Desember 2024. Jumlah total *record* data *loan* sebelum proses *data cleansing* adalah 276.320 baris, sedangkan data *payment* sebanyak 267.542. Berdasarkan hasil wawancara dengan direktur perusahaan, teridentifikasi sejumlah kebutuhan informasi bisnis yang perlu dianalisis dan divisualisasikan untuk mendukung pengambilan keputusan.

Informasi yang diperlukan meliputi rincian mengenai jumlah pokok *disbursement* berdasarkan waktu, serta data mengenai OBA (*Outstanding Balance Amount*) aktif dalam rentang waktu tertentu. Selain itu, informasi mengenai jumlah transaksi pinjaman yang terjadi juga dibutuhkan untuk melihat tren aktivitas pinjaman selama periode analisis. Perusahaan juga ingin mengetahui besaran *platform fee* yang diperoleh, serta menghitung *basket size* atau rata-rata nilai pinjaman per transaksi dari waktu ke waktu.

Dari sisi pembayaran, informasi yang diinginkan mencakup total nominal pembayaran yang masuk, serta jumlah transaksi pembayaran berdasarkan waktu. Keterlambatan pembayaran menjadi fokus penting, sehingga diperlukan data mengenai nilai keterlambatan dalam bentuk nominal, serta total nominal penalti yang dibayarkan oleh peminjam. Untuk memahami preferensi nasabah, data kategori metode pembayaran dan kategori *bank payment* (IDS) juga menjadi bagian dari analisis.

Lebih lanjut, perusahaan ingin mengetahui kategori keterlambatan pembayaran yang terjadi, distribusi durasi keterlambatan, serta perbandingan jumlah pelunasan yang dilakukan tepat waktu dan yang tidak. Identifikasi terhadap OBA dengan durasi keterlambatan terlama juga menjadi salah satu kebutuhan penting untuk membantu mitigasi risiko ke depan. Semua informasi ini diharapkan dapat ditampilkan dalam bentuk *dashboard* yang interaktif, sehingga memudahkan pihak manajemen dalam mengambil keputusan berbasis data yang akurat dan *real-time*. Pemrosesan data akan dilakukan dengan menggunakan python, Microsoft SQL *Server Management Studio* (SSMS), dan power BI.

c. Business Analysis and Design

Pada penelitian ini perancangan *data mart* dibuat menggunakan pendekatan *four-step methodology* yang merupakan penyederhanaan *nine-step methodology* oleh Ralph Kimball dalam bukunya yang berjudul *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (3rd Edition)*. Terdapat 4 proses utama pemodelan yaitu penentuan proses bisnis, penentuan grain, identifikasi serta penataan elemen dimensi, dan penentuan tabel fakta [20].

1) Penentuan Proses Bisnis

Kebutuhan informasi diambil dari 2 proses bisnis utama di PT. XYZ yaitu pemberian pinjaman dan juga pembayaran pinjaman. Tabel 1 menyajikan penjelasan mengenai masing-masing proses bisnis.

Tabel 1 Penentuan proses bisnis

Proses Bisnis	Keterangan
Pemberian pinjaman (<i>loan</i>)	Transaksi pengajuan dan pencairan pinjaman kepada peminjam, dicatat dalam tabel <i>loan</i> .
Pembayaran pinjaman (<i>payment</i>)	Transaksi pembayaran atas pinjaman oleh peminjam, termasuk pembayaran pokok dan denda, dicatat dalam tabel <i>payment</i> .

2) Penentuan *Grain*

Grain menentukan apa yang sebenarnya digambarkan dari tabel fakta dan mengacu pada tingkat detail informasi yang disusun untuk keperluan analisis data. Menentukan *grain*, dapat menjadi alat untuk memahami metrik apa yang ditampilkan dan bagaimana informasi tersebut dihubungkan dengan dimensi. Informasi yang disusun mencakup berbagai aspek transaksi dan pembayaran, sebagian besar dianalisis berdasarkan waktu. Data mencakup nilai pokok disbursement, OBA aktif, jumlah transaksi pinjaman, dan *platform fee*. Selain itu, ada informasi tentang *basket size*, jumlah pembayaran yang masuk, total nominal yang dibayarkan, total penalty dan rata-rata penalti. Untuk analisis keterlambatan, tersedia data nilai keterlambatan, serta kategori keterlambatan dan durasinya. Kategori metode dan jenis pembayaran bank juga dicatat, masing-masing dikaitkan dengan dimensi metode dan bank.

Beberapa metrik fokus pada evaluasi performa pembayaran, seperti perbandingan pelunasan tepat waktu dan tidak tepat waktu, serta OBA dengan keterlambatan paling tinggi. Seluruh metrik ini dirangkum dalam *dashboard* interaktif yang memungkinkan pengguna mengeksplorasi data berdasarkan berbagai dimensi.

3) Identifikasi serta Penataan Elemen Dimensi

Tabel dimensi merupakan tabel yang berfungsi untuk menyimpan informasi deskriptif atau atribut-atribut pendukung dari suatu peristiwa atau transaksi yang tercatat dalam tabel fakta. Dalam konteks data mart, biasanya dibutuhkan beberapa tabel dimensi untuk memberikan konteks yang lebih jelas terhadap data yang dianalisis. Tabel dimensi yang dibutuhkan diterangkan pada tabel 2.

Tabel 2 Identifikasi dimensi

Dimensi	Keterangan
<i>Account</i>	Menyimpan informasi mengenai peminjam, seperti ID akun dan nama peminjam.
<i>Method</i>	Menyimpan informasi mengenai metode pembayaran yang digunakan.
<i>Bank</i>	Menyimpan informasi mengenai bank yang digunakan dalam transaksi pembayaran IDS.
<i>Date</i>	Menyimpan informasi waktu seperti tanggal, bulan, dan tahun berbagai aktivitas.

