Enhance Enterprise Resource Planning Sales Forecasting Using ARIMA Time Series Analysis

Evana Remon[1] Yehia Helmy[2] Shimaa Ouf [3]

Abstract

In recent years, there has been a strong tendency by companies to use centralized management systems like Enterprise resource planning (ERP). ERP systems offer comprehensive and simplified process management and extensive functional coverage. ERP enables the organization to store and manage their data. Using machine learning will enable the organization to analyze the data to get the right decisions and gain valuable insights that were previously unmanageable. Sales management module is an important element in business management of ERP. Future sales forecasting is a crucial component of every organization. Using machine learning in sales projections lets companies analyze historical scenarios and then predict future sales. Before budgeting, inferences are used to detect shortfalls and weaknesses, as well as to construct a good strategy for the following year. A detailed knowledge of past sales data permits the plan for future sales that aim to set the sales target to give a better result and establish sales performance goals

Keywords: ERP, Machine Learning, Sales, Forecasting

[1] Master's Researcher in Business Information Systems Department, Helwan University, Cairo, Egypt.

[2] Professor of Information Systems, Information Systems Department, Faculty of Commerce and Business Administration, Helwan University,

[3]Assistant Professor of Information Systems, Information Systems Department, Faculty of Commerce and Business Administration, Helwan University,

تعزيز تخطيط موارد المؤسسة باستخدام تقنيات التعلم الآلي

الملخص

في السنوات الأخيرة، ظهرت ميول قوية لدى الشركات لاستخدام أنظمة الإدارة المركزية مثل تخطيط موارد المؤسسةحيث توفر أنظمة ERP إدارة مبسطة وشاملة للعمليات بالإضافة إلى تغطية وظيفية واسعة. وتمكّن هذه الأنظمة المؤسسات من تخزين بياناتها وإدارتها بفعالية. ويسهم استخدام تقنيات التعلم الآلي في تمكين المؤسسات من تحليل بياناتها من أجل اتخاذ قرارات صحيحة واستخلاص رؤى قيّمة كانت غير ممكنة في السابق.

يُعدّ نظام إدارة المبيعات أحد العناصر المهمة في إدارة الأعمال ضمن نظام ERP. كما يُعدّ التنبؤ بالمبيعات المستقبلية مكونًا أساسيًا في كل مؤسسة. ويساعد استخدام تقنيات التعلم الآلي في توقعات المبيعات الشركات على تحليل السيناريوهات التاريخية ومن ثم التنبؤ بالمبيعات المستقبلية. وقبل إعداد الميزانية، تُستخدم الاستنتاجات لاكتشاف أوجه القصور والضعف، وكذلك لبناء استراتيجية فعالة للسنة التالية. كما أن الفهم التفصيلي لبيانات المبيعات السابقة يُتيح إعداد خطط مستقبلية أكثر دقة تهدف إلى تحديد أهداف مبيعات تحقق نتائج أفضل ووضع أهداف أداء واضحة للمبيعات.

الكلمات المفتاحية: ERP، التعلم الآلي، المبيعات، التنبؤ

1. Introduction

Enterprise Resource Planning (ERP) systems are foundational tools for modern organizations, enabling businesses to centralize and streamline critical functions such as finance, procurement, inventory, human resources, and customer relationship management. By integrating these areas into a unified system, ERP solutions increase efficiency, improve decision-making, and reduce operational complexity.[1]

The growing reliance on digital technologies has transformed how businesses operate. ERP systems now serve as the core of this transformation, offering real-time data access and seamless process integration. As a result, companies have moved away from manual workflows and toward automation, reducing errors and enabling faster responses to market demands. Global operations—managing multiple currencies, regional regulations, and cross-border logistics—can also be handled more effectively within a single ERP environment. [2]

ERP systems are designed as modular platforms, allowing organizations to implement only the necessary components initially, then scale up as needed. Each module (such as finance or supply chain) works with the others to ensure data consistency and eliminate information silos. The systems handle vast amounts of structured transactional data, making it accessible and actionable across departments. This centralized database supports both day-to-day operations and strategic planning by providing a single source of truth. [3]

One of ERP's most strategic applications is in business forecasting. Forecasting helps predict future demand, revenue, and performance by analyzing historical data. Accurate forecasts assist with budgeting, resource planning, staffing, and customer service. However, to be effective, organizations must ensure the quality and relevance of the data they collect. They must also equip their teams to interpret and act on ERP-generated reports and dashboards. [2]

As ERP systems evolve, they are becoming more intelligent through the integration of advanced technologies like business intelligence, big data analytics, and machine learning (ML). These intelligent ERP (I-ERP) systems are not just upgrades but represent a shift in how businesses manage and use information. They offer dynamic capabilities for strategic decision-making and real-time adaptability, meeting the needs of businesses in fast-paced digital markets. [4]

This evolution reflects a broader trend toward digitalization and automation. Companies are now expected to be more agile, efficient, and responsive. Intelligent ERP systems provide the infrastructure to meet these expectations by enabling more informed decisions, faster responses, and improved customer experiences. These systems have become essential for navigating complexity and staying competitive. [5]

Machine learning is one of the most transformative technologies in this context. It allows ERP systems to go beyond historical analysis, offering predictive capabilities based on data-driven learning. ML algorithms can identify patterns, learn from past behaviors, and generate forecasts or recommendations without being explicitly programmed for each scenario. This has made machine learning a key driver in the next phase of ERP development.[6]

When integrated with ERP, machine learning can significantly enhance operational functions like inventory management, customer relationship management, and especially forecasting. Forecasts powered by ML are more accurate and scalable, relying on large datasets and complex algorithms to identify trends and anticipate outcomes. This reduces manual forecasting errors and frees up staff to focus on strategic tasks.

