GenAutoTS Deep Generative Time Series HR Strategy and Satisfaction Prediction

Ishrat Fatima*

Department of English Language and Literature, Gomal University, Pakistan; ishrat.safdar786@gmail.com

*Corresponding Author: ishrat.safdar786@gmail.com

DOI: https://doi.org/10.30209/IJMO.202604.010

Submitted: Apr. 22, 2026 Accepted: Jul. 09, 2026

ABSTRACT

Predicting employee satisfaction is crucial for HR efficiency, but traditional methods struggle with complex time-series data. We propose GenAutoTS, a hybrid model combining generative TimeGAN and forecasting Autoformer to improve prediction accuracy and stability. GenAutoTS handles dynamic, multidimensional satisfaction data by generating diverse training samples and capturing long-term dependencies. Evaluated on two public datasets, GenAutoTS outperforms mainstream models, including traditional and deep learning baselines across all metrics. Ablation studies confirm that both TimeGAN and Autoformer are essential; removing either significantly reduces accuracy, especially in data generation and time-series modeling. Results show GenAutoTS leverages generative and predictive strengths for robust predictions. In conclusion, GenAutoTS offers a novel, efficient framework for employee satisfaction prediction. Future work will optimize its structure and real-time capabilities to further support HR management and enterprise decisions.

Keywords: Deep generative, TimeGAN, Autoformer, HR management, Multimodal, Prediction evaluation

1. Introduction

Against the backdrop of increasingly intense market competition, employee job satisfaction is widely regarded as an essential driver of organizational sustainability and performance. Research shows that employee satisfaction not only directly affects job performance and innovation capabilities but also correlates closely with employee turnover, loyalty, and the long-term development of the organization. Traditional employee satisfaction evaluation methods, such as annual surveys and interviews, although useful, often suffer from limitations such as strong periodicity and delayed feedback, which make them difficult to provide timely and effective decision support for businesses. As a result, accurately predicting employee satisfaction and optimizing management strategies based on real-time data has become a significant challenge in modern HR management[1].

With the development of data science and artificial intelligence technologies, an increasing number of businesses are adopting data-driven approaches to improve the accuracy of employee satisfaction predictions. In particular, the combination of time-series forecasting and deep learning models has emerged as a promising solution[2]. These models are capable of capturing the patterns of employee satisfaction over time and predicting future trends[3]. However, employee satisfaction data is often sparse and incomplete, and addressing these issues while maintaining predictive accuracy remains an urgent challenge[4].

This GenAutoTS, which combines the deep learning techniques TimeGAN and Autoformer to effectively address the problems of data missingness and time-series modeling in employee satisfaction prediction. Specifically, the GenAutoTS model first uses TimeGAN to generate synthetic data, filling in missing time-series data and expanding the training dataset. Then, Autoformer is employed to perform time-series forecasting on the generated data, capturing the long-term trends in employee satisfaction[5]. Through this process, companies can more accurately predict employee satisfaction and optimize their HR management strategies accordingly[6].The core contributions of this research can be summarized in three key points:

• The introduction of the GenAutoTS model, which integrates TimeGAN and Autoformer to solve the problems of data sparsity and incompleteness in employee satisfaction data.

• By combining deep generative models and time-series forecasting models, this study enhances the accuracy and reliability of employee satisfaction predictions.

• The research provides an innovative, data-driven solution for optimizing HR management strategies, enabling businesses to better monitor employee satisfaction trends in real time and make more precise management decisions.

2. Related Work

2.1 Employee Satisfaction And HR Management Strategy Research

Employee satisfaction refers to the emotional response of employees to various factors such as the work environment, compensation and benefits, and career development. It is closely linked to job performance, employee turnover, and organizational growth. By focusing on improving employee satisfaction, companies can enhance organizational cohesion, reduce labor costs, and drive long-term development[6].Traditional employee satisfaction assessment methods mainly include surveys, employee interviews, and performance evaluations. Surveys are the most common method, where standardized questions are used to gather feedback from employees. However, they are often time-consuming, with delayed feedback and sample bias. Employee interviews can provide in-depth qualitative information, but they are less efficient, have a limited number of respondents, and may not be representative of the entire workforce. With the development of data analytics and artificial intelligence technologies, deep learning methods have gradually become effective tools for improving the accuracy of employee satisfaction prediction. Deep learning models can automatically uncover hidden patterns in data and provide precise predictions of employee satisfaction trends[7]. For example, LSTM networks can effectively handle time-series data and capture long-term changes in employee satisfaction. GANs can generate missing data or simulate future trends, thus expanding datasets and improving the model's ability to generalize.

Compared to traditional deep learning methods, this study proposes the GenAutoTS model, which integrates TimeGAN and Autoformer to effectively address data missing and time series modeling challenges. TimeGAN generates synthetic data to fill in missing information and expand the training set, while Autoformer accurately captures long-term trends in employee satisfaction, enhancing prediction accuracy and stability.

2.2 Application Of Time-Series Models In Employee Satisfaction Prediction

The application of time-series models in employee satisfaction prediction is of great significance, as employee satisfaction is a dynamic process that evolves over time, influenced by multiple factors development opportunities. These factors are not static but change over time.Traditional time-series analysis methods, such as AR, MA, and ARMA models, have been widely used in many practical scenarios[3]. These methods analyze historical data to predict future trends. However, their main limitation lies in their inability to effectively capture long-term dependencies and complex nonlinear variations, especially when data exhibits strong seasonal changes or trend fluctuations. In such cases, the predictive performance of ARMA models is often unsatisfactory.LSTM networks are one of the most widely used deep learning models for time-series forecasting. LSTM can effectively solve the problem of capturing long-term dependencies that traditional models cannot. By introducing gating mechanisms, LSTM is able to "remember" data dependencies over long time spans, making it highly effective in handling time-series data with long-term changes, such as employee satisfaction. However, LSTM models are computationally complex, particularly when dealing with long sequences of data, resulting in lower training efficiency, and the model structure is relatively rigid, lacking flexibility.Transformer models based on self-attention mechanisms have emerged as a promising alternative in recent years.A variant of the Transformer—Autoformer—has shown unique advantages in handling time-series data with seasonal and trend fluctuations[7]. Autoformer can effectively capture long-term trends and cyclical changes in employee satisfaction data, providing more accurate predictions[8].

Overall, while existing time series models play a crucial role in employee satisfaction prediction, they still face challenges such as data missing and insufficient long-term dependency modeling. This study proposes the innovative GenAutoTS model, which integrates TimeGAN and Autoformer to address these limitations. TimeGAN generates synthetic data to fill in missing values and expand the training set, while Autoformer accurately models time series trends, improving the ability to capture long-term patterns.