4) Penentuan *Fact Table*

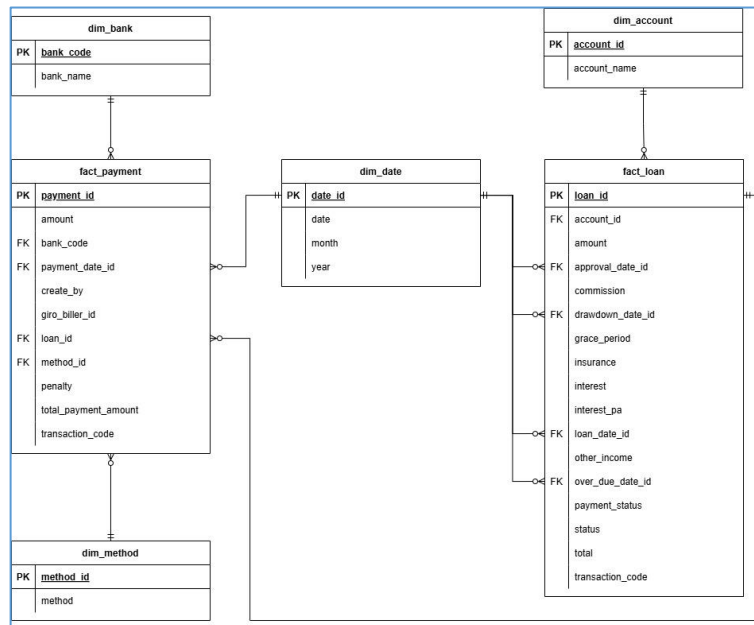
Fact table merupakan tabel yang berisi *primary key*, *foreign key* dari tabel dimensi serta data metrik atau pengukuran numerik yang dapat dihitung dan dianalisis. Dalam penelitian ini, terdapat 2 tabel fakta yang terbentuk, meliputi *fact_loan* dan *fact_payment*. Tabel *fact_loan* berfungsi untuk merekam data transaksi pinjaman yang mencakup informasi seperti jumlah pinjaman, tenor, dan tanggal pencairan. Sementara itu, tabel *fact_payment* digunakan untuk mencatat data transaksi pembayaran oleh peminjam, seperti jumlah pembayaran, metode pembayaran, dan tanggal pembayaran yang berguna untuk analisis performa pelunasan.

5) Desain Skema *Data Mart*

Pada tahap ini, perancangan skema *data mart* dilakukan dengan pendekatan skema *galaxy (fact constellation schema)* karena terdapat dua tabel fakta dari tabel *loan* dan *payment*. Kedua tabel fakta ini terhubung dengan beberapa tabel dimensi yang saling

<http://sistemasi.ftik.unisi.ac.id>

berbagi, seperti dimensi *date*, *method*, *bank* dan *account*. Skema *data mart* digambarkan pada Gambar 2.



Gambar 2 Skema data mart

d. Construction

1) Proses ETL (Extract, Transform, Load)

Proses ETL (Extract, Transform, Load) pada proyek ini dilakukan menggunakan Python untuk mengelola data pinjaman dan pembayaran. Tahap *extract* dilakukan dengan membaca data mentah menggunakan *library* pandas yang mana pada tahap ini dilakukan *extract data* dari csv ke python, ditunjukkan pada Gambar 3. Lalu diikuti oleh tahap *transform* yang mencakup pembersihan data seperti menghapus duplikasi, menangani nilai kosong, serta menyesuaikan format tanggal dan tipe data agar siap untuk dimuat ke *database*. Hasil dari *transform* data *loan* ditunjukkan pada Gambar 4, sedangkan data *payment* pada Gambar 5.

```

import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
loan = pd.read_csv("loan_2023_2024.csv", delimiter=';')
payment = pd.read_csv("payment_2023_2024.csv", delimiter=';', low_memory=False)
    
```

Gambar 3 Proses extract data CSV ke python

loan_id	account_name	amount	approval_date	approval_date_time	commission	drawdown_date	drawdown_date_time	grace_period	interest	interest_pa	loan_date	loan_date_time	other_income	over_due_date	payment_status	status	total	transaction_code
0	1003451	28014	2023-01-02	2023-01-02 08:18:49	4898.0	2023-01-02	2023-01-02 08:30:37	2	49.0	10	2023-01-01	2023-01-01 08:13:25	51.0	2023-01-09	CLOSED	APPROVED	746075	LN6482578961
1	1003452	87145	2023-01-02	2023-01-02 08:44:13	4898.0	2023-01-02	2023-01-02 08:30:35	2	49.0	10	2023-01-01	2023-01-01 08:44:03	49.0	2023-01-09	CLOSED	APPROVED	2954800	LN7543427222
2	1003453	33713	2023-01-02	2023-01-02 08:27:41	4898.0	2023-01-02	2023-01-02 08:36:37	2	98.0	10	2023-01-01	2023-01-01 10:45:54	0.0	2023-01-16	CLOSED	APPROVED	739225	LN8253246737
3	1003454	57753	2023-01-02	2023-01-02 09:05:38	4898.0	2023-01-02	2023-01-02 09:16:36	2	98.0	10	2023-01-01	2023-01-01 11:09:49	0.0	2023-01-16	CLOSED	APPROVED	378000	LN4255846750
4	1003455	68511	2023-01-02	2023-01-02 08:23:45	4898.0	2023-01-02	2023-01-02 08:30:37	2	98.0	10	2023-01-01	2023-01-01 11:12:21	0.0	2023-01-16	CLOSED	APPROVED	736475	LN6891238377

Gambar 4 Transform data loan

payment_id	amount	bank_id	bank_code	create_by	create_on	giro_billir_id	loan_id	method	note	paid_off_date	payment_date	payment_date_time	penalty	penalty_payment	status	total_payment_amount	transaction_code
0	938613.0	1486234.0	13.0	None	2023-01-01 07:32:30	1188.0	1000271.0	INTERNET_BANKING	Penunasan pinjaman (LN1958124183)	2023-01-01	2023-01-01	2023-01-01 07:32:30	1488.0	1488.0	CONFIRMED	1488013.0	RA230100000
1	938614.0	489200.0	13.0	None	SERVICES_HQH 2023-01-01 07:59:12	NaN	996145.0	INTERNET_BANKING	Penunasan pinjaman (LN1057502627 05 Dec 2022)	2023-01-01	2023-01-01	2023-01-01 07:32:12	6400.0	6400.0	CONFIRMED	495800.0	RA230100001
2	938615.0	378000.0	13.0	None	SERVICES_HQH 2023-01-01 08:08:47	NaN	996015.0	INTERNET_BANKING	Penunasan pinjaman (LN49497046270 16 Dec 2022)	2023-01-01	2023-01-01	2023-01-01 08:08:47	0.0	0.0	CONFIRMED	378000.0	RA230100002
3	938616.0	398900.0	13.0	None	SERVICES_HQH 2023-01-01 08:13:51	NaN	996630.0	INTERNET_BANKING	Penunasan pinjaman (LN2381334454 17 Dec 2022)	2023-01-01	2023-01-01	2023-01-01 08:13:51	0.0	0.0	CONFIRMED	398900.0	RA230100003
4	938617.0	725200.0	13.0	None	SERVICES_HQH 2023-01-01 08:53:09	NaN	999840.0	INTERNET_BANKING	Penunasan pinjaman (LN396411334 24 Dec 2022)	2023-01-01	2023-01-01	2023-01-01 08:53:09	0.0	0.0	CONFIRMED	725200.0	RA230100004