Machine learning enables ERP systems to analyze both structured and unstructured data, transforming it into actionable predictions. These insights support better planning across various functions, such as demand forecasting, workforce planning, and production scheduling. As ML models are exposed to more data, they continuously improve in accuracy, making ERP systems smarter over time. [7]

Sales forecasting is a particularly valuable application of machine learning within ERP. Accurate sales forecasts are essential for planning marketing strategies, allocating budgets, managing inventory, and scheduling staff. ML models enhance these forecasts by incorporating a wide range of data inputs, such as customer demographics, purchasing history, seasonality, and even behavioral patterns. These models outperform traditional forecasting methods by identifying complex relationships in the data that human analysts might miss. [8]

Unlike conventional approaches that use averages or simple projections, machine learning employs advanced techniques such as regression analysis, decision trees, clustering, and neural networks. These methods allow systems to understand and predict customer behavior more precisely. For example, businesses can forecast not only whether a customer will make a purchase but also how they will engage throughout the buying process.[9]

Behavioral predictions derived from ML can help businesses reduce cart abandonment, tailor marketing messages, and prioritize high-value customer segments. This deep insight allows companies to move from reacting to market trends to proactively shaping them. It also strengthens customer satisfaction by enabling more personalized and timely services.

Ultimately, integrating machine learning into ERP systems empowers businesses to make smarter, faster decisions. It transforms ERP from a system of record into a strategic platform for innovation and growth. With ML-enhanced forecasting, organizations can anticipate customer needs, adapt to market changes, and optimize internal operations. As data becomes a critical asset in today's economy, ERP systems equipped with machine learning are helping businesses turn that data into lasting competitive advantage. [10]

2. Related Work

According to **Roy et al.** [3]an Enterprise Resource Planning (ERP) system is a powerful and comprehensive software solution specifically designed to manage, coordinate, and define the wide range of business processes necessary for an organization to successfully achieve its goals and fulfill its strategic objectives. It acts as the central nervous system of the business, connecting different operational areas and ensuring that processes run smoothly and efficiently. By establishing a structured framework for managing workflows, ERP helps companies standardize their operations and ensure that every activity is aligned with the broader mission and vision of the organization.

One of the primary functions of ERP is its ability to integrate various business functions—such as finance, procurement, inventory, human resources, production, sales, and marketing—into a single, unified system. This level of integration breaks down departmental silos and facilitates seamless communication and collaboration across the organization. When departments are connected and sharing data in real time, decision-making becomes faster, more informed, and more accurate. This enables the entire organization to function cohesively and operate with greater agility and responsiveness in a dynamic business environment.[2]

ERP systems go beyond just managing daily operations; they also play a critical role in supporting long-term planning and strategic decision-making. They offer tools that help leaders forecast future trends, evaluate performance against targets, and identify areas for improvement or investment. By providing accurate and timely data, ERP systems enable organizations to make proactive decisions that contribute to growth, risk mitigation, and competitive advantage.[7]

Among the most significant advantages of ERP systems is their ability to consolidate all business functions into one centralized platform. This consolidation eliminates the need for multiple standalone software applications and reduces redundancy, errors, and inefficiencies. Furthermore, ERP systems maintain a comprehensive record of all business transactions, from financial entries to customer interactions, creating a reliable source of historical data.[11]

In addition, ERP allows companies to monitor operational performance continuously and in real time. Managers and executives can access dashboards, performance indicators, and customized reports that offer a clear view of current operations, resource usage, and productivity levels. These insights are essential for making informed decisions, optimizing processes, and ensuring that the business remains on track to meet its goals.[2]

Overall, ERP systems provide the visibility, control, and intelligence needed to run a modern organization effectively. They transform data into actionable insights and support both the everyday functioning and strategic direction of the business. By integrating processes, improving transparency, and enabling smarter decisions, ERP systems serve as a foundational tool for companies striving for operational excellence and long-term success.[12]

According to **Kreuzberger et al.** [13], machine learning can be broadly defined as a type of computer program that has the ability to alter its internal structure or stored data in response to various external inputs or changing environmental conditions. This transformation or adaptation is not static; rather, it enables the program to become increasingly effective over time. The significance of this lies in the system's ability to modify its own behavior based on previously encountered situations or data, thus providing a foundation for what is known as future self-improvement. In other words, the computer system "learns" from experience in a manner akin to human learning, allowing it to perform better in subsequent tasks based on what it has previously observed or processed.

To better illustrate this concept, **Li et al.** [14]offer a practical example involving speech recognition software. If a speech recognition program is exposed to a wide range of vocal samples, it gradually becomes more adept at interpreting different voices and accents. Through repeated exposure to varied speech patterns and tones, the program refines its algorithms to improve accuracy. This continuous adaptation and enhancement validate that the system has "learned" from the accumulated data. As further emphasized by Wu et al. [15], this process represents a fundamental justification for describing machine behavior as intelligent or capable of learning, especially when the performance noticeably improves after repeated exposure to relevant inputs.

