International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Airline Passenger Satisfaction Prediction by Genetic-Algorithm-Based Hybrid AutoEncoder and Machine Learning Models

Lee Ye Hean[1], Olanrewaju Victor Johnson[2], Chew XinYing*[3], Teoh Wei Lin[4], Chong Zhi Lin[5], Khaw Khai Wah[6]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: The evolving landscape of the Business-to-Client (B2C) model across the globe is reshaping service delivery paradigms and transforming consumer perceptions of service providers, thereby revolutionizing customer experiences. This paradigm shift directly impacts airline companies that offer multiple service tiers, necessitating ongoing promotional strategies to attract and retain customers. Moreover, in addition to attracting new passengers, it is equally vital for airlines to retain existing ones. Therefore, comprehensive research is imperative to comprehend customers' perceptions and conduct post-flight customer satisfaction surveys to delve into the factors influencing their decision-making processes. By gaining insights into these crucial causal factors, airlines can tailor their services to better meet customer expectations and enhance overall satisfaction levels. To address these challenges, this paper proposes a hybrid model comprising Deep Autoencoder (DAE) and Genetic Algorithm (GA) techniques for optimizing feature extraction. Utilizing eleven Machine Learning (ML) models as baseline predictors, the study endeavors to forecast passenger satisfaction levels. Furthermore, each ML model is intricately combined with the AE-GA optimization framework to conduct in-depth customer satisfaction experiments. Conducting a 5-fold cross-validation analysis in each experimental setup, the study highlights the efficacy of the proposed optimization strategy in significantly enhancing the predictive performance of ML methods in forecasting customer satisfaction levels.

Keywords: *Airline, Autoencoder, Genetic Algorithm, Customer Experience, Optimization.*

1. Introduction

Over the past decade, global economic growth and in- creased multiple interconnections have spurred the emergence of numerous airline companies worldwide. In the face of intense competition among both new entrants and established players, airlines are adopting innovative strategies to maintain their competitive edge. Among these strategies is a keen focus on understanding the factors that influence passenger satisfaction for pre-and-post flight cases. According to Leon and Martín [1], factors such as service levels and onboard facilities play a crucial role in gaining a competitive advantage in the airline industry.

Moreover, customer satisfaction is a multifaceted concept that revolves around two crucial pillars: customer loyalty and customer trust [2]. At its core, customer loyalty refers to a buyer's steadfast commitment to purchasing products, services, and brands from a particular firm over an extended period, irrespective of competitive alternatives. This unwavering loyalty is deeply intertwined with the attitudes and behaviors that define customer satisfaction. Understanding customer behavior, therefore, is key to deciphering the intricacies of customer satisfaction [3]. By delving into the dynamics of consumer decision-making processes, businesses can gain valuable insights into the factors that influence customer loyalty. Al-Mashraie et al. [4] advocate for ongoing research aimed at unraveling the complexities of customer behavior, identifying potential drivers of customer turnover, and exploring innovative strategies for customer retention.

To better comprehend passenger satisfaction, various Machine Learning (ML) techniques are being employed to analyze passenger datasets [5]. Feature selection and reduction with ML models have shown promising results in enhancing accuracy levels [6, 7]. By leveraging these insights, airlines can enhance the quality of their services, ultimately fostering customer loyalty. Furthermore, recent studies have extensively compared different ML algorithms to identify the most-performing models for customer satisfaction in airlines. For instance, Bhargav and Prabu [8] compared the novel hybrid Random Forest (RF) model with the K- Nearest Neighbour (KNN). Findings revealed a notable advantage for the hybrid RF model, showcasing superior ac- curacy compared to KNN. The result revealed an advantage for the hybrid RF model, showcasing superior accuracy compared to KNN. Furthermore, the findings underscore the importance of leveraging advanced analytical techniques to analyze airline operations and improve overall customer satisfaction. However, many of the MLs lack appropriate feature selection optimizations.

Meanwhile, conventional ML methods heavily rely on hand-crafted feature selection processes, which are time-

[1,2,3] School of Computer Science, USM, Gelugor–11800, Penang, Malaysia
[4]School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia– 62200, Putrajaya, Malaysia.
[5] Dept of Elect Engineering, Faculty of E&G Technology Universiti Tunku Abdul Rahman Jalan Universiti, Kampar, 31900, Perak, Malaysia
[6]School of Management, USM, Gelugor, 11800, Penang, Malaysia
** Corresponding Author Email: xinying@usm.my*

consuming, energy-intensive, and often require expert domain knowledge [9]. In recent times, Deep Learning (DL) models have emerged as a powerful alternative, capable of automatically extracting features during data training. DL has been effectively applied across various domains, demonstrating its versatility and efficiency. For example, it has been instrumental in imaging [10], medical diagnostics [11], precision agriculture [12], stock market analysis [13], climate science [14], banking [15], human resource management [7], and education [16]. The array of techniques used in these various applications consists of Multilayer Perceptron (MLP) [17], Convolutional Neural Network (CNN) [9], Recurrent Neural Network (RNN) [18], Long Short-Term Memory (LSTM) [13], Gated Recurrent Units (GRU) [19], and AutoEncoders (AEs) [20]. While these techniques have revolutionized the way data is processed and analyzed, leading to more efficient and accurate outcomes across various fields, AEs have had a significant impact. AEs are especially effective in addressing a wide range of challenges, including the curse of dimensionality [20], anomaly detection [21], and image enhancement [22]. Their ability to learn compact representations of data makes them invaluable for reducing dimensionality, thereby simplifying complex datasets without significant loss of information. Additionally, AEs excel in identifying anomalies by learning the normal patterns in data and flagging deviations, and they are instrumental in image correction and enhancement tasks by reconstructing images to remove noise and imperfections.

On the other hand, optimizing the feature extraction process for effective training of models has been a significant focus in recent advances in ML [23, 24]. It is believed that with the application of appropriate optimization techniques, models can be trained more effectively to achieve optimal performance. Various optimization techniques have been proposed in the literature, including ant colony optimization [25], firefly-spider optimization [23], arithmetic optimization [24], multi-objective rain optimization [6], Genetic Algorithm (GA) optimization [6], Brownian motion-based butterfly optimization [18], and fuzzy particle swarm optimization [6]. These methods aim to enhance the accuracy and efficiency of ML models by refining the feature selection process.