Gambar 5 Transform data payment

Setelah data dibersihkan, tahap *load* dilakukan dengan mentransfer data dari python ke Microsoft SQL Server Management Studio (SSMS) menggunakan koneksi ODBC melalui *library* pyodbc. Gambar 6 menunjukkan proses *load data* itu sendiri. Data yang

<http://sistemasi.ftik.unisi.ac.id>

telah dimuat ini kemudian digunakan sebagai dasar untuk membangun *data mart* dalam bentuk skema *galaxy (fact constellation schema)*, yang selanjutnya dimanfaatkan dalam pembuatan *dashboard* analitik menggunakan Power BI.

```
engine = create_engine(f"mssql+pyodbc:///?odbc_connect={params}")
existing_ids = pd.read_sql("SELECT loan_id FROM loan", con=engine)
filtered_df = loan_clean[~loan_clean['loan_id'].isin(existing_ids['loan_id'])]
filtered_df.to_sql(name='loan', con=engine, if_exists='append', index=False)

53

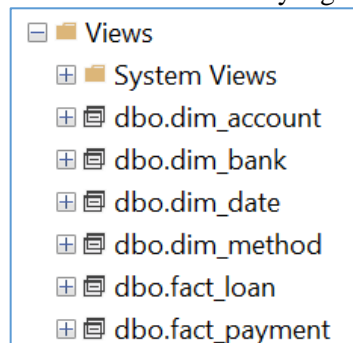
payment_clean.to_sql(name='payment', con=engine, if_exists='append', index=False)

46
```

Gambar 6 Proses load data dari python ke SSMS

2) Pembuatan Data Mart

Proses pembuatan *data mart* menggunakan Microsoft SQL Server Management Studio (SSMS), yang mana mengimplementasikan dari desain skema *galaxy* yang telah dibuat sebelumnya. Pembuatan menggunakan *query* dengan membuat *view* dari tabel utama yaitu *loan* dan *payment* yang telah di-load dari python. Pada Gambar 7 menunjukkan *view* dari tabel-tabel skema *data mart* yang telah dibuat.



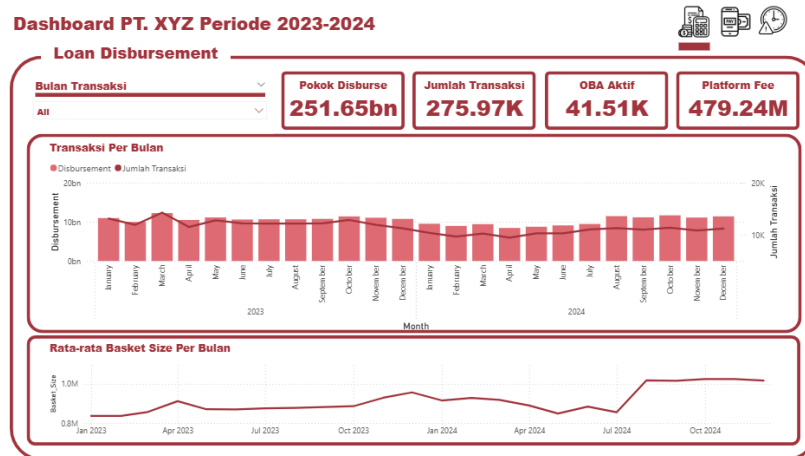
Gambar 7 Data Mart pada SSMS

3) Visualisasi Interaktif Berbasis OLAP

Dashboard dirancang sebagai sarana untuk menyatukan seluruh visualisasi data pinjaman dan pembayaran dalam satu tampilan, guna mempermudah pemantauan informasi yang relevan bagi pengambilan keputusan bisnis. Pembuatan *dashboard* menggunakan Microsoft Power BI, yang memungkinkan penyajian data secara dinamis dan interaktif. Data yang sebelumnya telah diolah dan disimpan di dalam *data mart* menggunakan Microsoft SQL Server Management Studio (SSMS) kemudian diimpor ke Power BI untuk divisualisasikan. Terdapat dua fokus utama dalam *dashboard* ini, yaitu *dashboard* performa penyaluran pinjaman (*loan disbursement*) dan *dashboard* performa pembayaran dan keterlambatan (*payment & delay performance*).

a) Loan Disbursement Dashboard

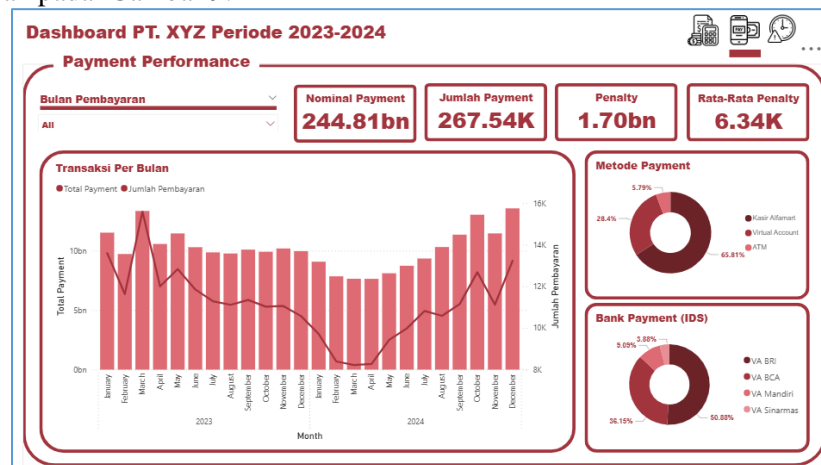
Dashboard ini menyajikan informasi terkait transaksi pinjaman yang terjadi di PT. XYZ selama periode 2023–2024. Informasi yang ditampilkan meliputi total pokok *disburse*, jumlah transaksi pinjaman, OBA aktif yang bertransaksi pada periode 2023-2024 serta *platform fee* yang dihasilkan. Selain itu, *dashboard* juga menampilkan tren transaksi per bulan dan rata-rata basket size untuk memberikan gambaran menyeluruh terhadap perilaku dan nilai penyaluran pinjaman dari waktu ke waktu. *Loan disbursement dashboard* dapat dilihat pada Gambar 8.



