Leveraging Artificial Intelligence and Machine Learning for Data-Driven Marketing Strategy: A Framework for Marketing Managers

VISHVESH SONI
Ecommerce & Digital Marketing Manager, Alliage Inc Los Angeles, California, United States

Abstract- In the ever-evolving landscape of modern business, harnessing the power of Artificial Intelligence and Machine Learning has become imperative for crafting effective data-driven marketing strategies. In this study marketing strategy with a thorough framework to help them maximize the potential of AI and ML tools to strengthen their data-driven marketing strategies. The framework includes important steps like skillful Data Collection, Data Preprocessing, which handles data cleansing, handling missing values, and normalization, then Feature Extraction, and unique Feature Selection using the Particle Enriched Pelican Optimizer. The integration of hybrid Deep Learning models such as Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) follows the use of Dimensionality Reduction as well. The combination of these factors yields much improved prediction accuracy, enabling marketing professionals to make more smart judgments in the fluid environment of modern marketing. The proposed model demonstrates exceptional performance across various metrics, outshining existing methods. With high acc (0.9763), pre (0.9766), and spec (0.9816), it excels in correctly identifying positive cases.

Indexed Terms- Artificial Intelligence, Particle Enriched Pelican Optimizer, Convolutional Neural Networks, Bidirectional Long Short-Term Memory.

I. INTRODUCTION

Artificial intelligence (AI) is a powerful technological force that is transforming and advancing civilizations by lowering costs and risks, boosting consistency and dependability, and offering fresh approaches to challenging issues [1]. All businesses and sectors now use AI systems and apps, which also provide a variety of potential for marketing strategy and activities as well as for improving consumer connections, engagement, and experience [2]. The capacity to develop customized and personalized offers as well as create and sustain responsive consumer engagements and relationships with experience value is made possible by the growing computing power, data availability and intensity, context awareness, and emotional-sensing capabilities of AI [3]. The enormous and expanding breadth of consumer data feeding AI systems, the level of emotional intelligence of AI, and the rise of AI-driven sales and consumption create ethical questions, issues, and concerns about the possible targeting or alienation of vulnerable consumer groups [4]. On a corporate level, market share concentration via AI-enabled e-commerce platforms and unequal representation on them might hurt certain enterprises while favoring others [5].

Despite having significant drawbacks, adopting machine learning algorithms for data-driven marketing is preferable to developing new AI [6]. The AI might maintain and even increase biases in decision-making processes if the training data used to create the AI model is not representative or has inherent biases [7]. The intended audiences might be impacted and societal disparities could be reinforced as a result, which would be unjust and discriminatory. A lack of transparency is also caused by the complexity of machine learning models, which makes them difficult to understand and analyze. Because of this opacity, stakeholders and marketers are unable to comprehend why the AI makes particular suggestions or forecasts, which undermines confidence and makes adoption difficult [8][9][10]. A sudden obsolescence of AI or incompatibility with developing marketing platforms and tactics might result from the rapid speed of technological innovation. Because of this, companies could spend a lot of money developing an AI system that quickly becomes obsolete. When decision-making becomes increasingly automated, an overreliance on AI may result in marketing techniques that are less imaginative and critical-thinking. In light of these disadvantages, it is crucial for organizations to approach the development and deployment of AI for data-driven marketing with a balanced perspective, incorporating robust bias mitigation techniques, ensuring transparency, and maintaining a harmonious human-AI collaboration [11].

Organizations should establish a comprehensive strategy to solve the problems associated with implementing machine learning in data-driven marketing while maximizing the benefits of AI [12]. This calls for the implementation of strong bias mitigation strategies to combat ingrained biases in training data, the maintenance of interpretable machine learning models or techniques, and the promotion of a symbiotic human-AI partnership to preserve critical thinking and creativity in marketing strategies. In order to increase the efficiency of AI-driven marketing, hybrid optimization algorithms (Pelican Optimization Algorithm and Particle Swarm Optimization) [13][14] like Particle Enriched Pelican Optimizer can help in choosing essential characteristics for the marketing challenge. A state-of-the-art approach incorporates a hybrid deep learning model that blends Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) [15] to create the answer. A better educated decision-making process and flexible tactics are eventually promoted by this unique technique, which improves feature extraction and makes it easier to identify sequential patterns in time-series marketing data.

The primary contribution in this study is:

- To recognize the essential ML and AI methods that may be used for data-driven marketing initiatives.
- To provide a workable framework that marketing managers can use to include AI and ML in their decision-making.
- To confirm the framework's efficacy via empirical research and real-world case studies.

The following is how the paper is set up: Using recent studies, Section 2 outlines the literature review. A comprehensive discussion of the proposed approach is given in Section 3. Section 4 discusses the results and Section 5 concludes the paper.

II. LITERATURE REVIEW

In 2020 Kharfan et al. [16] proposed a machine learning approach and identifying important predictor factors, fashion retailers may improve forecast accuracy by focusing on demand prediction from a data-driven viewpoint. The effectiveness of machine learning methods was demonstrated by comparing the results of the predictions. In order to predict the demand for just-introduced seasonal items without previous data, the suggested technique was used by a top fashion retail corporation.

In 2021 Wu et al. [17] proposed a unique bibliometric framework that mines scientific publications and patents for information. In order to statistically detect the evolutionary patterns and hierarchies at work in DT research, the framework integrates the hierarchical subject tree and scientific evolutionary pathways. Our findings, drawn from more than 10,179 scholarly articles on DT, include an extensive description of DT from a bibliometrics viewpoint and a methodical classification of the competencies needed to allow DT. The article also provides a case study of 9,454 patents that focus on artificial intelligence (AI), one of the rising technologies, in order to provide more practical insights on technical capabilities.