This paper, therefore, proposes a hybrid approach combining Deep-AE (DAE) and GA with various ML models, set out as a comparative study, to determine which model performs best for customer satisfaction in the airline industry. The AE is utilized for feature reduction, while the GA optimizes the feature reduction process for optimal results. The ML models serve as baseline models, which are later integrated with the AE-GA process. The integration of AE and GA significantly improves the performance of the ML models, enhancing the overall accuracy of customer satisfaction prediction in the airline industry.

2. Related Works

Ouf [26] conducted a study on the application of DL using the adaptive moment estimation (Adam) optimization algorithm to enhance classification performance. This approach was applied to the airline passenger satisfaction dataset from the Kaggle repository, addressing the often-overlooked issue of dataset quality. The study validated the proposed method by comparing it against Artificial Neural Networks (ANNs), RF, and Support Vector Machines (SVMs) applied to the same dataset. The experimental results demonstrated that the proposed method outperformed previous studies, achieving a remarkable accuracy of 99.3%. Guimarães et al. [27] introduced an ML-based decision support models for various stages of flight management, including strategic, pre-tactical, tactical, and post-operation phases. The research focused on predicting missed flight connections at an airline's hub airport using historical flight and passenger data, analyzing factors contributing to the predicted outcomes for each decision horizon. The dataset used was high-dimensional, heterogeneous, imbalanced, and noisy, lacking specific information about passenger transit times. Data balancing with Gaussian Mixture Models, and boosting were employed. Findings show that the model achieved an area under the curve (AUC) of the receiver operating characteristic (ROC) higher than 0.93.

Furthermore, Chen et al. [28] studied integrating multi-source big data to enrich the feature dimensions of airline passengers while ensuring the privacy of their information. The model aimed to create detailed user profiles to accurately identify high-value passengers. The study showed improved AUC and Kolmogorov-Smirnov (KS) values compared to the traditional models using single-source data. Wang et al. [29] tackled the inefficiency of random advertisements in predicting passenger willingness to pay for seat selection, which often resulted in user fatigue and decreased engagement. They proposed the Bagging in Certain Ratio Light Gradient Boosting Machine (BCR-LightGBM) model to address this issue. The experimental results indicated that BCR-LightGBM outperformed twelve comparison models in terms of ROC-AUC and F1-score. Similarly, Pranav and Gururaja [30] examined an ML approach to enhance customer experience with airlines using a dataset provided by a real but anonymized airline. A stacking ensemble model was employed, utilizing logistic regression (LR), RF, and DT classifiers as base learners, with XGBoost as the meta-classifier. The XGBoost outperforms the best base learner (RF) by 2.6%. Also, Mottini and Acuna-Agost [31] focused on modeling air passenger choices of flight itineraries, traditionally handled by Multinomial Logit model (MNL). A Pointer Network, which combines Recurrent Neural Networks with the Attention Mechanism was proposed to solve the challenge in MNL. The model was evaluated using a dataset combining online user search logs and airline flight bookings. Results indicated that the Pointer Network-based model outperformed the traditional MNL model across several metrics.

In a similar approach, Mirza et al. [20] explored deep-correlated AE model for competitive customer-dependent applications. The model specifically addresses the challenge of determining customer churn. The proposed technique uses the deer hunting optimization algorithm (DHOA) to optimize the feature selection process. The experimental evaluation conducted on a proprietary dataset demonstrated that the approach significantly outperforms existing methods, highlighting its effectiveness in enhancing customer churn prediction. Haridasan et al. [24] focused on developing an arithmetic optimization algorithm (AOA) integrated with a stacked bidirectional long short-term memory (SBLSTM) model to predict customer churn in the telecommunications industry. To improve the customer churn prediction (CCP) performance, the AOA is applied for optimal hyperparameter tuning of the SBLSTM model. Extensive simulations using a benchmark dataset comprising 3333 samples and 21 features demonstrated the superior performance of the AOA- SBLSTM model, showcasing its potential in effectively forecasting customer churn. Garimella et al. [23] proposed a CCP using an optimized DL classifier built within the Spark architecture to handle large-scale telecom data. The proposed model based on CNN uses the Firefly-Spider Optimization (FSO) algorithm to optimize the training process. The effectiveness of the prediction model was tested using the Churn in Telecom dataset, yielding impressive performance metrics, with a maximal dice coefficient, accuracy, and Jaccard coefficient of 0.9461, 0.9476, and 0.9480 respectively.

In summary, these studies collectively highlight the advancements in using ML techniques to enhance various aspects of airline operations, from improving passenger satisfaction to optimizing operational decision-making and marketing strategies.

3. Methodology

The section below discusses the approaches in this paper.

3.1. Framework of the proposed approach

The architectural framework of the proposed approach comprises three major phases. First, the dataset undergoes feature reduction using a DAE. This is achieved by designing the AE with a bottleneck at its midpoint, enabling the reconstruction of input data and effective feature reduction. The reduced features generated by the DAE are then utilized for optimization. The GA optimization process in the second phase involves three fundamental stages: selection, crossover, and mutation, as detailed in the subsequent section. The output from the GA optimization is then fed into various ML algorithms for classification and clustering. Among the classification and clustering algorithms, the one with the highest accuracy is selected as the primary algorithm. The primary algorithm is then hybridized with the remaining algorithms, as illustrated in Fig. 1. This approach ensures optimal performance in predicting customer satisfaction in the airline industry.

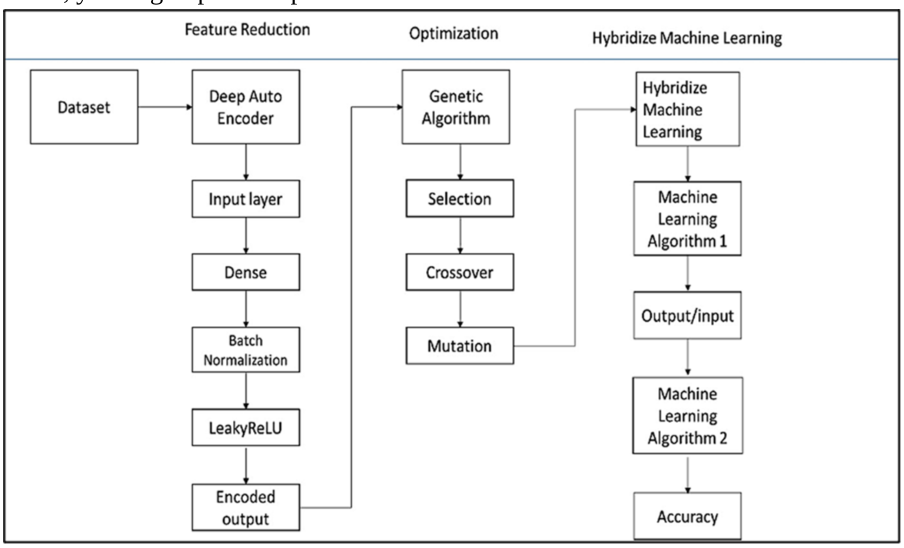

Fig. 1. Architectural Framework of the proposed approach.

