BAB II
KAJIAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Pada bagian ini akan diuraikan penelitian sebelumnya yang terkait pengembangan dan model prediksi pada sistem omnichannel.

2.1.1 Penelitian Sebelumnya Terkait Sistem *Omnichannel*

Saat ini cukup banyak penelitian sebelumnya yang terkait dengan model sistem *omnichannel*. Penelitian yang dilakukan adalah merumuskan dan mensimulasikan model analitis berbasis data *omnichannel* untuk menganalisis campuran produk konsumen dan mengelola pemilihan yang sesuai. Lebih lanjut lagi, fokus penelitian lebih kepada model optimasi yang mampu memaksimalkan pendapatan dan keuntungan. Alur penelitian menunjukkan bahwa model untuk mengoptimasi *omnichannel* yang dikembangkan berdasarkan data transaksi saja yang terekam dalam sistem. Data transaksi penjualan produk baik transfer, tunai, maupun jenis pembayaran lainnya yang terekam dalam database menjadi dasar optimasi pendapatn dan keuntungan (Srivastava dkk., 2022). Penelitian lain yang mirip adalah penelitian yang dilakukan oleh (Hendalianpour dkk., 2022) yang bertujuan untuk mengoptimalkan jaringan distribusi *omnichannel* multiproduk dan multilevel serta alur pengiriman produk dalam jaringan dalam kondisi yang tidak pasti. Model matematika multi-objektif dikembangkan yang meminimalkan biaya rantai pasokan sekaligus memaksimalkan kepuasan pelanggan dalam berbagai skenario. Gambaran model sistem distribusi *omnichannel* yang diperlihatkan mengacu pada data pengiriman produk. Penelitian lain yang menarik adalah penelitian yang dilakukan oleh (Suarez & Avila, 2019) di mana pendekatan yang dilakukan telah memungkinkan perusahaan untuk mengeksploitasi teknologi digital guna memperoleh keunggulan kompetitif, penerapannya melibatkan penyelarasan proses pemasaran, penjualan, pengiriman, dan layanan serta infrastruktur informasi dan teknologi yang mendasarinya untuk mendukung operasi perusahaan yang benar. Penelitian lain yang sejenis adalah

SEKOLAH PASCASARJAN

penelitian yang dilakukan oleh (Moreira & Santos, 2024) yang menggambarkan arsitektur sistem *omnichannel* lebih kepada monitoring proses aktivitas penjualan. Sistem mampu monitoring melalui visualisasi *open data* kesehatan yang terekam dalam sistem. Selain itu terdapat penelitian lain yang menarik yaitu penelitian yang dilakukan oleh (Lin dkk., 2020b) di mana sistem *omnichannel* yang diusulkan memiliki robot cerdas yang mampu membaca perilaku konsumen, namun teknologi robot cerdas yang diusulkan tidak disebutkan dengan detail. Penelitian yang sejenis juga mengusulkan adanya fitur analitik yang harus di tempatkan pada lapisan sistem *omnichannel*, namun juga tidak dijelaskan dengan detail dna sebatas sebagai usulan (Lin dkk., 2020b). Berdasarkan uraian terkait penelitian sebelumnya bahwa beberapa penelitian terkait pengembangan sistem *omnichannel* bersepakat bahwa perlu ada fitur analitik data pada sistem *omnichannel* yang dikembangkan. Namun secara umum tidak disebutkan fitur analitik seperti apa yang diperlukan berdasarkan karakteristik konsumen *omnichannel*.

2.1.2 Penelitian Sebelumnya Terkait Prediksi pada Sistem *Omnichannel*

Model prediksi penjualan telah banyak dilakukan sebagaimana pada model distribusi *singlechannel*, *multichannel* maupun *omnichannel*. Dalam Tujuh tahun terakhir terakhir terdapat banyak penelitian yang membahas terkait analisis dalam membaca pola perilaku konsumen. Namun analisis yang dilakukan mayoritas lebih banyak pada model layanan *singlechannel* dan berbasis pada dimensi tertentu yang menjadi variabel perilaku konsumen. Berbagai metode digunakan untuk melakukan analisis perilaku konsumen. Penelitian yang ada kebanyakan menggunakan metode yang relatif sederhana yaitu pendekatan deskriptif sebagaimana dilakukan oleh (Spáčil & Teichmannová, 2016) dan (Guthrie dkk., 2021). Kedua penelitian tersebut menggambarkan perilaku konsumen yang didasarkan pada data deskriptif yang diperoleh melalui *survey online* dan observasi di lapangan.

Penelitian lain dilakukan dengan pendekatan regresi sebagaimana dilakukan oleh (Escobar-Rodríguez & Bonsón-Fernández, 2017), penelitian ini

SEKOLAH PASCASARJANA

menjelaskan niat pembelian produk *fashion* secara *online*. Dari sudut pandang praktis, terdapat implikasi manajerial yang relevan dalam temuan kasus *fashion e-commerce*. Studi ini mengidentifikasi dua faktor utama yang mempengaruhi niat pembelian *fashion online* yang bisa dilakukan oleh manajer komunikasi dan pemasaran yaitu kepercayaan dan nilai yang dirasakan oleh konsumen. Pendekatan metode regresi juga digunakan dalam penelitian (Klein & Sharma, 2022) dengan hasil penelitiannya menunjukkan bahwa keterlibatan secara signifikan proses mediasi hubungan antara rekreasional, hedonistik, kesadaran pada harga, kebiasaan, kesetiaan pada merek, dan kebingungan pada pengambilan keputusan konsumen yang dikarenakan terlalu banyaknya pilihan.

Selain itu ada pula penelitian yang menggunakan pendekatan regresi berganda yang dilakukan oleh (Eger dkk., 2021), mekanisme penelitian tersebut adalah mencoba menghubungkan daya tarik rasa takut, seperti rasa ketakutan akan kesehatan dan ketakutan pada ekonomi. Rasa takut ini dikaitkan dengan perubahan perilaku konsumen dan pengaruh belanja tradisional dan *online* terkait COVID-19.

Hasil penelitian menunjukkan perbedaan yang signifikan serta persamaan perilaku konsumen antar generasi. Metode analisis faktor juga digunakan dalam menggambarkan perilaku konsumen pengguna *E-Commerce* pada penelitian yang dilakukan oleh (Huseynov & Özkan Yıldırım, 2019). Pada awalnya, penelitian ini dilakukan dengan menganalisis segmentasi pasar dari sudut pandang psikografis. Selanjutnya, ditemukan empat segmen konsumen *online* yang berbeda, dan kemudian perilaku belanja dari masing-masing segmen tersebut diidentifikasi dan dinilai dengan menggunakan model evaluasi perilaku yang telah dikembangkan sebelumnya. Hasil penelitian ini memberikan informasi penting bagi *e-retailer* tentang karakteristik perilaku masing-masing segmen konsumen. *E-retailer* dapat menggunakan temuan ini untuk mengatur sumber daya pemasaran mereka dan membuat bauran pemasaran yang lebih efektif untuk masing-masing segmen konsumen.

SEKOLAH PASCASARJANA

Analisis faktor juga digunakan dalam penelitian yang dilakukan oleh (Bucko dkk., 2018) di mana tujuan penelitiannya adalah untuk mengetahui faktor-faktor yang mempengaruhi keinginan konsumen dalam membeli produk dari toko *online*. Penelitian tersebut mengevaluasi kriteria berdasarkan bagaimana konsumen membuat keputusan saat membeli secara *online*. Penelitian ini mengeksplorasi tujuh faktor, di mana faktor yang paling menonjol adalah faktor harga yang menjelaskan bagian terbesar dari *varians* dalam data.

Penelitian dari (THAM dkk., 2019) juga menggunakan pendekatan analisis faktor untuk mengkaji dampak risiko keuangan, risiko kenyamanan, non-risiko tidak terkirim; risiko kebijakan pengembalian dan risiko produk pada perilaku konsumen *online* konsumen Malaysia. Hasil penelitian ini menunjukkan bahwa risiko produk, risiko kenyamanan, dan risiko kebijakan pengembalian memiliki pengaruh yang signifikan dan berdampak positif pada perilaku belanja *online*. Risiko keuangan ditemukan memiliki efek yang tidak signifikan dan negatif terhadap perilaku konsumen. Selain itu, risiko non-pengiriman ditemukan memiliki dampak yang signifikan dan negatif terhadap perilaku belanja *online*. Temuan penelitian ini memberikan model yang berguna untuk mengukur dan mengelola risiko yang dirasakan dalam belanja *online* yang dapat mengakibatkan peningkatan partisipasi konsumen.

Hal yang sama juga dilakukan oleh (De Canio & Fuentes-Blasco, 2021) di mana dalam penelitiannya mengusulkan segmentasi pengguna internet dengan menerapkan analisis faktor dan *cluster* untuk membagi pengguna menjadi tiga kelompok, berdasarkan tujuan penggunaan utama yang dilaporkan dari berbagai aplikasi dan alat internet pada waktu tertentu. Hasil analisis memperlihatkan penggalian faktor utama yang mencakup sikap yang berbeda terhadap penggunaan internet serta refleksi segmen diperoleh implikasinya untuk pengembangan manajemen *e-commerce*.

Pendekatan lain juga digunakan dalam menganalisis perilaku konsumen pada *multichannel* dengan pendekatan analisis jalur, pendekatan ini dilakukan oleh (Niemczyk, 2014). Hasil penelitian menunjukkan bahwa konsumen

SEKOLAH PASCASARJANA

dengan sifat *haptic* yang kuat lebih menyukai saluran fisik dan seluler. Dimensi *autotelic* adalah kunci dalam saluran *online*. Hasil temuan dalam penelitian mendukung penerapan strategi *multichannel* yang efektif di antara pengecer produk *high-haptic* serta menunjukkan bahwa media ponsel menjadi alternatif yang berharga untuk belanja di dalam toko.

Beberapa penelitian berkembang dengan menggunakan pendekatan *Structure Equation Model* (SEM-PLS) sebagaimana penelitian yang dilakukan oleh (Chauhan dkk., 2021) pada model *singlechannel*. Hasil penelitian tersebut menunjukkan pengaruh *Hedonic Shopping Value* (HSV) dan *Positive Emotions* (PE) yang signifikan dan positif terhadap impulse buying (IB), berbeda dengan *Fashion Involvement* (FI) dan *Sales Promotion* (SP) yang tidak menunjukkan pengaruh yang signifikan. Selain itu, PE adalah mediator yang signifikan dalam hubungan konstruksi. Penelitian sejenis juga dilakukan oleh (Lestari, 2019) di mana Analisis *Smart Partial Least Square* (Smart PLS) yang digunakan mengungkapkan hubungan yang kuat antara *self-efficacy*, kegunaan yang dirasakan, sikap, niat, dan adopsi *e-commerce*.

Masih pada bentuk *singlechannel* juga terdapat penelitian yang dilakukan oleh (Yang dkk., 2021) yang menggunakan pada SEM-PLS pada *mobile channel*. Hasil penelitian menunjukkan bahwa rangsangan lingkungan secara signifikan mempengaruhi nilai yang dirasakan konsumen yaitu, nilai utilitarian yang dirasakan dan nilai hedonis yang dirasakan, dan persepsi konsumen tentang nilai hedonis secara signifikan langsung berdampak pada *Impulse Buying Behavior* (IBB).

Penelitian dengan menggunakan SEM-PLS juga digunakan untuk menganalisis perilaku konsumen di *multichannel* sebagaimana penelitian yang dilakukan oleh (Nofrizal dkk., 2023). Menurut penelitiannya, duta merek tidak membuat konsumen setia, tetapi kualitas produk dan kepercayaan dapat mendorong loyalitas. Pada bentuk *omnichannel* terdapat juga penelitian dengan metode SEM-PLS yang dilakukan oleh (Cheah dkk., 2022) yang memperlihatkan bahwa persepsi konsumen tentang integrasi saluran,

SEKOLAH PASCASARJANA

pemberdayaan konsumen, dan kepercayaan berpengaruh signifikan terhadap kebiasaan dalam ritel *omnichannel*.

Pendekatan penelitian dengan metode UTAUT2 juga digunakan untuk menganalisis perilaku konsumen sebagaimana dilakukan oleh (Juaneda-Ayensa dkk., 2016) pada *omnichannel*. Hasil penelitian ini menunjukkan bahwa niat konsumen untuk membeli pada saluran toko yang tidak biasa dipengaruhi oleh inovasi pribadi, harapan, dan kinerja.

Selain penelitian di atas, metode yang digunakan dalam menganalisis perilaku konsumen juga berkembang dengan menggunakan pendekatan *Data mining*. Salah satunya adalah penelitian yang mengulas tentang dampak ledakan data terhadap prediksi produk dan bagaimana meningkatkannya dengan menggunakan pendekatan analisis konsumen, MAPE dan deret waktu sehingga menghasilkan kesimpulan bahwa ukuran dan konten yang tidak terstruktur dan sangat besar serta pengalaman konsumen dan rantai pasokan yang selalu berubah membuat perusahaan harus mengatasi tiga masalah yaitu privasi, keamanan dan tata kelola (Boone dkk., 2019).

Penelitian lain yang terkait adalah mengusulkan dan menguji beberapa metode baru untuk memperkirakan ketidakpastian, memanfaatkan perhitungan empiris, dan simulasi untuk menentukan kuantil guna mengoptimalkan tingkat perkiraan yang tepat dari distribusi penjualan. Pendekatan yang digunakan adalah *Simple Exponential Smoothing* (SES), *Exponential Smoothing Models* (ESM), *Autoregressive Integrated Moving Average models* (ARIMA), *Quantile Empirical Estimation* (QEE) sehingga disimpulkan bahwa metode yang relatif sederhana menghasilkan kinerja inventaris yang lebih baik daripada pendekatan yang lebih canggih dan intensif (Spiliotis dkk., 2021).

SEKOLAH PASCASARJANA

Pada penelitian lain menggunakan model prediksi (*forecasting)* pada *singlechannel* yang mengusulkan dua tahap pendekatan siklus hidup prediksi untuk produk teknologi konsumen dengan data penjualan yang terbatas yaitu *clustering* segmentasi produk berdasarkan market dan menerapkan model *Bass* untuk produk agregat dalam kelompok menggunakan penjualan periodik rata-rata (Li dkk., 2021). Beberapa penelitian model prediksi juga dilakukan dalam

pendekatan atau metode yang berbeda seperti penelitian yang menyajikan kerangka *meta-learning* berdasarkan jaringan saraf *convolutional*. Penelitian mempelajari representasi fitur dari data mentah deret waktu penjualan secara otomatis dan kemudian menautkan fitur yang dipelajari dengan serangkaian bobot yang digunakan untuk menggabungkan kumpulan metode prediksi dasar (Ma & Fildes, 2021).

Pendekatan lain juga dilakukan dengan mengadaptasi model campuran untuk memperkirakan transaksi konsumen dan meningkatkan prediksi dengan menyajikan metodologi *Bayesian* yang menggabungkan tren yang bervariasi waktu, musim, harga, promosi, efek acak. (Berry dkk., 2020), pendekatan perancangan model yang dapat secara akurat meramalkan penjualan rantai pasokan dengan menggunakan pendekatan *lightGBM* dan LSTM (Weng dkk., 2019), dan penelitian yang memecahkan masalah akurasi yang rendah dalam prediksi permintaan produk baru yang disebabkan oleh tidak adanya data historis dan pertimbangan yang tidak memadai dari faktor-faktor yang mempengaruhi. Pada penelitian ini menggunakan pendekatan *hybrid* dengan menggabungkan pengukuran kemiripan produk, bobot atribut kesamaan produk diwujudkan dengan menggunakan *fuzzy clustering-rough set* lalu kesalahan prediksi model *Bass* disesuaikan (Yin dkk., 2020).