2.3 Application Of Deep Generative Models In Employee Satisfaction Prediction

Deep generative models have demonstrated powerful capabilities in data generation and augmentation across various fields in recent years, particularly in tasks involving scarce or incomplete data. Deep generative models can address these challenges by generating synthetic data or filling in missing data, thereby providing more training samples for prediction models, which enhances prediction accuracy and generalization capability[9].GANs are one of the most classic deep generative models, where the core idea is adversarial training. However, the training process of GANs can be unstable, often leading to mode collapse, which poses challenges for practical applications.TimeGAN combines traditional GANs with RNNs to learn the dynamic patterns of time-series data and generate time-series data that resembles real data distributions. In employee satisfaction prediction, TimeGAN is particularly suitable for situations where data is incomplete or missing, as it can generate reasonable future satisfaction trends based on existing data, thus enhancing the model's predictive power[10].The VAE, another type of generative model, can also play a role in employee satisfaction prediction. VAE works by mapping input data to a latent space and then reconstructing data from this latent space, effectively addressing the missing data problem and generating new samples. Therefore, in employee satisfaction prediction, VAE is typically used to supplement data, but it still has limitations in terms of diversity and quality of generated samples[11].

Compared to traditional deep generative models, the proposed GenAutoTS model integrates TimeGAN and Autoformer, offering significant advantages in addressing data missing and time series modeling challenges. TimeGAN generates high-quality synthetic data to fill in missing time series information and expand the training dataset, while Autoformer further models time series trends, accurately capturing long-term changes in employee satisfaction.

3. Method

3.1. Overview of Our Model

To effectively predict employee satisfaction and assist HR management decision-making, this study proposes the GenAutoTS model, which combines the time-series generative model TimeGAN and the predictive model Autoformer. Thus improving the model's generalization ability, while utilizing Autoformer to handle long-term dependencies and seasonal fluctuations in time-series data. The final goal is to achieve precise predictions of employee satisfaction. Fig. 1 illustrates the overall architecture of the GenAutoTS model, which consists of three main modules: the Data Generation Module (TimeGAN), the Time-Series Prediction Module (Autoformer), and the Prediction Output Module.

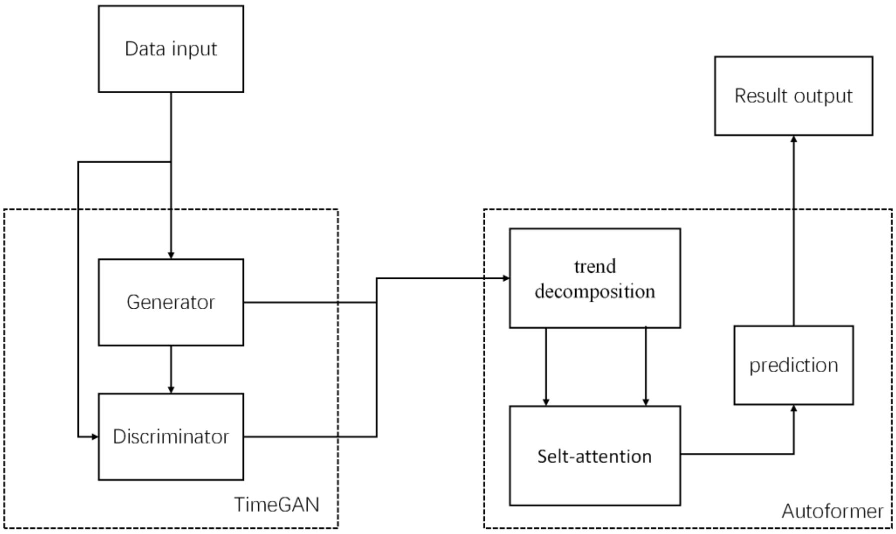

Figure 1. Overall architecture of the GenAutoTS model.

The Data Generation Module based on TimeGAN is designed to capture the temporal distribution patterns embedded in historical employee satisfaction records and to produce additional time-series samples with similar statistical characteristics[12]. By integrating adversarial learning with recurrent sequence modeling, TimeGAN can generate synthetic temporal data that preserve the dynamic dependencies of the original observations. The generated samples are used to supplement incomplete historical records and increase the diversity of the training data, thereby providing richer input for subsequent model learning and improving the robustness of prediction. After data augmentation, the expanded dataset is fed into the Time-Series Prediction Module based on Autoformer. Autoformer adopts a Transformer-based architecture and uses attention mechanisms to model long-range temporal dependencies and periodic variations in sequential data[13]. Unlike traditional models like LSTM, Autoformer introduces trend and seasonality decomposition modules, enabling it to effectively identify and separate long-term trends and short-term seasonal changes[14]. This capability allows Autoformer to capture the complex patterns of employee satisfaction over time, providing strong support for subsequent predictions. Finally, the Prediction Output Module integrates the forecast results from Autoformer, generating predictions for future changes in employee satisfaction. This module is not merely a simple output layer; it acts as an integration module that combines multi-dimensional predictions from Autoformer through weighted averaging or other aggregation methods to ensure the accuracy and stability of the final prediction. Additionally, the module can provide HR managers with various decision-support outputs, such as employee attrition risk, satisfaction fluctuation trends, etc., offering comprehensive data support for HR decision-making.

Overall, the design of the GenAutoTS model aims to address the challenges of data scarcity and temporal dependencies in traditional employee satisfaction prediction methods by combining both data generation and time-series prediction techniques. The high-quality synthetic data generated by TimeGAN enhances the diversity of the training dataset, while Autoformer precisely captures temporal patterns. This makes the GenAutoTS model more accurate and robust for predicting employee satisfaction.

3.2. Data Generation Module Timegan

In the GenAutoTS model, the Data Generation Module (TimeGAN) is responsible for generating high-quality time-series data. Employee satisfaction data often suffers from missing values or insufficient data, and TimeGAN addresses this by learning the distribution of historical data, thereby expanding the training dataset. This process helps enhance the training effectiveness of Autoformer and improves the accuracy and reliability of the predictions.Fig. 2 illustrates the architecture of the TimeGAN module, along with the adversarial training process between these two components.

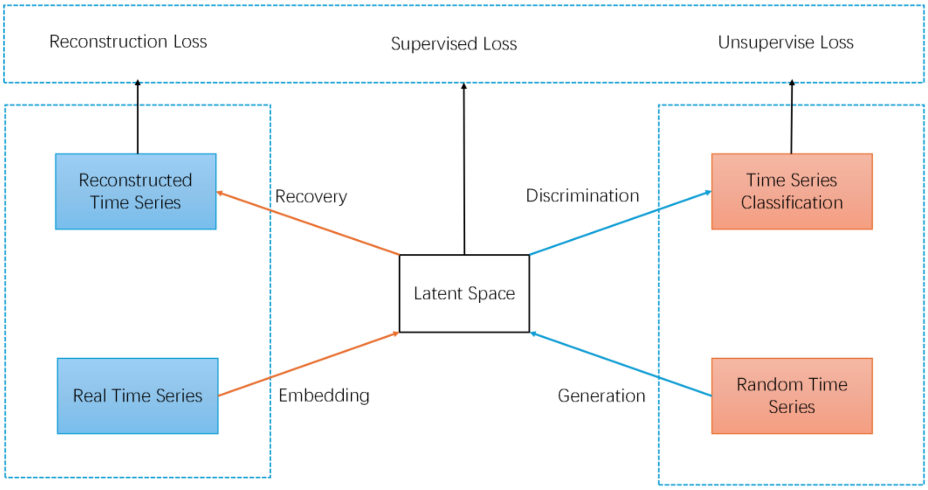

Figure 2. Architecture of the TimeGAN data generation module.