In 2021 Akter et al. [18] focused on digital behemoths like Amazon, Alibaba, Google, Apple, and Facebook benefit from long-term competitive advantages because to DDI. The existence of algorithmic biases that might affect the DDI process and lead to the generation of unfair, biased, or harmful data products is, however, still little understood. With the use of a thorough thematic analysis, and a case study on the Australian Robo-Debt programmed, this guest editorial tries to investigate the causes of algorithmic biases throughout the DDI process. According to the research, data bias, technique bias, and social prejudice are the three main causes of algorithmic bias.
In 2021 Linaza et al. [19] researched projects, which they executed and evaluated across numerous European nations, with the aim of showcasing the technical problems still at hand as well as the accomplishments previously attained. As a general conclusion, it should be noted that, even though they are still in the early stages in some cases, AI technologies help with decision-making at the farm

level by monitoring conditions and maximizing production to enable farmers to use the right number of inputs for each crop, increasing yields and lowering water use and greenhouse gas emissions.

In 2020 Zhou et al. [20] experimented to make semiconductor production smarter, they undertook an extensive study based on Evolutionary Computing and Deep Learning Techniques. We provide a dynamic method to handle a variety of issues and gather insightful information about semiconductor production processes. They describe how a Genetic technique and Neural Network are used to develop an intelligent feature selection technique.

In 2019 Tong et al. [21] combined the classic marketing mix model to create a foundation for customized mobile marketing initiatives. The architecture places customization at the core of mobile products, locations, prices, promotions, and predictions. In accordance with the suggested structure, recent studies on mobile marketing are evaluated, and potential areas of study for personalized mobile marketing are covered.

In 2022 Sabharwal et al. [22] provided a few recommendations on ways that businesses might enhance their marketing strategy. Digital marketing, a legitimate subject of marketing science, has been able to benefit organizations by enhancing value and boosting consumer engagement with online offerings. Industry monitoring processes, such as branding, marketing, advertising, production, channel distribution, etc., have benefited from the digital age. Business managers might make more precise and data-driven choices based on acquired data, interactive customer experience, and a digital perspective of processes and sales.

In 2020 Huang et al. [23] proposed the applications of AI in marketing research, strategy (including segmentation, targeting, and positioning, or STP), and actions. Mechanical AI may be utilized to gather data, thinking AI for market analysis, and feeling AI for consumer comprehension throughout the marketing research stage. Mechanical AI, thinking AI, and feeling AI may all be used to segmentation (segment recognition), targeting (segment suggestion), and positioning (segment resonance) at the stage of marketing strategy (STP). Mechanical AI, thinking AI, and feeling AI may all be employed at the marketing action stage to standardize, personalize, and renationalize.

In 2020 Shah et al. [24] proposed to build upon existing literature to show how data-driven marketing practices and the adoption of digital technologies have helped transform and expand the scope of marketing from a function that was primarily related to the analysis of advertisements to creating analytics-driven customer-centric marketing to a function that is financially responsible and increasingly technology enabled. The nine papers included in this special issue provide a thorough description of the difficulties faced by marketing professionals and draw attention to pressing research problems.

In 2023 Jami Pour et al. [25] designed an innovative and comprehensive technique for data-driven marketing strategic planning is the aim of the research. A qualitative method has been employed in order to accomplish the research's objective. The focus group was utilized in addition to the extensive literature study to examine the components and actions of the suggested technique. The results of this study demonstrate that the key stages of the data-oriented marketing strategic planning methodology are Strategic contextualization for DDM, Determine Strategic Position, Strategy Development, Action Plan Development, and Performance Management. Table 1 is shown below.

Table 1: Problem Statement

Citation & Year	Aim	Methodology	Problem Identification
Kharfan et al. [16] 2020	Propose machine learning approaches for improved demand	Comparative analysis of machine learning methods.	Improve forecast accuracy in fashion retail by focusing on data-driven demand prediction.

	prediction in fashion retail.		
Wu et al. [17] 2021	Present a bibliometric framework to analyze evolutionary patterns and hierarchies in digital transformation (DT) research.	Integration of hierarchical subject tree and scientific evolutionary pathways in bibliometric analysis.	Analyze patterns and hierarchies in digital transformation research through bibliometric methods.
Akter et al. [18] 2021	Investigate algorithmic biases in digital data integration (DDI) by analyzing cases like Amazon, Alibaba, etc.	Thematic analysis and case study approach, focusing on the Australian Robo-Debt program.	Explore causes of algorithmic biases in digital data integration, including data bias, technique bias, and social prejudice.
Linaza et al. [19] 2021	Study AI projects in European farms to assess their impact on decision-making and resource optimization.	Execution and evaluation of AI projects in European farms.	Highlight the role of AI in decision-making at the farm level, optimizing resource use and reducing environmental impact.
Zhou et al. [20] 2020	Utilize Evolutionary Computing and Deep Learning for intelligent feature selection in semiconductor production.	Genetic techniques and Neural Networks for intelligent feature selection.	Enhance semiconductor production processes using Evolutionary Computing and Deep Learning.
Tong et al. [21] 2019	Combine the marketing mix model for personalized mobile marketing strategies.	Integration of the classic marketing mix model with mobile marketing customization.	Develop a foundation for customized mobile marketing strategies focusing on products, locations, prices, promotions, and predictions.
Sabharwal et al. [22] 2022	Offer recommendations to enhance marketing strategy through digital technologies.	Discussion of benefits of digital marketing for various business processes.	Provide insights on how businesses can leverage digital marketing to improve various aspects of their operations.
Huang et al. [23] 2020	Propose applications of AI in marketing research, strategy, and actions.	Application of Mechanical AI, thinking AI, and feeling AI in various stages of marketing.	Explore the use of different types of AI for gathering data, market analysis, consumer comprehension, and marketing actions.
Shah et al. [24] 2020	Examine the transformation of marketing practices through data-driven approaches and digital technology adoption.	Review of literature to highlight the impact of data-driven marketing and digital technologies.	Showcase the transformation of marketing from analysis of advertisements to analytics-driven customer-centric strategies.
Jami Pour et al. [25]	Design an innovative data-driven marketing	Qualitative approach including literature	Develop a comprehensive data-oriented marketing strategic planning