Penelitian (Androniceanu dkk., 2020) melakukan analisis *K-Means Clustering* untuk mengetahui dan menganalisis perilaku konsumen Uni Eropa dalam lingkungan *online*. Hasil penelitian mengarah pada identifikasi lima *cluster* di mana negara-negara Uni Eropa dikelompokkan sesuai dengan karakteristik dan perilaku konsumen di lingkungan *online*. Pendekatan *clustering* juga dilakukan oleh (Kondo & Okubo, 2022) yang mengidentifikasi perilaku konsumen pada *multichannel*. Studi ini mengeksplorasi bagaimana konsumen memanfaatkan beberapa saluran dalam lingkungan ritel. Penelitian ini menyediakan kerangka kerja yang kompatibel untuk memperoleh segmentasi konsumen pada keseluruhan produk dan kategori produk utama yang berfokus secara *online* yang mencerminkan kebutuhan pembelian yang

SEKOLAH PASCASARJANA

dinamis di berbagai saluran, menggunakan analisis klaster kelas laten dan berfokus pada karakteristik demografis.

Berdasarkan literatur yang berhasil dikumpulkan, justru yang paling menarik adalah penelitian yang dilakukan oleh (Hernandez dkk., 2017), di penelitiannya menggunakan pendekatan *process mining* pada perilaku konsumen pada *Web E-commerce*. Penelitian difokuskan pada rekaman kejadian konsumen dalam penggunaan Web. Hasil penelitiannya memperlihatkan bahwa rekaman proses yang dilakukan oleh konsumen dapat menjadi masukan bagi pemilik *Web E-commerce* untuk memperbaiki fiturnya.

Tidak banyak penelitian model prediksi pada distribusi *multichannel* dan *omnichannel* dalam 5 tahun terakhir. Penelitian terkait model prediksi pada *multichannel* salah satunya adalah mengusulkan kerangka kerja prediksi lintas-temporal baru untuk menghasilkan prakiraan yang koheren di semua tingkat rantai pasokan retail pada *multichannel*. Penelitian ini menghasilkan prakiraan *bottom-up* yang lebih akurat daripada prakiraan *top-down* ketika data penjualan digunakan untuk prediksi dalam rantai pasok ritel *online* dan *offline* (Punia dkk., 2020).

Sementara model prediksi pada *omnichannel* sebagaimana penelitian yang dilakukan oleh (Hense & Hübner, 2022) yaitu melakukan analisis masalah yang terdapat pada penjualan di sistem *omnichannel* dengan menitikberatkan pada interaksi permintaan konsumen di seluruh saluran menggunakan pendekatan *NP-Hard* dan non-linear *multi-knapsack*, sehingga dihasilkan kesimpulan yaitu mampu mewujudkan peningkatan keuntungan tergantung pada besarnya tarif substitusi saluran yang berpotensi dan bukan membagi biaya sama rata ke seluruh saluran. Terdapat juga penelitian terkait prediksi dalam pengembangan *omnichannel* sebagaimana penelitian yang dilakukan oleh (Byrne, 2016). Penelitian yang dilakukan menitikberatkan pada atribut pembelian pada saluran online. Atribut-atribut tersebut digunakan sebagai variabel prediktor dengan pendekatan regresi yang difokus pada data empiris menggunakan data penjualan dan keranjang yang terkait dengan

SEKOLAH PASCASARJANA

berbagai macam 24.000 produk dari pengecer kosmetik besar di Perancis yang dijual melalui saluran ritel *online* dan fisik.

Selain itu pula terdapat penelitian yang mengusulkan pendekatan prediktif untuk menyelesaikan tantangan bagi *retailer* dalam meningkatkan akurasi prediksi permintaan saluran *offline* dan *online* serta memprediksi permintaan di berbagai saluran sehingga bisa mengurangi ketidakpastian rantai pasok pada penjualan *omnichannel*. Metode yang digunakan yaitu menggabungkan *clustering* dengan jaringan saraf tiruan untuk menangani ketidakpastian permintaan (Pereira & Frazzon, 2019).

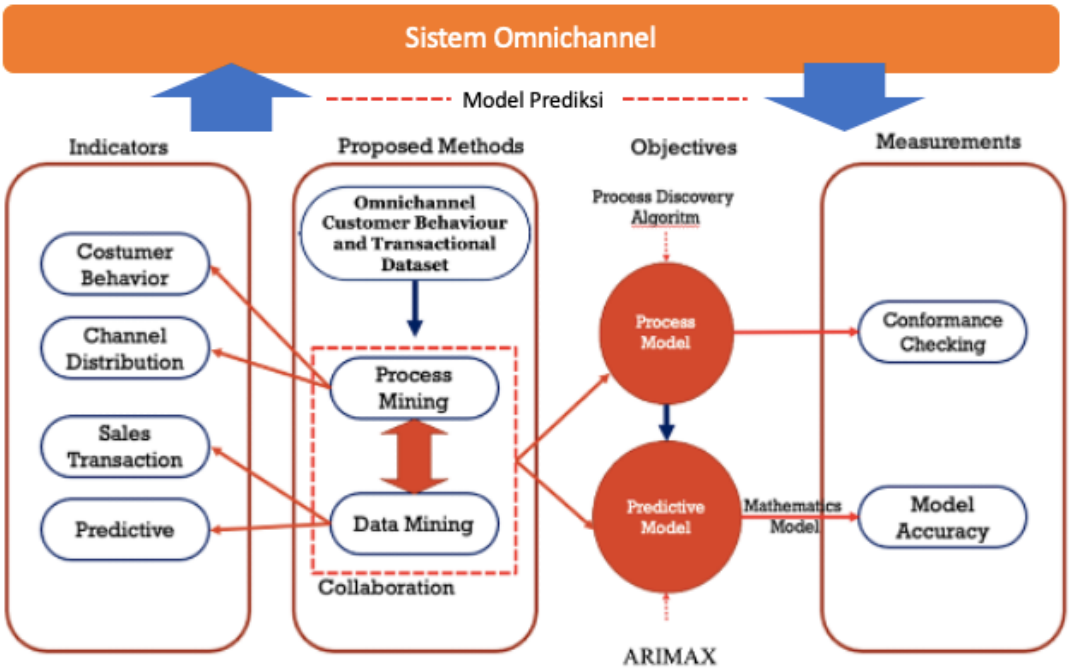
Berdasarkan pada penelitian-penelitian sebelumnya, penelitian terkait dengan pengembangan sistem *omnichannel* dan model prediksi pada *omnichannel* berdasarkan perilaku konsumen sudah ada namun sangat terbatas. Kebanyakan konsep yang digunakan hanya berdasar pada data transaksi penjualan semata dan perlunya fitur analitik data dalam *omnichannel* namun fitur analitik yang disebutkan tidak spesifik menyebutkan metode yang digunakan. Hal inilah yang menjadi dasar usulan penelitian disertasi ini yaitu mengembangkan model prediksi berdasarkan perilaku konsumen yang dinamis pada sistem *omnichannel* menggunakan metode *hybrid* yaitu metode *process mining* dengan menggunakan *Process Discovery Algorithm* dan metode *Data Mining* dengan menggunakan ARIMAX.

2.2 Keaslian Penelitian

Mayoritas penelitian terkait dengan analisis perilaku konsumen menggunakan survei opini. Hal ini menunjukkan bahwa pola perilaku konsumen yang dihasilkan sangat tergantung pada kondisi responden, latar belakang responden serta aspek psikologis saat proses pengisian survei. Sedangkan pada penelitian ini, analisis perilaku konsumen didasarkan pada aktifitas konsumen yang terekam langsung pada platform sistem *omnichannel* dalam bentuk *event log*. Penelitian ini, merupakan penelitian yang mengembangkan metode *hybrid* yaitu *process discovery* dari *process mining* dan pendekatan model ARIMAX untuk memprediksi penjualan berdasarkan perilaku konsumen pada sistem distribusi

omnichannel.

Untuk memudahkan dalam pemaparan konsep-konsep pada penelitian ini diperlukan adanya kerangka berpikir yang divisualisasikan dalam bentuk bagan atau juga skema. Kerangka berpikir ini, juga dikenal sebagai model konseptual, dibangun berdasarkan pertanyaan penelitian. Pendekatan ini menggambarkan kumpulan konsep serta hubungan antara konsep-konsep tersebut (Polančič dkk., 2009). Model konseptual usulan penelitian ini sebagaimana diperlihatkan dalam Gambar 2.1.



Gambar 2.1 Model Konseptual Penelitian

Sebagaimana diperlihatkan pada Gambar 2.1 bahwa Model Konseptual Penelitian ini terbagi menjadi 5 skema, yaitu : Sistem *Omnichannel*, *Indicators*, *Proposed Methods*, *Objectives* dan *Measurements.* Pada penelitian ini, sistem *omnichannel* merupakan model utama pada skema prediksi yang merupakan bagian dari sistem. Sistem *omnichannel* yang meliputi berbagai saluran di dalamnya menjadi tempat konsumen melakukan aktivitas di mana segala aktivitas tersebut terekam dalam sistem yang berbentuk *dataset*. *Dataset* yang

digunakan adalah *dataset* perilaku perpindahan konsumen antar saluran dan transaksi yang dilakukan konsumen pada sistem *omnichannel*. Metode prediksi dalam sistem *omnichannel* yang dikembangkan sebagaimana terlihat pada skema obyektif adalah penggunaan metode *hybrid* antara *Process Mining* dan *Data Mining* yang nantinya masing-masing akan menghasilkan indikator penilaian. Dalam penggunaan metode *Process Mining* akan menghasilkan pola perpindahan konsumen dari saluran satu ke saluran yang lain atau disebut sebagai *Customer Behaviour* dan juga bisa dilihat saluran mana yang paling optimal dan banyak digunakan oleh konsumen. Sedangkan pada penggunaan metode *Data Mining* menghasilkan pola transaksi konsumen yang di mana dapat dilihat pola penjualan yang nantinya bisa digunakan untuk memprediksi nilai penjualan di masa yang akan datang.

Setiap model prediksi yang dihasilkan diperlukan adanya pengukuran. Hal itu tergambar pada skema *Measurement* yaitu pengukuran pada *Process Discovery* menggunakan *Conformance Checking* sedangkan pada model prediksi yang berupa model matematis pengukurannya menggunakan *Model Accuracy.*

2.3 Landasan Teori

Berikut ini adalah beberapa landasan teori yang terkait dengan penelitian yang dilakukan.

2.3.1 Saluran Distribusi

Kebutuhan pemilik bisnis terkait dengan penjualan sangat beragam dan dapat bervariasi tergantung pada jenis bisnis, skala bisnis, dan tujuan yang ingin dicapai. Namun, secara umum kebutuhan utama pemilik bisnis terkait dengan penjualan adalah peningkatan penjualan itu sendiri**.** Tujuan utama setiap bisnis adalah menghasilkan keuntungan, dan penjualan adalah kunci untuk mencapai tujuan tersebut. Hal tersebut akan dapat tercapai bila pemilik bisnis dapat memahami konsumen. Pemahaman terkait kebutuhan, keinginan, dan perilaku konsumen adalah kunci untuk mengembangkan produk atau jasa yang relevan dan efektif. Kemudian yang tidak kalah penting adalah membangun hubungan

SEMARANG
SEKOLAH PASCASARJANA

yang kuat dengan konsumen. Membangun hubungan yang kuat dengan konsumen akan menciptakan loyalitas dan meningkatkan kemungkinan terjadinya pembelian berulang (*repeat order*).

Terkait dengan hal tersebut, pendekatan saat ini yang dilakukan adalah dengan membangun saluran distribusi. Saluran distribusi (*Distribution Channel*) merupakan saluran yang saling terhubung satu sama lain dalam membantu produk atau jasa yang bisa didapatkan atau dikonsumsi oleh konsumen (Kotler, 2008). Secara umum, saluran distribusi melakukan pembagian atau penyaluran produk ke pihak-pihak terkait seperti distributor, agen, *reseller* atau konsumen akhir langsung yang di mana dalam penelitian ini berfokus pada penyaluran produk ke konsumen akhir langsung.

Pada prinsipnya saluran distribusi bukan sekedar dipandang sebagai sarana penjualan, tetapi sebagai pintu terdepan dalam membangun keterlibatan konsumen melalui saluran-saluran distribusi ini, sehingga pelaku bisnis dapat menggali kebutuhan dan keinginan konsumen serta membangun hubungan kedekatan dengan konsumen (*Customer Relationship Management*). Untuk memastikan bahwa produk dikirim pada waktu yang tepat, dalam kuantitas yang tepat, dengan kualitas yang tepat, dan ke lokasi yang tepat dengan biaya yang paling hemat, pengelolaan logistik yang andal diperlukan selama proses pergerakan produk melalui saluran distribusi (Kembro dkk., 2018).

Pendekatan modern memperlihatkan bahwa terdapat 2 (dua) model saluran distribusi yang populer saat ini yaitu *Multichannel* dan *Omnichannel* (Melacini dkk., 2018). *Multichannel* adalah ritel atau pengecer yang menjual barang dagangan atau jasa melalui lebih dari satu saluran dari *online store* maupun *offline store* (toko fisik), tetapi setiap saluran berdiri sendiri dengan aturan dan operasionalnya masing-masing (Levy dan Weitz, 2013). Sedangkan *omnichannel* adalah suatu transformasi model saluran *multichannel* yang saling mengintegrasikan setiap saluran yang dimiliki oleh pelaku bisnis dan setiap salurannya dikelola dalam manajemen operasional yang menyatu dengan tujuan untuk memberikan pengalaman terbaik kepada konsumen (Cao & Li, 2018).

SEKOLAH PASCASARJANA

Tabel 2.1 Perbedaan antara *Multichannel* dan *Omnichannel* berdasarkan Sumber Daya Teknologi Informasi (Tridalestari, Mustafid, & Jie, 2022).

Sumber Daya Teknologi Informasi	***Multichannel Framework***	***Omnichannel Framework***
People	Pengguna dalam hal ini konsumen sangat tergantung dari segmentasi saluran distribusi. Pengelola membutuhkan a*gent* yang relatif lebih banyak.	Segmentasi konsumen lebih luas karena saluran distribusi yang terintegrasi. Jumlah *agent* lebih sedikit karena efisiensi dan efektifitas saluran.
Infrastruktur	Infrastruktur relatif lebih murah dalam implementasinya karena menambah saluran baru cenderung tidak mengganggu saluran lainnya	Infrastruktur sangat kompleks, biaya akan sangat mahal di awal karena mengintegrasikan saluran yang akan digunakan dalam manajemen rantai pasok.
Aplikasi	Pembangunan aplikasi pada model ini relatif lebih mudah karena sangat tergantung pada jenis saluran yang disediakan	Pembangunan aplikasi pada model *omnichannel* lebih kompleks karena saluran yang disediakan harus terintegrasi
Informasi	Informasi yang disediakan tidak terintegrasi karena saluran yang disediakan bekerja sesuai fungsi dan segmentasinya masing-masing.	Informasi yang dikelola terintegrasi sehingga lebih mudah dalam hal monitoring data dan informasi konsumen.