Gambar 8 Loan disbursement dashboard

b) *Payment Performance Dashboard*

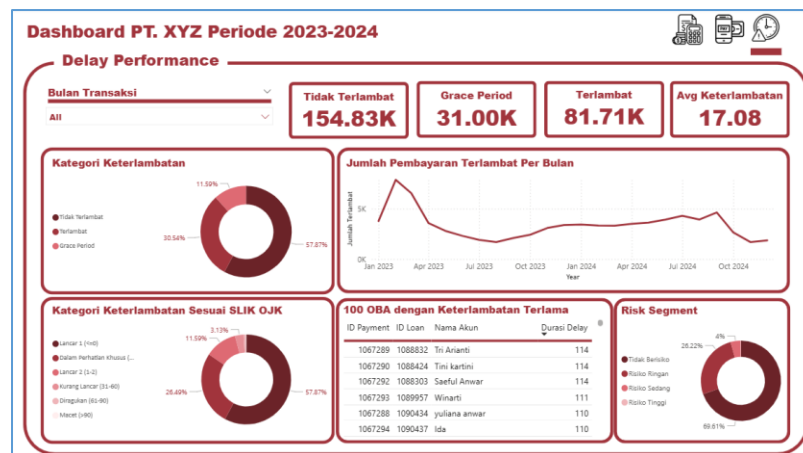
Dashboard ini menampilkan rangkuman performa pembayaran di PT. XYZ selama periode 2023–2024. Informasi utama mencakup total nominal pembayaran yang berhasil diterima, jumlah transaksi pembayaran, total denda yang dikenakan, serta nilai rata-rata denda per transaksi. Visualisasi tren pembayaran bulanan disajikan untuk menggambarkan dinamika volume dan frekuensi pembayaran dari waktu ke waktu. Selain itu, terdapat persentase metode pembayaran serta penggunaan bank *virtual account* (VA) yang memberikan wawasan mengenai perilaku pengguna dalam melakukan pembayaran. *Payment performance dashboard* disajikan pada Gambar 9.



Gambar 9 Payment performance dashboard

c) *Delay Performance Dashboard*

Dashboard ini menggambarkan performa keterlambatan pembayaran di PT. XYZ selama periode 2023–2024. Di dalamnya tercantum jumlah pembayaran yang dilakukan tepat waktu, *grace period*, maupun yang terlambat, lengkap dengan rata-rata hari keterlambatannya. Grafik tren bulanan menunjukkan fluktuasi jumlah keterlambatan dari waktu ke waktu. Untuk memperjelas tingkat keparahan, keterlambatan juga dikelompokkan berdasarkan kategori umum dan sesuai klasifikasi SLIK OJK. Selain itu, *dashboard* ini menampilkan daftar 100 akun OBA dengan durasi keterlambatan tertinggi serta segmentasi risiko pengguna berdasarkan tingkat risiko gagal bayar. *Delay performance dashboard* ditampilkan pada Gambar 10.



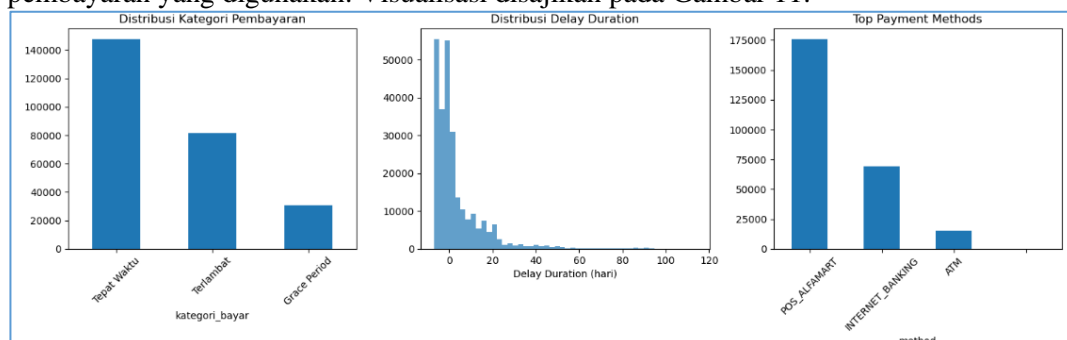
Gambar 10 Delay performance dashboard

e. Analisis Prediktif

Pada penelitian ini, tahapan analisis prediktif dilakukan dengan menggunakan dua pendekatan utama, yaitu *Random Forest Classifier* untuk mengidentifikasi faktor-faktor yang paling berkontribusi terhadap keterlambatan pembayaran pinjaman di PT. XYZ dan *ARIMA (AutoRegressive Integrated Moving Average)* untuk peramalan jumlah keterlambatan pembayaran di masa mendatang. Namun sebelum membangun model prediktif, dilakukan tahap awal berupa *Exploratory Data Analysis (EDA)* yang berfungsi untuk memahami struktur dan karakteristik data, mengidentifikasi pola, mengenali *outlier*, dan memastikan bahwa data yang digunakan layak dan relevan untuk proses pemodelan prediktif.