3.2. Deep AutoEncoder

The DAE consists of two symmetrical deep neural networks, typically comprising n layers, based on the neural network theorem. The process begins with the input vector, which is compressed into a smaller dimension through the encoder layer, featuring fewer neurons than the input layer. This compressed representation is then decompressed through the decoder layer to reconstruct the input data [32]. DAEs are commonly used for feature reduction or feature extraction. The DAE architecture includes three standard layers: the input layer, the hidden layer, and the output layer, as shown in Fig. 2.

In this paper, the DAE is employed for feature reduction. The encoder part of the DAE learns to interpret the input data and compress it into a reduced feature set, defined by the bottleneck layer. The decoder then takes this compressed representation from the encoder and attempts to reconstruct the original input data. Once the DAE is trained, the decoder is discarded, and the encoder is used to compress the input dataset into a reduced vector output via the bottleneck layer. An AE is defined with the encoding and decoding parts as in (1) and (2) respectively:

$$x_n = \sum_{j=1}^{n} E_n \left(W_{E_n} x_0 + b_{E_n}\right) \quad (1)$$

$$x' = \sum_{j=1}^{n} D_n\left(w_{D_n} x_n + b_{D_n}\right) \quad (2)$$

where E_n and D_n are the encoding and decoding functions of the hidden layer $1 \dots n$, w represents the weights, and b the bias.

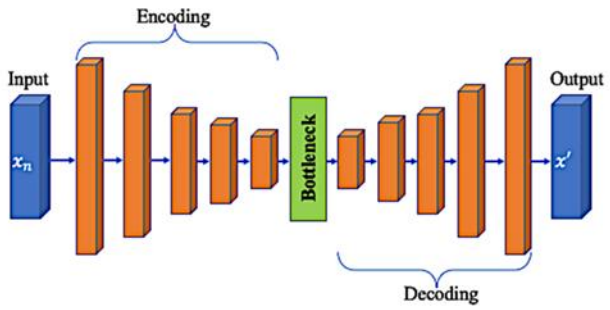

Fig. 2. The design of the DAE network.

3.3. Genetic Algorithms

GAs are stochastic global search optimization algorithms developed in recent years with inspiration based on the biological theory of evolution through natural selection [33]. The theory suggests that traits which enhance survival are more likely to be passed on to subsequent generations. GA incorporates principles of genetics and natural selection to find optimal or near-optimal solutions to complex problems.

In GAs, key concepts such as "Population," "Chromosome," and "Gene" are employed. While the population represents a subset of all potential solutions, comprising a set of chromosomes, each chromosome, in turn, is composed of a series of genes, as illustrated in Fig. 3. Accordingly, optimizing a model using GA involves three fundamental operators, including selection, cross-over, and mutation (For further reading see [33]).

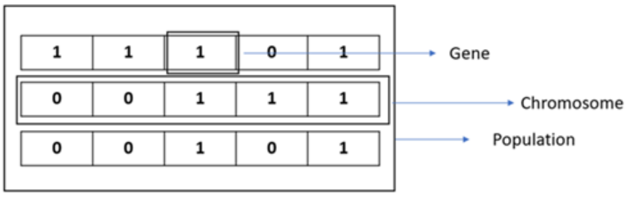

Fig. 3. Structure of genetic algorithm.

3.4. Hybrid Approaches in ML

Generally, the hybrid ML techniques are based on the combination of different approaches of classifier and clustering techniques [34], including i.) classifier-to-cluster shown, ii.) cluster-to-classifier, iii.) cluster-to-cluster, and iv.) classifier-to-classifier, as shown in Fig. 4 (a) to (d), respectively.

The idea behind the hybrid most often is that the first algorithm serves the purpose of "pre-processing", feature selection, or reduction. The result from the first algorithm then being replaced on the original dataset is then used to train the second algorithm which could provide better results than a single individual algorithm. In this paper, the ML algorithm with the highest baseline accuracy is used as the first ML algorithm in the hybrid techniques. The process, therefore, involves the combination of classifier and classifier or classifier and cluster as the experimental results reveal.

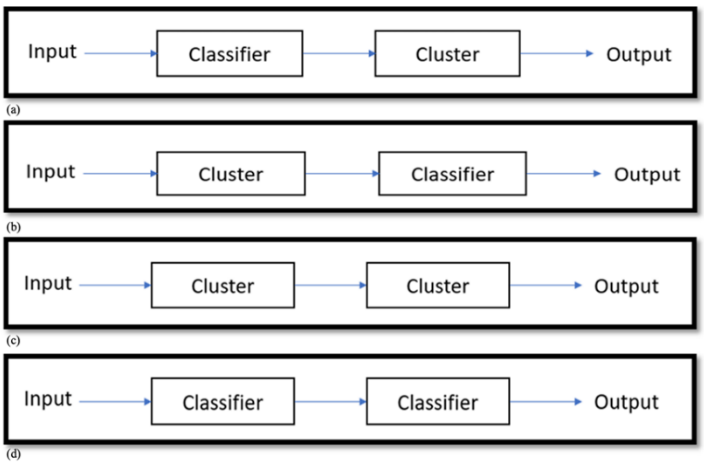

Fig. 4. Hybrid model mechanism (a) classifier with cluster, (b) cluster with classifier, (c) cluster with cluster, and (d) classifier with classifier.

3.5. ML Classification Approaches

The following ML models considered in this paper include SVM, DT, KNN, MLP, LR, Birch, Multinomial NB, Agglomerative clustering, and Kmeans.

3.5.1. SVM

SVM is a versatile classification technique that can be used to address both classification and regression problems. It is renowned for its strong theoretical foundations, excellent generalization performance, and flexibility in handling high-dimensional data. A critical component of SVM is the hyperplane, which serves as the decision boundary that distinguishes between different classes. The dimensionality of the hyperplane is determined by the input features of the data. Formally, the hyperplane is defined by (3):

$$(w.x) + b = 0 \quad (3)$$

where $w \in \mathbb{R}^N$ and $b \in \mathbb{R}$

The decision function for classification in SVM is then expressed in (4) as:

$$f(x) = sign((w.x) + b) \quad (4)$$

To determine the optimal hyperplane, SVM seeks the hyperplane that maximizes the margin between the two classes. This is achieved by solving a constrained quadratic optimization problem. The solution, w, can be expanded in (5) as:

$$w = \sum_i \alpha_i x_i \quad (5)$$

where α_i are the Lagrange multipliers corresponding to the training patterns that x_i lie on the margin. The C and the γ are key parameters for SVM optimization, aside from the kernel functions (rbf, poly, linear, sigmoid) [35].