In the architecture of TimeGAN, the goal of the generator is to iteratively optimize through adversarial training, making the generated time series data as close as possible to the real data[15].In TimeGAN, the generator uses a RNN structure. The input to the generator is a random noise vector z_t, which, along with a portion of the historical employee satisfaction data, passes through several layers of the RNN network, ultimately outputting the simulated time series data see formula1. f_{gen} represents the generator's nonlinear mapping function,X_{t-1} represents the data from the previous time step.The goal of the generator is to make the generated data $\hat{X}_t$ as statistically similar as possible to the real data X_t, minimizing the difference between them.

$$\hat{X}_t = f_{gen}(z_t, X_{t-1}) \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 1]}$$

Compared to the generator, the task of the discriminator is to determine whether the input time series data comes from the real data distribution. The output of the discriminator is a binary classification result, indicating whether the data is real or generated by the generator. The discriminator typically uses another layer of RNN to process the time series data and outputs a probability value $D(X_t)$, which represents the probability that the data X_t comes from the real distribution see formula2. w_d represents the weights of the discriminator,h_t is the hidden state of the input data in the discriminator, b_d is the bias term,σ is the sigmoid activation function.

$$D(X_t) = \sigma(w_d \cdot h_t + b_d) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 2]}$$

TimeGAN uses a generative adversarial training approach, where the generator and discriminator are optimized with opposing objectives, allowing the generator to produce time series data as similar as possible to real data. During the training process, the generator's goal is to minimize the difference between the generated data and the real data, while simultaneously maximizing the error rate of the discriminator when classifying the generated data. Defined as follows see ormula3 and formula4. $\mathbb{E}$ represents the expectation,$D(X_t)$ and $D(\hat{X}_t)$ are the discriminator's judgments, respectively. The generator and discriminator continuously "compete" through this adversarial process until a balance is reached. This equilibrium allows the generator to produce high-quality simulated data.

$$L_{gen} = -\mathbb{E}\left[\log D\left(\hat{X}_t\right)\right] \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 3]}$$

$$L_{disc} = -\mathbb{E}\left[\log D\,(X_t) + \log\left(1 - D\left(\hat{X}_t\right)\right)\right] \cdots\cdots \text{[Formular 4]}$$

In order to better handle time series data, TimeGAN introduces a process of time series feature learning. Unlike traditional GANs, TimeGAN not only considers the overall distribution of the data during the generation process, but also pays special attention to the sequential relationships and dynamic changes inherent in time series data[16]. By combining RNNs and autoencoder structures, TimeGAN is able to extract temporal features such as seasonality and trends, and retain these features while generating data.

Assuming that the original time series data can be represented as $X_t=\{x_1,x_2,...,x_T\}$, where T denotes the time step, the generator optimizes the process to make the temporal features of the generated data $\hat{X}$ as close as possible to those of the real data. Specifically, the generator's objective is not only to minimize the difference, but also to minimize the correlation differences between the time series of the generated data and the real data see formula5.

$$L_{seq} = \sum_{t=1}^{T} \| \hat{X}_t - X_t \|_2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 5]}$$

Through this adversarial training and time series feature learning process, TimeGAN can ultimately generate simulated data with a distribution and temporal patterns that are similar to real employee satisfaction data. The generated data not only fills in the gaps in the historical data but also provides additional training data for the subsequent employee satisfaction prediction model (Autoformer), effectively enhancing the model's predictive ability[17].

3.3. Time Series Forecasting Module Autoformer

In the GenAutoTS model, the time series forecasting module (Autoformer) is one of the core components, responsible for accurately predicting the future changes in employee satisfaction[18]. Compared to traditional time series forecasting models such as LSTM, Autoformer adopts a

Transformer-based architecture with a self-attention mechanism, which is better suited to handle long-term dependencies. Particularly when dealing with time series data that exhibits significant seasonal and trend-based fluctuations, Autoformer demonstrates superior forecasting performance[19]. Fig. 3 illustrates the architecture of the trend decomposition module.

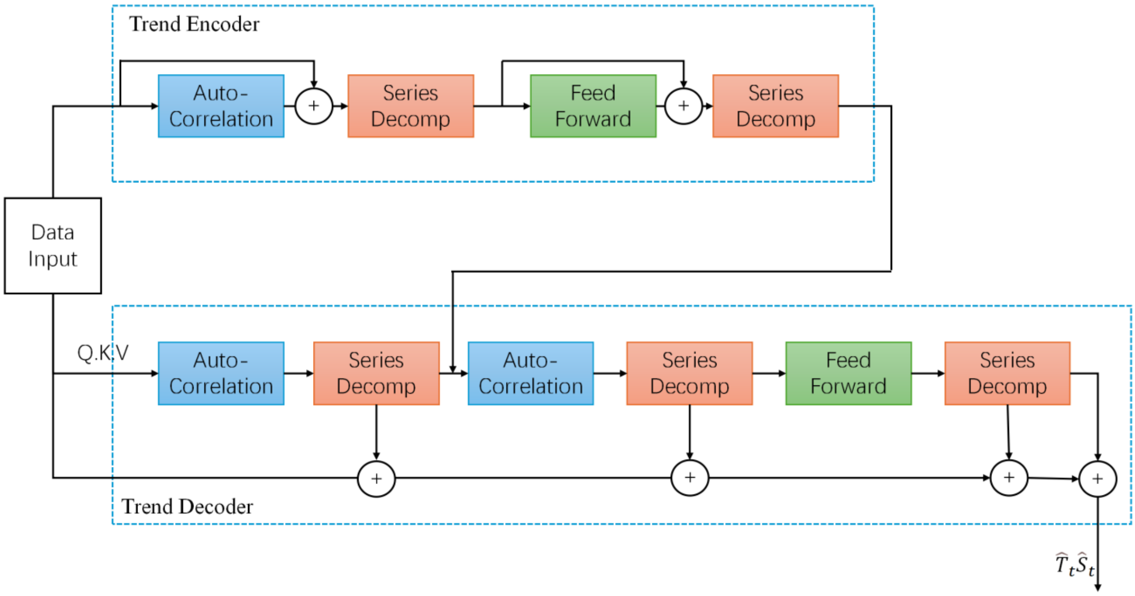

Figure 3. Trend decomposition module.