2.3.2 Sistem *Omnichannel*

Sistem *omnichannel* adalah strategi bisnis yang mengintegrasikan berbagai saluran komunikasi dan distribusi perusahaan untuk menciptakan pengalaman konsumen yang konsisten dan terpadu. Dalam konteks ini, "*omnichannel*" mengacu pada semua saluran (*online, offline, mobile*, dan lain-lain) yang digunakan perusahaan untuk berinteraksi dengan konsumen.

Berikut adalah beberapa karakteristik utama dari sistem *omnichannel* (Hübner dkk., 2016b):

1. Integrasi saluran, semua saluran, termasuk toko fisik, situs web, aplikasi mobile, media sosial, dan layanan konsumen, terhubung satu sama lain. Informasi tentang produk, stok, dan konsumen dapat diakses dan diperbarui secara *real-time* di semua saluran.
2. Pengalaman konsumen yang konsisten, konsumen mendapatkan pengalaman yang konsisten dan mulus di semua saluran. Mereka dapat memulai interaksi di satu saluran dan melanjutkannya di saluran lain tanpa kehilangan informasi atau kualitas layanan.
3. Personalisasi, sistem ini memungkinkan personalisasi yang lebih baik karena data dari berbagai saluran dikumpulkan dan dianalisis untuk memahami preferensi dan perilaku konsumen. Hal ini memungkinkan perusahaan untuk memberikan rekomendasi produk yang lebih relevan dan layanan yang lebih baik.
4. Fleksibilitas dalam pembelian, konsumen memiliki berbagai opsi untuk membeli produk, seperti membeli online dan mengambil di toko (*click-and-collect*), membeli di toko dan mengatur pengiriman ke rumah, atau mengembalikan barang yang dibeli secara online ke toko fisik.
5. Efisiensi operasional, dengan integrasi saluran dan manajemen inventaris yang lebih baik, perusahaan dapat meningkatkan efisiensi operasional, mengurangi biaya, dan meningkatkan kepuasan konsumen.

Implementasi sistem *omnichannel* memerlukan investasi dalam teknologi dan perubahan proses bisnis, tetapi dapat memberikan keuntungan kompetitif yang signifikan dengan meningkatkan loyalitas konsumen dan meningkatkan pendapatan. Gambaran besar lingkungan sistem *omnichannel* dapat diperlihatkan pada Gambar 2.2.

SEKOLAH PASCASARJANA

Gambar 2.2 Contoh Gambaran Besar Sistem *Omnichannel*

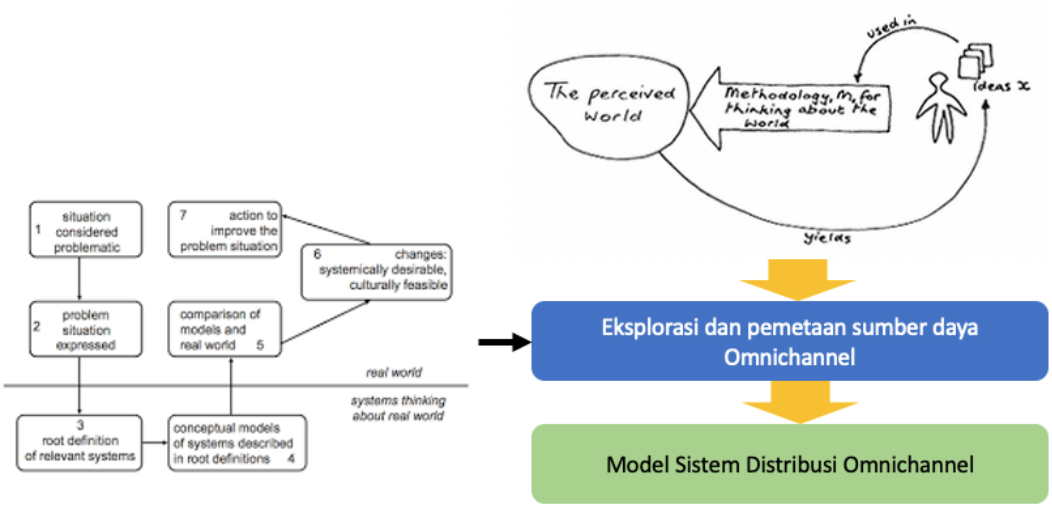
Gambar 2.2 memperlihatkan bahwa sistem *omnichannel* memiliki komponen yaitu sistem, perilaku konsumen, dan saluran komunikasi dan distribusi. Pada komponen sistem *omnichannel* dibutuhkan beberapa sumber daya yang dijadikan inputan dalam menjalankan proses sistem ini yaitu aplikasi-aplikasi saluran distribusi yang dimiliki oleh perusahaan dan ranah operasional berupa data transaksi serta infrastruktur teknologi informasi yang dimiliki oleh perusahaan. Adapun saluran pada sistem *omnichannel* yang menjadi fokus pada penelitian ini adalah saluran digital yaitu *Marketplace*, Media Pesan (*Messenger Media*), media sosial (*Social Media*), *Social Media Shop*, dan *Webstore.*

2.3.3 Skema Pemodelan Sistem Distribusi *Omnichannel* dengan Konsep *Soft Systems Methodology* (SSM)

Pada pertengahan tahun 1970-an, Checkland dan rekan-rekannya mengembangkan pendekatan yang disebut pendekatan *Soft System Methodology* (SSM), yang dikenal sebagai SSM, karena fokusnya pada pembelajaran dan inovasi pada masalah yang dihadapi (Checkland & Poulter,

SEKOLAH PASCASARJANA

2020). Fokus SSM adalah untuk menciptakan sistem aktivitas dan hubungan manusia dalam sebuah organisasi atau grup dalam rangka mencapai tujuan bersama, yang merupakan sebuah metodologi partisipatori yang dapat membantu memahami perspektif masing-masing stakeholder.



Gambar 2.3. Cara Pandang Membangun SSM pada Pengembangan Model Sistem *Omnichannel*

SSM bergantung pada partisipasi pengguna dalam proses untuk memahami masalah dan cara memperbaikinya, sehingga setiap orang yang terlibat merasa memiliki masalah tersebut dan berkomitmen untuk melakukan perubahan (Al-Harrasi, 2015). SSM adalah metodologi yang tepat untuk membantu suatu organisasi atau perusahaan menjelaskan sekaligus merancang sistem aktivitas manusia untuk mencapai tujuan tersebut. Tahapan proses SSM dimulai dengan memberikan klarifikasi terhadap situasi masalah yang tidak terstruktur melalui perancangan sistem aktivitas manusia. Setelah menghasilkan model konseptual, model tersebut kemudian dibandingkan dengan situasi masalah untuk menemukan perubahan yang diperlukan. Pendekatan SSM yang dilakukan dalam penelitian ini digunakan sebagai langkah pengembangan model sistem *omnichannel* yang akan dikembangkan sesuai kebutuhan berdasarkan masalah yang terjadi di lapangan saat ini.

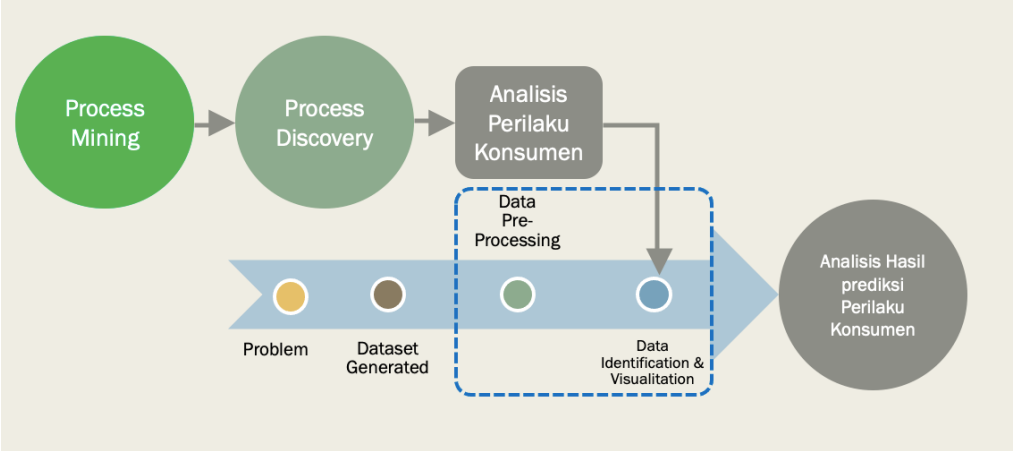
SEMARANG
SEKOLAH PASCASARJANA

2.3.4 Skema Model Perilaku Konsumen pada Sistem Distribusi *Omnichannel*

Proses dan tindakan individu atau kelompok dalam mempertimbangkan, memilih, membeli, memanfaatkan, dan mengevaluasi sebuah produk untuk memenuhi kebutuhan mereka dikenal sebagai perilaku konsumen (Kotler & Keller, 2016). Perilaku konsumen juga membahas bagaimana pelanggan membuat keputusan tentang produk, mulai dari menerima, membeli, memanfaatkan, dan menentukan barang dan jasa yang digunakan (Donavan dkk., 2016). Oleh karena itu, perilaku konsumen dapat didefinisikan sebagai kondisi di mana konsumen mencari, memilih, membeli, menggunakan, dan mengevaluasi barang dan jasa untuk memenuhi kebutuhan dan keinginan mereka, yang pada akhirnya menyebabkan keputusan untuk membeli barang tersebut.

Perkembangan digital saat ini telah mengubah cara orang berbelanja di toko online. Konsumen tidak hanya berbelanja di *desktop* mereka tetapi juga di berbagai perangkat elektronik. Mereka bahkan terhubung dengan merek melalui berbagai saluran, seperti toko fisik, situs web, dan media sosial (Carnein dkk., 2017). Dikarenakan kemajuan dalam teknologi dan jaringan internet membuat konsumen mudah berpindah saluran untuk memenuhi keinginan dan kebutuhan akan barang yang diinginkan, sehingga konsumen yang menggunakan teknologi ini sangat luas dan dapat mengakses situs web toko online melalui tablet, mengakses halaman merek melalui Facebook, dan menggunakan *smartphone* untuk mendaftarkan email melalui promosi yang ditemukan di mesin pencari. Oleh karena itu, layanan sistem *omnichannel* memainkan peran penting bagi merek dan bisnis untuk memungkinkan pengalaman berbelanja pelanggan yang lebih baik. Hal ini disebabkan oleh kemajuan teknologi yang memudahkan navigasi konsumen untuk mendapatkan pengalaman belanja yang integrasi. Beberapa faktor yang mempengaruhi konsumen *omnichannel* termasuk kemudahan yang didapat, harga yang bersaing, dan pengalaman belanja yang didapat oleh pelanggan. Saat ini, layanan sistem *omnichannel* memiliki peran yang sangat penting bagi merek dan bisnis untuk menyediakan pengalaman (Chopra, 2016).

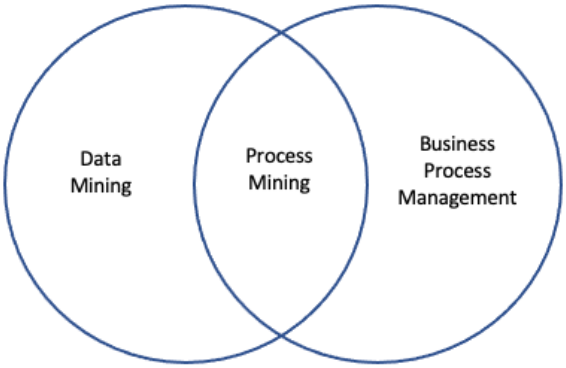
SEKOLAH PASCASARJANA



Gambar 2.4 Skema Analisis Perilaku Konsumen dengan Menggunakan *Process Discovery* dari *Process Mining*

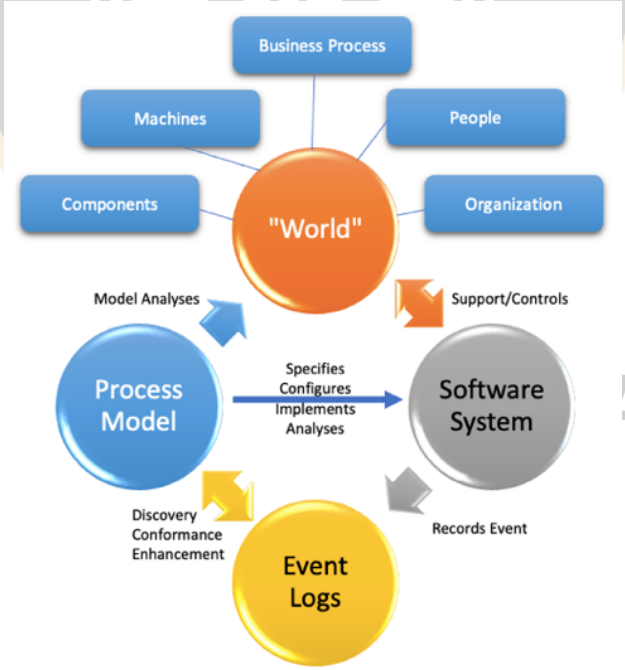
Gambar 2.4 memperlihatkan pendekatan yang dilakukan dalam penelitian ini untuk menganalisis perilaku konsumen pada sistem distribusi *omnichannel* dengan menggunakan *process discovery* dari *process mining*. *Process discovery* pada dasarnya digunakan sebagai teknik untuk menemukan dan memvisualisasikan proses bisnis yang sebenarnya terjadi berdasarkan data *event log* yang ada. Dalam konteks perilaku konsumen pada sistem distribusi *omnichannel*, data *event log* ini bisa berupa data transaksi, data kunjungan website, data interaksi aplikasi, atau data lainnya yang merekam aktivitas konsumen.

Pada dasarnya *process mining* merupakan teknik yang membantu organisasi menganalisis dan memantau proses internal untuk mengidentifikasi aktivitas maupun r*esources* yang dimiliki (vom Brocke dkk., 2021). Di masa lalu, organisasi akan mengelola proses ini melalui serangkaian wawancara yang menghasilkan ringkasan bisnis yang ideal secara keseluruhan. Dengan *proses mining*, perusahaan dapat secara otomatis memantau dan melacak proses bisnis secara *real-time* tanpa campur tangan manusia. Metode *proses mining* menggunakan berbagai teknologi, seperti *data mining*, analitik proses mendalam, dan kecerdasan bisnis berbasis AI (Okoye dkk., 2018). Untuk melihat hubungan antara *process mining* dan *data mining* sebagaimana diperlihatkan dalam Gambar 2.5.



Gambar 2.5 hubungan antara *process mining* and *data mining*

Gambar 2.5 memperlihatkan bahwa *process mining* dan *data mining* terhubung melalui konsep eksplorasi manajemen proses bisnis (Bogarín dkk., 2018). *Process mining* diciptakan sebagai jembatan solusi antara konsep *data mining* dan manajemen proses bisnis. Jadi dalam hal ini penggunaan *process mining* dan *data mining* bisa dilakukan secara bersamaan untuk menemukan analisis yang lebih holistik daripada hanya menggunakan salah satu saja (Tridalestari, Warsito, dkk., 2022). Secara umum konsep *process mining* sebagaimana diperlihatkan dalam Gambar 2.6.