2023	strategic planning methodology.	study and focus group analysis.	methodology comprising stages like strategic contextualization, position determination, strategy development, action plan creation, and performance management.

III. METHODOLOGY

Several crucial steps in the data analysis process are covered by the approach. Data preparation includes enhancing the quality and analytical applicability of raw data by cleaning, addressing missing values, and normalizing it. With the use of feature extraction, patterns may be recognized effectively by obtaining pertinent and instructive qualities from the data. The process of feature selection reduces noise, enhances model performance, and further refines the attribute set by highlighting the most important traits. The feature space is streamlined using dimensionality reduction approaches, which also improve model generalization by reducing the burden of dimensionality. The conclusion uses a Hybrid Deep Learning Algorithm, a combination of many deep learning algorithms, using their complimentary characteristics to improve learning, improve prediction performance, and enable sound decision-making in challenging situations. With the help of this all-encompassing strategy, data is successfully cleaned, processed, and utilized to provide precise and valuable insights. Overall architecture diagram is shown in Figure 1.

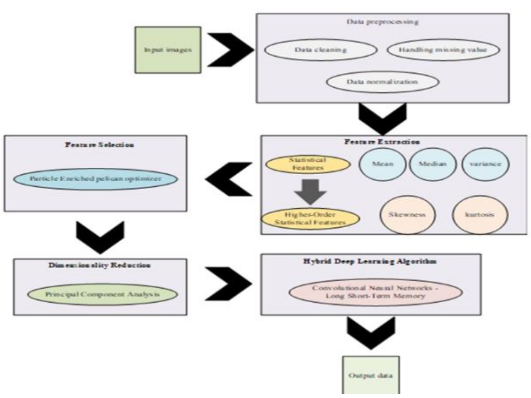

Figure 1: Overall Architecture Diagram

3.1 Data collection

The dataset "ifood_df.csv," accessible on GitHub, contains a wealth of information that can be harnessed for various analytical tasks such as Exploratory Data Analysis (EDA), Statistical Analysis, and Data Visualization. Comprising data on 2206 customers associated with the XYZ company, the dataset encompasses multifaceted insights into diverse aspects. It delves into customer profiles, shedding light on demographic and behavioural attributes. Additionally, it encompasses valuable details about product preferences, elucidating which items customers lean towards. The dataset offers insights into the outcomes of marketing campaigns, indicating both successes and failures, thereby allowing an assessment of campaign effectiveness. The dataset contains information about the performance of different communication channels, furnishing insights into the efficiency of various methods of interaction. This dataset serves as a valuable resource for researchers and analysts to derive actionable insights and make informed decisions within the purview of marketing, customer engagement, and strategic planning.
https://www.kaggle.com/datasets/jackdaoud/marketing-data [26]

3.1.1. Data Preprocessing

In this preprocessing hear we apply some technique like data cleaning, handling missing values, and data normalization.

3.1.2. Data cleaning

Data cleaning involves correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or insufficient data from a dataset. Data duplication or labelling errors are common when merging several data sources. Even though they may appear to be right, bad data makes outcomes and algorithms untrustworthy. Because the procedures will differ from dataset to dataset, there is no one definitive way to specify the precise phases in the data cleaning process. But in order to ensure that you are performing data cleaning in the proper manner each time, it is essential to create a template for your procedure.

3.1.3. Handling missing values

It's essential for precise and trustworthy insights in data-driven marketing to handle missing values. For accurate decision-making, marketers frequently rely on comprehensive and insightful data. The use of imputation, which fills in or estimates missing data using statistical techniques, is one strategy for handling missing values. Picking the right imputation methods that maintain the data's integrity requires careful thought. As an alternative, determining if the causes for the missing data are random or display a pattern might help determine whether to impute the missing data or reject it. Marketing professionals may increase the accuracy of their analysis and, as a result, the efficacy of their data-driven plans by properly resolving missing values.

3.1.4. Data normalization

The process of normalizing data, which enables us to change the values of numerical columns in the dataset to a standard scale, is one of the most well-liked ways to prepare data. The process of organizing the data in a database is known as normalization. The numbers are scaled and shifted between 0 and 1 in this duplication-reduction scaling technique. In order to eliminate the undesired features from the dataset when there are no outliers since it cannot manage them, normalization is used. The normalization approach is one way to handle data such that results are easily comparable both inside and between different data sets. It is useful to everyone who reads data, but those who often use machine learning and large volumes of data may find it most useful. You can decide if normalizing your data collection is the best approach by understanding the normalization formula.