1) Penerapan *Exploratory Data Analysis (EDA)*

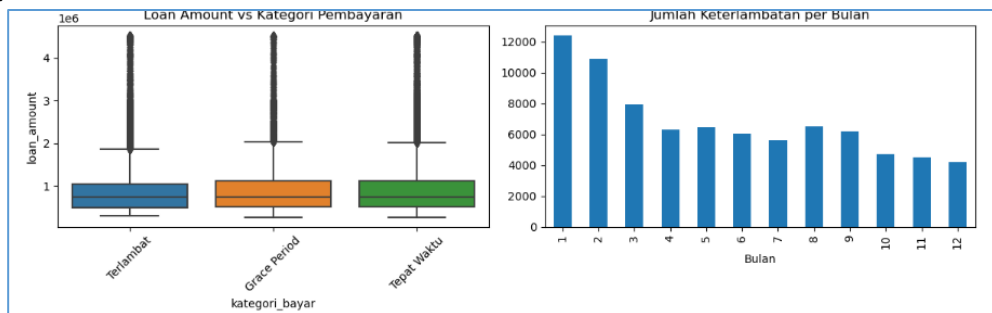
Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik dasar dari data pembayaran pinjaman yang dimiliki PT. XYZ yang diimplementasikan dengan menggunakan python. Terdapat visualisasi distribusi yang menunjukkan bahwa mayoritas pembayaran dilakukan tepat waktu, diikuti oleh kategori terlambat dan *grace period*. Hal ini mengindikasikan bahwa sebagian besar peminjam masih mematuhi tenggat pembayaran, namun tetap terdapat proporsi signifikan yang mengalami keterlambatan. Histogram distribusi *delay duration* memperlihatkan pola yang sangat condong ke kiri (*right-skewed*), di mana sebagian besar keterlambatan hanya terjadi dalam beberapa hari pertama, dan hanya sebagian kecil pinjaman yang mengalami keterlambatan lebih dari 20 hari. Sementara itu, pada grafik metode pembayaran, metode POS_ALFAMART mendominasi sebagai cara paling umum yang digunakan oleh peminjam. Informasi ini dapat memberikan pertimbangan dalam evaluasi risiko keterlambatan berdasarkan metode pembayaran yang digunakan. Visualisasi disajikan pada Gambar 11.



Gambar 11 Visualisasi exploratory data analysis 1

Visualisasi selanjutnya dilakukan untuk menggambarkan hubungan antara karakteristik pinjaman dan perilaku pembayaran. *Boxplot* pertama menunjukkan bahwa distribusi jumlah pinjaman (*loan amount*) relatif serupa di antara kategori pembayaran baik terlambat, *grace period*, maupun tepat waktu, dengan median dan persebaran nilai yang

tidak berbeda signifikan, meskipun terdapat outlier bernilai besar pada semua kategori. Sementara itu, grafik batang pada sisi kanan memperlihatkan tren jumlah keterlambatan pembayaran yang menurun dari bulan ke bulan, dengan puncaknya terjadi pada Januari dan jumlah terendah pada Desember. Penurunan ini dapat disebabkan oleh berbagai faktor musiman atau perbaikan dalam pengelolaan pembayaran seiring waktu. Analisis ini memberikan gambaran awal mengenai pola keterlambatan dan potensi risiko yang perlu ditindaklanjuti pada tahapan modeling selanjutnya. Visualisasi digambarkan pada Gambar 12.



Gambar 12. Visualisasi *exploratory data analysis* 2

2) Penerapan *Random Forest Classifier* untuk Analisis Prediktif

Analisis prediktif keterlambatan pembayaran dilakukan menggunakan model *Random Forest Classifier*, dengan visualisasi akhir yang dibangun menggunakan python. Tujuan dari analisis ini adalah untuk mengetahui apa saja yang paling berkontribusi dalam memengaruhi keterlambatan pembayaran pinjaman pada PT. XYZ. Berdasarkan hasil evaluasi, model *Random Forest Classifier* menghasilkan akurasi sebesar 61,47%, *precision* 38,33%, *recall* 36,89%, F1-score 37,60%, dan AUC-ROC 0,5881 yang mana berarti peformanya masih terbatas, metrik *recall* dan AUC-ROC yang relatif rendah menunjukkan bahwa model belum optimal dalam mendeteksi kategori terlambat. Hal ini dapat disebabkan oleh ketidakseimbangan jumlah data antara kelas tepat waktu dan terlambat, serta keterbatasan informasi pada variabel input. Hasil dapat dilihat dilihat pada Gambar 13.

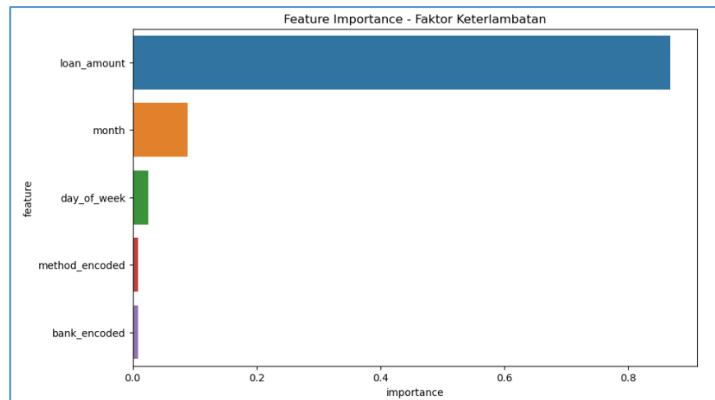
```

=== Random Forest Classifier Performance ===
Akurasi      : 0.6147
Precision    : 0.3833
Recall       : 0.3689
F1-Score     : 0.3760
AUC-ROC      : 0.5881
    
```

Gambar 13 *Random forest classifier performance*

Hasil visualisasi *feature importance* menunjukkan bahwa *loan_amount* merupakan fitur paling dominan dengan bobot sekitar 0.867269, menandakan bahwa semakin besar jumlah pinjaman, semakin tinggi kemungkinan terjadi keterlambatan pembayaran. *Month* memiliki kontribusi sebesar 0.089250, menunjukkan adanya pola musiman yang memengaruhi perilaku bayar OBA.

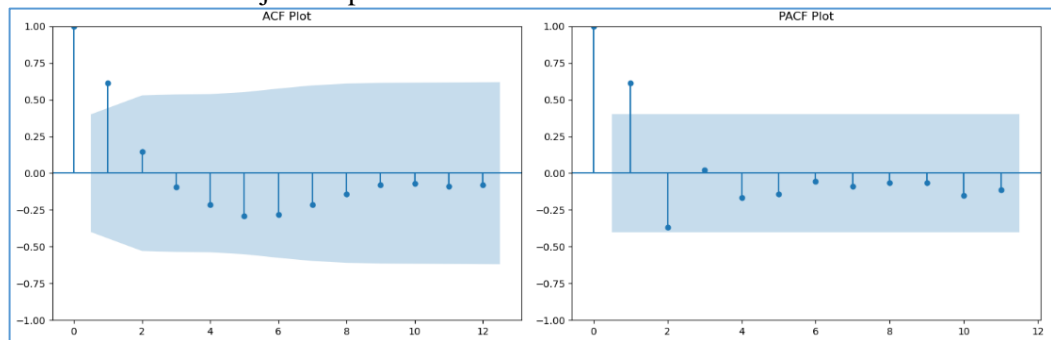
Fitur lain seperti *day_of_week* (0.025047), *method_encoded* (0.009365), dan *bank_encoded* (0.009070) memiliki pengaruh yang jauh lebih kecil. Visualisasi dilakukan menggunakan *bar chart*, yang secara langsung menampilkan *feature importance* hasil pelatihan model yang disajikan pada Gambar 14. Meskipun performa model belum tinggi, hasil prediksi tetap dapat memberikan gambaran awal terkait faktor-faktor yang berpengaruh terhadap keterlambatan. Informasi ini dapat digunakan sebagai dasar pengambilan keputusan awal atau sebagai *screening tool* untuk mengidentifikasi kasus yang berpotensi terlambat.