Gambar 2.6 *Process Mining Framework* (Prasetyo dkk., 2021)

Gambar 2.6 memperlihatkan bahwa mekanisme *process mining* diawali dengan menangkap *event log* yang diperoleh dari sistem sehari-hari yang dijalankan oleh perusahaan, kemudian pada *event log* tersebut dilakukan mekanisme *process mining* yang terdiri dari 3 tahap yaitu *Process Discovery*, *Conformance checking*, dan *enhancement* (W. Van Der Aalst, 2012).

Tiga tahapan tersebut dijalankan sesuai kebutuhan analisis tanpa perlu berurutan namun bisa saja terjadi secara random. Pada langkah selanjutnya adalah melakukan analisis terhadap model proses yang dihasilkan dari tahapan *proses mining*. Dalam usulan penelitian ini, mekanisme *Process Mining* yang digunakan adalah dengan *Process Discovery*. *Process Mining* pada prinsipnya adalah menangkap data proses yang terekam dalam sistem dalam bentuk dataset *event log*. Hal ini sebagaimana diperlihatkan dalam Gambar 2.7:

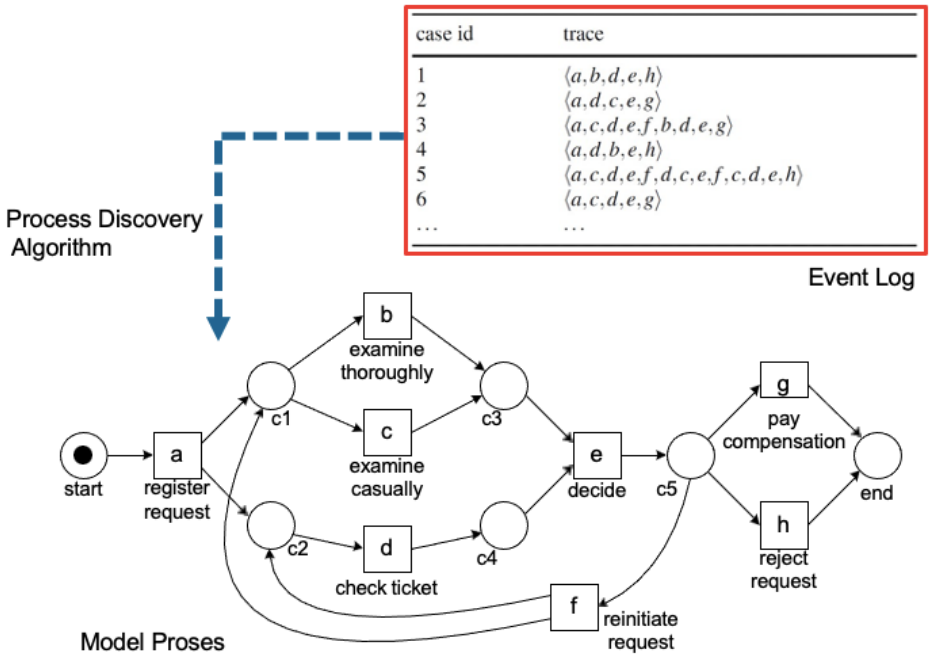
case id	event id	properties				
		timestamp	activity	resource	cost	…
1	35654423	30-12-2010:11.02	register request	Pete	50	…
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	…
	35654425	05-01-2011:15.12	check ticket	Mike	100	…
	35654426	06-01-2011:11.18	decide	Sara	200	…
	35654427	07-01-2011:14.24	reject request	Pete	200	…
2	35654483	30-12-2010:11.32	register request	Mike	50	…
	35654485	30-12-2010:12.12	check ticket	Mike	100	…
	35654487	30-12-2010:14.16	examine casually	Pete	400	…
	35654488	05-01-2011:11.22	decide	Sara	200	…
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	…

UNIVERSITAS DIPONEGORO PASCASARJANA

Gambar 2.7 Contoh Data Aktivitas dalam sistem yang tertangkap sebagai dataset event log

Dataset *event log* ini berisi data rekaman aktivitas dalam sebuah proses yang akan digunakan dalam mekanisme proses. Setiap aktivitas yang dilakukan

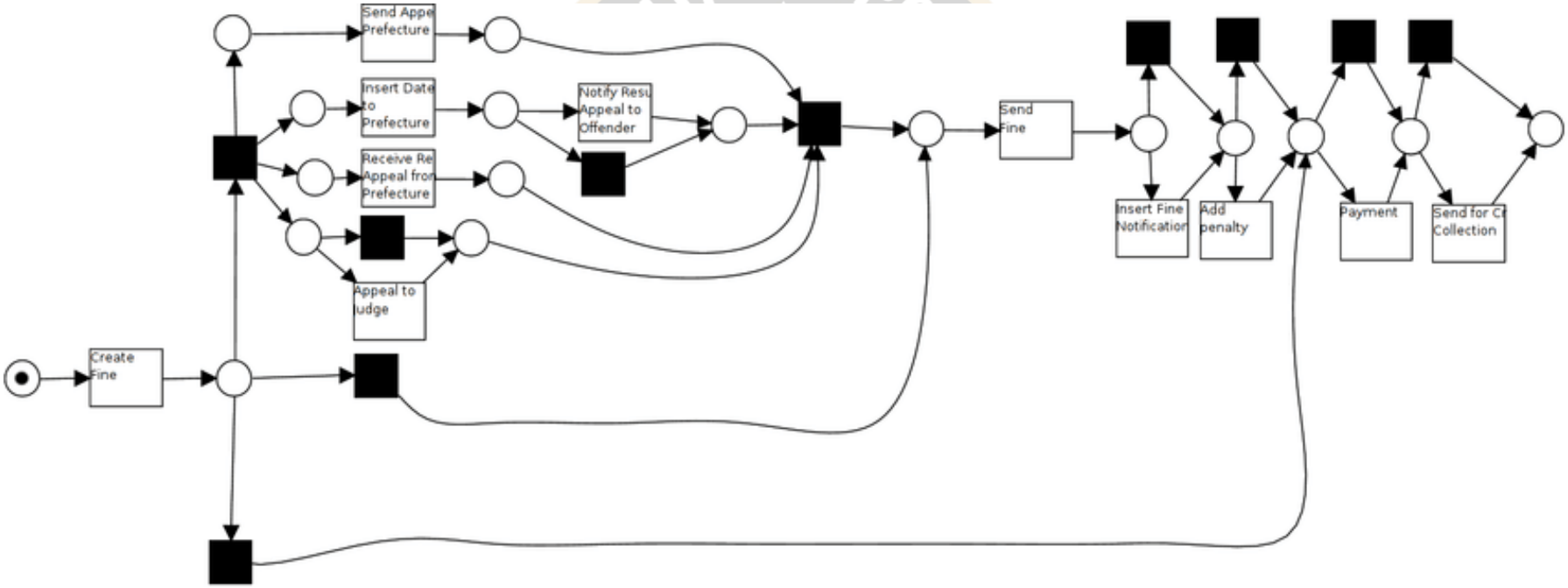
oleh individu atau *user* memiliki aktivitas yang bisa saja berbeda. Pola urutan yang terjadi dalam sebuah *event log* dapat digambarkan dalam model proses. Hal ini sebagaimana diperlihatkan dalam Gambar 2.8.



Gambar 2.8 Model Proses yang diperoleh melalui *Process Discovery Algorithm* berdasarkan dataset Event log

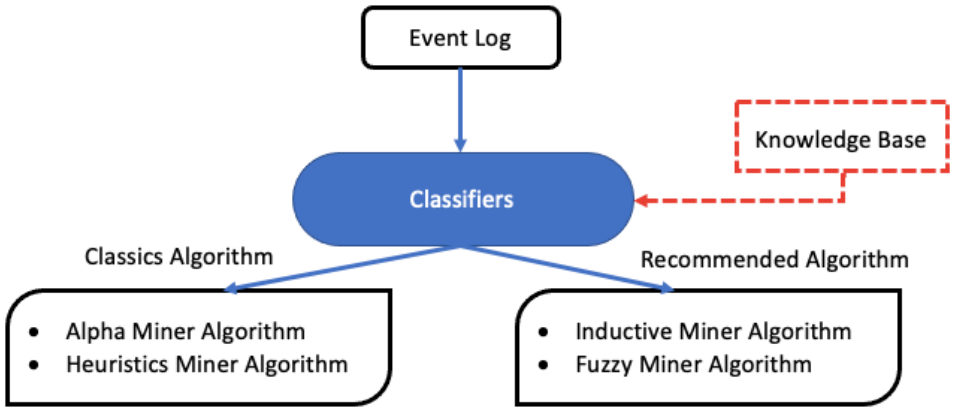
Untuk memvisualisasikan perilaku konsumen maka pendekatan yang dilakukan adalah menggunakan *process discovery*. *Process discovery* merupakan teknik untuk mengidentifikasi, menganalisis, dan memvisualisasikan proses bisnis berdasarkan log data atau jejak digital yang terekam di sistem informasi (W. M. P. van der Aalst, 2022). Proses ini dimulai dengan mengumpulkan data aktivitas yang telah terjadi dalam sistem, seperti timestamp, identitas aktivitas, dan identitas kasus atau transaksi. Dari data ini, algoritma *process discovery* akan mencoba membangun model proses yang merepresentasikan urutan dan aliran aktivitas yang terjadi. Gambar 2.9 memperlihatkan contoh model proses yang dihasilkan dari suatu algoritma *Process Discovery* dalam bentuk *Petri Net* berdasarkan data *event log* suatu proses yang terjadi.

SEMARANG
SEKOLAH PASCASARJANA



SEKOLAH PASCASARJANA

Saat ini algoritma *process discovery* terus berkembang, untuk pertama kali dan menjadi dasar algoritma *process mining* untuk pemodelan proses adalah Algoritma *Alpha Miner*. Berdasarkan Gambar 2.10, penelitian ini menggunakan 4 (empat) algoritma penemuan proses (*Process Discovery*) yaitu *Alpha Miner* dan *Heuristics Miner* dalam hal ini Algoritma klasik dan *Induktif Miner* serta *Fuzzy Miner* sebagai perwakilan algoritma rekomendasi untuk menggambarkan model proses perilaku konsumen dalam penelitian ini.



Gambar 2.10 *Process Discovery Algorithm* klasik dan rekomendasi

SEKOLAH PASCASARJANA

Berikut ini adalah penjelasan terkait algoritma *process discovery* yang digunakan pada penelitian ini:

1. Algoritma *Alpha Miner*

   Algoritma *Alpha Miner* dalam *process mining* merupakan salah satu algoritma pertama yang dikembangkan untuk mengidentifikasi model proses dari *event log*. *Alpha Miner* bekerja dengan mengidentifikasi urutan langsung dan hubungan sebab-akibat antara aktivitas dalam event log untuk membangun model proses (W. M. P. Van Der Aalst dkk., 2007). Adapun Algoritma *Alpha Miner* adalah

   a. Identifikasi Urutan Langsung, misalkan `A` dan `B` adalah dua aktivitas dalam log event. misalkan `A` → `B` jika ada instance dalam log di mana aktivitas `A` langsung diikuti oleh aktivitas `B`. kemudian Catat semua pasangan aktivitas yang memiliki hubungan ini.

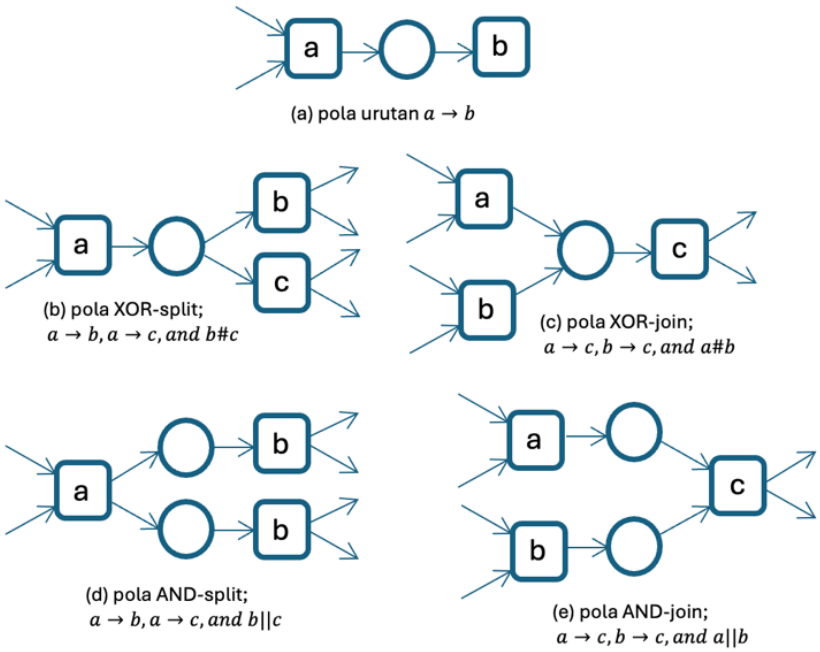
b. Identifikasi Hubungan kausal, hubungan kausal antara A dan B (A→B) ditentukan jika A sering diikuti oleh B dalam urutan yang konsisten. Jika B tidak pernah mendahului A, maka kita menyatakan bahwa A menyebabkan B.
c. Identifikasi Hubungan Paralel, Jika A dan B kadang-kadang muncul dalam urutan A→B dan kadang-kadang dalam urutan B→A, maka keduanya dianggap berjalan paralel. Hubungan paralel (A||B) menunjukkan bahwa aktivitas tersebut dapat terjadi bersamaan dalam proses.
d. Identifikasi Hubungan Eksklusif, Jika setelah A, hanya salah satu dari B atau C yang bisa terjadi (bukan keduanya), maka B dan C dianggap sebagai pilihan eksklusif. Ini berarti proses dapat bercabang, memilih satu di antara beberapa jalur.
e. Konstruksi tempat, tempat (places) ditambahkan untuk setiap hubungan yang diidentifikasi dalam langkah-langkah sebelumnya. Kemudian places digunakan untuk menyatukan aliran aktivitas dalam petri net, dengan kondisi sebagai penghubung antara aktivitas. Kemudian penambahan *Start* dan *End Places*, *start place* ditambahkan untuk menunjukkan titik awal dari proses, dan tempat akhir (*end place*) untuk menunjukkan titik akhirnya. Places ini menghubungkan aktivitas pertama dan terakhir dalam proses.
f. Kemudian konstruksi Petri Net dengan cara mengkombinasikan semua tempat, transisi (aktivitas), dan aliran untuk membentuk petri net lengkap. Petri net ini menunjukkan model proses yang telah diidentifikasi dari log event.

SEKOLAH PASCASARJANA

Tabel 2.2 Contoh *event log* sederhana

Case id	**Trace**
1	$< a, b, d, e, h >$
2	$< a, d, c, e, g >$
3	$< a, c, d, e, f, b, d, e, g >$
4	$< a, d, b, e, h >$
5	$< a, c, d, e, f, d, c, e, f, c, d, e, h >$
…	…

Selanjutnya, relasi urutan *events* tersebut akan digambarkan dalam notasi Petri net. Gambar 2.11 menunjukkan pola-pola sederhana yang dapat digunakan untuk menggambarkan pola urutan *events*.