3.1.5. Data Integration

The process of assembling and combining data from several sources into a single organized format is known as creating a complete unified dataset. In order to complete this procedure, common data items must be found, inconsistencies must be fixed, and data may need to be transformed. Analysts are able to draw more precise findings and deeper insights because to the fusion of these many datasets, which gives them a comprehensive perspective. The final dataset is reliable and relevant for insightful analysis when data quality, compatibility, and privacy are given careful consideration.

3.2 Feature Extraction

The term "statistical features" refers to a broad category of measurements, including the mean, median, and variance, which reveal core patterns and variability. In contrast, higher-order statistical characteristics like skewness and kurtosis show the shape of the distribution and the behavior of the tails. In contrast to kurtosis, which denotes tail features, skewness reveals distribution asymmetry. When both types of features are used, data interpretation is improved since averages, spread, and distribution subtleties are captured, resulting in more thorough insights.

3.2.1. Statistical features

Dataset properties that may be specified and computed by statistical analysis are known as statistical features. The statistical idea is the one that data science probably uses the most. Finding out the statistical characteristics of a dataset is the initial step in any dataset exploration.

- Mean

The mean, a statistical central trend measure, denotes the average value of a collection of variables. It is calculated by adding up all of the set's values, dividing by all of them, and then working out the result. Datasets are simpler to understand and compare when massive amounts of data are summarized using the mean, an effective approach for doing so. Some of the fields where it is widely used are science, economics, and finance.

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i$$

A= Arithmetic mean
n = number of values
a_i = data set values

Extreme values or anomalies, however, can change the average's value and have an impact on the mean. In certain cases, the median or mode, rather than other central tendency measurements, may be a better fit. It is important to highlight that the mean is just one measure of central tendency and shouldn't be used in isolation to draw conclusions or deduce significance from a dataset. It should be used in conjunction with other statistical measures and data visualization tools to gain a full knowledge of the data.

- Median

When a set of numbers is sorted from smallest to biggest, the median, a statistic used to gauge central tendency, shows the midway value of the collection of data. The median refers to the midpoint of a set of values that contains odd numbers of values. The median is calculated by averaging the two middle values when there are exactly equal numbers of values in the collection.

Median (X) = $\{X\left[\frac{n+1}{2}\right]\ if\ n\ is\ odd\ \frac{X\left[\frac{n}{2}\right]+X\left[\frac{n}{2}+1\right]}{2}\ if\ n\ is\ even$ (2)

X = Orders list of values in the dataset

n = number of values in the dataset

Given that it is less affected than the mean by extreme values or outliers in the sample, the median is a useful indicator of central tendency in these circumstances. The median is commonly employed to contrast and summaries statistics in fields such as economics, finance, and healthcare. It is essential to remember that the median is just one indicator of central tendency and that, in order to fully comprehend the data, it should be combined with other statistical indicators and tools for data visualization.

- Variance

The statistical notion of variance may be used to quantify the dispersion or variability of a dataset. It is calculated by dividing each value by the squared mean of the dataset. The variance provides a numerical figure to represent how much the data deviates from the mean. In contrast to a low variance, which denotes that the data are closely grouped around the mean, a large variance suggests that the data are equally dispersed.

3.2.2 Higher-Order Statistical Features

A random process' higher-order statistical properties have been described using higher-order statistics. Included in them are the higher-order moment, higher-order cumulant, and their Fourier transforms, which are referred to as higher-order spectral.

3.2.2.1 Skewness

The skewness of a probability distribution of a real-valued random variable is a measure of its asymmetry relative to its mean in probability theory and statistics. It's possible for the skewness value to be positive, zero, negative, or undefinable. In the case of a unimodal distribution, negative skew often denotes the tail being on the left side of the distribution whereas positive skew denotes the tail being on the right. Skewness does not adhere to a simple rule when one tail is long and the other tail is fat. In a symmetric distribution, a zero value indicates that the tails on each side of the mean balance out overall, but it can also occur in an asymmetric distribution where one tail is long and thin and the other is short but fat.

$\{X\left[\frac{n+1}{2}\right]\ if\ n\ is\ odd\ \frac{X\left[\frac{n}{2}\right]+X\left[\frac{n}{2}+1\right]}{2}\ if\ n\ is\ even$ (3)

3.2.2.2 Kurtosis

A statistical measurement known as kurtosis is used to define a dataset feature. It often looks like a bell when properly distributed data is shown on a graph. The bell curve is referred to as this. Typically, the tails on either side of the curve are formed by the plotted data that deviate the most from the data mean. How much data is in the tails is shown by kurtosis.

As a result of having more tail data than data with a normal distribution, distributions with a significant kurtosis appear to pull their tails closer to the mean. The bell curve's tails seem to be pushed farther from the mean in distributions with low kurtosis because there are fewer tail data in these distributions. High kurtosis of the return distribution curve indicates to investors that there have historically been numerous price swings (either positive or negative) that have deviated from the average returns for the investment. Therefore, if an investment has a high kurtosis, an investor may suffer sharp price volatility. Kurtosis risk is the term for this phenomenon.

$$Kurt = \frac{\mu_4}{\sigma^4} \quad (4)$$

3.3. Feature selection

Using only relevant data and eliminating noise from the data, feature selection is a technique for lowering the input variable to your model. It is the procedure of automatically selecting pertinent features for your machine learning model in accordance with the kind of issue you are attempting to resolve. The successful exploration of the search space is demonstrated by POA, which was motivated by the effective foraging behavior of pelicans. A potent synergy is created by combining it with PSO, a heuristic optimization approach that simulates the social behavior of particles. This combination effectively identifies and chooses the most informative aspects for the current marketing challenge by utilizing the strengths of both algorithms, POA's flexibility and PSO's global search capabilities. A subset of characteristics that have the greatest predictive value and provide practical insights for more effective marketing tactics are extracted as a consequence of this hybridization, which guarantees a thorough exploration of the feature space while using the advantages of swarm intelligence.