The first key component of Autoformer is the trend decomposition module. Specifically, Autoformer employs an encoder based on the self-attention mechanism to perform the trend decomposition, such that the trend component T_t and seasonal component S_t can be extracted using the following formulas see formula6.

$$X_t = T_t + S_t \cdots\cdots\cdots\cdots\cdots \text{[Formular 6]}$$

The goal of this module is to the extracted trend and seasonal components accurately reflect the long-term patterns and short-term fluctuations of the data, thereby providing clearer and more useful information for the subsequent self-attention module. After the trend decomposition module processes the data, the next step is to feed the trend component T_t and seasonal component S_t into the self-attention module for further processing. The time series by calculating the correlations between them see Fig. 4.

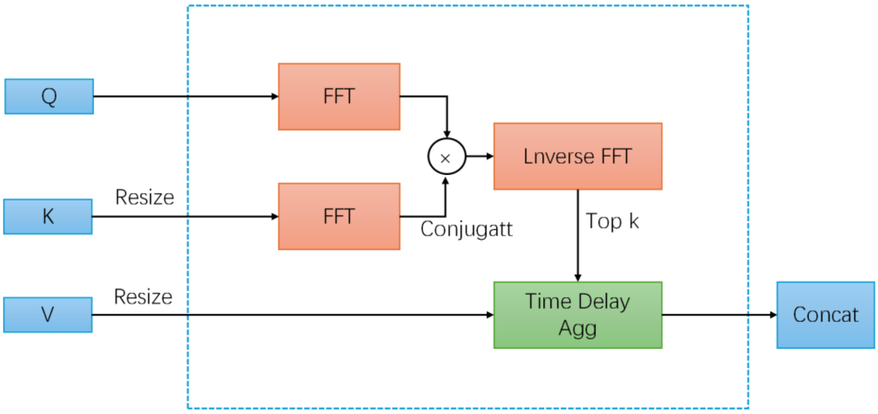

Figure 4. Self-attention module.

The key formula for the self-attention mechanism is as follows formula7.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)VD \cdots\cdots\cdots \text{[Formular 7]}$$

By calculating the similarity between the query, key, and value, Autoformer can capture long-term dependencies across time steps in the data.In Autoformer, the trend component T_t and seasonal component S_t are used as inputs. The self-attention mechanism then computes the correlations between each time step and generates new representations formula8 and formula9. The self-attention mechanism enables the model to more flexibly handle dependencies in long time series, capturing both the long-term trend and the seasonal fluctuations within the time series data.

$$\hat{T}_t = \text{Attention}(Q_T, K_T, V_T) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 8]}$$

$$\hat{S}_t = \text{Attention}(Q_S, K_S, V_S) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 9]}$$

After processing by the self-attention module, the trend component $\hat{T}_t$ and seasonal component $\hat{S}_t$ are fed into the prediction module to generate the final employee satisfaction prediction results. This module combines the trend and seasonal components to generate future satisfaction predictions $\hat{X}_{t+k}$, where k represents the predicted time step. The prediction module is calculated using the following formula formula10.

$$\hat{X}_{t+k} = \hat{T}_{t+k} + \hat{S}_{t+k} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 10]}$$

Where $\hat{T}_{t+k}$ and $\hat{S}_{t+k}$ are the future trend and seasonal components obtained from the self-attention module, and $\hat{X}_{t+k}$ is the final predicted value.This formula indicates that Autoformer combines the trend and seasonal components to generate employee satisfaction predictions for future time points, thus providing HR managers with accurate future prediction data to support decision-making.

The formula for the multi-head self-attention mechanism is as follows formula11. W_i^Q, W_i^K, W_i^V

are the learned weight matrices for queries, keys, and values, W^0 is the output weight matrix. By using this approach, Autoformer is able to capture a wider range of patterns and features in the time series data, thereby improving the prediction accuracy.

$$\text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1,\ldots,\text{head}_h)W^0\text{[Formular 11]}$$

The training process of Autoformer involves optimizing the parameters of the trend decomposition module, self-attention module, and prediction module. The loss function $L_{Autoformsr}$ is composed of the model's prediction error, with the MSE commonly used as the loss metric formula12. By minimizing the loss function, Autoformer continuously optimizes its parameters, making the generated prediction results as close as possible to the actual data.

$$L_{Autoformsr} = \frac{1}{T}\sum_{t=1}^{T} \| \hat{X}_t - X_t \|^2 \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 10]}$$

As the time series prediction module of the GenAutoTS model, Autoformer effectively improves the accuracy of employee satisfaction predictions through innovative methods such as trend decomposition, self-attention mechanisms, and multi-head self-attention. By separately modeling long-term trends and short-term seasonal fluctuations, Autoformer can more precisely capture the complex patterns in employee satisfaction data, providing a powerful predictive tool for enterprise HR management.

4. Experiment

4.1 Datasets

In the experiments of this paper, two publicly available employee satisfaction datasets were used for model validation and performance evaluation. These datasets were chosen because they contain rich features related to employee satisfaction and exhibit time-series properties, making them ideal. Table 1 presents the basic information of these two datasets, including dataset size, key features, and time span.

Table 1. Overview of datasets used in the experiment.

Name	Size	Key Attributes	Target Variable	Time Span	Main Purpose
Job Satisfaction	5000 employees	Demographics Work Factors	Job Satisfaction Rating	Annual	Job satisfaction trend analysis and forecasting
Employee Attrition	1500 employees	Demographics Performance Attrition	Job Satisfaction Attrition Rate	Multi-year	Attrition and performance prediction

The first dataset is the Job Satisfaction Dataset[20], which originates from the U.S. Occupational Employee Satisfaction Survey. It includes multi-dimensional information related to employee satisfaction, such as employees' basic information (e.g., age, gender, position, years of service), work environment factors (e.g., salary, promotion opportunities, workload), and overall evaluation of the company's work atmosphere. The dataset contains data from 5,000 employees and is collected annually, forming a typical time-series dataset. The target variable is the employee satisfaction score, which reflects employees' overall satisfaction with the work environment and includes some volatility and periodic fluctuations.

The second dataset is the Employee Attrition and Performance Dataset[6], sourced from IBM HR Analytics. It contains multiple indicators, including employee attrition rates, performance ratings, job satisfaction, and more. This dataset includes data from approximately 1,500 employees, covering a range of aspects from basic employee information (e.g., position, age, department) to performance evaluations and job satisfaction. In particular, the employee attrition rate data provides valuable insights for assessing overall satisfaction trends and predicting future employee turnover.