Gambar 2.11 Pola urutan *events* dalam Petri net

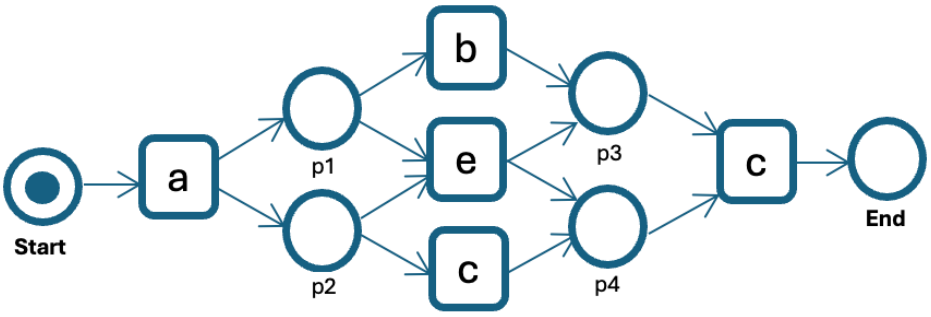
Pada Tabel 2.3 menunjukkan *footprint* dari *event log* yang ditampilkan dalam bentuk matriks.

Tabel 2.3 *Footprint* dari *Event Log*

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

SEKOLAH PASCASARJANA

Selanjutnya, matriks *footprint* tersebut dapat digunakan untuk menampilkan model proses dalam format Petri net (Gambar 2.12).



Gambar 2.12 Sebuah pola Petri net

2. Algoritma Heuristics Miner

Algoritma *Heuristics Miner* bekerja dengan menghitung frekuensi urutan aktivitas dalam event log dan menentukan kekuatan hubungan antar aktivitas berdasarkan bobot yang dihitung (W. M. P. Van Der Aalst dkk., 2007). Algoritma ini menggunakan aturan heuristik untuk menyaring hubungan yang kurang signifikan, sehingga menghasilkan model proses yang lebih realistis. Adapun langkah-langkah Algoritma *Heuristics Miner*:

a. Pengumpulan data frekuensi urutan aktivitas, diawali dengan menghitung berapa kali satu aktivitas diikuti oleh aktivitas lain dalam log event. Misalnya, jika aktivitas `A` diikuti oleh `B` sebanyak 10 kali, hubungan `A`→`B` akan memiliki bobot yang lebih tinggi dibandingkan dengan hubungan `A`→`C` yang mungkin hanya terjadi 2 kali.

b. Penentuan hubungan kausal dengan Bobot, dari data frekuensi, tentukan hubungan kausal antar aktivitas dengan menghitung bobot untuk setiap urutan langsung. Bobot ini menunjukkan seberapa kuat hubungan tersebut dalam log event, dengan hubungan yang lebih sering terjadi memiliki bobot yang lebih tinggi. Rumus Sederhana untuk Bobot Causal Relation:

   1) Bobot (`A`→`B`) = (Jumlah kejadian `A`→`B`) - (Jumlah kejadian `B`→`A`)
   2) Jika hasilnya positif dan signifikan, `A` dianggap menyebabkan `B`.

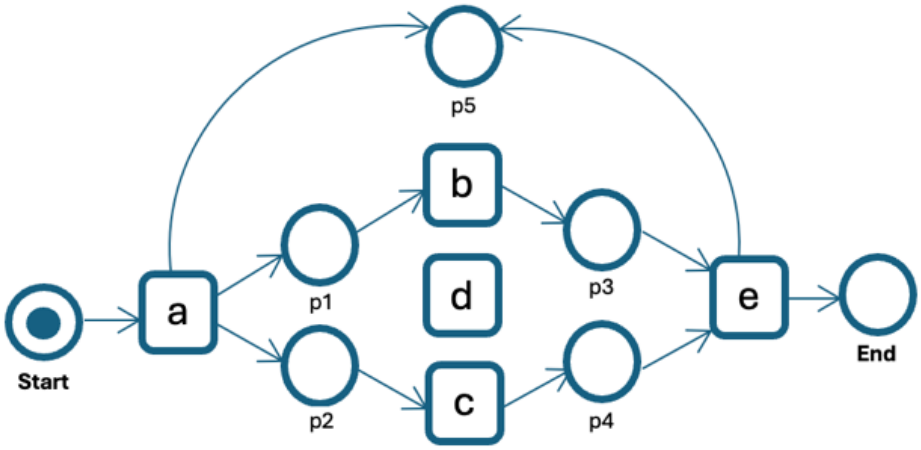
c. Identifikasi Hubungan Paralel dan Eksklusif, algoritma kemudian mengidentifikasi aktivitas yang berjalan paralel (`A||B`) atau eksklusif

SEKOLAH PASCASARJANA

(A→ (B XOR C)). Hubungan paralel diidentifikasi jika dua aktivitas sering muncul dalam urutan yang berbeda-beda, sedangkan hubungan eksklusif diidentifikasi jika setelah A, hanya satu dari beberapa aktivitas yang terjadi.

d. Penerapan *threshold* untuk penyaringan, hanya hubungan dengan bobot yang cukup tinggi yang akan dimasukkan dalam model akhir, membantu mengurangi noise dan menyederhanakan model.

e. Penyederhanaan dan agregasi model dengan cara menggabungkan aktivitas yang serupa untuk menyederhanakan model. Ini mencakup penggabungan aktivitas yang berfungsi sebagai langkah antara dalam proses yang lebih besar. Konstruksi model proses dengan menggunakan hubungan yang telah disaring dan disederhanakan untuk membangun model proses. Gambar 2.13 menunjukkan Petri net yang dihasilkan dengan algoritma *Alpha Miner* sebagai pembanding.



Gambar 2.13 Petri net dari event log dengan pembanding algoritma Alpha

SEKOLAH PASCASARJANA

3. Algoritma *Inductive Miner*

Inductive Miner adalah algoritma *process discovery* yang lebih maju, dirancang untuk menghasilkan model proses yang selalu berfungsi dengan baik tanpa *deadlock* atau anomali lain. *Inductive Miner* menghasilkan model proses yang dalam bentuk *Process Trees* dan dapat diubah menjadi *Petri Net*. Berikut ini adalah algoritma *Inductive Miner*:

a. Penentuan *Base Case*, Jika log event hanya berisi satu jenis aktivitas (misalnya, semua aktivitas adalah A), maka proses dianggap sebagai sekuensial tunggal. Kemudian lakukan dekomposisi, algoritma *Inductive Miner* bekerja dengan memecah event log menjadi bagian-bagian yang lebih kecil berdasarkan pola dalam data. Pemecahan dilakukan berdasaarkan operator logika seperti *sequence, choice, parallelism,* dan *loops.*

b. Identifikasi Operator,

   1) *Sequence* (→): Jika suatu urutan aktivitas selalu diikuti oleh aktivitas lain, ini diidentifikasi sebagai *sequence*. Misalnya, jika log event menunjukkan A→B→C, maka A diikuti oleh B, lalu C.

   2) *Parallelism* (AND), Jika dua atau lebih aktivitas dapat berjalan bersamaan dalam log, maka mereka diidentifikasi sebagai *parallelism*. Misalnya, A dan B bisa terjadi secara paralel.

   3) *Choice* (XOR), Jika setelah suatu aktivitas ada beberapa jalur yang berbeda yang dapat diambil, itu diidentifikasi sebagai *choice*. Misalnya, setelah A, proses bisa menuju B atau C, tetapi tidak keduanya.

   4) *Loop*, jika ada aktivitas atau serangkaian aktivitas yang bisa diulang, itu diidentifikasi sebagai *loop*. Misalnya, jika setelah A → B, proses kembali ke A, ini adalah loop.

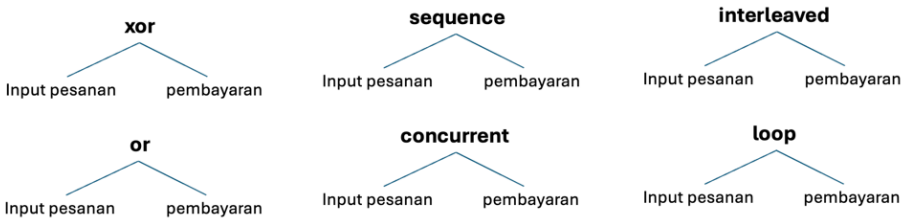
c. *Recursive Discovery*, setelah *event log* dipecah sesuai dengan operator yang diidentifikasi, algoritma *Inductive Miner* secara rekursif memproses setiap bagian dari log yang terpecah. Setiap bagian diperlakukan sebagai log baru, dan proses pemecahan dan identifikasi operator diulangi.

d. Kontruksi *Process Tree*, setelah pemecahan selesai, hasilnya disusun menjadi sebuah *process tree* yang mewakili seluruh model proses. *Process Tree* adalah struktur pohon di mana setiap node mewakili operator (*sequence, parallel, choice, loop*) dan anak-anak node

SEKOLAH PASCASARJANA

mewakili aktivitas atau subproses yang terkait sebagaimana Gambar 2.14.

e. Konversi menjadi Petri Net, *process tree* kemudian dikonversi menjadi *Petri Net* untuk visualisasi dan analisis lebih lanjut.



Gambar 2.14 Notasi tipe nodes

4. Algoritma *Fuzzy Miner*

Fuzzy Miner adalah algoritma yang sangat efektif untuk *process discovery* untuk *log event* yang besar dan kompleks, serta mengandung banyak *noise*. Algoritma ini memprioritaskan keterbacaan dan penyederhanaan model proses dengan menyaring aktivitas dan hubungan yang tidak signifikan, sehingga menghasilkan model yang lebih mudah dipahami. Adapun Algoritma *Fuzzy Miner*:

a. Pra-pemrosesan, *event log* diproses untuk menghitung berbagai metrik penting, seperti frekuensi aktivitas dan hubungan antar aktivitas. *Fuzzy Miner* menghitung dua metrik utama yaitu signifikan dan korelasi.

1) Signifikan, mengukur pentingnya suatu aktivitas atau hubungan berdasarkan frekuensi dan relevansinya dalam log event.
2) Korelasi, mengukur kekuatan hubungan antar aktivitas, terutama seberapa sering dua aktivitas muncul bersamaan atau dalam urutan tertentu. Aturan ini dapat mengakomodasi kelas event yang terjadi secara bersamaan berdasarkan intervalnya dari yang terbaru yaitu *Timestamp* akhir aktivitas sebelumnya dan *Timestamp* mulai aktivitas saat ini. Signifikansi interval memberikan nilai signifikansi = 1 pada hubungan dengan interval

SEKOLAH PASCASARJANA

rata-rata terkecil. Interval hubungan lain yang signifikan dapat diukur secara relatif terhadap interval terkecil, yang dijelaskan oleh persamaan (2.1).

$$Int.Sig. = \frac{1}{relation\ interval - least\ occuring\ interval} \quad (2.1)$$

b. Lakukan agregasi, setelah menghitung metrik *significance* dan *correlation*, aktivitas dan hubungan antar aktivitas di agregasi berdasarkan nilai-nilai tersebut. Aktivitas yang jarang terjadi atau hubungan yang tidak signifikan akan memiliki nilai yang lebih rendah dan bisa diabaikan dalam model proses.
c. Lakukan *Filtering*, terapkan aturan penyaringan untuk menghapus atau menyederhanakan elemen-elemen yang memiliki nilai *significance* dan *correlation* di bawah *threshold* tertentu. Aktivitas dan hubungan yang tidak signifikan atau tidak cukup kuat akan disembunyikan atau digabungkan dengan aktivitas lain untuk membuat model lebih mudah dibaca
d. Lakukan *Clustering, Fuzzy Miner* mengelompokkan aktivitas yang berkaitan erat satu sama lain menjadi satu entitas atau *cluster*. Pengelompokan ini membantu menyederhanakan model dengan mengurangi jumlah node yang harus ditampilkan, menjadikannya lebih mudah dipahami.
e. Model proses yang dihasilkan divisualisasikan dengan menampilkan hanya aktivitas dan hubungan yang telah disaring, serta klaster aktivitas yang berkaitan. Hubungan yang lebih kuat atau lebih signifikan digambarkan dengan garis yang lebih tebal atau warna yang lebih mencolok.
f. Penyempurnaan Iteratif, pengguna dapat iteratif mengatur threshold penyaringan atau mengelompokkan ulang aktivitas untuk menghasilkan model yang paling sesuai dengan kebutuhan analisis.

Pendekatan *process discovery* dalam penelitian ini digunakan untuk menggambarkan model proses perilaku konsumen berdasarkan data aktivitas

SEKOLAH PASCASARJANA

konsumen yang terekam dalam sistem *omnichannel*. Langkah berikutnya yang juga sangat penting adalah melakukan analisis Conformance checking pada model proses yang dihasilkan oleh algoritma process discovery. Conformance checking dalam process mining merupakan teknik yang digunakan untuk membandingkan model proses yang dihasilkan (atau yang sudah ada) dengan data aktual dari event log untuk mengevaluasi sejauh mana perilaku yang diamati sesuai dengan model yang diharapkan. Terdapat 2 alat ukur yang biasanya digunakan yaitu nilai Fitness dan nilai presisi. Nilai Fitness (FV) dapat diukur dengan Persamaan (2.2).

$$FV = \frac{Case\ Captured}{Cases\ on\ Actual\ event\ Logs} \tag{2.2}$$

Pengukuran nilai untuk fitness berkisar dari 0 sampai 1, di mana 0 berarti tidak ditemukan kasus sama sekali pada proses bisnis, sedangkan 1 berarti semua kasus dapat ditemukan oleh algoritma dan ditampilkan pada model proses bisnis, sedangkan *precision* merupakan nilai yang menunjukkan keakuratan jejak yang ditangkap oleh model proses bisnis (Buijs dkk., 2012). Nilai presisi (PV) dihitung dengan Persamaan (2.3).

$$PV = 1 - \frac{sum\ of\ bias\ traces}{traces\ from\ logs} \tag{2.3}$$

Nilai presisi berkisar dari 0 hingga 1, di mana 0 berarti tidak ada kasus yang ditangkap dari *event log* oleh algoritma, sedangkan 1 berarti semua kasus dapat ditangkap dalam model proses. Nilai Fitness dan presisi dalam penelitian digunakan untuk mengukur model proses yang dihasilkan dari *algoritma process discovery*.

SEKOLAH PASCASARJANA

2.3.5 Skema Model Prediksi Penjualan Berdasarkan Perilaku Konsumen pada Sistem Distribusi *Omnichannel*

Pada penelitian ini akan dikembangkan model prediksi penjualan pada sistem distribusi *omnichannel* berdasarkan perilaku konsumen. Pendekatan yang dilakukan adalah menggunakan model prediksi dengan pendekatan *data*

mining. Model prediksi merupakan hasil dari proses *data mining* yang bertujuan untuk menemukan pola dan hubungan dalam data. Setelah pola ini ditemukan, model tersebut dapat digunakan untuk memprediksi nilai atau kejadian di masa depan. Dengan menggunakan algoritma tertentu, model ini belajar dari data historis dan menerapkan pola-pola tersebut pada data baru untuk menghasilkan prediksi. Fungsi prediktif ini bertugas memproyeksikan bagaimana proses akan mendeteksi pola tertentu dalam data berdasarkan berbagai variabel yang ada. Selain itu, model juga diharapkan mampu memahami perilaku data, yang nantinya dapat membantu dalam mengenali karakteristik data tersebut.

Perilaku konsumen pada sistem distribusi *omnichannel* dapat diperlihatkan sebagaimana Gambar 2.15.