Gambar 14 Visualisasi feature importance faktor keterlambatan

3) Penerapan ARIMA untuk Analisis Prediktif

Proses analisis prediktif *time series* dengan menerapkan ARIMA untuk prediksi keterlambatan kedepannya dengan menggunakan python. Pada tahap awal peramalan jumlah keterlambatan pembayaran, dilakukan analisis stasioneritas dan identifikasi pola menggunakan grafik ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*). Grafik ACF menunjukkan bahwa data memiliki nilai autokorelasi yang signifikan pada beberapa lag, yang mengindikasikan adanya pola musiman atau ketergantungan antar waktu. Hal ini menandakan bahwa data masih mengandung unsur non-stasioner. Sementara itu, grafik PACF menunjukkan nilai autokorelasi parsial yang signifikan pada lag ke-1 dan ke-2, sehingga memberikan indikasi bahwa komponen *autoregressive* (AR) dalam model memiliki orde 2. Grafik ACF dan PACF ditunjukkan pada Gambar 15.



Gambar 15 Grafik ACF dan PACF

Untuk memastikan tingkat stasioneritas data, dilakukan Uji Augmented Dickey-Fuller (ADF) yang menghasilkan ADF *Statistic* sebesar -2.0375 dan nilai p-value sebesar 0.2703, yang lebih besar dari 0.05. Berdasarkan hasil tersebut, disimpulkan bahwa data belum stasioner secara statistik, sehingga diperlukan proses *differencing* satu kali untuk menghilangkan tren. Hasil ditunjukkan pada Gambar 16.

```

ADF Statistic: -2.0375679250659937
p-value: 0.2703475726749444
Critical Values:
    1%: -3.7529275211638033
    5%: -2.998499866852963
    10%: -2.6389669754253307
Data belum stasioner, perlu differencing
    
```

Gambar 16 Hasil augmented dickey-fuller

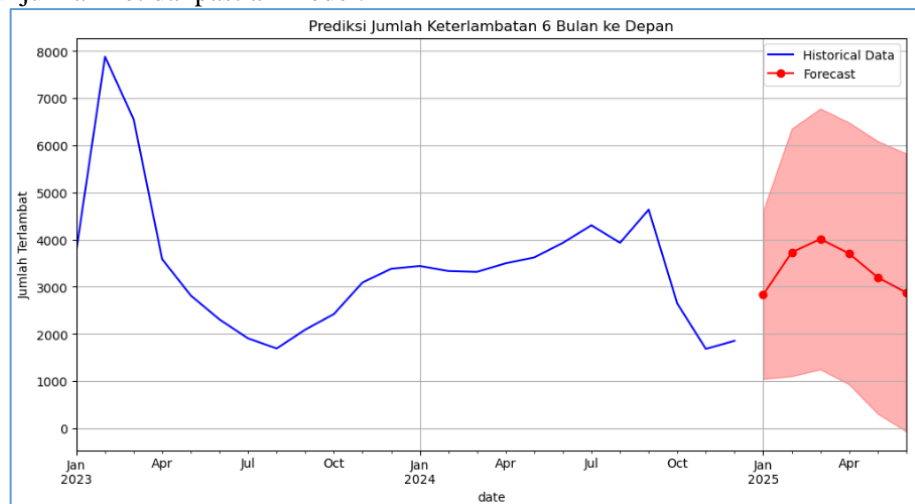
Setelah data menjadi stasioner, dilakukan pemilihan parameter model ARIMA dengan pendekatan kombinasi berbagai nilai orde AR (p), *differencing* (d), dan *moving*

average (q). Berdasarkan nilai AIC (*Akaike Information Criterion*) terendah dengan hasil 387.33, model terbaik yang dipilih adalah ARIMA(2,1,1), dengan $p = 2$ (dari PACF), $d = 1$ (dari hasil uji stasioneritas), dan $q = 1$ (dari ACF). Model ini dipandang paling optimal untuk melakukan peramalan terhadap jumlah keterlambatan pembayaran di masa mendatang. Hasil modelling ARIMA ditampilkan pada Gambar 17.

SARIMAX Results						
Dep. Variable:	jumlah_telat	No. Observations:	24			
Model:	ARIMA(2, 1, 1)	Log Likelihood	-189.665			
Date:	Fri, 01 Aug 2025	AIC	387.329			
Time:	22:30:18	BIC	391.871			
Sample:	01-01-2023	HQIC	388.471			
	- 12-01-2024					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	1.0388	0.137	7.558	0.000	0.769	1.308
ar.L2	-0.6671	0.133	-5.012	0.000	-0.928	-0.406
ma.L1	-1.0000	0.386	-2.593	0.010	-1.756	-0.244
sigma2	7.955e+05	4.85e-07	1.64e+12	0.000	7.96e+05	7.96e+05
Ljung-Box (L1) (Q):			1.55	Jarque-Bera (JB):	1.36	
Prob(Q):			0.21	Prob(JB):	0.51	
Heteroskedasticity (H):			0.46	Skew:	0.42	
Prob(H) (two-sided):			0.29	Kurtosis:	3.85	