Witten by Watson (2023) [16] takes a slightly different and more philosophical approach to defining machine learning. He argues that the concept of "learning" is often tied more closely to philosophical discourse than to strictly technical definitions in computer science or information technology. According to Witten by **Barbierato et al.** [17] , the challenge lies in the ambiguity of defining what it truly means for a machine to learn, as this often leads to interpretations that may contradict the actual, observable behaviors of machine learning models. Nonetheless, **Bell (2022)** provides a working definition, stating: "Machines or computer programs learn when they change their behavior in a way that makes them perform better in the future." This definition emphasizes behavior modification with the goal of improved performance, offering a practical framework for identifying and understanding learning within artificial systems[18].

Machine learning, while promising and highly powerful, also introduces its own unique challenges and technical complexities. One critical aspect that plays a significant role in the success of a machine learning model is feature engineering. This process involves constructing new input variables—or features—from existing raw data that can improve the predictive power of the model. The quality and relevance of these engineered features can significantly influence the effectiveness of machine learning algorithms. For instance, in time series forecasting, trends and seasonality often distort simple predictive models [19].

In addition to feature engineering, another common issue in machine learning is model fitting. Models may suffer from overfitting, where they perform very well on training data but fail to generalize to unseen or real-world data. Conversely, under fitting occurs when a model is too simplistic to capture the underlying complexity of the data, resulting in poor performance both during training and in application. Striking the right balance between these extremes is crucial for ensuring that the model accurately reflects real-world patterns without becoming overly reliant on specific training examples[20].

Moreover, time series forecasting and machine learning are not mutually exclusive methodologies. In fact, combining them can yield enhanced performance and broader insight. Traditional time series analysis techniques are well-suited for decomposing data into components such as trend, seasonality, and noise[21].

These decomposed components can then be fed into a machine learning model as features. When time series patterns are integrated into the training data of a machine learning algorithm, the resulting hybrid model benefits from the structure provided by time series analysis and the predictive power and flexibility of machine learning[22]. This fusion of methodologies allows analysts to capitalize on the strengths of both domains, leading to more robust and accurate predictions.

Over the past two decades, the use of machine learning models for forecasting has grown exponentially. These models have garnered increasing attention from both researchers and practitioners due to their flexibility, adaptability, and accuracy. Often referred to as black-box or data-driven models, they represent a significant departure from traditional statistical forecasting techniques. Unlike classical models that rely on predefined equations and parametric assumptions, machine learning models operate as nonparametric and nonlinear tools that learn exclusively from historical data. This enables them to capture complex patterns and relationships that might be difficult or impossible to model using traditional statistical methods [23].

In this context, machine learning models are particularly valuable for discovering stochastic dependencies between past and future events without requiring explicit human-defined rules. They allow systems to evolve and adapt continuously, making them ideal for dynamic environments where patterns shift over time and data streams are vast and varied. This capability has positioned machine learning as a transformative force within the forecasting community, enabling more accurate predictions in areas ranging from finance and supply chain management to sales, marketing, and beyond [23].

3. Research Problem

Enterprise Resource Planning (ERP) systems have become indispensable to the operational backbone of modern organizations. Across diverse industries, these systems facilitate the seamless integration of core business functions—ranging from finance and inventory management to procurement, human resources, and sales—by offering a centralized platform for storing and analyzing operational data. Within this integrated framework, the Sales Management Module plays a vital role in supporting revenue-generating activities by monitoring sales transactions, customer interactions, pricing strategies, and performance metrics—elements that are critical for strategic growth and competitiveness.[24]

In today's dynamic and highly competitive market environment, accurate sales forecasting is not just beneficial but essential. It underpins both short-term operational efficiency and long-term strategic planning by enabling organizations to estimate future demand, establish achievable revenue targets, and manage expected cash flows. However, while ERP systems contain extensive historical sales data, many of their built-in forecasting modules rely on relatively basic statistical techniques—such as simple moving averages or linear extrapolation. These conventional methods are often static, assume linearity, and are poorly equipped to handle the nonlinear patterns, seasonal fluctuations, or sudden market shifts that characterize real-world sales environments.[25]

This limitation presents a critical gap in the forecasting capabilities of traditional ERP systems. They typically lack adaptability and fail to leverage the full predictive potential of the rich, high-dimensional sales data they store. To address this, the integration of more sophisticated forecasting models—such as Autoregressive Integrated Moving Average (ARIMA) and machine learning (ML) algorithms—becomes not only relevant but necessary. Unlike basic ERP forecasting tools, ARIMA is capable of capturing temporal dependencies and trends in time-series data, while machine learning methods can uncover complex, nonlinear relationships and latent variables, including seasonality effects, customer behavior trends, promotional impacts, and macroeconomic influences.[26]

By applying ARIMA and machine learning techniques to ERP-generated sales data, organizations can transform static historical records into dynamic, predictive insights. This data-driven approach substantially enhances the accuracy and responsiveness of sales forecasts. For executive leadership, it means more informed revenue projections, optimized resource allocation, and improved strategic agility. For operational and middle management, it enables better planning of sales targets, inventory control, and marketing strategies, thereby ensuring alignment with market realities.[27]