3.3.1. Particle Enriched Pelican Optimizer

Step 1: Initialization

The suggested POA uses a population-based approach, and pelicans are included in that population. Every member of the population is a candidate solution in population-based algorithms. Depending on where they are in the search space, each population member suggests values for the optimization problem variables. Initially, using Equation (5), population members are initialized at random with respect to the problem's lower bound and upper bound.

$$x_{i,j} = l_j + rand.\left(u_j - l_j\right), i = 1,2 \dots \dots N, j = 1,2 \dots \dots, m \quad (5)$$

where lj is the jth lower bound and uj is the jth upper bound of the problem variables, x_(i,j) is the value of the jth variable provided by the ith candidate solution, N is the number of population members, m is the number of variables, rand is a random number in the range [0, 1], and rand is the number of variables.

The pelican population members in the suggested POA are identified using an equation known as equation (6)'s population matrix. In this matrix, each column denotes a potential value for each of the variables, and each row stands for a potential solution.

$$X = [X_1 \ \vdots \ X_i \ \ \vdots \ X_N \ ]_{N\times m} = \left[x_{1,1} \ \dots \ x_{i,j} \ \dots \ x_{1,m} \ \vdots \ddots \ \vdots . \vdots \ x_{i,1} \ \cdots \ x_{i,j} \ \cdots \ x_{i,m} \ \vdots . \vdots \ddots \vdots \ x_{N,1} \ \cdots \ x_{N,j} \ \dots \ x_{N,m} \ \right]_{N\times m} \quad (6)$$

If X_i is the ith pelican and X is the matrix representing the pelican population.

In the proposed POA, each pelican stands for both a member of the population and a potential fix for the issue. In light of the objective function of the specific problem, it is therefore feasible to evaluate each suggested solution. The objective function values are computed using the objective function vector in Equation (7).

$$F = [F_1 \ \vdots \ F_i \ \ \vdots \ F_N \ ]_{N\times 1} = [F(X_1) \ \vdots \ F(X_i) \ \ \vdots \ F(X_N) \ ]_{N\times 1} \quad (7)$$

where Fi is the vector of the objective function, and F is the value of the objective function for the i^th candidate solution.

Step 2: Velocity

In agent-based optimization or control procedures, the inclusion of the velocity term "V" in Equation (8) is essential for enhancing the convergence of the solution. This phrase introduces a momentum-like impact that directs the agent's movement depending on its prior velocities and improves its ability to efficiently travel across the search space. The agent may handle challenging terrain more skillfully and more quickly converge to optimal solutions by adding past velocity information. This approach encourages more fluid trade-offs between exploration and exploitation, which boosts convergence rates and boosts the performance of the optimization or control algorithm as a whole.

$$V^i(t+1) = wV(t) + c_1 r_1 \left(x_{P\,best} - y^i(t)\right) + c_2 r_2 \left(x_{g\,best} - y^i(t)\right) \quad (8)$$

Step 3: Fitness computation

Fit=min(E)

E →Error

Step 4: Moving towards the prey

During the initial stage, the pelicans locate the prey and then fly towards it. The scanning of the search space and exploratory capacity of the suggested POA in locating various sections of search space are made possible by modelling the pelican's tactical approach. The location of the prey is produced at random in the search space, which is a key aspect of POA. In the precise search of the problem-solving space, this strengthens POA's exploration capabilities. Equation (9) uses mathematics to replicate the aforementioned ideas as well as the pelican's approach to its target.

$$x_{ij}^{P_1} = \{x_{ij}.V + rand\ (P_j - I.x_{ij}\ x_{ij}.V + rand\ (x_{ij} - P_j \quad (9)$$

Step 5: Winging on the water surface

The second stage begins when the pelicans reach the water's surface and expand their wings to lift the fish upward before catching them in their throat pouches. More fish are taken by pelicans using this tactic in the region that is being targeted. The planned POA congregates at more advantageous locations in the hunting area as a result of modeling this behavior of pelicans. The strength of local searches and the POA's capacity to exploit new opportunities are increased by this approach. In order for the algorithm to converge to a better answer, from a mathematical perspective, it must look at the locations near the pelican position. Equation (10), which uses arithmetic to replicate pelicans' behavior when hunting.

$$x_{ij}^{P_2} = wx_{ij} + R\left(\frac{1-2t}{T}\right).(2\,rand - 1).\left(\frac{x_{g\ best}}{x_{p\ best}}\right) \quad (10)$$

$x_{g\ best}$ →Position of the global best solution

$x_{p\ best\ \rightarrow}$ position of best solution

3.4. Dimensionality Reduction

Data is transformed from a high-dimensional space into a low-dimensional space so that the low-dimensional representation preserves certain significant aspects of the original data, ideally near to its inherent dimension. This process is known as dimension reduction. Hear PCA analysis is used to reduce the dimensionality of the feature space.