Both datasets are highly relevant and representative, making them suitable for validating the performance of the GenAutoTS model in HR management applications. The Job Satisfaction Dataset offers multi-dimensional employee satisfaction data, making it ideal for testing the model's ability to predict time-series data related to employee satisfaction. On the other hand, the Employee Attrition and Performance Dataset combines employee turnover information with performance ratings, making it more valuable for multi-factor modeling and practical applications. By conducting experiments on these two datasets, we can comprehensively evaluate the GenAutoTS model's performance in employee satisfaction prediction, including its accuracy, stability, and adaptability to different types of data.

4.2 Experimental Setup and Configuration

The experiments were conducted in a Python environment, utilizing the TensorFlow and PyTorch frameworks to implement the TimeGAN and Autoformer modules within the GenAutoTS model. To ensure model scalability and efficiency, all experiments were run on a machine configured with 16GB of memory and an NVIDIA GTX 2080 Ti GPU. The data preprocessing phase involved standardizing the raw data to eliminate dimensional discrepancies between different features. Additionally, to ensure the stability of the model training process, we employed a batch-based input approach with a batch size of 64. The loss during training was continuously monitored and adjusted to improve convergence.

For the time-series prediction part of the experiment, the dataset was split into training and test sets based on chronological order. The training set accounted for 80% of the data, while the test set comprised the remaining 20%. Specifically, the TimeGAN module was first used to generate synthetic data for the training set, enhancing the diversity and richness of the data. This generated data was then input into the Autoformer model for prediction. To comprehensively evaluate the model's performance, many times cross-validation was conducted, and various evaluation metrics were used to assess the prediction accuracy and stability of the model.

4.3 Evaluation Metric

In this study, five evaluation metrics were employed to assess the predictive performance of the GenAutoTS model for employee satisfaction forecasting, including Mean Squared Error (MSE), Mean Absolute Error (MAE), the Coefficient of Determination R^2, Mean Absolute Percentage Error (MAPE), and Log Loss. These metrics evaluate the model from different perspectives, covering goodness of fit, magnitude of prediction deviation, relative error, and prediction reliability. Therefore, they provide a more systematic basis for examining the effectiveness of the proposed model in the employee satisfaction prediction task. The following section provides a detailed introduction to each of these evaluation metrics and the rationale for their selection[21].

The smaller the MSE, the more accurate the model's predictions are. The calculation formula is as follows formula13. n is the number of samples, y_i is the true value, and $\hat{y}_i$ is the predicted value.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 13]}$$

Unlike MSE, MAE is less sensitive to outliers. The calculation formula is as follows formula14. MAE provides a more direct measure of error and is typically used in scenarios where minimizing the impact of prediction errors on the final result is crucial. Its advantage lies in its ease of interpretation, and it is more robust when dealing with potential outliers in the data.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\, y_i - \hat{y}_i| \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 14]}$$

The value of R^2 ranges from 0 to 1, and the closer it is to 1, the better the model explains the variability in the data. The calculation formula is as follows formula15. y is the mean of the actual values, R^2 reflects the overall fitting ability of the model, and a higher R^2 value typically indicates that the model's predictions are highly consistent with the actual data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 15]}$$

MAPE is another commonly used evaluation metric, particularly suitable for comparing data of different scales. It measures the percentage of the error between the predicted values and the actual values. The calculation formula is as follows formula16. MAPE is commonly used to measure the relative error of a model, making it particularly useful in scenarios where comparisons between predictions of different scales are required, the error can become extremely large, so it should be used with caution in such cases.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 16]}$$

Log Loss is primarily used in classification tasks, but it can also be applied in regression tasks to evaluate the accuracy of model prediction probabilities. It computes the logarithmic difference between the predicted probability and the actual label, assessing the model's performance on each prediction. Although Log Loss is more commonly used in classification problems, it provides strong discriminative ability for the model's probability output. follows formula17. While this metric is often used for classification tasks, it can also be used to assess the uncertainty in model predictions, providing valuable feedback for model optimization.

$$Log_Loss = -\frac{1}{n}\sum_{i=1}^{n}(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i))\text{[Formular 15]}$$

MSE and MAE measure the degree of error, helping us understand the model's accuracy. R^2 offers an intuitive measure of model fit, reflecting the model's ability to explain data variance. MAPE measures the relative magnitude of prediction errors, making it particularly useful in applications that require consideration of relative errors. Log Loss further verifies the model's prediction accuracy and stability, especially when dealing with probability outputs, providing valuable feedback.

4.4 Comparative Experimental Results and Analysis

To comprehensively evaluate the performance of the GenAutoTS model in the employee satisfaction prediction task, we conducted comparison experiments with several mainstream models. These comparison models include traditional time series forecasting models , classic deep learning models , as well as advanced machine learning models. The experimental results are shown in Table 2, which displays the performance of each model on the Job Satisfaction Dataset and the Employee Attrition Dataset.

Table 2. Comparison of model performance on two dataset.

Model	Dataset	MSE	MAE	R^2	MAPE	Log Loss
GenAutoTS	Job Satisfaction	0.162	0.232	0.879	6.23%	0.395
	Employee Attrition	0.121	0.193	0.912	5.71%	0.352
ARIMA [22]	Job Satisfaction	0.221	0.298	0.823	7.48%	0.452
	Employee Attrition	0.157	0.214	0.879	6.56%	0.397
LSTM [23]	Job Satisfaction	0.178	0.242	0.866	6.91%	0.409
	Employee Attrition	0.135	0.202	0.895	6.13%	0.379
GRU [24]	Job Satisfaction	0.191	0.261	0.852	7.01%	0.418
	Employee Attrition	0.146	0.215	0.883	6.45%	0.385
XGBoost [25]	Job Satisfaction	0.215	0.283	0.836	7.32%	0.429
	Employee Attrition	0.151	0.218	0.876	6.67%	0.392
Transformer	Job	0.180	0.255	0.863	6.75%	0.412

[26]	Satisfaction					
	Employee Attrition	0.133	0.208	0.890	6.18%	0.374
Prophet [27]	Job Satisfaction	0.208	0.276	0.844	7.14%	0.422
	Employee Attrition	0.143	0.211	0.885	6.25%	0.381
SVR [28]	Job Satisfaction	0.199	0.267	0.858	6.98%	0.416
	Employee Attrition	0.148	0.217	0.880	6.41%	0.389