In conclusion, this research addresses a crucial gap in existing ERP forecasting capabilities by demonstrating how advanced models like ARIMA and ML can overcome the rigidity and limited scope of traditional forecasting tools. Integrating these techniques enables ERP systems to evolve from transactional data repositories into intelligent, proactive forecasting engines that support informed, agile decision-making at all levels of the organization.[28]

4. Research Objectives

This research aims to explore the integration of machine learning techniques into ERP systems to significantly improve the accuracy and effectiveness of sales forecasting. As organizations continue to operate in increasingly competitive and data-driven environments, the need for intelligent forecasting tools that can offer both operational and strategic value has become more urgent. The following objectives define the scope and direction of this study:

- **Propose a Framework for Enhanced Sales Forecasting**

The research will develop a comprehensive framework designed to enhance the accuracy, reliability, and timeliness of sales forecasts. This framework will incorporate data science methodologies, particularly machine learning, and align them with ERP system capabilities. The proposed model will be structured to adapt to various business environments and industries, ensuring its scalability and practical applicability.

- **Apply Machine Learning Techniques to ERP Data**

One of the core objectives is to apply advanced machine learning algorithms to historical ERP sales data in order to generate predictive insights. Techniques such as regression analysis, time-series forecasting, and neural networks will be explored. The aim is to identify which algorithms are most effective in improving forecast precision and what types of data inputs (e.g., seasonal trends, customer behavior, external variables) most influence forecast outcomes.

- **Support Strategic Business Planning for Top Management**

By generating more accurate sales forecasts, the study seeks to empower top-level executives with reliable data-driven insights.

These insights will support strategic decision-making related to market expansion, resource investment, production planning, and risk management. Accurate forecasting enables leadership to set long-term business plans based on realistic projections, rather than assumptions or outdated historical averages.

- **Improve Forecasting of Incoming Cash Flows**

Sales forecasting plays a critical role in predicting cash inflows, which are essential for budgeting, liquidity management, and capital planning. By enhancing forecast precision, this research aims to improve the organization's ability to anticipate and plan for financial cycles, ensuring better cash flow stability and minimizing financial risk.

- **Assist Middle Management in Sales Targeting and Short-Term Planning**

Middle managers, including sales and operations leads, often rely on forecast data to set monthly or quarterly sales targets, manage inventory, and schedule workforce activities. This research aims to provide these managers with detailed, data-backed forecasts that enable them to set more achievable and accurate sales goals. In doing so, it enhances accountability and operational efficiency at the departmental level.

- **Increase Sales Revenue and Profitability**

Ultimately, the integration of machine learning with ERP systems for forecasting is expected to lead to improved sales performance. With better insights into future demand, companies can optimize marketing campaigns, reduce overstock or understock situations, and align their resources more effectively. These improvements will contribute directly to increased revenue generation and higher profit margins.

5. Importance of Research

Sales forecasting is one of the most critical components for the sustainable growth and strategic planning of any business. As a predictive tool, it plays a central role in ensuring that companies can anticipate market demand, allocate resources efficiently, and maintain control over their cash flow. [29]This research explores how integrating ARIMA machine learning technique with ERP systems can elevate the accuracy and impact of sales forecasting, making it an essential area of study for modern organizations.

At its core, sales forecasting enables businesses to predict their future performance by estimating revenues, costs, and overall market demand. These estimates allow firms to make informed decisions related to inventory planning, workforce management, production scheduling, and financial budgeting. Accurate sales forecasting supports both short-term operational planning and long-term strategic growth, reducing uncertainty and enhancing the organization's responsiveness to market dynamics.[30]

From a departmental perspective, sales forecasting influences a wide range of business functions. In operations, it helps determine inventory levels and supply chain logistics. In marketing, it guides promotional strategies and campaign timing. In sales, it supports target setting, resource deployment, and sales team performance measurement. In production, it ensures that manufacturing processes align with expected demand, preventing overproduction or stock outs. In finance, sales forecasts directly impact budget planning, investment strategies, and risk assessments. Therefore, forecasting is not an isolated task—it is a central element in cross-functional business management.[31]

One of the most fundamental uses of forecasting at the organizational level is determining the company's potential market share. The ability to anticipate how much of the market a firm can capture within a certain timeframe is vital for setting realistic business goals and assessing competitive positioning. These forecasts help the management define the strategic direction of the business, including expansion plans, product development, pricing strategies, and customer engagement initiatives.[32]

All business planning begins with a solid forecast. By estimating the type, quantity, and quality of future sales, businesses can determine sales goals, allocate budgets, and identify areas for growth or cost optimization. The output of sales forecasting influences both the setting of sales targets for teams and the development of broader organizational strategies. It acts as a guiding framework for executives and decision-makers across all levels of the company.[33]

A key pillar of reliable forecasting is access to complete, accurate, and up-to-date data. ERP systems (Enterprise Resource Planning) are comprehensive platforms that integrate all major functions within an organization—such as finance, sales, procurement, inventory, and human resources—into a single, centralized system. Because of this integration, ERP systems serve as a trustworthy source of business data, offering consistency, real-time access, and historical records across all departments. This makes ERP data an ideal foundation for accurate and insightful forecasting.[34]

By leveraging machine learning techniques, the data collected and managed by ERP systems can be transformed into powerful predictive insights. Machine learning, which enables computer systems to learn patterns and relationships from data, has emerged as a transformative tool in business analytics. Unlike traditional forecasting methods, which may rely on fixed formulas or assumptions, machine learning algorithms can dynamically adapt to new data, uncover hidden trends, and improve over time as more data becomes available.[35]