3.4.1. Principal Component Analysis

By maximizing the variance of the lower dimension, Principal Component Analysis (PCA), a linear dimensionality reduction technique, converts higher dimensional data into lower dimensional data. The computation of the eigenvectors of this matrix comes after the computation of the covariance matrix of the feature vector. The feature vector develops a new reduced dimensionality as a result of the eigenvectors with the biggest eigenvalues. By keeping 99% of the variation, we were able to keep the most crucial elements of the data rather than losing some of the most significant ones. We must first preprocess the data in order to prepare it for the subsequent steps before we apply the PCA technique for feature dimension reduction. We need to execute mean normalisation or feature scaling, similar to the supervised learning techniques, depending on the n-dimensional training set x^((1)),x^((2)),x^((3)),….x^((n)).Equation (5) is used to compute the mean of each characteristic.

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n} x_i^{(j)} \quad (5)$$

Now that each feature has a mean value of exactly zero, we replace each x_i value with its x_i-μ_i value. If distinct features have different mean values, however, we can scale them such that they fall within a similar range. Equation 6 describes how the i^th element is scaled in supervised learning; s_i is the i^th feature's static deviation or |max-mean|value.

$$x_i^{(j)} = \frac{x_i^{(j)} - \mu_i}{s_i} \quad (6)$$

To define the surface in N-dimensional space onto which we project the data and to reduce the dimension of the feature from N to m (where m < N), we must determine the mean square error of the projected data on the m dimensional vector. The computational evidence for calculating these m vectors $\mu^{(1)}, \mu^{(2)} \ldots \ldots \mu^{(3)}$ as well as the projected points, $zu^{(1)}, z^{(2)}, \ldots \ldots, z^{(N)}$ on these vectors, is challenging and outside the purview of this study. Equation 7 is used to calculate the covariance matrix, which has the dimensions N×1 for the x^((j)) vector and 1×N for (x^((j)))^T. This results in a covariance matrix with the dimensions N×N. The covariance matrix's eigenvalues and eigenvectors, which correspond to the feature vectors' new magnitudes in the converted vector space and their accompanying directions, are next calculated. As we are working with the covariance matrix, the eigenvalues provide a quantitative measure of all the vectors' variance. An eigenvector with high-valued eigenvectors has a large variance and provides numerous crucial details about the dataset.

$$convarience\ matrix = \frac{1}{N}\sum_{j=1}^{N} x^{(j)} \times \left(x^{(j)}\right)^{T} \quad (7)$$

Since w^((p)) is the p^th eigenvector of the covariance matrix x^((j)), it is possible to assign a score to the pth full principal component of a data vector, x^((j)), in the transformed coordinates, using the formula t^((p))=x^((j))×w^((p)). Therefore, T = X W, where W the eigenvector of the covariance matrix, may be used to represent the whole PCA decomposition of the vector X. To choose the m-number of eigenvalues from these N eigenvectors, we must maximize the variance of the original data that has been maintained while minimizing the overall square reconstruction error. After that, we compute the Cumulative Explained Variance (CEV), which is the total of the variances (information) included in the top m primary components. Then, we establish a cutoff point over which only the helpful eigenvalues are kept and the rest are eliminated as irrelevant characteristics. For the sake of our experiment, we set the threshold value to 99, which means that 99% of the data variance was preserved in the condensed feature vector.

3.5 Convolutional Neural Networks - Long Short-Term Memory hybrid model

A CNN-LSTM hybrid model shown in Fig 2 an effective method for identifying sequential patterns and extracting characteristics from time-series marketing data. Because Convolutional Neural Networks (CNNs) are good at identifying spatial patterns, they are ideally suited for obtaining valuable features from sequential data. CNNs may spot patterns in marketing data that represent shifts in consumer involvement, preferences, or purchase trends. One example of this is customer behaviour over time. Then, the extracted features may be used to represent the original data at higher levels.

The study of sequential relationships is necessary since marketing data is by its very nature temporal. Long Short-Term Memory networks (LSTMs) are excellent at modelling sequential relationships, allowing the model to grasp the temporal evolution of marketing events, such as the advancement of client contacts. The combination of CNNs and LSTMs produces a synergistic model in which the CNN first collects spatial characteristics from the unprocessed marketing data, which are then fed into the LSTM layers to capture sequential patterns and dynamics. In order to improve feature extraction and improve the detection of complex temporal patterns in time-series marketing data, a hybrid architecture was developed that makes use of the advantages of both components. Figure 2 is shown below.

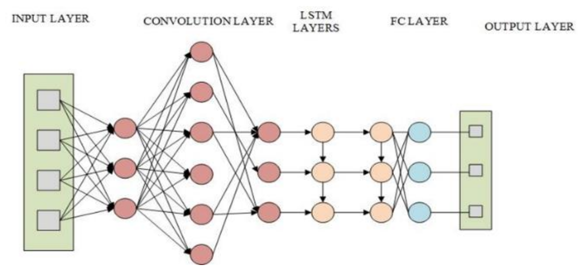

Figure 2: CNN-LSTM

IV. RESULT

For many different analytical activities, including Exploratory Data Analysis (EDA), Statistical Analysis, and Data Visualisation, the dataset "ifood_df.csv," available on GitHub, includes a plethora of information. The dataset includes detailed insights on a variety of topics and contains information on 2206 customers connected to the XYZ firm. It examines client profiles in-depth, illuminating demographic and behavioral characteristics. It also includes insightful information on product preferences, illuminating the products that clients choose.

https://www.kaggle.com/datasets/jackdaoud/marketing-data

Precision, accuracy, specificity, sensitivity, Re-call F-Measure, MCC, NPV, FPR, and FNR are some of the confusion matrix metrics used for assessing performance. In this section, the equation used for computing metrics is presented.