3.5.2. DT

DT is a type of ML algorithm that iteratively splits the data based on specific parameters to create a model that predicts the value of a target variable. The structure of a DT consists of two main entities: decision nodes and leaves. Decision nodes represent the points where the data is split, while leaves represent the final decision or outcome of the tree. DT operates by selecting the best attribute to split the data at each node, based on a criterion that maximizes the separation of the classes [36]. This pro- cess continues recursively until a stopping criterion is met, such as a maximum depth or a minimum number of samples per leaf. According to Mitrofanov and Semenkin [37], five common parameters define the DT, including i.) types of predicates at the vertices ii.) quality functional $\mathrm{Q(X, j, s)}$ of the split, iii.) stopping criterion, iv.) missing values processing method, and iv.) the pruning method to prevent overfitting. Methods include cost complexity pruning and reduced error pruning.

3.5.3. KNN

KNN is a supervised learning method used for both regression and classification problems. In this paper, KNN is employed to address classification problems. The algorithm operates on the principle of proximity, making classifications by grouping individual data points based on their similarities. For example, as depicted in Fig. 5, if there are four squares and two triangles near a new data point, the algorithm classifies the new point as a square due to its proximity to the majority category [38].

The symbol "K" in KNN represents the number of nearest neighbors considered when making a classification decision. The choice of K is crucial as it determines the number of neighbors that influence the classification of a new data point. As a lazy algorithm, it stores all the training data and makes decisions only at the time of classification. Using a distance metric such as the Euclidean distance [3, 39], KNN is expressed in (6) as:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (6)$$

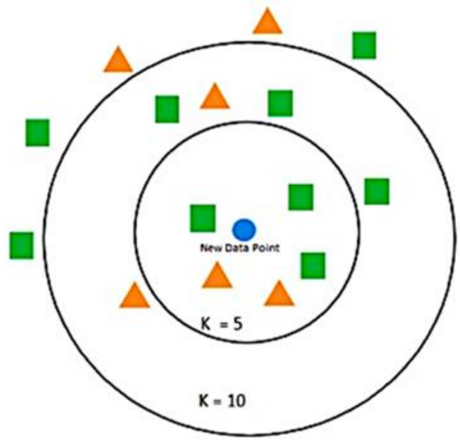

Fig. 5. KNN model using proximity method [38]

3.5.4. MLP

MLP is a widely known network structure utilized for solving classification and regression tasks. In a feed-forward network of MLP, data moves in a forward direction from the input to the output layer, trained by backpropagation utilizing multiple layers of nodes interconnected with unidirectional connections [40]. Each node's outputs consist of weighted units along with a nonlinear activation function to differentiate non-linearly separable data expressed in (7) as:

$$y_j^k = \sigma\left(\sum_i w_i^{kj}\, x_i^{k-1} + b_j^k\right) \quad (7)$$

The $\sigma(x)$ is defined as the activation function, while y_j^k represents output, w_i^{kj} is the weight, and b_j^k is the "bias."

3.5.5. LR

LR is a supervised learning algorithm useful to conduct classification tasks when the dependent variable is binary.

LR bounden data coded with only 2 possible outcomes between 0 and 1 or yes and no. This algorithm estimates the probability of an event occurring based on the given dataset of independent variables. LR predicts $P(Y = 1)$ as a function of X given the set of inputs and the output is obtained by a function [41]. In LR, the output can be obtained through the equation [42]. It is defined as the probability estimation of an event occurring or not as follows in (8):

$$prob(y) = \begin{cases} 1 \\ 0 \end{cases} = \frac{e^{\alpha_0 + \sum_{k=1}^{K} \alpha_k f_k}}{1 - e^{\alpha_0 + \sum_{k=1}^{K} \alpha_k f_k}} \tag{8}$$

where y is a binary target variable depicting the occurrence of the event (e.g., $y = 1$ if an event occurs and 0 otherwise), $f_1, f_2, \ldots, f_k$ are the explanatory features as inputs into the model while the regression coefficients estimated by the maximum likelihood based on the training data available are$\alpha_1, \alpha_2, \ldots, \alpha_k$.

3.5.6. NB

NB-algorithm utilizes the Bayes theorem for classification tasks with a strong assumption that the attributes are conditionally independent, given the class [38]. As a result, the likelihood or probability density of characteristics X (feature matrix) given class Y (response vector) forms the foundation of NB classification. The posterior probability, P(c|x), can be calculated using the Bayes theorem method expressed in (9) as:

$$P(c|x) = \frac{P(x|C)P(C)}{P(x)} \tag{9}$$

where P(c|x) is the posterior probability of the class (target) given predictor (attribute), P(c) is the class prior probability, P(x|c) is the likelihood which is the probability of the predictor given class, and P(x) is the predictor prior probability. The posterior probability is further defined in (10) as:

$$p(Y = y_k | X_1 \ldots Xn) = \frac{P(Y = y_k)P(X_1 \ldots X_n | Y = y_k)}{\Sigma_j P\big(Y = y_j\big)P(X_1, \ldots Xn | Y = y_k)}$$

$$= \frac{P(Y=y_k)\Pi_i P(X_i|Y=y_k)}{\Sigma_j P(Y=y_j)\Pi_j P(X_i|Y=y_j)} \tag{10}$$

where X is the input vector, $(X_1, X_2, \cdots, X_n)$ and Y is the output category. NB uses Bernoulli, Gaussian, or Multinomial parameters depending on the task or data type [42, 43].

3.6. Clustering Approaches

3.6.1. KMeans

KMeans is an unsupervised ML technique used to identify clusters within a dataset. This algorithm operates through two main processes: expectation and maximization. During the expectation step, each data point is assigned to the nearest centroid, whereas, in the maximization step, the mean of all points in each cluster is computed and set as the new centroid. These steps are repeated iteratively until the centroids no longer change significantly, indicating that the algorithm has converged.