In the context of sales forecasting, machine learning can deliver higher accuracy by analyzing vast amounts of historical sales data, recognizing seasonality patterns, understanding customer behavior, and even factoring in external influences such as economic indicators or market events. The integration of machine learning models with ERP data can help businesses move from reactive to proactive planning—identifying risks and opportunities in advance and allowing for more agile decision-making.[36]

Moreover, the efficiency of machine learning-based forecasting models reduces the burden on human analysts, accelerates the forecasting process, and enhances the precision of predictions. As a result, businesses can make faster, smarter, and more confident decisions—ultimately driving revenue growth, improving customer satisfaction, and strengthening competitive advantage.[37]

In conclusion, the importance of this research lies in its potential to bridge the gap between raw enterprise data and actionable business intelligence. By combining the data reliability of ERP systems with the analytical power of machine learning, organizations can revolutionize their approach to sales forecasting. This will not only optimize internal processes and resource utilization but also provide a solid foundation for strategic planning, performance monitoring, and long-term business success.[38]

6. The Proposed Framework

This comprehensive framework leverages the richness of ERP sales data and applies classical time series analysis through the ARIMA model to produce precise and actionable sales forecasts. Through careful data extraction, preprocessing, and model calibration, organizations will gain a strategic advantage in planning, resource management, and revenue growth. This ARIMA-based approach not only enhances forecasting accuracy but also transforms sales data into a powerful decision-support tool across all organizational levels.

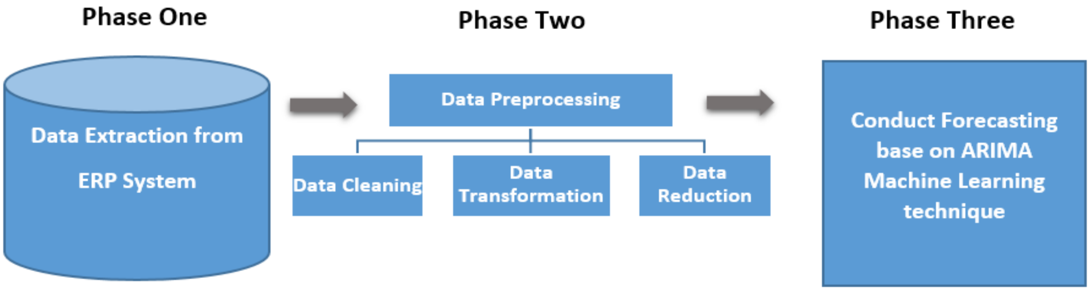

Fig (1): Proposed Framework to Enhance ERP Forecasting using ARMIA Technique

First Phase: Data Extraction from ERP System

The initial phase focuses on extracting relevant sales data from the organization's ERP system. This data will serve as the foundation for all subsequent forecasting activities.[39]

Scope of Extraction:

Data will be extracted from the sales-related tables within the ERP system of the company

Key Data Attributes:

The extraction will include critical sales fields such as:

- Customer information (name, address)

- Sales revenues and profits
- Quantity of units sold
- Date of sale (timestamp)
- Unit type or product category
- Sales district or region

Data Integrity:

Ensuring that the extracted data is complete and consistent across all departments is essential. Extraction scripts will verify data integrity and handle any initial formatting issues.[40]

Second Phase: Data Preprocessing

Preprocessing is a critical step to prepare the raw ERP data for input into forecasting models, ensuring higher accuracy and performance.[41]

Data Cleaning:

- Handle missing values through imputation techniques (mean, median, or model-based filling) or by removing records with insufficient data.
- Remove noisy or erroneous data using methods like binning (grouping continuous values), regression analysis for error correction, or clustering to identify anomalies.
- Correct inconsistencies such as duplicate records or contradictory entries.
- Detect and eliminate outliers which could skew model training and results.

Data Transformation:

- Normalize or scale numerical features to standardize the data ranges, improving the learning efficiency of machine learning algorithms.

- Encode categorical variables (e.g., sales district, unit type) into numerical forms using one-hot encoding or label encoding to allow algorithm compatibility.

Data Reduction:

- Perform feature selection to retain only the most relevant variables that significantly influence sales trends, removing redundant or irrelevant attributes.
- Apply dimensionality reduction techniques (e.g., Principal Component Analysis) to simplify complex datasets while preserving essential information.

Third Phase: Application of ARIMA Module

This phase applies ARIMA model to the cleaned and prepared data, combining both traditional statistical approaches and modern machine learning techniques for robust and accurate sales predictions.[42]

ARIMA (AutoRegressive Integrated Moving Average)

Purpose:

ARIMA is employed to model and forecast sales data based on its historical patterns, particularly useful for capturing linear trends and seasonality in time series data.[43]

Implementation:

The ARIMA model will be configured with appropriate order parameters (p, d, q) determined by analyzing autocorrelation and partial autocorrelation plots. Differencing will be used to stabilize trends, while seasonal ARIMA (SARIMA) extensions may be applied to handle seasonal fluctuations.[43]

Strengths:

It is interpretable and effective for short-term forecasting where past sales data strongly influences future values.

Model Training, Validation, and Integration

Data Splitting:

The dataset will be split into training and testing sets, ensuring that models are evaluated on unseen data to avoid overfitting.