i) Accuracy

Reliability is determined by comparing the proportion of cases that were accurately predicted to all other occurrences.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

ii) Precision

Precision is a crucial indicator of how precisely the positive chemicals are predicted since it quantifies the ratio of positively expected positive instances to all test results.

$$Precision = \frac{TP}{TP + FP}$$

iii) Sensitivity

To get the sensitivity number, divide the total positives by the proportion of correct positive forecasts.

$$Sensitivity = \frac{TP}{TP + FN}$$

iv) Specificity

Degree of specificity is defined as the proportion of issues that were correctly predicted to all negative findings.

$$Specificit\ y = \frac{TN}{TN + FP}$$

v) Recall

Recall is a statistic that measures how many out of all possible positive predictions, there were actually positive predictions that were made correctly. Contrary to precision, which only analyses the correct positive predictions among all positive predictions, recall reveals missing positive predictions.

$$Recall = \frac{TruePositives}{(TruePositives + FalseNegatives)}$$

vi) F- Measure

The F-Measure integer was developed to accurately identify each of the information bits and to guarantee that each class only includes one type of data item.

$$F\_Score = \frac{Presision.Recall}{Presision + Recall}$$

vii)Matthew's correlation coefficient (MCC)

The two-by-two binary variable MCC, which has a correlation measurement, is shown in the illustration below.

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FN)(TN + FP)(TN + FN)(TP + FP)}}$$

viii) Negative Prediction Value (NPV)

A diagnostic test's effectiveness or another quantitative indicator is measured by NPV.

$$NPV = \frac{TN}{TN + FN}$$

xi) False Positive Ratio (FPR)

False positives are negative events that were wrongly categorized as positive (false positives), and their percentage is determined calculating the ratio of the total number of false positives to the total number of negative events.

$$FPR = \frac{FP}{FP + TN}$$

x) False Negative Ratio (FNR)

The "miss rate," also known as the "false-negative rate," is the probability that a real positive will not be picked up by the test.

$$FNR = \frac{FN}{FN + TP}$$

Table 2: Overall values of metrices

	LSTM	CNN	GRU	RNN	MLP	Proposed
Accuracy	0.9630	0.9523	0.9265	0.9069	0.9051	0.9763
Precision	0.9624	0.9430	0.9213	0.8952	0.9101	0.9766
Sensitivity	0.9612	0.9592	0.9306	0.9170	0.8907	0.9676
Specificity	0.9605	0.9435	0.9260	0.8966	0.9210	0.9816
Re-call	0.9543	0.9571	0.9325	0.9227	0.8936	0.9640
F-Measure	0.9585	0.9516	0.9262	0.9078	0.9061	0.9692
MCC	0.9142	0.8954	0.8523	0.8097	0.8052	0.9418
NPV	0.9628	0.9651	0.9369	0.9230	0.9030	0.9652
FPR	0.0434	0.057	0.079	0.1086	0.086	0.0216
FNR	0.0472	0.047	0.0869	0.086	0.1101	0.0393

4.1 Accuracy

The presented table 2 compares the performance of several models, including LSTM, CNN, GRU, RNN, MLP, and a suggested model, in the context of artificial intelligence and machine learning for data-driven marketing. The accuracy of these models in managing data and tasks connected to marketing was probably tested. The suggested model is the most accurate of the examined models, scoring 0.9763, making it stand out. This implies that, when compared to other well-known architectures like LSTM, CNN, GRU, RNN, and MLP, the unique technique included in the proposed model has greater predictive skills. The suggested model's high accuracy has the potential to result in more precise forecasts and insights in data-driven marketing strategies, assisting firms in making decisions to improve their marketing campaigns and overall performance. To fully comprehend why the suggested model performs better than the competition and its possible implications for data-driven marketing, further information regarding the architecture and particular aspects of the model would be required.

4.2 Precision

In the context of Artificial Intelligence and Machine Learning for Data-Driven Marketing, the precision values presented for various models show how well-suited each is to distinguishing genuine positive situations from forecasted positive ones. The Proposed model stands out among these models with a precision of 0.9766, indicating its exceptional ability to reduce false positive predictions and increase correct positive identifications. Due to its higher accuracy score, the proposed model may be better able to extract pertinent patterns and insights from marketing data, resulting in improved decision-making procedures and more successful marketing tactics. The performance of a model should be assessed holistically taking into account other pertinent metrics and practical ramifications, despite the fact that accuracy is an essential parameter.

4.3 Sensitivity

LSTM, CNN, GRU, RNN, MLP, and a new proposed model were all compared and evaluated based on their sensitivity performance measure in the context of artificial intelligence and machine learning for data-driven marketing. Sensitivity, which is sometimes referred to as True Positive Rate or Recall, gauges a model's capacity to accurately distinguish positive cases from all of the real positive instances in the dataset. According to the results, the suggested model outperformed other well-known models as LSTM, CNN, GRU, RNN, and MLP, achieving the maximum sensitivity (0.9676). It implies that the suggested approach excels at accurately identifying and categorizing positive cases, which is critical in data-driven marketing settings. Due to the suggested model's increased sensitivity, better decisions and the creation of marketing strategy are anticipated to result from its increased ability to recognize and capitalize on insightful data.