3.6.2. Birch

Birch (Balanced Iterative Reducing and Clustering using Hierarchies) is a clustering algorithm designed to efficiently handle large datasets by first summarizing the data into smaller clusters, which can then be further clustered. Additionally, these small summaries can be clustered using other algorithms after the initial summarization. Birch is exceptionally fast because it requires only a single scan of the dataset, making it ideal for large datasets. The algorithm con- structs a Clustering Features (CF) tree, which compresses and summarizes the data, using less memory and improving performance [44].

3.6.3. AC

Agglomerative Clustering (AC) as an unsupervised learner organizes data into several clusters, ensuring that data points within the same cluster are similar and close to each other, while data points in different clusters are distinct and far apart. This algorithm is a type of hierarchical clustering, employing a bottom-up approach. It begins by partitioning the dataset into individual singleton nodes and progressively combines them with the nearest data points to form new nodes. This process continues until no final node remains or the specified stopping conditions are met [45]. The distance between two clusters is measured using Euclidean Distance, which is calculated using (11) [46].

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots \big(x_p - y_P\big)^2} \tag{11}$$

3.7. Dataset

The original dataset consists of 25977 survey entries from a US airline [47]. There are 24 feature columns with 1 binary label column. The description of each feature is presented in Table 1. The description of each feature is presented in Table 1.

Table 1. The description of the airline dataset

Feature	Description
Counter	The counter of each row
ID	The ID of the passenger
Gender	Gender of the passengers (Female, Male)
Customer Type	The customer type (Loyal customer, disloyal customer)
Age	The actual age of the passengers
Type of Travel	Purpose of the flight of the

	passengers (Personal Travel, Business Travel)
Class	Travel class in the plane of the passengers (Business, Eco, Eco Plus)
Flight Distance	The flight distance of this journey
Inflight wifi service	Satisfaction level of the inflight wifi service (0: Not Applicable;1-5)
Departure/Arrival time convenient	Satisfaction level of Departure/Arrival time convenient
Ease of Online booking	Satisfaction level of online booking
Gate location	Satisfaction level of Gate location
Food and drink	Satisfaction level of Food and drink
Online boarding	Satisfaction level of online boarding
Seat comfort	Satisfaction level of Seat comfort
Inflight entertainment	Satisfaction level of inflight entertainment
On-board service	Satisfaction level of On-board service
Legroom service	Satisfaction level of Legroom service
Baggage handling	Satisfaction level of baggage handling
Check-in service	Satisfaction level of Check-in service
Inflight service	Satisfaction level of inflight service
Cleanliness	Satisfaction level of Cleanliness
Departure Delay in Minutes	Minutes delayed when departure
Arrival Delay in Minutes	Minutes delayed when Arrival
Satisfaction	Airline satisfaction level (Satisfaction, neutral or dissatisfaction)

3.8. Data Preprocessing

In the pre-processing stage, features, including ID and counter were removed since they do not contribute to the model training. A missing value process was also performed ensuring no missing instances in the dataset

3.8.1. One Hot encoding

One hot encoding [48] was used to transform features, including "Gender", "Customer Type", "Type of Travel", "Class" and "Satisfaction."

3.8.2. Normalization

Following [49], data normalization was carried out on the numeric feature in the dataset to ensure quality data is fed into the model. The Z-score normalization in (12) was employed in this instance, which is adaptive to outlier resolution.

$$Z - score = \frac{x-\mu}{\sigma} \quad (12)$$

3.9. Model Development

Three model development approaches were adopted in this paper. They include models through ML classifiers, models via clustering, and models with the highest accuracy hybridization. The first two approaches were set up as a baseline, while the third approach uses the proposed model in the approach to form the hybrid model. For the baseline model, the following classifiers were used, including SVM, LR, DT, KNN, MLP, Multinomial-(MNB), and Gaussian-(GNB), while for the clustering approach, Kmeans, Birch, Meanshift, and AC were used. A 5-fold cross-validation (CV) was used for the classifiers to ensure training is pre-vented from overfitting. The result obtained with the highest accuracy together with the result from the clustering technique was used as a baseline model to hybridize with the proposed DAE+GA.

4. Results

4.1. Baseline Models

The baseline models were trained on the airline dataset using a 5-fold CV to split the data, effectively preventing overfitting. Default parameters were used for each baseline ML model. Accuracy was used as the basis for evaluation, and the performance of the baseline models is depicted in Fig. 6. The results show that the DT achieved the highest accuracy score of 85.54%, followed by GNB. In contrast, K-means exhibited the lowest accuracy score of 16.18%, followed closely by Birch.

It was generally observed that the clustering methods performed less than the supervised ones. The poor performance observed might result from clustering algorithms struggling in high-dimensional spaces, where distances lose meaning. Moreover, both K-means and Birch assume data is linearly separable, a condition that may not hold for the airline dataset.

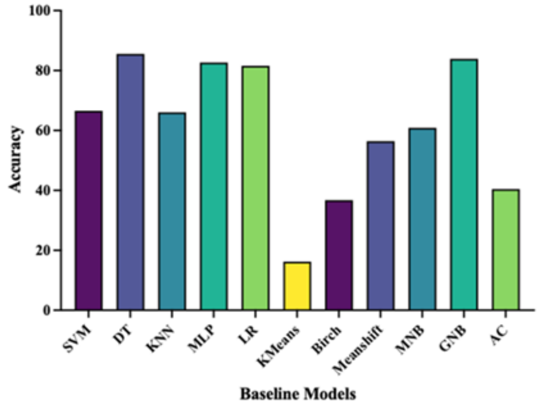

Fig. 6. Performance comparison of baseline models

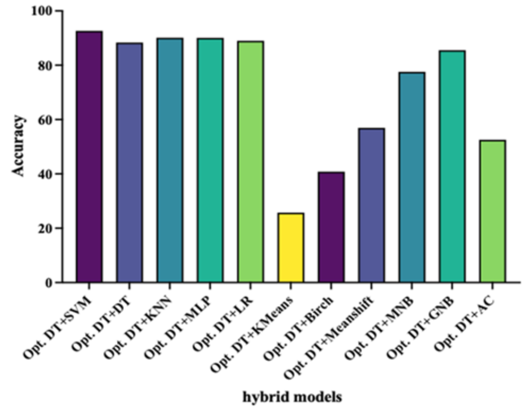

Fig. 7. Performance comparison of hybrid models

4.2. Hybrid model

In the hybrid model experimentation, the DT model with the highest accuracy serves as the primary algorithm in the optimization process of DAE+GA. The resulting output is then utilized in the secondary algorithm using each of the baseline ML models again. The hybridization concept is illustrated in Table 2, while the performance of these hybrid models is depicted in Fig. 7. Employing the same 5-fold cross-validation to split the data, it was noted that the hybrid SVM achieved the highest accuracy score at 92.68%, closely trailed by KNN. Conversely, K-means and Birch yielded notably lower scores of 25.73% and 40.83%, respectively. In general, the clustering methods exhibited inferior performance compared to the supervised approach.