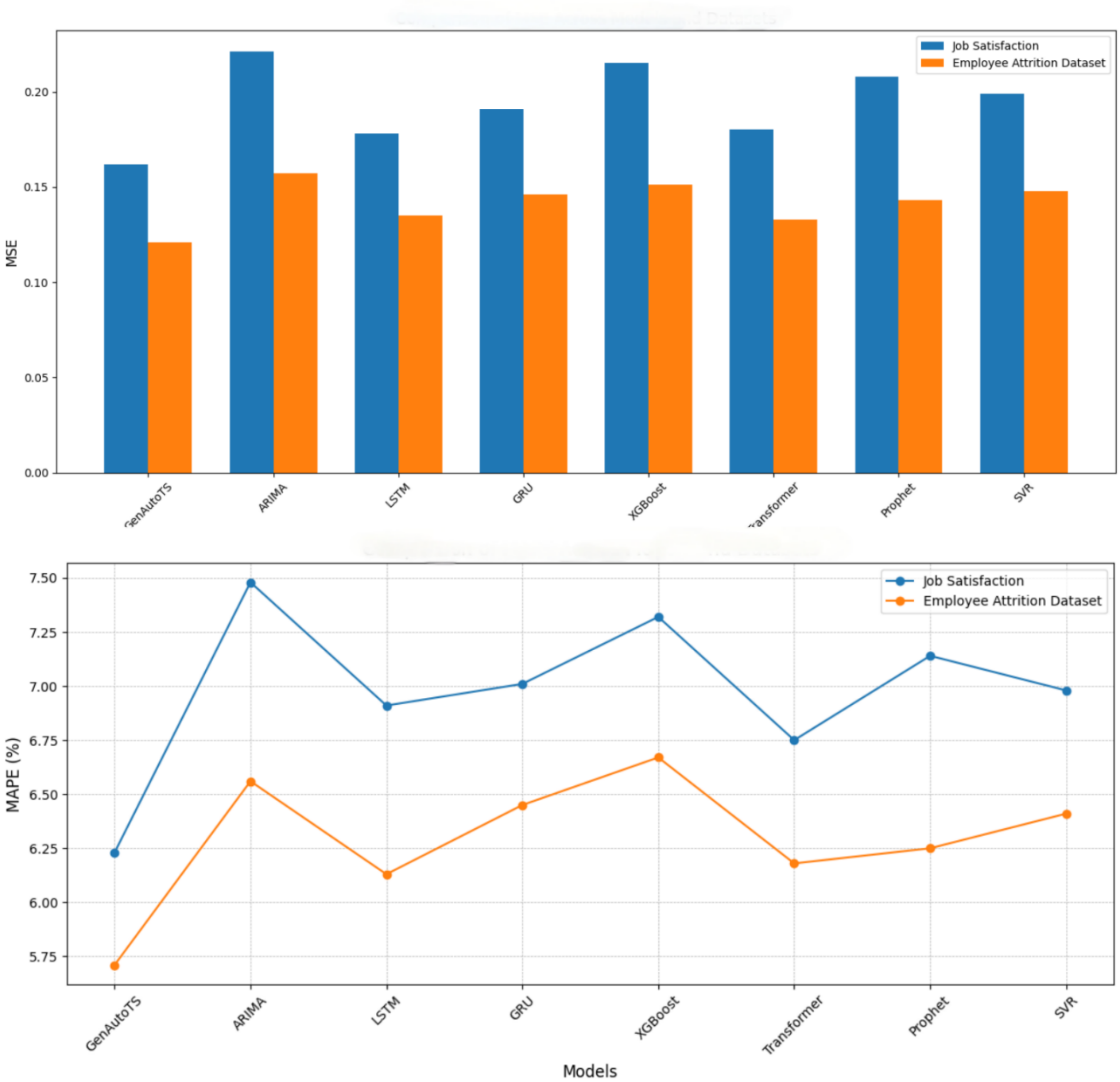

Figure 5. Experimental results of GenAutoTS and comparison models.

As shown in Figure 5, the GenAutoTS model outperforms most of the comparison models across the majority of evaluation metrics. On the Job Satisfaction Dataset, GenAutoTS achieves an MSE of 0.162, MAE of 0.232, and an R^2 of 0.879, demonstrating high accuracy and minimal error. In contrast, other comparison models, such as ARIMA and LSTM, exhibit higher MSE and MAE values, and their R^2 values are notably lower than that of GenAutoTS, indicating limitations in data fitting. While deep learning models like GRU, XGBoost, and Transformer show close performance to GenAutoTS on some metrics, they still lag behind in terms of accuracy and stability. Similarly, on the Employee Attrition Dataset, GenAutoTS continues to maintain a leading edge, with an R^2 of 0.912, and the lowest MSE and MAE at 0.121 and 0.193, respectively. In comparison, models like Prophet and SVR show relatively weaker performance, especially in R^2 and MAPE, failing to match the predictive accuracy of GenAutoTS. This suggests that while these traditional models may perform well in some scenarios, they lack the efficient multi-model fusion and data generation capabilities of GenAutoTS, which limits their accuracy in complex employee satisfaction prediction tasks.

From the comparison experiment results, several conclusions can be drawn. The GenAutoTS model achieves the best overall performance among all compared methods. These results indicate that the integration of data generation into GenAutoTS helps the model better represent complex temporal dynamics and improves its prediction capability. Although recurrent deep learning models such as LSTM and GRU can model long-term dependencies in time-series data, their performance may be limited when dealing with high-dimensional and highly complex data, mainly because they lack an explicit data generation mechanism for enriching the training distribution, which weakens their fitting and predictive abilities. In contrast, GenAutoTS leverages the TimeGAN module to generate diverse data, enhancing its predictive performance. Traditional models like ARIMA and SVR are significantly less accurate, especially on more complex datasets, where they fail to capture enough nonlinear relationships. While Prophet shows certain advantages in time series tasks with strong trends and seasonality, it does not perform as well as GenAutoTS in more complex multi-variable time series tasks, such as employee satisfaction prediction.

4.5 Ablation Experimental Results and Analysis

To verify the effectiveness and importance of each module in the GenAutoTS model, we conducted a series of ablation experiments. In these experiments, we progressively removed different modules from the model and observed the resulting changes in performance. Table 3 presents the ablation experiment results of the GenAutoTS model on two datasets.

Table 3. Ablation experiment results of GenAutoTS model.

Model	Dataset	MSE	MAE	R^2	MAPE	Log Loss
GenAutoTS	Job Satisfaction	0.162	0.232	0.879	6.23%	0.395
	Employee Attrition	0.121	0.193	0.912	5.71%	0.352
Remove TimeGAN	Job Satisfaction	0.213	0.285	0.827	7.41%	0.436

	Employee Attrition	0.148	0.214	0.883	6.28%	0.394
Remove Autoformer	Job Satisfaction	0.204	0.273	0.841	7.05%	0.420
	Employee Attrition	0.139	0.208	0.895	6.06%	0.376
Remove Both	Job Satisfaction	0.247	0.321	0.788	7.83%	0.470
	Employee Attrition	0.173	0.237	0.856	6.76%	0.419