Gambar 2.15 Aktivitas Perilaku Konsumen yang Dapat Berpindah-pindah pada Sistem Distribusi *Omnichannel*

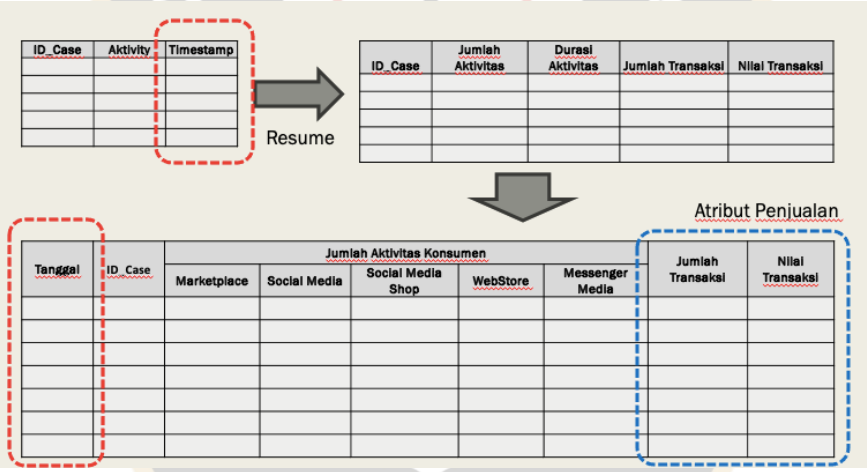
SEKOLAH PASCASARJANA

Aktivitas konsumen yang berpindah-pindah pada sistem distribusi *omnichannel* dapat ditangkap atau direkam oleh sistem melalui *event log*. Model data *event log* aktivitas konsumen sebagaimana diperlihatkan pada Tabel 2.4 yang mengacu pada Gambar 2.7 sebagaimana pada pembahasan *process discovery*.

Tabel 2.4 Data *Event Log* aktivitas konsumen yang terekam pada sistem distribusi *omnichannel*

Case_ID	Aktivitas	Timestamp	Saluran
001	Aktivitas A	dd/mm/yy hh:mm	Saluran 1
001	Aktivitas B	dd/mm/yy hh:mm	Saluran 1
001	Aktivitas C	dd/mm/yy hh:mm	Saluran 2
…	…	…	…
002	Aktivitas B	dd/mm/yy hh:mm	Saluran 1
002	Aktivitas C	dd/mm/yy hh:mm	Saluran 2
…	…	…	…

Untuk melakukan prediksi pada sistem distribusi *omnichannel* tersebut diperlukan turunan dataset *event log* aktivitas perilaku konsumen pada setiap saluran di sistem distribusi *omnichannel* (saluran *Marketplace*, *Messenger media*, media sosial, media sosial shop, dan *Webstore*). Adapun turunan dataset *event log* sebagaimana diperlihatkan pada Gambar 2.16.



Gambar 2.16 Konversi *Event Log* menjadi *Dataset* untuk Kebutuhan Prediksi

Gambar 2.16 memperlihatkan konversi proses pengolahan data dari bentuk data *event log* menjadi data transaksi atau aktivitas konsumen. Berikut ini adalah penjelasan detail:

1. Bagian atas kiri (input awal):
   a. Ada tabel berisi tiga kolom: ID_Case, Activity, dan Timestamp.
   b. Tabel tersebut adalah data *event log* aktivitas awal yang mengandung identifikasi kasus (ID_Case), aktivitas apa yang terjadi, dan kapan (Timestamp) aktivitas itu dilakukan. Data ini digunakan dalam analisis *process mining* untuk melacak aliran proses.

SEKOLAH PASCASARJANA

2. Bagian tengah (proses ringkasan/resume):

   Tabel ini menunjukkan bahwa data awal diringkas menjadi beberapa informasi kunci: Jumlah_Aktivitas, Durasi_Aktivitas, Jumlah_Transaksi, dan Nilai_Transaksi. Ini adalah hasil dari agregasi data aktivitas per kasus, yang diubah menjadi fitur yang lebih ringkas untuk dianalisis lebih lanjut.

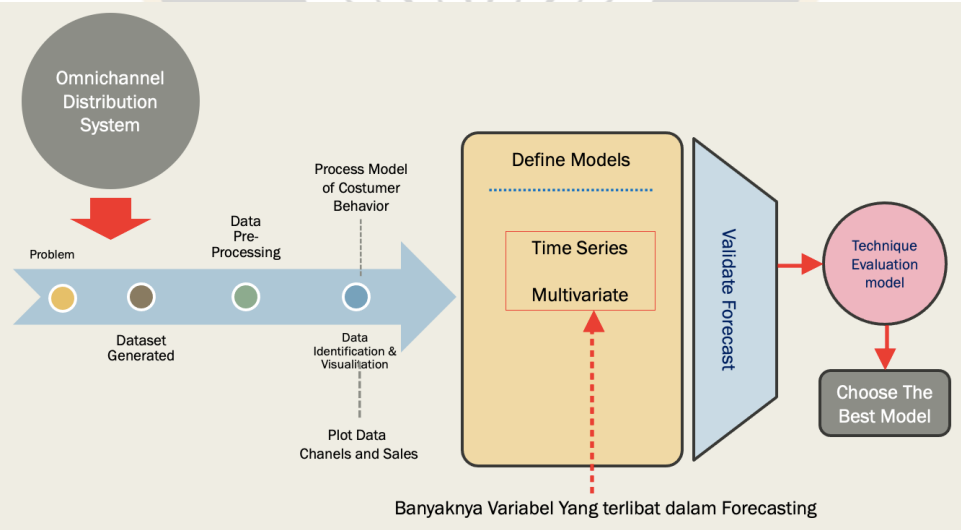
3. Bagian bawah (atribut penjualan lengkap):

   a. Tabel di bagian bawah berisi data yang lebih rinci. Terdapat kolom Tanggal, ID_Case, dan jumlah aktivitas konsumen di berbagai saluran, seperti *Marketplace*, media sosial, *Social Media Shop*, *WebStore*, dan *Messenger Media*.

   b. Di sisi kanan, terdapat kolom Jumlah_Transaksi dan Nilai_Transaksi, yang menambahkan informasi tentang jumlah dan nilai dari transaksi yang dihasilkan.

   c. Ini menunjukkan bahwa data aktivitas dari berbagai saluran e-commerce diintegrasikan dengan data transaksi, sehingga memberikan pandangan komprehensif tentang perilaku konsumen di seluruh saluran disertai hasil penjualannya.

Adapun skema pemodelan prediksi sebagaimana diperlihatkan pada gambar 2.17.



Gambar 2.17 Pengembangan Model Sistem Distribusi *Omnichannel* pada Prediksi Penjualan berdasarkan Perilaku Konsumen

Gambar 2.17 memperlihatkan skema prediksi penjualan berdasarkan perilaku konsumen pada sistem distribusi *omnichannel*. Menurut (Makridakis, 1994) terdapat dua jenis pendekatan dalam melakukan prediksi, yaitu kualitatif dan kuantitatif. Sebagaimana dapat dilihat pada skema Gambar 2.17 bahwa pendekatan prediksi yang dilakukan pada penelitian ini adalah pendekatan kuantitatif. Prediksi kuantitatif pada pengembangan model sistem distribusi *omnichannel* pada prediksi penjualan didasarkan pada analisis data numerik perilaku konsumen. Metode ini menggunakan data historis aktivitas konsumen yang berkunjung pada saluran sistem distribusi *omnichannel* untuk membangun model matematis yang dapat memproyeksikan nilai penjualan di masa depan. Data kunjungan perilaku konsumen pada berbagai saluran di sistem distribusi *omnichannel* bersifat multivariat. Dasar multivariat digunakan ketika menganalisis lebih dari satu variabel secara simultan. Hal ini berguna karena dalam konteks fenomena dalam penelitian ini melibatkan beberapa variabel yang saling berhubungan yaitu kunjungan konsumen pada berbagai saluran pada sistem distribusi *omnichannel*. Melalui analisis multivariat, dapat dilihat bagaimana variabel-variabel saling berkaitan dan hubungan antara aktivitas di berbagai saluran (*Marketplace*, Media Sosial, *Webstore*) terhadap penjualan dapat dianalisis. Dengan demikian penggunaan prediksi kuantitatif lebih objektif dibandingkan dengan prediksi kualitatif karena bergantung pada data yang dapat diukur dan diuji. Pada prinsipnya prediksi kuantitatif terdiri dari dua pendekatan, yaitu:

1. Analisis Kausal (*Causal Methods*)

   Metode prediksi yang didasarkan pada analisis hubungan sebab-akibat antara variabel-variabel. Dalam prediksi kausal, diasumsikan bahwa terdapat variabel *independent* (penyebab) yang secara langsung mempengaruhi variabel *dependent* (hasil) yang ingin diprediksi. Pada metode kausal terdapat tiga metode yang sering digunakan:

   a. Regresi Berganda (*Multiple Regression*): Menggunakan beberapa variabel independen yang dianggap sebagai penyebab untuk memprediksi variabel dependen. Misalnya, menggunakan faktor-

SEKOLAH PASCASARJANA

faktor seperti iklan, harga, dan pendapatan konsumen untuk memprediksi penjualan.

b. Model Struktural (*Structural Equation Modeling*, SEM): Memodelkan hubungan kausal antara variabel yang berbeda dalam suatu sistem kompleks, sering digunakan dalam ilmu sosial dan ekonomi.

c. Model *Econometric*: Seperti *Vector Autoregression* (VAR) atau *Vector Error Correction Model* (VECM), yang memodelkan hubungan kausal antara beberapa variabel ekonomi.

2. Analisis Deret Waktu (*Time Series* Analysis)

Tujuan utama dari analisis deret waktu ini adalah untuk membangun model prediksi nilai yang akan datang berdasarkan pola data historis. Terdapat 3 langkah yang dilakukan dalam analisis deret waktu pada penelitian ini yaitu:

a. Identifikasi Model (*Model Identification*)

Tahap ini melibatkan pemilihan model yang tepat berdasarkan karakteristik data yang tersedia. Proses ini dimulai dengan menganalisis data deret waktu untuk memahami pola seperti tren, musiman, dan siklus, serta menentukan apakah data bersifat stasioner atau non-stasioner.

b. Estimasi Parameter (*Parameter Estimation*)

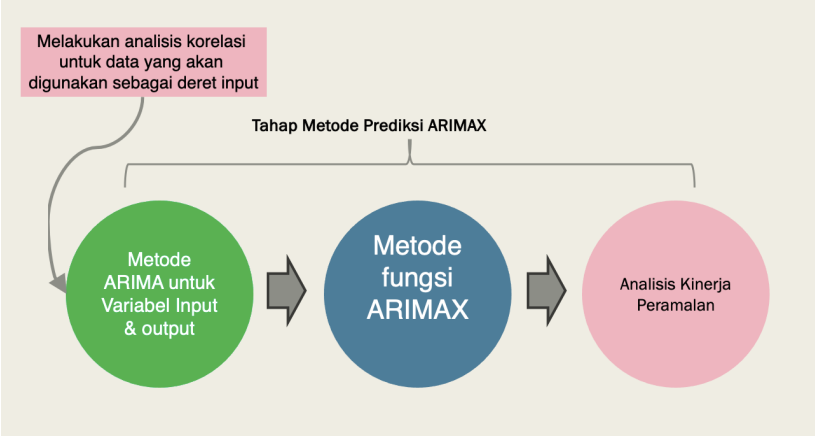
Setelah model dipilih, tahap berikutnya adalah mengestimasi parameter model tersebut. Parameter adalah nilai-nilai dalam model yang harus ditentukan agar model dapat secara akurat mencerminkan data.

SEKOLAH PASCASARJANA

c. Validasi dan Evaluasi (*Model Validation and Evaluation*)

Tahap akhir melibatkan pengujian model yang telah dibangun untuk memastikan bahwa model tersebut dapat memberikan prediksi yang akurat dan dapat diandalkan. Evaluasi dilakukan dengan menggunakan data yang tidak digunakan dalam proses estimasi untuk menguji kemampuan prediksi model.

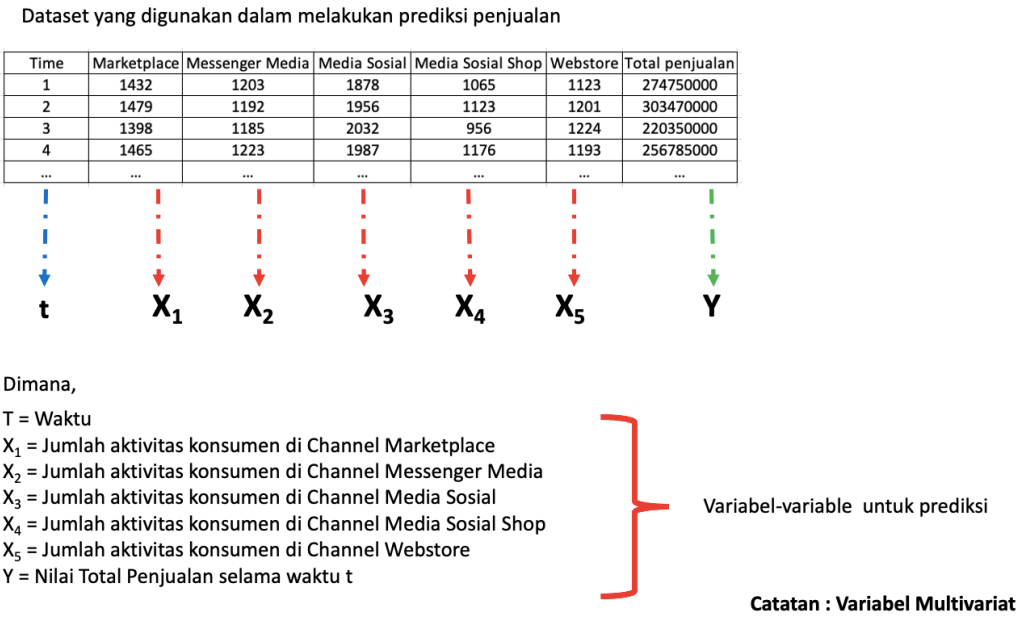
Terdapat dua jenis deret waktu yaitu deret waktu *univariate* dan *multivariate*. Analisis deret waktu *univariate* menggunakan *Autoregressive Integrated Moving Average* (ARIMA), sedangkan analisis runtun waktu *multivariate* menggunakan ARIMAX. Adapun metode prediksi yang digunakan dalam penelitian ini adalah analisis deret waktu ARIMAX. Penggunaan ARIMAX didasarkan pada karakteristik dataset multivariat perilaku konsumen dan nilai penjualan. Adapun skema prediksi dengan menggunakan ARIMAX pada penelitian ini sebagaimana diperlihatkan pada Gambar 2.18.



Gambar 2.18 Model kinerja ARIMAX pada penelitian yang dilakukan

Metode fungsi ARIMAX merupakan perluasan dari ARIMA, yang menggabungkan komponen autoregresif (AR), integrasi (I), dan *moving average* (MA), namun ARIMAX juga memasukkan variabel eksogen yang berperan dalam mempengaruhi nilai variabel utama yang diprediksi. Model prediksi menjadi menjadi lebih akurat dengan memasukkan variabel eksogen yang relevan, sehingga model dapat menangkap dinamika kompleks yang mempengaruhi deret waktu. Untuk mendukung metode tersebut maka diperlukan dataset yang berasal dari saluran *omnichannel*. Model *dataset* yang sesuai dengan pendekatan ARIMAX yaitu memiliki variable eksogen dan endogen dikembangkan dari Gambar 2.16. Adapun bentuk *dataset* sebagaimana diperlihatkan pada Gambar 2.19.