Table 2. Proposed hybrid model

Best Method in Base + Optimization Technique	Algorithm 2: Classification Task
DT+DAE-GA optimization	SVM DT KNN MLP LR K-means Birch Meanshift MNB GNB AC

Additionally, the findings from the analysis underscore the notable improvements achieved by employing the DAE+GA optimization process, particularly in enhancing the accuracy scores of the baseline DT model as the first algorithm. As illustrated in Fig. 8, the observed enhancements signify the efficacy of this approach, with significant implications for the training process, especially concerning the second algorithm.

Furthermore, the percentage (%) performance change analysis, detailed in Table 3, provides further insights into the magnitude of improvement across various ML algorithms when comparing the baseline models to the proposed approach. Notably, the Kmeans algorithm stands out with the highest observed improvement of 59.02%, showcasing the substantial impact of the DAE+GA optimization. This significant enhancement underscores the effectiveness of incorporating the proposed approach, particularly for algorithms like Kmeans that traditionally face challenges in certain datasets.

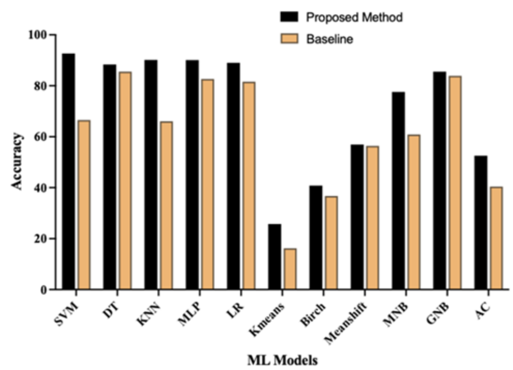

Fig. 8. Performance comparison across all models

Conversely, while the Meanshift algorithm exhibited the lowest improvement at only 0.94%, its inclusion in the analysis underscores the comprehensive evaluation of various clustering techniques. Even though clustering methods generally showed lower performance compared to supervised approaches, this detailed examination provides valuable in- sights into their potential for improvement through optimization processes like DAE+GA. The variation in models' performance is presented in Table 3.

Table 3. Percentage Improvement

Models	Percentage (%)
SVM	39.30%
DT	3.31%
KNN	36.51%
MLP	8.94%
LR	9.07%
K-means	59.02%
Birch	11.07%
Meanshift	0.94%
MNB	27.44%
GNB	1.93%
AC	29.92%

5. Discussion

This study investigated a hybrid model combining DAE and GA with selected ML models, including both classifiers and clustering techniques. The ML models were initially used as baseline learners to establish a performance benchmark. The first round of experiments identified DT as the top-performing baseline model, achieving an accuracy of 85.54%. This established DT as a robust foundation for further optimization. In the hybrid approach, DT was integrated with DAE-GA to create an optimized model, referred to as the first algorithm. This optimized model was then combined with various second algorithm ML models to generate final predictions. Notably, the hybrid model DAE-GA-DT+SVM achieved the highest accuracy of 92.68%, representing a significant improvement over the baseline DT model.

The integration of the DAE-GA-DT optimized model with various ML models resulted in notable improvements in accuracy across all second algorithm models. The percentage changes in accuracy for each model are summarized in Table 3. Of particular interest is the significant improvement observed in the SVM algorithm, which recorded a 39.30% increase in accuracy. This substantial enhancement underscores the effectiveness of the DAE+GA optimization approach, positioning SVM as a standout performer among the evaluated models. This finding aligns with [35], who also reported significant performance gains through optimization techniques. The K-Nearest Neighbors (KNN) algorithm also showed remarkable improvement with a 36.51% increase in accuracy, demonstrating the broad applicability and effectiveness of the hybrid model. Similarly, the Adaptive Clustering (AC) algorithm exhibited a notable enhancement of 29.92%.

Furthermore, the clustering techniques, while typically lagging behind supervised approaches in standalone performance, benefited significantly from the hybrid optimization. For instance, K-means showed a substantial 59.02% improvement, highlighting the potential of DAE+GA in enhancing clustering algorithms' performance. Birch and Mean Shift algorithms also demonstrated improvements of 11.07% and 0.94%, respectively, indicating varied yet positive impacts of the optimization process.

The application of heuristic algorithms for optimizing ML models in this study is consistent with the methodologies used in existing research [6, 25, 41]. The proposed approach shows competitive performance when compared with [3], who achieved 89.20% accuracy with Random Forest (RF) on the same airline dataset. This comparison underscores the competitive edge and robustness of the hybrid DAE+GA optimization approach.

The findings from Noviantoro and Huang [38] emphasize the importance of prioritizing specific service types that airlines need to focus on, based on big data analysis of passengers on full-service airlines. Their study aims to enhance passenger satisfaction and loyalty by identifying and improving key service areas. In contrast, our study was oriented towards providing an optimal prediction of customer satisfaction through the application of advanced ML models and hybrid optimization techniques. By leveraging the DAE-GA optimization approach, our study not only identifies key predictors of satisfaction but also significantly improves the predictive accuracy of these models.

Overall, the results suggest that while clustering techniques may still trail behind supervised methods in raw performance, the application of DAE+GA optimization demonstrates considerable promise in bridging this gap. The observed improvements across various algorithms, particularly in SVM, KNN, and AC, underscore the potential of heuristic optimization in enhancing ML model efficacy. The hybrid model approach not only enhances performance metrics but also provides a robust framework for integrating different ML techniques for superior predictive accuracy [15, 20, 26, 30].

6. Conclusion

This paper has presented an experiment aimed at assessing the predictive capabilities of a hybrid model combining DAE, GA, and ML algorithms. The objective was to determine the effectiveness of the hybrid approach compared to traditional algorithms. The results obtained from the experiment demonstrate that indeed, the hybrid model surpassed the performance of conventional methods.

The implications of the findings in this paper are twofold. Firstly, it underscores the effectiveness of the hybrid model approach in leveraging diverse data sources and advanced optimization techniques to enhance predictive accuracy. Secondly, it highlights the evolving landscape of feature importance within the airline industry, emphasizing the need for adaptive modeling approaches to capture changing passenger preferences and industry dynamics. Moving forward, the integration of hybrid modeling methodologies and a dynamic feature selection process will be instrumental in developing robust predictive models capable of addressing the evolving needs of the airline sector in a post-pandemic world.