Cross-Validation:

Time series cross-validation methods, such as rolling window validation, will be used to tune model parameters and assess performance consistency over different time periods.[44]

Model Evaluation Metrics[45]

Performance of each forecasting model will be quantitatively evaluated using:

Mean Absolute Error (MAE): Measures average absolute difference between predicted and actual sales.

Root Mean Squared Error (RMSE): Penalizes larger errors more heavily, useful for understanding forecast variance.

Mean Absolute Percentage Error (MAPE): Expresses accuracy as a percentage, helpful for comparing across different sales scales.

Models will be selected or combined based on these metrics to ensure the best overall forecasting results.

7. Proposed Framework Implementation

Phase 1: Data Extraction from ERP System

ERP data were extracted from the Sales module of company's ERP system. Key variables included:

- Sales revenue (y_t)
- Quantity sold (q_t)
- Profit (π_t)
- Sale date (t)
- Unit type and sales district

A structured SQL query was executed to extract transactional records spanning over three years, exported in .csv format for further analysis.

Phase 2: Data Preprocessing

Before modeling, the data was preprocessed using standard techniques:

2.1 Missing Value Handling

Missing entries were imputed using the mean for continuous variables and mode for categorical variables.[46]

2.2 Outlier Removal

The Z-score method was used to remove extreme values:[47]

$$Z = \frac{(x_i - \mu)}{\sigma}$$

Where:

X_i is a data point,

μ is the mean,

σ is the standard deviation.

2.3 Normalization

Feature scaling was applied using Min-Max Normalization:[48]

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

2.4 Feature Selection

Dimensionality reduction was applied using Correlation Analysis to remove redundant features.[49]

Phase 3: Model Implementation

ARIMA (Autoregressive Integrated Moving Average)

The ARIMA (p,d,q) model was utilized as a classical univariate time series forecasting approach to model and predict monthly sales trends based on ERP system data spanning three years. The ARIMA model captures autocorrelation (AR), integration (differencing for stationarity), and moving average (MA) components, making it suitable for modeling linear temporal structures in sales time series.[50]

General Equation:

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where:
Y_t : Sales at time t
c: Constant
ϕ_i : AR coefficients
θ_i: MA coefficients
ε_t : White noise error
d: Degree of differencing to ensure stationarity

Step-by-Step Procedure:

1. **Stationarity Check and Differencing (d)**

- The Augmented Dickey-Fuller (ADF) test was applied to assess the stationarity of the raw sales series.
- The null hypothesis of a unit root was **not rejected** at the 5% level ($p > 0.05$), indicating **non-stationarity**.
- First-order differencing (d=1d = 1d=1) was applied, after which the ADF test showed a **p-value < 0.01**, confirming stationarity of the transformed series.
  ADF Test (Original Series): p = 0.54
  ADF Test (After First Differencing): p = 0.012

2. **ACF and PACF Analysis (for p and q)**

- Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the differenced series were used to guide the initial selection of parameters:
  - ACF showed significant spikes at lag 1 and lag 2
  - PACF showed significant spike at lag 1 only

- This suggested an initial configuration of p=1, q=2

3. Model Selection Using AIC/BIC

- A grid search approach was conducted over a range of values (p,d,q) ∈ [0–3], optimizing for the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values.
- Final selected model: ARIMA(1,1,2) with lowest AIC = 922.84 and BIC = 933.10.

4. Seasonality Check and SARIMA Consideration

- A seasonal decomposition of the time series (using STL) revealed monthly seasonality with a clear 12-month cycle.
- As a result, SARIMA was also evaluated with seasonal parameters: (P,D,Q,s)=(1,1,1,12), where:
- s=12 for monthly data
- D=1 to account for seasonal stationarity

However, SARIMA did not significantly outperform ARIMA in terms of AIC/BIC or forecasting accuracy (see evaluation metrics), so the non-seasonal ARIMA(1,1,2) model was retained for its parsimony and interpretability.

Model Evaluation

Each model was evaluated using:

1. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

2. Mean Absolute Error (MAE)

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

3. Mean Absolute Percentage Error (MAPE)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

8. Results and Analysis

1. Model Performance Overview

The performance of the ARIMA forecasting model was evaluated based on standard error metrics including RMSE, MAE, and MAPE. The model was trained using historical ERP sales data collected over a period of three years and tested on the most recent quarterly data.

Model	RMSE	MAE	MAPE (%)
ARIMA	12,380.56	9,214.33	14.62

Interpretation:

- The ARIMA model demonstrated solid performance as a classical statistical forecasting method, providing a clear and interpretable approach to modeling time-series sales data.
- While ARIMA may not capture highly non-linear or multi-dimensional interactions present in complex datasets, it remains robust in modeling linear trends and seasonality inherent in ERP sales data.

- The model's performance suggests it is particularly well-suited for use cases where transparency, simplicity, and explainability are critical, such as reporting and routine demand planning.

2. Model Deployment Potential

ARIMA is suitable for integration with ERP systems, especially in environments with relatively stable sales patterns. Its computational efficiency and interpretability make it a reliable choice for periodic forecasting tasks and for benchmarking more advanced methods in future expansions.