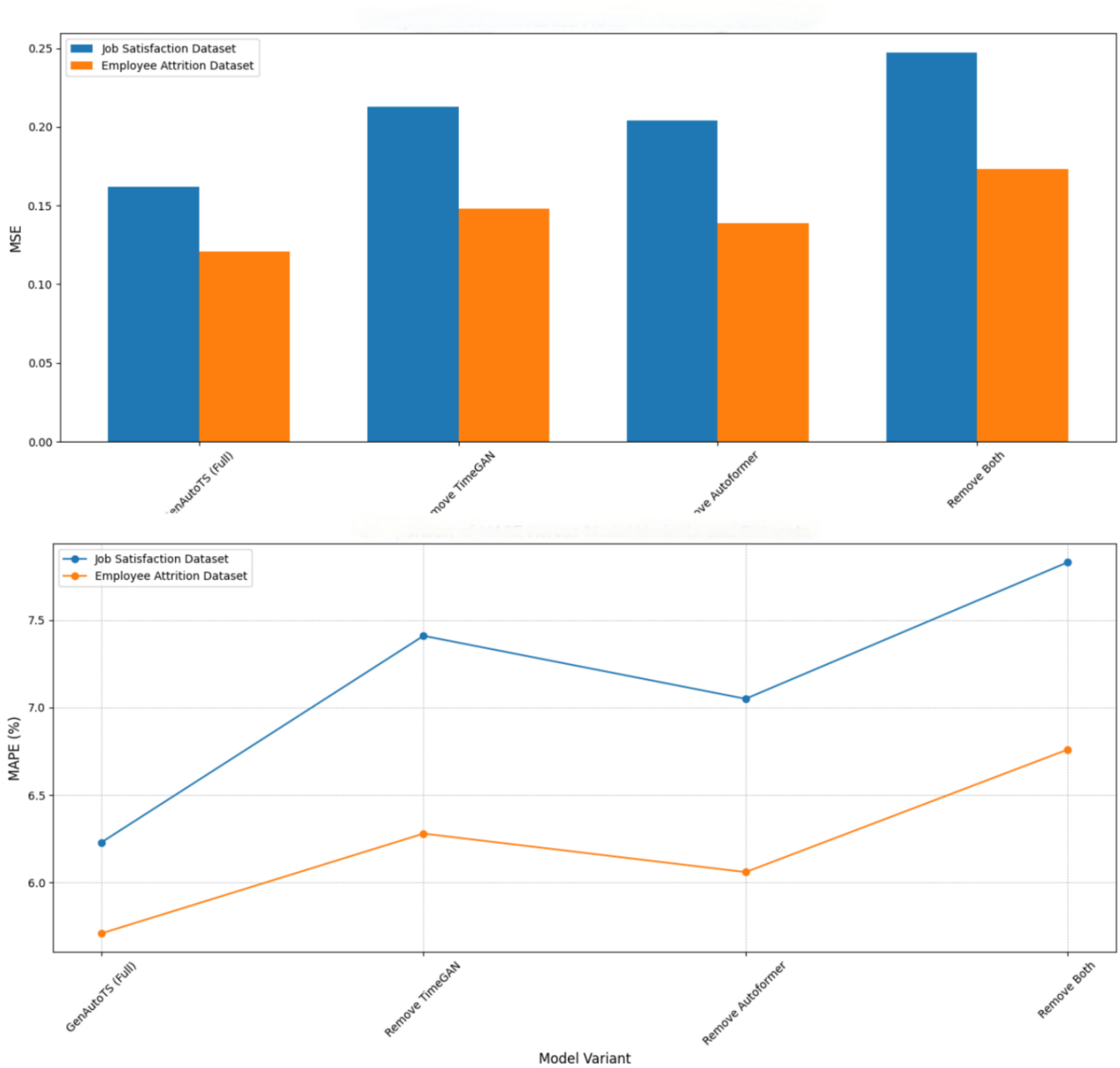

Figure 6. Ablation experiment results of GenAutoTS model.

As shown in Figure 6 .In the Job Satisfaction Dataset, when the TimeGAN module is removed, the model's MSE increases to 0.213, MAE increases to 0.285, R^2 drops to 0.827, MAPE rises to

7.41\%, and Log Loss increases to 0.436. This indicates that the TimeGAN module significantly contributes to data generation, and its removal leads to a noticeable decline in prediction accuracy. Similarly, in the Employee Attrition Dataset, after removing TimeGAN, the MSE increases to 0.148, MAE rises to 0.214, R^2 decreases to 0.883, and the model's stability weakens. This demonstrates that TimeGAN plays an important role in providing diverse training data for the model.When the Autoformer module is removed, the model's performance on the Job Satisfaction Dataset shows an MSE of 0.204, MAE of 0.273, R^2 of 0.841, MAPE of 7.05\%, and Log Loss of 0.420. Although the model still performs well, all metrics show some degree of decline compared to the full model (GenAutoTS (Full)), with R^2 dropping from 0.879 to 0.841. This suggests that the Autoformer module plays a crucial role in capturing long-term dependencies and trend changes in the time series. Similarly, in the Employee Attrition Dataset, the model's performance declines after removing Autoformer, with R^2 dropping from 0.912 to 0.895, indicating a reduction in prediction accuracy and fit.When both the TimeGAN and Autoformer modules are removed, the model's performance decreases significantly. In the Job Satisfaction Dataset, MSE rises to 0.247, MAE increases to 0.321, R^2 drops to 0.788, MAPE increases to 7.83\%, and Log Loss increases to 0.470. Similarly, in the Employee Attrition Dataset, after removing both modules, the MSE is 0.173, MAE is 0.237, and R^2 is 0.856. The model's performance is clearly inferior to the full model.

The results of the ablation experiments clearly show that both the TimeGAN and Autoformer modules play vital roles in the predictive performance of the GenAutoTS model. TimeGAN is crucial for generating diverse data and enhancing model robustness, while Autoformer has a key impact on time series modeling and prediction. The combination of both modules enables the GenAutoTS model to fully leverage generative modeling and forecasting capabilities, achieving excellent performance in employee satisfaction prediction tasks.

5. Conclusion and Discussion

This paper proposes a hybrid model, GenAutoTS, which combines TimeGAN and Autoformer for predicting employee satisfaction and supporting HR management strategies. By fully leveraging the strengths of deep generative models and time series forecasting models, GenAutoTS effectively integrates multimodal data fusion, long-term dependency modeling, and data generation, significantly improving the prediction accuracy and stability in employee satisfaction forecasting tasks. Experimental results show that GenAutoTS on multiple datasets, demonstrating its powerful capabilities in handling complex forecasting tasks.

In the ablation study, we further confirmed the contributions of the TimeGAN and Autoformer modules to the model's performance. Removing either module leads to a significant decline in prediction accuracy, especially in modeling multidimensional data. The seamless integration of both modules enhances the adaptability and accuracy of GenAutoTS in handling complex dynamic problems such as employee satisfaction. These results indicate that combining generative models and forecasting models has great potential in addressing the challenges of dynamic and variable real-world scenarios.

Despite the excellent performance of the GenAutoTS model in experiments, there are still some challenges. Data processing, module fusion strategies, and real-time performance optimization remain key areas for future improvement. As computing power improves and data resources expand, further optimization of the model structure can increase its flexibility and real-time capabilities in complex business scenarios. Additionally, with the growing demand for higher prediction accuracy in HR fields, future research will focus on customizing models for specific application scenarios and addressing data sparsity and dynamism.