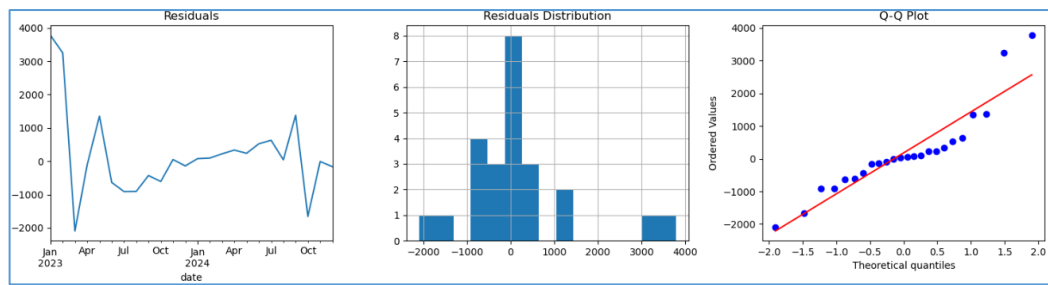
Gambar 17 Hasil modelling ARIMA

Grafik yang disajikan pada Gambar 18 menunjukkan prediksi jumlah keterlambatan pembayaran selama enam bulan ke depan menggunakan model ARIMA. Garis biru menggambarkan data historis, sedangkan garis merah menunjukkan hasil forecast Januari–Juni 2025. Prediksi menunjukkan tren peningkatan hingga Maret, lalu menurun di bulan-bulan berikutnya. Area merah muda merupakan interval kepercayaan 95% yang menunjukkan ketidakpastian model.



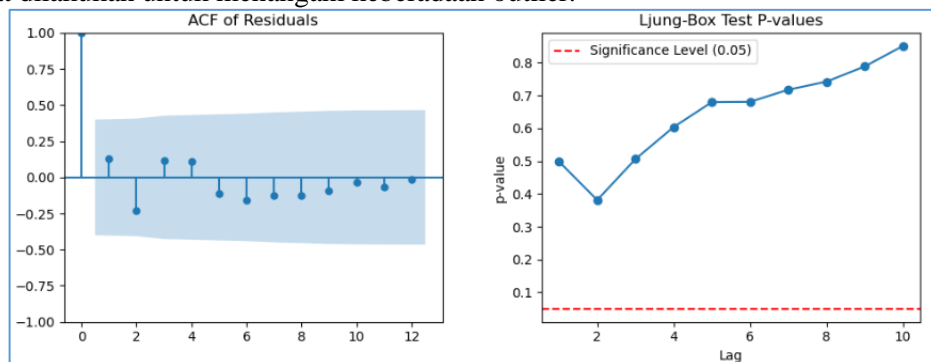
Gambar 18 Grafik prediksi keterlambatan 6 bulan ke depan

Analisis residual dari model ARIMA menunjukkan bahwa model cukup baik dalam menangkap pola data keterlambatan pembayaran. Plot residual terlihat menyebar acak di sekitar nol tanpa pola khusus, meskipun masih terdapat beberapa outlier. Histogram menunjukkan distribusi mendekati normal namun sedikit skew ke kanan, dan Q-Q plot menunjukkan sebagian besar titik mengikuti garis lurus, dengan sedikit penyimpangan di ujung. Hasil analisis digambarkan pada Gambar 19.



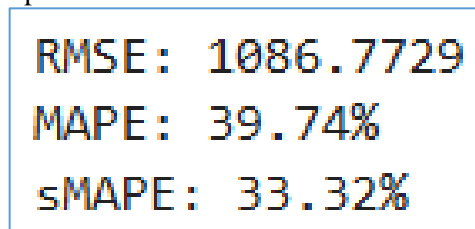
Gambar 19 Analisis residual, residual distribution dan Q-Q Plot

Hasil uji Ljung-Box untuk lag 1 hingga 5 memberikan p-value berturut-turut sebesar 0.4985, 0.3812, 0.5061, 0.6037, dan 0.6797. Seluruh nilai tersebut jauh di atas 0.05, yang menandakan bahwa tidak ada autokorelasi signifikan dalam residual. Temuan ini juga didukung oleh plot ACF residual yang seluruh titiknya berada dalam batas kepercayaan. Hasil pengujian ditampilkan pada Gambar 20. Dengan demikian, model ARIMA yang digunakan dapat dianggap valid secara statistik, meskipun peningkatan dapat dilakukan untuk menangani keberadaan outlier.



Gambar 20 Plot ACF residual dan hasil uji ljung-box

Model ARIMA yang digunakan pada penelitian ini menunjukkan tingkat akurasi cukup memadai. Hasil evaluasi memberikan nilai MAPE sebesar 39.74%, sMAPE sebesar 33.32%, dan RMSE sebesar 1086.7729. Nilai tersebut mengindikasikan bahwa meskipun model mampu menangkap pola umum dan tren data secara keseluruhan, kemampuan dalam memprediksi fluktuasi detail masih terbatas karena rata-rata kesalahan prediksi masih relatif besar. Oleh karena itu, model ini masih dapat dimanfaatkan untuk perkiraan jangka pendek atau sebagai indikasi arah tren, namun kurang direkomendasikan untuk estimasi yang memerlukan tingkat presisi tinggi. Hasil evaluasi dapat disajikan pada Gambar 21.



Gambar 21 Hasil Evaluasi model ARIMA

5 Kesimpulan

Penelitian ini mengimplementasikan *Business Intelligence* berbasis OLAP untuk menganalisis data pinjaman dan pembayaran pada perusahaan *fintech* P2P lending. *Data mart* dibangun dengan dua tabel fakta (*fact_loan* dan *fact_payment*) serta beberapa tabel dimensi. Tiga *dashboard* utama, *loan disbursement*, *payment performance*, dan *delay performance* memungkinkan pemantauan kredit dan deteksi keterlambatan secara *real time*. Analisis menunjukkan tingkat keterlambatan tinggi (31,39%), didominasi keterlambatan pendek, dengan metode POS_ALFAMART menjadi saluran paling

bermasalah. *Random Forest* mengidentifikasi *loan_amount* sebagai faktor utama keterlambatan, namun akurasi model (61,47%), *precision* (38,33%), dan *recall* (36,89%) menunjukkan performa yang masih terbatas dalam mendeteksi keterlambatan. Model ARIMA (2,1,1) memproyeksikan jumlah keterlambatan dalam enam bulan ke depan yang mana terjadi peningkatan hingga Maret, lalu menurun di bulan-bulan berikutnya. Evaluasi menunjukkan MAPE (39.74%), sMAPE (33.32%), dan RMSE (1086.77), menandakan akurasi prediksi yang cukup memadai namun tingkat kesalahan masih relatif tinggi. Secara strategis, PT. XYZ disarankan untuk menerapkan insentif bagi OBA dengan performa pembayaran yang baik serta membatasi plafon pinjaman berdasarkan skor risiko otomatis. Penelitian ini memiliki keterbatasan karena data yang digunakan hanya berasal dari dua tabel, sehingga analisis kurang maksimal dan performa *Random Forest* kurang akurat dalam prediksi. Sebagai pengembangan selanjutnya, disarankan memperkaya data dengan variabel tambahan seperti informasi detail OBA, histori pendaftaran, dan cabang. Selain itu, peningkatan model prediksi dapat dilakukan dengan menggunakan teknik *oversampling* seperti SMOTE untuk menangani data *imbalance*, dan *hyperparameter tuning* untuk memaksimalkan akurasi dengan kombinasi parameter terbaik.