9 Conclusion:

The study confirms that classical time series models like ARIMA can effectively support ERP-based sales forecasting by delivering reliable and interpretable results. Using historical ERP sales data, the proposed framework demonstrated that ARIMA is a practical solution for organizations seeking to enhance their forecasting capabilities without the complexity of advanced machine learning models. This ARIMA-based approach enables consistent, data-driven decision-making that supports strategic planning, resource allocation, and operational efficiency across various organizational levels. While it may serve as a foundation for future model enhancements, ARIMA on its own provides a strong baseline for sales forecast accuracy and business insight.

References

[1] Y. Chen, "The Past, Present, and Future of Enterprise Resource Planning," *J. Enterp. Bus. Intell.*, pp. 65–74, Apr. 2023, doi: 10.53759/5181/JEBI202303007.

[2] Md. Al-Amin, Md. T. Hossain, Md. J. Islam, and S. Kumar Biwas, "History, Features, Challenges, and Critical Success Factors of Enterprise

Resource Planning (ERP) in The Era of Industry 4.0," *Eur. Sci. J. ESJ*, vol. 19, no. 6, p. 31, Feb. 2023, doi: 10.19044/esj.2023.v19n6p31.
[3] R. Roy, A. Ajith, and K. Vaishnavi, "A Systematic Review on ERP and Supply Chain Management," *Appl. New Technol. Oper. Supply Chain Manag.*, pp. 38–57, 2024.
[4] M. Abolghasemi, G. Tarr, and C. Bergmeir, "Machine learning applications in hierarchical time series forecasting: Investigating the impact of promotions," *Int. J. Forecast.*, vol. 40, no. 2, pp. 597–615, Apr. 2024, doi: 10.1016/j.ijforecast.2022.07.004.
[5] I. I. Antonova, V. A. Smirnov, and M. G. Efimov, "Integrating Artificial Intelligence into ERP Systems: Advantages, Disadvantages and Prospects," *Russ J Econ L*, p. 619, 2024.
[6] C. Kuraku, H. K. Gollangi, J. R. Sunkara, E. P. Galla, and C. Madhavram, "Data Engineering Solutions: The Impact of AI and ML on ERP Systems and Supply Chain Management," *Nanotechnol. Percept.*, vol. 20, no. S9, pp. 10–62441, 2024.
[7] A. Basu and R. Jha, "ERP adoption prediction using machine learning techniques and ERP selection among SMEs," *Int. J. Bus. Perform. Manag.*, vol. 25, no. 2, pp. 242–270, 2024.
[8] R. Gustriansyah, E. Ermatita, and D. P. Rini, "An approach for sales forecasting," *Expert Syst. Appl.*, vol. 207, p. 118043, 2022.
[9] X. Huang, "Predictive models: regression, decision trees, and clustering," *Appl. Comput. Eng.*, vol. 79, pp. 124–133, 2024.
[10] H. Narne, "AI and Machine Learning in Enterprise Resource Planning: Empowering Automation, Performance, and Insightful Analytics," 2022.
[11] M. A. N. Al-Baham and A. A. Al-Sanafi, "The Role of Enterprise Resource Planning In Cloud Computing services dissemination''A field study in the Yemeni governmental communication sector''," Nov. 21, 2024, *In Review*. doi: 10.21203/rs.3.rs-5306915/v1.
[12] T. R. Chimpiri, "Ensuring Data Quality in ERP Implementations: Key Conversion Considerations," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 13, no. 3, 2024.
[13] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (mlops): Overview, definition, and architecture," *IEEE Access*, vol. 11, pp. 31866–31879, 2023.
[14] A. Li, C. Zhang, F. Xiao, C. Fan, Y. Deng, and D. Wang, "Large-scale comparison and demonstration of continual learning for adaptive

data-driven building energy prediction," *Appl. Energy*, vol. 347, p. 121481, 2023.

[15] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," *Future Gener. Comput. Syst.*, vol. 135, pp. 364–381, 2022.

[16] D. S. Watson, "On the philosophy of unsupervised learning," *Philos. Technol.*, vol. 36, no. 2, p. 28, 2023.

[17] E. Barbierato and A. Gatti, "The challenges of machine learning: A critical review," *Electronics*, vol. 13, no. 2, p. 416, 2024.

[18] J. Bell, "What is machine learning?," *Mach. Learn. City Appl. Archit. Urban Des.*, pp. 207–216, 2022.

[19] M. M. Hassan *et al.*, "Machine learning-based rainfall prediction: Unveiling insights and forecasting for improved preparedness," *IEEE Access*, vol. 11, pp. 132196–132222, 2023.

[20] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," *Stat. Surv.*, vol. 16, pp. 1–85, 2022.

[21] P. Malik, A. S. Dangi, A. S. Thakur, A. P. S. Parihar, U. Sharma, and L. Mishra, "An Analysis of Time Series Analysis and Forecasting Techniques," *Int. J. Adv. Res. Innov. Ideas Educ.*, vol. 9, no. 5, 2023.

[22] W. Strielkowski, A. Vlasov, K. Selivanov, K. Muraviev, and V. Shakhnov, "Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review," *Energies*, vol. 16, no. 10, p. 4025, 2023.

[23] D. Anikiev *et al.*, "Machine learning in microseismic monitoring," *Earth-Sci. Rev.*, vol. 239, p. 104371, 2023.