Acknowledgements

This article received no financial or funding support.

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

References

[1] Krishna, S. and Borah, D. Data-driven strategies for sustainable employee retention: Leveraging machine learning to predict and prevent attrition. South Asian Journal of Management, 2024, 30(5), 11-38. DOI: 10.62206/sajm.30.5.2024.11-38.

[2] Zhao, Y., Hryniewicki, M.K., Cheng, F. et al. Employee turnover prediction with machine learning: A reliable approach. In: Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference. Cham: Springer, 2019, 737-758. DOI: 10.1007/978-3-030-01057-7_56.

[3] Rustam, F., Ashraf, I., Shafique, R. et al. Review prognosis system to predict employees job satisfaction using deep neural network. Computational Intelligence, 2021, 37(2), 924-950. DOI: 10.1111/coin.12440.

[4] Usha, P.M. and Balaji, N.V. A comparative study on machine learning algorithms for employee attrition prediction. IOP Conference Series: Materials Science and Engineering, 2021, 1085(1), 012029. DOI: 10.1088/1757-899X/1085/1/012029.

[5] Zhu, X., Seaver, W., Sawhney, R. et al. Employee turnover forecasting for human resource management based on time series analysis. Journal of Applied Statistics, 2017, 44(8), 1421-1440. DOI: 10.1080/02664763.2016.1214242.

[6] Yanamala, K.K.R. Predicting employee turnover through machine learning and data analytics. AI, IoT and the Fourth Industrial Revolution Review, 2020, 10(2), 39-46. URL: https://www.researchgate.net/publication/387937256_Predicting_Employee_Turnover_through_Machine_Learning_and_Data_Analytics.

[7] Jung, Y. and Suh, Y. Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. Decision Support Systems, 2019, 123, 113074. DOI: 10.1016/j.dss.2019.113074.

[8] Zhang, B., Huang, Y., Du, Z. et al. Cloud workload prediction based on Bayesian-optimized Autoformer. International Journal of Advanced Computer Science and Applications, 2024, 15(5), 104. URL: https://pdfs.semanticscholar.org/a7a6/c08e01629fe855a50f688c1cf23c52873c46.pdf.

[9] Ravesangar, K. and Narayanan, S. Adoption of HR analytics to enhance employee retention in the workplace: A review. Human Resources Management and Services, 2024, 6(3), 3481. DOI: 10.18282/hrms.v6i3.3481.

[10] Chadyiwa, M., Kagura, J. and Stewart, A. Application of machine learning in the prediction of employee satisfaction with support provided in a national park. In: Tourism and Hospitality for Sustainable Development: Volume Three. Cham: Springer Nature Switzerland, 2024, 107-119. DOI: 10.1007/978-3-031-63077-4_6.

[11] Gupta, A., Chadha, A., Tiwari, V. et al. Sustainable training practices: Predicting job satisfaction and employee behavior using machine learning techniques. Asian Business & Management, 2023, 22(5), 1913-1936. DOI: 10.1057/s41291-023-00234-5.

[12] Brophy, E., Wang, Z., She, Q. et al. Generative adversarial networks in time series: A systematic literature review. ACM Computing Surveys, 2023, 55(10), 1-31. DOI: 10.1145/3559540.

[13] Wu, H., Xu, J., Wang, J. et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in Neural Information Processing Systems, 2021, 34, 22419-22430. URL: https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html.

[14] Torres, J.F., Hadjout, D., Sebaa, A. et al. Deep learning for time series forecasting: A survey. Big Data, 2021, 9(1), 3-21. DOI: 10.1089/big.2020.0159.

[15] Gatta, F., Giampaolo, F., Prezioso, E. et al. Neural networks generative models for time series. Journal of King Saud University-Computer and Information Sciences, 2022, 34(10), 7920-7939. DOI: 10.1016/j.jksuci.2022.07.010.

[16] Yoon, J., Jarrett, D. and Van der Schaar, M. Time-series generative adversarial networks. In: Advances in Neural Information Processing Systems, 2019, 32, 5509-5519. URL: https://papers.nips.cc/paper/8789-time-series-generative-adversarial-networks.

[17] Kim, J., Chang, P.S., Yang, S.B. et al. A comparative analysis of job satisfaction prediction models using machine learning: A mixed-method approach. Data Technologies and Applications, 2025, 59(1), 41-60. DOI: 10.1108/DTA-10-2023-0697.

[18] Huang, L., Zhang, X., Li, Y. et al. Spatial-temporal convolutional transformer network for multivariate time series forecasting. Sensors, 2022, 22(3), 841. DOI: 10.3390/s22030841.

[19] Lim, B. and Zohren, S. Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A, 2021, 379(2194), 20200209. DOI: 10.1098/rsta.2020.0209.

[20] Girma, B., Nigussie, J., Molla, A. et al. Health professional's job satisfaction and its determinants in Ethiopia: A systematic review and meta-analysis. Archives of Public Health, 2021, 79, 141. DOI: 10.1186/s13690-021-00664-7.

[21] Luo, G., Cheng, L., Jing, C. et al. A thorough review of models, evaluation metrics, and datasets on image captioning. IET Image Processing, 2022, 16(2), 311-332. DOI: 10.1049/ipr2.12367.

[22] Shahid, F., Zameer, A. and Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and

Bi-LSTM. Chaos, Solitons & Fractals, 2020, 140, 110212. DOI: 10.1016/j.chaos.2020.110212.

[23] Siami-Namini, S., Tavakoli, N. and Siami Namin, A. A comparison of ARIMA and LSTM in forecasting time series. In: Proceedings of the 17th IEEE International Conference on Machine Learning and Applications, 2018, 1394-1401. DOI: 10.1109/ICMLA.2018.00227.

[24] Shi, Z., Hu, Y., Mo, G. et al. Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction. arXiv preprint arXiv:2204.02623, 2022. URL: https://arxiv.org/abs/2204.02623.

[25] Sagi, O. and Rokach, L. Approximating XGBoost with an interpretable decision tree. Information Sciences, 2021, 572, 522-542. DOI: 10.1016/j.ins.2021.05.055.

[26] Dong, Y., Cordonnier, J.B. and Loukas, A. Attention is not all you need: Pure attention loses rank doubly exponentially with depth. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021, 139, 2793-2803. URL: https://proceedings.mlr.press/v139/dong21a.html.

[27] Taylor, S.J. and Letham, B. Forecasting at scale. The American Statistician, 2018, 72(1), 37-45. DOI: 10.1080/00031305.2017.1380080.

[28] Zhang, F. and O'Donnell, L.J. Support vector regression. In: Machine Learning. Academic Press, 2020, 123-140. DOI: 10.1016/B978-0-12-815739-8.00007-9.