The emergence of the COVID-19 pandemic in 2019 has significantly altered the landscape of airline operations, introducing a heightened emphasis on adhering to Standard Operation Procedures (SOPs) to ensure passenger safety. Consequently, factors related to SOPs have become pivotal in assessing passenger satisfaction. As passenger priorities continue to evolve, there is a noticeable shift towards placing greater importance on in-flight hygiene measures over traditional amenities such as in-flight entertainment.

In future research work, it is imperative to incorporate SOPs-related factors into the framework for evaluating airline customer satisfaction. Furthermore, future work should explore the effectiveness of alternative optimization algorithms in comparison to GA. Experimenting with different optimization techniques can provide valuable insights into their respective strengths and weaknesses in enhancing the performance of hybrid models.

Conflicts of interest

The authors declare no competing interests

Acknowledgements

This work is funded by Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme [Grant Number: FRGS/1/2022/STG06/USM/02/4], for the Project entitled “Efficient Joint Process Monitoring using a New Robust Variable Sample Size and Sampling Interval Run Sum Scheme”.

References

[1] S. Leon and J. C. Martín, “A fuzzy segmentation analysis of airline passengers in the U.S. based on service satisfaction,” *Res. Transp. Bus. Manag.*, 2020.

[2] V. Leninkumar, “The relationship between customer satisfaction and customer trust on customer loyalty,” *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 7, no. 4, 2017.

[3] A. C. Y. Hong, K. W. Khaw, C. XinYing, and W. C. Yeong, “Prediction of US airline passenger satisfaction using machine learning algorithms,” *Data Anal. Manag.*, vol. 4, no. 1, 2023.

[4] M. Al-Mashraie, S. H. Chung, and H. W. Jeon, “Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach,” *Comput. Ind. Eng.*, vol. 144, Apr. 2020, 10.1016/j.cie.2020.106476.

[5] L. Hibović, S. Smajić, and E. Yaman, “Predicting satisfaction of airline passengers using classification,” *International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2022, 10.1109/ISMSIT56059.2022.9932850.

[6] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, M. Elhoseny, and K. Shankar, “Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector,” *Complex Intell. Syst.*, 2021, 10.1007/s40747-021-00353-6.

[7] A. C. Kae, C. XinYing, O. V. Johnson, and K. W. Khaw, “Employee Turnover Prediction by Machine Learning Techniques,” *J. Telecommun. Electron. Comput. Eng.*, vol. 13, no. 4, pp. 49-56, 2021. [Online]. Available: https://jtec.utem.edu.my/jtec/article/view/6148/4063.

[8] B. Bhargav and R. T. Prabu, “Airline passenger satisfaction prediction using novel hybrid random forest model comparison with k-nearest neighbour model,” *Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 2023.

[9] S. Li, G. Xia, and X. Zhang, “Customer Churn Combination Prediction Model Based on Convolutional Neural Network and Gradient Boosting Decision Tree,” *ACM Int. Conf. Proc. Ser.*, 2022, 10.1145/3579654.3579666.

[10] J. A. Nichols, H. W. Chan, and M. A. B. Baker, “Machine learning: applications of artificial intelligence to imaging and diagnosis,” *Biophys. Rev.*, vol. 11, no. 1, pp. 111-118, 2019, 10.1007/s12551-018-0449-9.

[11] O. V. Johnson, C. XinYing, K. W. Khaw, and L. H. Ming, “A Cost-Based Dual ConvNet-Attention Transfer Learning Model for ECG Heartbeat Classification,” *J. Inform. Web Eng.*, vol. 2, no. 2, pp. 90-110, 2023, 10.33093/jiwe.2023.2.2.7.

[12] F. Balducci, D. Impedovo, and G. Pirlo, “Machine learning applications on agricultural datasets for smart farm enhancement,” *Machines*, vol. 6, no. 3, 2018, 10.3390/machines6030038.

[13] B. K. Chan, O. V. Johnson, X. Chew, K. W. Khaw, M. H. Lee, and A. Alnoor, “Proposed Bayesian optimization based LSTM-CNN model for stock trend prediction,” *Comput. Inform.*, vol. 43, no. 1, pp. 38–63, 2024, 10.31577/cai_2024_1_38.

[14] M. Pesaresi, V. Syrris, and A. Julea, “A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning,” *Remote Sens.*, vol. 8, no. 5, p. 399, 2016, 10.3390/rs8050399.

[15] E. Domingos, B. Ojeme, and O. Daramola, “Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector,” *Computation*, vol. 9, no. 3, 2021,