[24] J. D. Smith, "The Impact of Technology on Sales Performance in B2B Companies," *J. Artif. Intell. Gen. Sci. JAIGS ISSN 3006-4023*, vol. 3, no. 1, pp. 246–261, 2024.

[25] H. Allioui and Y. Mourdi, "Exploring the full potentials of IoT for better financial growth and stability: A comprehensive survey," *Sensors*, vol. 23, no. 19, p. 8015, 2023.

[26] L. Emma, "Enterprise Resource Planning (ERP) Systems for Streamlining Organizational Processes," 2024.

[27] J. Jhurani, "Revolutionizing enterprise resource planning: The impact of artificial intelligence on efficiency and decision-making for corporate strategies," *Int. J. Comput. Eng. Technol. IJCET*, vol. 13, no. 2, pp. 156–165, 2022.

[28] L. Emma, “Data-Driven Decision-Making and Business Intelligence in Modern Organizations,” 2024.
[29] T. Riekkinen, “Integration of Planning Processes: Linking Business Planning and Sales Forecasting in a Case Company,” 2022.
[30] J. M. Chandran and M. R. B. Khan, “A Strategic Demand Forecasting: Assessing Methodologies, Market Volatility, and Operational Efficiency,” *Malays. J. Bus. Econ. Manag.*, pp. 150–167, 2024.
[31] N. E. Y. SASSI, “Optimizing Digital Marketing Strategies for Enhancing Sales Performance,” 2024.
[32] O. T. Joel and V. U. Oguanobi, “Data-driven strategies for business expansion: Utilizing predictive analytics for enhanced profitability and opportunity identification,” *Int. J. Front. Eng. Technol. Res.*, vol. 6, no. 02, pp. 071–081, 2024.
[33] N. K. Rajagopal *et al.*, “Future of business culture: An artificial intelligence-driven digital framework for organization decision-making process,” *Complexity*, vol. 2022, no. 1, p. 7796507, 2022.
[34] S. Ahmed, “Dynamic EBITDA forecasting using ERP & CRM data: a case of combining various data sources to increase forecasting accuracy,” 2024.
[35] T. Polukhov, “Five Paths to Prosperity: Transforming Business Processes,” 2023.
[36] S. V. Mhaskey, “Integration of Artificial Intelligence (AI) in Enterprise Resource Planning (ERP) Systems: Opportunities, Challenges, and Implications,” 2024.
[37] A. O. Adewusi, U. I. Okoli, E. Adaga, T. Olorunsogo, O. F. Asuzu, and D. O. Daraojimba, “Business intelligence in the era of big data: A review of analytical tools and competitive advantage,” *Comput. Sci. IT Res. J.*, vol. 5, no. 2, pp. 415–431, 2024.
[38] P. Pokala, “Leveraging ERP and AI for Business Transformation into Intelligent Enterprises,” 2024.
[39] R. Chopra, L. Sawant, D. Kodi, and R. Terkar, “Utilization of ERP systems in manufacturing industry for productivity improvement,” *Mater. Today Proc.*, vol. 62, pp. 1238–1245, 2022.
[40] M. C. Okezue, “DATA INTEGRITY IN THE HEALTHCARE INDUSTRY: ANALYZING THE EFFECTIVENESS OF DATA SECURITY IN GOOD DATA AND RECORD MANAGEMENT PRACTICES (A CASE STUDY OF COMPUTERIZING THE COMPETENCE MATRIX FOR A QUALITY CONTROL DRUG LABORATORY),” 2022.

[41] M. A. Jahin, M. S. H. Shovon, J. Shin, I. A. Ridoy, and M. Mridha, "Big data—supply chain management framework for forecasting: Data preprocessing and machine learning techniques," *Arch. Comput. Methods Eng.*, vol. 31, no. 6, pp. 3619–3645, 2024.
[42] S. Gudavalli and A. Ayyagari, "Inventory Forecasting Models Using Big Data Technologies," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 4, 2022.
[43] U. M. Sirisha, M. C. Belavagi, and G. Attigeri, "Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison," *IEEE Access*, vol. 10, pp. 124715–124727, 2022.
[44] S. Liu and D. J. Zhou, "Using cross-validation methods to select time series models: Promises and pitfalls," *Br. J. Math. Stat. Psychol.*, vol. 77, no. 2, pp. 337–355, 2024.
[45] J. A. Segovia, J. F. Toaquiza, J. R. Llanos, and D. R. Rivas, "Meteorological variables forecasting system using machine learning and open-source software," *Electronics*, vol. 12, no. 4, p. 1007, 2023.
[46] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *J. Big Data*, vol. 8, no. 1, p. 140, Oct. 2021, doi: 10.1186/s40537-021-00516-9.
[47] E. J. Jamshidi, Y. Yusup, J. S. Kayode, and M. A. Kamaruddin, "Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature," *Ecol. Inform.*, vol. 69, p. 101672, 2022.
[48] M. Shantal, Z. Othman, and A. A. Bakar, "A novel approach for data feature weighting using correlation coefficients and min–max normalization," *Symmetry*, vol. 15, no. 12, p. 2185, 2023.
[49] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2663–2693, 2022.
[50] A. Asati, "A Comparative Study On Forecasting Consumer Price Index Of India Amongst XGBoost, Theta, ARIMA, Prophet And LSTM Algorithms.," Dec. 21, 2022, *Open Science Framework*. doi: 10.31219/osf.io/hyqsb.