SEKOLAH PASCASARJANA



Gambar 2.19 Dataset untuk Prediksi Penjualan berdasarkan Perilaku Konsumen

Gambar 2.19 menunjukkan *dataset* yang digunakan untuk melakukan prediksi penjualan dengan beberapa variable input, yang cocok untuk analisis ARIMAX (*AutoRegressive Integrated Moving Average with Exogenous Variables*). Berikut adalah penjelasan terkait elemen-elemen pada Gambar 2.19 untuk analisis ARIMAX:

1. Variabel t (*Time*): Ini adalah variabel waktu, yang merupakan dasar dari model ARIMAX. Dalam analisis deret waktu, model ARIMAX memerlukan urutan waktu sebagai landasan untuk menangkap pola temporal dalam data penjualan (Y).
2. Variabel X_1, X_2, X_3, X_4, X_5: Variabel-variabel ini adalah variabel eksogen (exogenous variables) yang merepresentasikan aktivitas konsumen di berbagai saluran:
   a. X_1 (*Social Media*): Aktivitas konsumen di saluran media sosial.
   b. X_2 (*Marketplace*): Aktivitas konsumen di saluran marketplace.
   c. X_3 (*Messenger Media*): Aktivitas di saluran media pesan.
   d. X_4 (*Social Media Shop*): Aktivitas di saluran belanja melalui media sosial.
   e. X_5 (*Webstore*): Aktivitas konsumen di saluran toko online.

SEMARANG
SEKOLAH PASCASARJANA

3. Variabel Y (Total Penjualan): Ini adalah variabel dependen atau target yang ingin diprediksi. Nilai total penjualan selama waktu *t* dipengaruhi oleh kombinasi dari aktivitas-aktivitas di berbagai channel yang direpresentasikan oleh variabel X.

Untuk memahami metode ARIMAX diperlukan pemahaman mendasar terkait konsep ARIMAX secara keseluruhan. Adapun perkembangan model AR menjadi ARIMAX berdasarkan formulasi adalah sebagai berikut:

1. Model *AutoRegressive* (AR), memodelkan nilai saat ini dari deret waktu sebagai fungsi linear dari nilai-nilai masa lalu pada lag tertentu. Model AR hanya mempertimbangkan ketergantungan linier pada nilai masa lalu. AR berguna untuk data yang menunjukkan ketergantungan pada nilai masa lalu, tetapi tidak mempertimbangkan fluktuasi acak (Box et al., 2015). Persamaan umumnya adalah

$$x_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \emptyset_3 x_{t-3} + \cdots + \emptyset_p x_{t-p} + e_t \quad (2.4)$$

dengan,

$\emptyset_1, \emptyset_1, \ldots, \emptyset_p$: parameter model,

p : orde model AR, dan

e_t : error

2. Model *Moving Average* (MA), MA memodelkan nilai saat ini dari deret waktu sebagai fungsi linear dari gangguan masa lalu pada lag tertentu. Model MA hanya mempertimbangkan ketergantungan linier pada gangguan masa lalu. MA berguna untuk data di mana fluktuasi acak dari masa lalu mempengaruhi nilai saat ini (Box et al., 2015). Adapun formulanya adalah

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \cdots - \theta_k e_{t-q} \quad (2.5)$$

SEKOLAH PASCASARJANA

dengan,

$\theta_1, \theta_2, \ldots, \theta_q$: parameter model,

q : orde model MA,

μ : rata-rata, dan

e_t : gangguan/kesalahan.

3. Model *AutoRegressive Moving Average* (ARMA), model ARMA menggabungkan komponen AR dan MA untuk memodelkan data deret waktu. Ini menggabungkan ketergantungan linier pada nilai masa lalu (AR) dan ketergantungan linier pada gangguan masa lalu (MA) (Box et al., 2015). Adapun persamaan umumnya adalah

$$x_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \cdots + \emptyset_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_1 e_{t-2} - \cdots - \theta_k e_{t-q} \quad (2.6)$$

dengan,

$\emptyset$: parameter AR

θ : parameter MA.

x_{t-p} : nilai masa lalu ke-p

e_{t-q} : nilai kesalahan masa lalu ke-q

e_t : error

ARMA berguna untuk data yang non-stasioner dengan tren yang dapat menjadi stasioner setelah *differencing*. Ideal untuk data yang menunjukkan pola tren atau siklus yang harus diperhitungkan. ARMA berguna untuk data yang menunjukkan ketergantungan baik pada nilai masa lalu maupun fluktuasi acak masa lalu, tetapi memerlukan data yang stasioner (tanpa tren atau musiman). Untuk mengetahui nilai p dan q yang akan digunakan oleh model dapat diidentifikasi dengan melihat autokorelasi dan parsial autokorelasi dari data deret waktu yang ada. Secara umum fungsi autokorelasi dan model yang sesuai dirangkum dalam Tabel 2.5 (Montgomery, 1998):

SEKOLAH PASCASARJANA

Tabel 2.5 Bentuk fungsi autokorelasi dan fungsi autokorelasi parsial

Model	**Fungsi Autokorelasi/ACF**	**Fungsi Autokorelasi Parsial/PACF**
AR (p)	Naik/Turun secara eksponensial	Terpotong pada lag q
MA (q)	Terpotong pada lag q	Naik/turun secara eksponensial
ARMA (p,q)	Naik/turun secara eksponensial	Naik/turun secara eksponensial

4. Model *AutoRegressive Integrated Moving Average* (ARIMA), model ARIMA adalah ekstensi dari model ARMA yang juga mencakup *differencing* untuk menangani data non-stasioner dengan tren. Ini menggabungkan komponen AR, MA, dan *differencing* (integrasi). Komponen dasar ARIMA berdasarkan (Hamilton, 2020) adalah:
    a. AR (*AutoRegressive*) adalah komponen regresi otomatis, yang berarti bahwa nilai masa lalu dari deret waktu digunakan untuk memprediksi nilai masa depan.
    b. I (*Integrated*) berkaitan dengan perbedaan deret waktu untuk membuatnya stasioner (*stationary*). Stasioner berarti bahwa statistik deret waktu (seperti *mean* dan *varians*) tidak berubah seiring waktu. Parameter *d* menunjukkan tingkat *differencing* yang diperlukan untuk mencapai stasioneritas.
    c. MA (*Moving Average*) adalah komponen rata-rata bergerak, yang berarti bahwa kesalahan prediksi masa lalu digunakan untuk memprediksi nilai masa depan. Model MA diwakili oleh parameter q, yang menunjukkan jumlah lag dari kesalahan prediksi yang akan digunakan dalam model. Model ARIMA diidentifikasi dengan tiga parameter yaitu:
        - p: Jumlah lag dari komponen *autoregressive* (AR).
        - d: Tingkat *differencing* yang diperlukan untuk mencapai stasioneritas.
        - q: Jumlah lag dari komponen *moving average* (MA).

        Sehingga model ARIMA ditulis sebagai ARIMA(p, d, q).

Model ARIMA yang akurat adalah yang memiliki kesalahan (*error*) yang kecil. Dalam ARIMA terdapat empat proses penting menurut (Montgomery & Johnson, 1988) yaitu

1. Identifikasi Model yaitu analisis data untuk menentukan nilai yang tepat untuk p, d, dan q. Ini biasanya dilakukan dengan menggunakan *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF).

UNIVERSITAS DIPONEGORO SEMARANG SEKOLAH PASCASARJANA

2. Estimasi Parameter, setelah model diidentifikasi, parameter model kemudian diestimasi.
3. Uji diagnostik, setelah model dibangun, perlu dilakukan uji diagnostik untuk memastikan bahwa residu dari model tidak berkorelasi (*white noise*) yang menunjukkan bahwa model tersebut cukup sesuai.
4. *Forecasting*, kemudian menggunakan model yang dibangun untuk memprediksi nilai masa depan dari deret waktu.

Sebagaimana penjelasan sebelumnya bahwa model *Autoregressive Moving Average* (ARMA) adalah model yang digunakan untuk menganalisis deret waktu stasioner, yang dapat diandalkan untuk memprediksi nilai di masa depan dari deret waktu yang sangat bervariasi. Prinsip utama ARMA yaitu korelasi dan stasioneritas yang berarti bahwa model ini meramalkan nilai masa depan dengan memperhitungkan hubungan antara nilai-nilai masa lalu. Data stasioner adalah data yang memiliki rata-rata dan varians konstan sepanjang waktu. Jika data tidak stasioner terhadap rata-rata, maka dilakukan proses *differencing* untuk mengubahnya menjadi stasioner. Sementara itu, data yang tidak stasioner terhadap varians dapat ditransformasi agar sesuai dengan model (Osborne, 2010; Zhang & Yang, 2017).

Untuk menunjukkan keadaan data secara keseluruhan, seperti data tanpa tren, musiman, dan sebagainya, keadaan stasioner diperlukan; namun, model ARMA dapat digunakan untuk berbagai jenis deret waktu. Adapun pemeriksaan stasioneritas terhadap varian dapat dilihat dengan menggunakan plot apabila fluktuasi variannya tidak terlalu besar. Dalam pengujian stasioneritasnya dapat menggunakan transformasi *Box-Cox* yang ditunjukkan pada persamaan berikut (Box dkk., 2016).

SEKOLAH PASCASARJANA

$$W = y_t^{(\lambda)} = \begin{cases} \frac{y_t^{\lambda}-1}{\lambda}, \lambda \neq 0 \\ \ln y_t, \lambda = 0 \end{cases} \quad (2.7)$$

dengan,

$y_t^{(\lambda)}$: data transformasi

y_t : pengamatan pada waktu ke-t

λ : parameter transformasi

Bentuk transformasi *Box-Cox* untuk beberapa nilai estimasi λ yang sering digunakan dapat dilihat pada Tabel 2.6 (Wei, 2018)

Tabel 2.6 Nilai Transformasi *Box-Cox*

λ	Transformasi
1	Stasioner
0,5	$\sqrt{Y_t}$
0	$\ln Y_t$
-0,5	$1/\sqrt{Y_t}$
-1,0	$1/Y_t$

Dalam pengujian stasioneritas terhadap varian dikatakan telah stasioner apabila nilai batas bawah dan batas atas λ dari data bernilai satu.

5. Model *Auto Regressive Integrated Moving Average* with *Exogenous Variables* (ARIMAX)

   Model ARIMAX adalah perpanjangan dari ARIMA yang memasukkan variabel eksogen (variabel luar) sebagai input tambahan untuk memodelkan data deret waktu. Ini memungkinkan pengaruh variabel eksternal pada deret waktu yang sedang dianalisis. Bentuk umum model ARIMAX untuk multi-input berdasarkan (Makridakis et al., 2008) dan (Box et al., 2015) adalah sebagai berikut:

   $$y_t = v_1(B)x_{1,t} + \cdots + v_n(B)x_{m,t} + e_t \qquad (2.8)$$

   y_t : nilai variabel endogen (*dependent*) pada waktu t,

   $x_{m,t}$: variabel eksogen (*independent*) ke-m pada waktu t, huruf m mengindikasikan bahwa bisa terdapat beberapa variabel eksogen, yaitu $x_1, x_2, \ldots, x_m$

   $v(B)$: polinomial dalam operator *backshift* B,

   e_t : *noise* atau *error* term yang dimodelkan pada ARIMA(p,d,q)

   Persamaan (2.8) mengekspresikan hubungan antara variabel dependen y_t dan beberapa variabel independen (*exogenous*) $x_1, x_2, \ldots, x_m$ dengan gangguan kesalahan (*error*) e_t . Berikut ini adalah penjelasan masing-masing komponen dari persamaan tersebut:

SEKOLAH PASCASARJANA

1. y_t : variabel terikat atau variabel yang diprediksi (*dependent variable*) pada waktu t.
2. $v_i(B)x_{i,t}$: komponen regresi dari variabel eksogen (independent variables). Di sini, B adalah operator lag, sehingga $v_i(B)$ adalah fungsi polinomial lag untuk setiap variabel eksogen $x_{i,t}$ yang menunjukkan bagaimana variabel eksogen $x_{i,t}$ pada beberapa periode waktu sebelumnya mempengaruhi variabel terikat y_t.
3. e_t: Komponen kesalahan (error term), yang biasanya dianggap sebagai proses ARMA, yang terdiri dari bagian autoregresif (AR) dan *moving average* (MA) untuk menangani korelasi dalam error residual.

Adapun prosedur penentuan parameter pada persamaan ARIMAX adalah sebagai berikut (Makridakis et al., 2008):

1. Identifikasi variabel eksogen $x_{i,t}$

   Pilih variabel-variabel independen yang mungkin mempengaruhi variabel terikat y_t. Variabel ini berasal dari faktor eksternal yang diyakini mempengaruhi sistem (dalam konteks penelitian ini adalah variabel kunjungan perilaku konsumen pada saluran *omnichannel* seperti *marketplace*, Media Messenger, Media Sosial, dll.).
2. Uji stasioneritas

   Lakukan uji stasioneritas (seperti *Augmented Dickey-Fuller test*) untuk memastikan apakah data harus di-*differencing* (integrasi) atau tidak. Jika tidak stasioner, tentukan derajat *differencing* (parameter I) yang diperlukan agar data menjadi stasioner.
3. Menentukan parameter pada formula 2.8 (p, d, q):
   a. p (*Autoregressive order*): Menentukan orde autoregresif dari variabel dependen y_t dengan menggunakan *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF).
   b. d: Derajat *differencing* untuk membuat deret menjadi stasioner.
   c. q (*Moving Average order*): Menentukan orde MA dari error e_t, juga berdasarkan ACF dan PACF.

SEKOLAH PASCASARJANA

4. Menentukan lag dari variabel eksogen (*exogenous variables*)

   Pemilihan lag yang tepat untuk setiap variabel eksogen $x_{i,t}$ dilakukan dengan menggunakan ACF atau uji statistik seperti *Cross-Correlation* antara variabel eksogen dan residual dari model ARMA pada y_t.
5. Estimasi parameter model.

   Gunakan metode *Least Squares* untuk mengestimasi koefisien polinomial lag $v_i(B)$ dari variabel eksogen dan koefisien ARMA dari error e_t.
6. Evaluasi model

   Evaluasi performa model menggunakan kriteria evaluasi dengan MAPE. Lakukan uji diagnostik pada residual model (contohnya, uji Ljung-Box) untuk memastikan residual tidak berkorelasi (*white noise*).

Pada persamaan serentak ARIMAX yaitu persamaan (2.8) terdapat error e_t yang merupakan error atau residual dari model. Dalam konteks ARIMAX, asumsi terhadap error (e_t) biasanya mengikuti beberapa prinsip penting dalam analisis deret waktu yaitu:

1. Distribusi Normal: Error (e_t) diasumsikan berdistribusi normal.
2. *White Noise*: Error juga dianggap sebagai *white noise*, yang berarti bahwa tidak ada pola yang dapat ditangkap dalam residual. Error tidak memiliki autokorelasi (keterkaitan dengan nilai error sebelumnya)
3. Tidak Terkorelasi dengan Variabel Eksogen: Error (e_t) juga diasumsikan tidak berkorelasi dengan variabel eksogen $x_{m,t}$ dalam model. Ini penting untuk memastikan bahwa kesalahan dalam prediksi tidak bergantung pada variabel penjelas yang dimasukkan ke dalam model.