Referensi

- [1] E. Budiayanti, "Mengatasi Kredit Macet Fintech Peer To Peer (P2P) Lending," 2024. [Online]. Available: <https://pusaka.dpr.go.id>
- [2] D. K. Turban, Efraim, Ramesh Sharda, Dursun Delen, *Business Intelligence: A Managerial Approach*. Prentice Hall, 2011.
- [3] D. Loshin, *Business Intelligence: The Savvy Manager's Guide*. Waltham: Morgan Kaufmann, 2013. doi: <https://doi.org/10.1016/C2010-0-67240-3>.
- [4] A. Maulana and D. A. N. Wulandari, "Business Intelligence Implementation To Analyze Perfect Store Data Using the OLAP Method," *J. Publ. Informatics Eng. Res.*, vol. 3, no. 2, pp. 104–111, 2019, doi: 10.33395/sinkron.v3i2.10036.
- [5] S. A. S. Mola, N. D. Rumlaklak, and D. P. N. Polly, *Analisis Sentimen dengan Metode Random Forest*. Bandung: Kaizen Media Publishing, 2024. doi: <https://kaizenpublisher.co.id/portfolio/view/87>.
- [6] Aswi and Sukarna, *Analisis Deret Waktu : Teori dan Aplikasi*. Makassar: Andira Publisher, 2006. doi: <http://eprints.unm.ac.id/id/eprint/21944>.
- [7] W. Sulistyoningsih, I. N. Y. A. Wijaya, and H. S. Alam, "Penerapan Model Business Intelligence Pada Perusahaan Retail XLT Untuk Meningkatkan Strategi Pemasaran," *J. Ilm. Teknol. Inf. Asia*, vol. 17, no. 1, p. 33, 2023, doi: 10.32815/jitika.v17i1.893.
- [8] A. R. I. Sipayung, N. E. Zandrato, T. A. Marbun, J. N. Telaumbanua, and O. Sihombing, "Penerapan Teknologi Business Intelligence Dalam Meningkatkan Strategi Penjualan Dengan Metode OLAP Pada Cafe Le Kahve," *J. TEKINKOM*, vol. 7, no. 2, pp. 628–636, 2024, doi: 10.37600/tekinkom.v7i2.1529.
- [9] M. Anis, "Perancangan Sistem Business Intelligence Menggunakan Olap Sebagai Decision Financial Support Berbasis Web dengan Metode Waterfall," *J. Sos. Teknol.*, vol. 1, no. 10, pp. 131–144, 2021, doi: 10.59188/jurnalsostech.v1i10.217.
- [10] M. Mahendra Alvanof and R. Kesuma Dinata, "Penerapan Algoritma Random Forest dalam Deteksi dan Klasifikasi Ransomware," *J. Elektron. dan Teknol. Inf.*, vol. 5, no. 2, pp. 2721–9380, 2024, doi: <https://doi.org/10.5201/jet.v5i2.488>.
- [11] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 58–64, 2023, doi: 10.52158/jacost.v4i1.491.
- [12] M. M. Rofi, F. A. Setiawan, and F. Riana, "Perbandingan Metode K-NN Dan Random Forest Pada Klasifikasi Mahasiswa Berpotensi Dropout," *INFOTECH J.*, vol. 10, no. 1, pp. 84–89, 2024, doi: 10.31949/infotech.v10i1.8856.
- [13] H. Zhang *et al.*, "Application of Random Forest Classifier in Loan Default Forecast," in *Communications in Computer and Information Science*, 2020, pp. 410–420. doi: 10.1007/978-981-15-8101-4_37.
- [14] R. C. Putri and L. Junaedi, "Penerapan Metode Peramalan Autoregressive Integrated Moving

- Average pada Sistem Informasi Pengendalian Persediaan Bahan Baku (Studi Kasus : Toko Kue Onde-Onde Surabaya),” *J. Ilmu Komput. dan Bisnis*, vol. XIII, no. 1, pp. 164–173, 2022, doi: 10.47927/jikb.v13i1.293.
- [15] M. S. Fahmuddin, Ruliana, and S. M. Sri Mustika, “Perbandingan Metode ARIMA dan Single Exponential Smoothing dalam Peramalan Nilai Ekspor Kakao Indonesia,” *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 5, no. 3, pp. 163–176, 2023, doi: 10.35580/variasiunm193.
- [16] A. R. Pratama and Firdaus, “Perbandingan Metode ARIMA dengan Fuzzy Time Series Model Chen pada Peramalan Curah Hujan di Kota Bengkulu,” *J. Math-UMB.EDU*, vol. 11, no. 3, pp. 154–166, 2024, doi: 10.36085/mathumbedu.v11i3.6480.
- [17] S. Hanifah, F. Akbar, and R. P. Santi, “Implementasi Business Intelligence dan Prediksi Menggunakan Regresi Linear pada Data Penjualan dan Breakage di PT XYZ,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 3, pp. 144–152, 2022, doi: 10.25077/teknosi.v8i3.2022.144-152.
- [18] S. A. Moss, Larissa Terpeluk, *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison-Wesley Professional, 2003. doi: <https://dl.acm.org/doi/10.5555/773287>.
- [19] J. C. J. Martilla, John A., “Importance-Performance Analysis,” *J. Mark.*, vol. 41, no. 1, pp. 77–79, 1977, doi: <https://doi.org/10.2307/1250495>.
- [20] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (3rd Edition)*. John Wiley & Sons, 2013. doi: <https://dl.acm.org/doi/10.5555/2543973>.