10.3390/computation9030034.
[16] O. M. Ajinaja, O. S. Egwuche, and O. V. Johnson, “Prediction of Graduating Students for Tertiary Institutions Using Data Mining Technique,” *2nd International Conference on Education and Development*, 2019, pp. 58-62.
[17] H. Ramchoun, M. Amine, J. Idrissi, Y. Ghanou, and M. Ettaouil, “Multilayer Perceptron: Architecture Optimization and Training,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 1, p. 26, 2016, 10.9781/ijimai.2016.415.
[18] R. Sudharsan and E. N. Ganesh, “A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy,” *Connect. Sci.*, vol. 34, no. 1, pp. 1855-1876, 2022, 10.1080/09540091.2022.2083584.
[19] S. J. Haddadi, O. M. Mohammadi, M. Bahrami, E. Khoeini, M. Beygi, and M. H. Khoshkar, “Customer churn prediction in the Iranian banking sector,” *International Conference on Applied Artificial Intelligence*, 2022, pp. 1-6, 10.1109/ICAPAI55158.2022.9801574.
[20] O. M. Mirza, G. J. Moses, R. Rajender, E. L. Lydia, S. Kadry, C. Me-Ead, and O. Thinnukool, “Optimal Deep Canonically Correlated Autoencoder-Enabled Prediction Model for Customer Churn Prediction,” *Comput. Mater. Continua*, vol. 73, no. 2, pp. 3757-3769, 2022, 10.32604/cmc.2022.030428.
[21] H. Gao, B. Qiu, R. J. D. Barroso, W. Hussain, Y. Xu, and X. Wang, “TSMAE: A novel anomaly detection approach for Internet of Things time series data using memory-augmented autoencoder,” *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 2978-2990, 2023, 10.1109/TNSE.2022.3163144.
[22] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” *Pattern Recognit.*, vol. 61, pp. 650-662, 2017, 10.1016/j.patcog.2016.06.008.
[23] B. Garimella, G. V. S. N. R. V. Prasad, and M. H. M. K. Prasad, “Churn prediction using optimized deep learning classifier on huge telecom data,” *J. Ambient Intell. Human Comput.*, vol. 14, no. 3, pp. 2007-2028, 2023, 10.1007/s12652-021-03413-4.
[24] V. Haridasan, K. Muthukumaran, and K. Hariharanath, “Arithmetic optimization with deep learning enabled churn prediction model for telecommunication industries,” *Intell. Autom. Soft Comput.*, vol. 35, no. 3, pp. 3531-3544, 2023, 10.32604/iasc.2023.030628.
[25] C. K. Praseeda and B. L. Shivakumar, “Fuzzy particle swarm optimization (FPSO) based feature selection and hybrid kernel distance-based possibilistic fuzzy local information C-means (HKD-PFLICM) clustering for churn prediction in telecom industry,” *SN Appl. Sci.*, vol. 3, no. 6, 2021, 10.1007/s42452-021-04576-7.
[26] S. Ouf, “An optimized deep learning approach for improving airline services,” *Comput. Mater. Continua*, vol. 75, no. 1, pp. 1213-1233, 2023, 10.32604/cmc.2023.037276.
[27] M. Guimarães, C. Soares, and R. Ventura, “Decision support models for predicting and explaining airport passenger connectivity from data,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16005-16015, 2022, 10.1109/TITS.2022.3148830.
[28] S. Chen, D. L. Xu, and W. Jiang, “High value passenger identification research based on federated learning,” *12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2020, vol. 1, pp. 107-110.
[29] Z. Wang, X. Han, Y. Chen, X. Ye, K. Hu, and D. Yu, “Prediction of willingness to pay for airline seat selection based on improved ensemble learning,” *Aerospace*, vol. 9, no. 2, p. 47, 2022, 10.3390/aerospace9020047.
[30] R. Pranav and H. S. Gururaja, “Explainable stacking machine learning ensemble for predicting airline customer satisfaction,” *Congress on Intelligent Systems*, 2022, pp. 41-56, Springer Nature Singapore, 10.1007/978-981-19-3812-4_5.
[31] A. Mottini and R. Acuna-Agost, “Deep choice model using pointer networks for airline itinerary prediction,” *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 1575-1583, https://doi.org/10.1145/3097983.3098094.
[32] I. Rahmany, N. Dhahri, T. Moulahi, and A. Alabdulatif, “Optimized stacked auto-encoder for unnecessary data reduction in cloud of things,” *International Wireless Communications and Mobile Computing (IWCMC)*, 2022, pp. 110-115, https://doi.org/10.1109/IWCMC55113.2022.9824846 .
[33] A. Lambora, K. Gupta, and K. Chopra, “Genetic algorithm - a literature review,” *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 380-384.
[34] S. Ardabili, A. Mosavi, and A. R. Várkonyi-Kóczy, “Advances in machine learning modeling reviewing hybrid and ensemble methods,” *International Conference on Global Research and Education*, 2019, pp. 215-227, Springer International Publishing, https://doi.org/10.1007/978-3-030-20393-1_20.
[35] H. H. Huang, Z. Wang, and W. Chung, “Efficient parameter selection for SVM: The case of business intelligence categorization,” *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2017, pp. 158-160, https://doi.org/10.1109/ISI.2017.8004895.
[36] W. Baswardono, D. Kurniadi, A. Mulyani, and D. M.

Arifin, "Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification," *J. Phys. Conf. Ser.*, vol. 1402, p. 066055, 2019, https://doi.org/10.1088/1742-6596/1402/6/066055.

[37] S. Mitrofanov and E. Semenkin, "An approach to training decision trees with the relearning of nodes," *2021 International Conference on Information Technologies (InfoTech)*, 2021, pp. 1-5, https://doi.org/10.1109/InfoTech52442.2021.9586108

[38] T. Noviantoro and J. P. Huang, "Investigating airline passenger satisfaction: Data mining method," *Res. Transp. Bus. Manag.*, vol. 43, p. 100726, 2022, https://doi.org/10.1016/j.rtbm.2022.100726.

[39] S. Khan, Z. A. Khan, Z. Noshad, S. Javaid, and N. Javaid, "Short term load and price forecasting using tuned parameters for k-nearest neighbors," *Sixth HCT Information Technology Trends (ITT)*, 2019, pp. 89-93.

[40] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 133-144, 2018, https://doi.org/10.1016/j.isprsjprs.2017.07.014.

[41] X. Zou, Y. Hu, Z. Tian, and K. Shen, "Logistic regression model optimization and case analysis," *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019, pp. 135-139, https://doi.org/10.1109/ICCSNT47585.2019.8962464.

[42] X. Jiang, Y. Zhang, and Y. Li, "Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model," *Sci. Rep.*, vol. 12, p. 11174, 2022, https://doi.org/10.1038/s41598-022-15341-6.

[43] B. Spoorthi, S. S. Kumar, A. P. Rodrigues, R. Fernandes, and N. Balaji, "Comparative analysis of bank loan defaulter prediction using machine learning techniques," *IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 2021, pp. 24-29, https://doi.org/10.1109/DISCOVER52568.2021.9624056.

[44] T. Wahyuningrum, S. Khomsah, S. Suyanto, S. Meliana, P. E. Yunanto, and W. F. Al Maki, "Improving clustering method performance using k-means, mini batch k-means, birch and spectral," *4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2021, pp. 206-210, https://doi.org/10.1109/ISRITI54043.2021.9702677.

[45] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011, https://doi.org/10.48550/arXiv.1109.2378.

[46] A. Nugraha, M. A. H. Perdana, H. A. Santoso, J. Zeniarja, A. Luthfiarta, and A. Pertiwi, "Determining the senior high school major using agglomerative hierarchical clustering algorithm," *2018 International Seminar on Application for Technology of Information and Communication*, 2018, pp. 225-228, https://doi.org/10.1109/iSemantic.2018.8579985.

[47] D. John, "Passenger satisfaction," *US airline passenger satisfaction*. Available at: https://www.kaggle.com/datasets/johnddddd/customer-satisfaction. Accessed 21 Nov 2023.

[48] K. Pordar, T. S. Pardawala, and C. D. Pai, "A comparative study of categorical variable encoding techniques for neural network classifiers," *Int. J. Comput. Appl.*, vol. 175, no. 4, pp. 7-9, 2017, https://doi.org/10.5120/ijca2017914706.

[49] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, p. 105524, 2020, https://doi.org/10.1016/j.asoc.2020.105524.