Sedangkan hubungan antara error (e_t) , variabel independen $x_{m,t}$ dan variabel dependen y_t dalam model ARIMAX dapat dijelaskan sebagai berikut:

1. Hubungan antara error (e_t) dan Variabel Dependen (y_t):

SEKOLAH PASCASARJANA

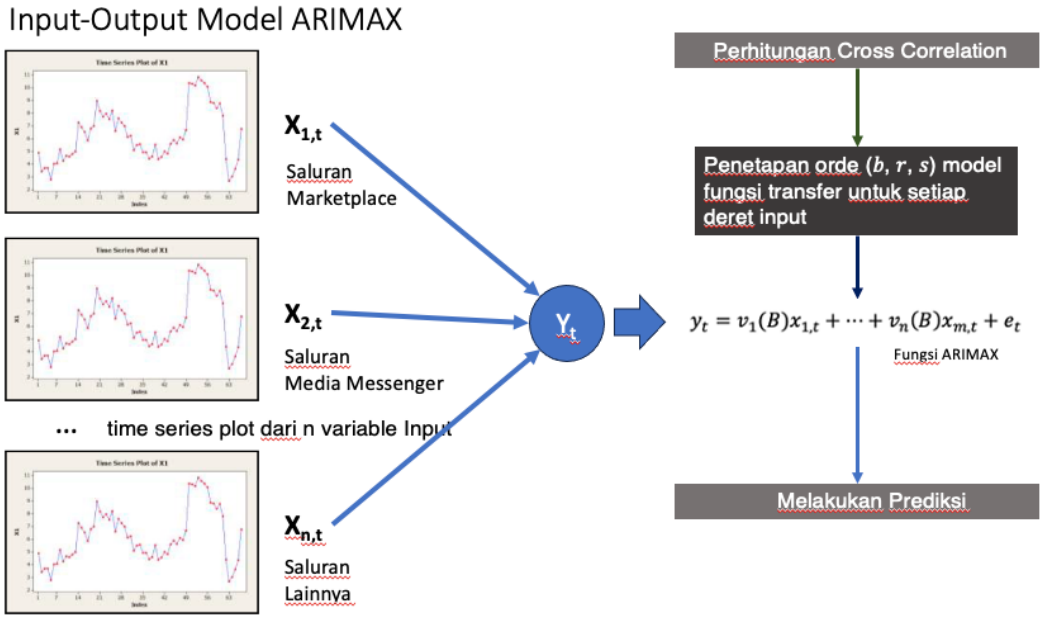
- Nilai y_t yang diprediksi oleh model adalah kombinasi dari pengaruh variabel independen $x_{m,t}$ melalui koefisien $v(B)$, dan error (e_t) adalah sisa atau residual yang tidak dapat dijelaskan oleh variabel-variabel independen tersebut.
- Dengan kata lain, *error* menggambarkan bagian dari variabel dependen y_t yang tidak dapat dijelaskan oleh variabel independen $x_{m,t}$ dan dinamika masa lalu dari y_t.

2. Hubungan antara Error (e_t) dan Variabel Independen $(x_{m,t})$:
   - Asumsi dalam ARIMAX adalah bahwa error (e_t) tidak berkorelasi dengan variabel independen $x_{m,t}$. Ini berarti tidak ada hubungan linier antara error dan variabel independen. Jika error berkorelasi dengan $x_{m,t}$, hal tersebut mengindikasikan bahwa model belum menangkap sepenuhnya hubungan antara variabel independen dan dependen, dan model mungkin bias atau tidak sesuai.
   - Eksogenitas mengasumsikan bahwa variabel independen $x_{m,t}$ mempengaruhi variabel dependen y_t, tetapi error (e_t) tidak mempengaruhi $x_{m,t}$. Jika asumsi ini dilanggar (misalnya, jika ada korelasi antara (e_t) dan $x_{m,t}$), model prediksi dapat menjadi tidak akurat.

3. Peran Error (e_t) dalam Prediksi y_t:
   - Error (e_t) memainkan peran penting dalam mengevaluasi kualitas model. Jika model ARIMAX sudah cukup baik, error (e_t) akan acak dan tidak memiliki pola yang dapat diidentifikasi, menunjukkan bahwa variabel independen $x_{m,t}$ sudah menjelaskan sebagian besar variasi dalam y_t.
   - Jika error (e_t) mengandung pola (misalnya, autokorelasi atau heteroskedastisitas), hal tersebut menunjukkan bahwa ada informasi yang tidak terwakili dengan baik oleh variabel independen $x_{m,t}$, atau ada faktor lain yang harus dimodelkan (misalnya, variabel independen tambahan atau transformasi).

SEKOLAH PASCASARJANA

Secara keseluruhan, proses pemodelan ARIMAX melibatkan beberapa tahap, mulai dari uji asumsi stasioneritas, identifikasi model, estimasi parameter, hingga uji diagnostik dan evaluasi model. Secara umum model ARIMAX dalam penelitian ini diperlihatkan pada Gambar 2.20.



Gambar 2.20 Gambaran umum Model ARIMAX

Adapun langkah-langkah pemodelan ARIMAX dengan disertai asumsi data dalam penelitian ini adalah sebagai berikut:

1. Uji Stasioneritas untuk Data Endogen

   Model ARIMAX memerlukan data yang bersifat stasioner, khususnya pada bagian residualnya. Gunakan uji stasioneritas berikut:

   a. *Dickey-Fuller Test* (ADF Test)

      Untuk menguji apakah data y_t dan variabel eksogen x_t stasioner atau tidak.

      Adapun hipotesis yang digunakan adalah

      - H_0: Data non-stasioner
      - H_1: Data stasioner

      Jika hasil uji menunjukkan bahwa data tidak stasioner (p-value > 0.05), lakukan diferensiasi (*differencing*) hingga data menjadi stasioner.

SEKOLAH PASCASARJANA

b. Transformasi *Differencing*

Jika data tidak stasioner, lakukan *differencing* pada variabel dependen y_t:

$$y_t' = y_t - y_{t-1} \quad (2.9)$$

Ulangi proses ini hingga data menjadi stasioner.

2. Uji Stasioneritas untuk Data Eksogen

Sama seperti variabel dependen, pastikan variabel eksogen x_t juga stasioner. Jika tidak stasioner maka dilakukan diferensiasi pada variabel eksogen:

$$x_t' = x_t - x_{t-1} \quad (2.10)$$

3. Identifikasi Orde ARIMA (p, d, q)

Setelah data y_t menjadi stasioner, gunakan plot ACF (*Autocorrelation Function)* dan PACF (*Partial Autocorrelation Function*) untuk menentukan parameter ARIMA, yaitu:

- AR (p): Berdasarkan PACF, menunjukkan jumlah *lag* untuk komponen autoregresif.
- I (d): Jumlah *differencing* yang diperlukan untuk membuat data stasioner.
- MA (q): Berdasarkan ACF, menunjukkan jumlah *lag* untuk komponen *moving average*.

4. Pada model ARIMAX (*AutoRegressive Integrated Moving Average with Exogenous Variables*), penentuan parameter (r, b, s) memiliki dasar teoritis yang serupa dengan model ARIMA, tetapi dengan penambahan variabel eksogen (X) yang memengaruhi hasil prediksi. Adapun penjelasan untuk masing-masing parameter adalah sebagai berikut:

a. Parameter r (*AutoRegressive* - AR)

1) Teori dasar dari Komponen AR(r) dalam model ARIMA adalah merepresentasikan hubungan linear antara nilai saat ini dengan nilai sebelumnya hingga lag r. Artinya, model akan mencoba menangkap

SEKOLAH PASCASARJANA

pola dalam data historis, di mana nilai saat ini dipengaruhi oleh nilai masa lalu pada lag 1, lag 2, hingga lag ke-r.

2) Cara penentuan nilai r ditentukan dengan menggunakan *partial autocorrelation function* (PACF). Jika plot PACF menunjukkan cut off setelah lag tertentu, nilai tersebut dianggap sebagai r (lag AR).

b. Parameter b (*Moving Average* - MA)

1) Komponen MA(b) menggambarkan model yang didasarkan pada kesalahan residu dari proses sebelumnya. Dengan kata lain, nilai sekarang dipengaruhi oleh kesalahan model pada periode sebelumnya hingga lag b.

2) Cara penentuan nilai b dapat ditentukan dengan menggunakan *autocorrelation function* (ACF). Jika plot ACF menunjukkan *cut-off* setelah beberapa lag, nilai lag tersebut menjadi kandidat untuk b.

c. Parameter s (*Lag Exogenous Variables*)

1) Pada model ARIMAX, terdapat variabel eksogen (X), yaitu variabel yang dianggap memiliki pengaruh terhadap variabel dependen, tetapi tidak dimodelkan secara endogen (di dalam sistem ARIMA). Parameter s adalah lag dari variabel eksogen tersebut, yang digunakan untuk menangkap pengaruh dari variabel eksogen di periode sebelumnya. Cara penentuan lag variabel eksogen dapat dilakukan dengan cara yang mirip dengan AR atau MA, yaitu melihat bagaimana variabel eksogen berpengaruh terhadap variabel dependen di berbagai lag waktu. Dalam beberapa kasus, metode seleksi model otomatis seperti *Cross-Correlation. Cross-Correlation Function* (CCF) dalam konteks ARIMAX adalah alat statistik yang digunakan untuk mengukur dan menganalisis keterkaitan atau korelasi antara dua time series yang terjadi pada berbagai lag. Secara khusus, dalam model ARIMAX (*AutoRegressive Integrated Moving Average with Exogenous Variables*), CCF digunakan untuk mengevaluasi hubungan

SEKOLAH PASCASARJANA

antara variabel independen (input/exogenous) dan variabel dependen (output) pada berbagai waktu sebelumnya (lag). Adapun peran dalam ARIMAX untuk melihat apakah perubahan dalam variabel eksogen x_t memengaruhi y_t (output) pada saat itu atau pada waktu-waktu berikutnya (dengan lag). Dengan menggunakan CCF, dapat menemukan lag optimal di mana variabel eksogen memiliki efek paling signifikan pada variabel dependen.

5. Pemodelan ARIMAX

   Pada tahap ini melakukan pemodelan ARIMAX dengan menggunakan persamaan (2.8)

6. Uji Diagnostik Residual

   Setelah model diestimasi, lakukan uji diagnostik untuk memeriksa apakah asumsi model terpenuhi yaitu dengan

   a. Uji Autokorelasi Residual (*Ljung-Box Test*)

      Untuk memeriksa apakah residual e_t adalah *white noise* (tidak berkorelasi).

      Adapun hipotesis yang digunakan adalah

      - H_0: Residual tidak berkorelasi (*white noise*)
      - H_1: Residual berkorelasi

      Jika p-value > 0.05, maka residual tidak berkorelasi dan model dianggap baik.

   b. Uji Normalitas Residual

      Untuk menguji normalitas digunakan pendekatan Kolmogorov-Smirnov. Kolmogorov-Smirnov untuk menguji apakah residual model mengikuti distribusi normal.

      Adapun hipotesis yang digunakan adalah

      - H_0: Residual mengikuti distribusi normal.
      - H_1: Residual tidak berdistribusi normal.

      Jika nilai statistik K-S lebih besar dari nilai kritis pada tingkat signifikansi yang dipilih (p-value < 0.05) maka tolak hipotesis nol. Hal ini berarti bahwa residual tidak berdistribusi normal.

SEKOLAH PASCASARJANA

7. Evaluasi Model

   Evaluasi MAPE digunakan untuk melihat akurasi model ARIMAX

8. Prediksi dengan Model ARIMAX

   Model ARIMAX yang sudah diestimasi kemudian digunakan untuk memprediksi nilai y_t di masa mendatang berdasarkan variabel eksogen x_t yang diketahui.

2.3.6 Pengukuran Kinerja prediksi dan Kesalahan Prediksi

Untuk menilai seberapa baik suatu model prediksi bekerja, digunakan ukuran kinerja yang dikenal sebagai *performance metrics*. Menurut (Montgomery, 2007) *performance metrics* yang paling baik untuk mengukur kinerja prediksi adalah *forecasting error* atau kesalahan prediksi. *Error* dalam prediksi didefinisikan sebagai perbedaan antara nilai yang diprediksi (nilai estimasi yang diwakili oleh $\hat{y}$) dan nilai aktual yang terjadi (dilambangkan dengan *y*). Jadi, *forecasting* error menunjukkan seberapa jauh hasil prediksi menyimpang dari kenyataan. *Error* dalam prediksi didefinisikan sebagai nilai penyimpangan dari nilai estimasi $\hat{y}$ terhadap nilai aktual Prediksi y.

$$e_t = y_t - \hat{y}_t \tag{2.11}$$

Kesalahan dalam prediksi mempengaruhi keputusan melalui dua cara yaitu kesalahan dalam memilih teknik prediksi dan kesalahan dalam mengevaluasi keberhasilan penggunaan teknik prediksi bagi pemakai prediksi, ketepatan ramalan yang akan datang adalah yang paling penting (Herjanto. E, 2004). Banyak pendekatan telah dikembangkan untuk mengukur *forecasting error*. Sebagian besar dari pendekatan tersebut dijelaskan dalam (Adhikari & Agrawal, 2013) seperti *Mean Forecast Error* (MFE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE) dan sebagainya. Pendekatan yang digunakan dalam penelitian ini untuk evaluasi model ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) adalah *Mean Absolute Percentage Error* (MAPE) dengan persamaan (2.12).

SEKOLAH PASCASARJANA

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \times 100\% \quad (2.12)$$

dengan,

e_t: $y_t - \hat{y}_t$

y_t: nilai Aktual variabel endogen

$\hat{y}_t$: nilai prediksi variabel endogen

n : jumlah total observasi (periode waktu yang dipertimbangkan)

Adapun langkah untuk menghitung *Mean Absolute Percentage Error* (MAPE) adalah

1. Menghitung error absolut.
2. Menghitung error absolut persentase.
3. Menghitung rata-rata error persentase.
4. Interprestasi hasil.

Nilai MAPE memberikan persentase kesalahan rata-rata antara prediksi dan nilai aktual. Semakin kecil nilai MAPE, semakin baik akurasi model. Penggunaan MAPE didasarkan bahwa MAPE mengukur kesalahan dalam bentuk persentase, sehingga lebih mudah dipahami oleh pengguna non-teknis. Sebagai contoh, jika MAPE adalah 5%, berarti prediksi model rata-rata meleset 5% dari nilai sebenarnya. Kemudian secara skalabilitas MAPE dinyatakan dalam persentase, hasil evaluasinya tidak bergantung pada skala data. Ini memungkinkan perbandingan performa model di berbagai dataset atau unit pengukuran yang berbeda. Selain itu MAPE hanya melihat besarnya kesalahan (absolute error), sehingga tidak terjadi pembatalan antara kesalahan positif dan negatif, seperti yang mungkin terjadi pada evaluasi dengan MAE (Mean Absolute Error). MAPE memiliki kelemahan yaitu apabila terdapat nilai 0 pada variable yang akan diprediksi akan menyebabkan terjadi pembagian oleh 0. Namun dalam konteks penelitian yang dilakukan dan didasarkan pada karakterisik dataset yang digunakan yaitu selalu terjadi transaksi penjualan dan kunjungan konsumen pada sistem distribusi omnichannel maka MAPE dapat digunakan dengan baik.

SEKOLAH PASCASARJANA