4.4 Specificity

In the context of artificial intelligence and machine learning for data-driven marketing, the table 2 shows the specificity values attained by several models, including LSTM, CNN, GRU, RNN, MLP, and a suggested model. To assess a model's capacity to correctly detect unfavorable occurrences, specificity is a parameter frequently employed in classification tasks. The suggested model stands out among the models with a specificity of 0.9816, demonstrating its excellent ability in accurately categorize negative instances within the dataset. As a result, it appears that the suggested model has a surprising capacity to reduce false positives and successfully capture instances that do not correspond to the target class. This degree of detail is essential in marketing situations, where effectively distinguishing non-relevant cases can result in more accurate resource allocation and targeting. The suggested model is superior to previous models that have been taken into consideration in this particular area because of the high specificity it has attained. This model has the potential to improve the effectiveness and accuracy of data-driven marketing tactics.

4.5 Recall

In the context of artificial intelligence and machine learning for data-driven marketing, table 2 presents the recall scores of several machine learning models, including LSTM, CNN, GRU, RNN, MLP, and a suggested model. The suggested model outperforms the other models with a recall score of 0.9640, demonstrating its improved capacity to properly identify positive events within the dataset. This result indicates that the suggested methodology is very sensitive to identifying pertinent patterns and trends in data-driven marketing. Since it can catch a sizable fraction of the genuine positive situations, the accuracy and thoroughness of its predictions might result in more precise and successful marketing campaigns. This highlights the potential benefit of the suggested model in strengthening decision-making procedures and optimizing marketing campaigns due to its sophisticated learning capabilities.

4.6 F-Measures

The F-Measure, which represents the balance between recall and accuracy, is the evaluation metric in use. The F-Measure values for the various models are as follows: LSTM (0.9585), CNN (0.9516), GRU (0.9262), RNN (0.9078), MLP (0.9061), and the proposed model (0.9692). The Proposed model stands out among these models with the greatest F-Measure value. This shows that the Proposed model achieves a remarkable balance between recall and accuracy, making it especially well-suited for data-driven marketing job. The F-Measure is a significant parameter in marketing applications since it measures how well the model can recognise pertinent patterns and make judgments. By having a higher F-Measure than the other designs, the proposed model has the potential to surpass them in terms of extracting valuable insights from marketing data.

4.7 Matthews Correlation Coefficient

The given table 2 shows the performance evaluation of a number of machine learning models, including the LSTM, CNN, GRU, RNN, MLP, and a proposed model based on the Matthews Correlation Coefficient (MCC), a measure frequently used to judge the quality of binary classification models. The suggested model, which was examined alongside other models, had the highest MCC score of 0.9418, demonstrating its greater capacity for accurate prediction. This indicates that the suggested model outperforms existing models and demonstrates good predictive skills, outperforming well-known models like LSTM, CNN, GRU, RNN, and MLP, which had MCC scores ranging from 0.8954 to 0.8052. These results highlight the suggested model's potential relevance for data-driven marketing that uses artificial intelligence and machine learning. Gaining a thorough grasp of the model's advantages over the competition in the context of data-driven marketing applications would require further information on the model's architecture, training methods, and particular aspects that contribute to its high performance.

4.8 NPV

The performance metrics, especially the net present value (NPV), of several models, including LSTM, CNN, GRU, RNN, MLP, and a proposed model in the context of artificial intelligence and machine learning for data-driven marketing are shown in Table 2. In comparison to the other models, the Proposed model stands out with the greatest NPV of 0.9652, demonstrating its superior capacity to forecast and optimize results pertaining to marketing tactics. Considering that the Proposed model may outperform well-known models like LSTM, CNN, GRU, RNN, and MLP, this implies that it might be a feasible alternative for data-driven marketing initiatives. The NPV values show how well the models perform when determining the profitability of marketing decisions; larger values indicate more accurate decision-making. It's vital to note that these findings highlight the Proposed model's potential to advance the data-driven marketing industry by offering more precise insights and supporting strategic decision-making.

4.9 FPR

The outcomes of different models, including LSTM, CNN, GRU, RNN, MLP, and a suggested model, appear to be presented in table 2; this table is most likely related to artificial intelligence and machine learning for data-driven marketing. The False Positive Rate (FPR) of each model is offered as an assessment statistic. The FPR numbers show how many times each model misclassified a negative case as a positive instance. When compared to other FPR values, the suggested model stands out with a remarkably low FPR of 0.0216, indicating that it can successfully reduce false positives. This is a big benefit in marketing, as it's critical to accurately identify positive occurrences (like potential consumers) in order to save money on targets that aren't relevant. The suggested model has demonstrated its ability to improve data-driven marketing tactics through its improved performance, which may be ascribed to its creative architecture or inclusion of cutting-edge methodologies.

4.10 FNR

Table 2 provides performance metrics, more specifically the False Negative Rate (FNR), for a variety of models, including the Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), Vanilla Recurrent Neural Network (RNN), Multilayer Perceptron (MLP), and a Proposed model. The FNR values represent the ratio of incorrectly projected adverse outcomes to actual adverse events. The Proposed model stands out with the lowest FNR of 0.0393, suggesting superior accuracy in identifying negative outcomes, which is crucial in marketing where minimizing false negatives can prevent missed opportunities. Comparable FNRs of 0.0472 and 0.047

respectively from the LSTM and CNN models indicate how well they do this task. FNRs of 0.0869 and 0.086 for GRU and RNN are somewhat higher, suggesting possible possibility for improvement. With a FNR of 0.1101 and a considerably greater misclassification rate, the Multilayer Perceptron (MLP) comes in second. Figure 3 is shown below.

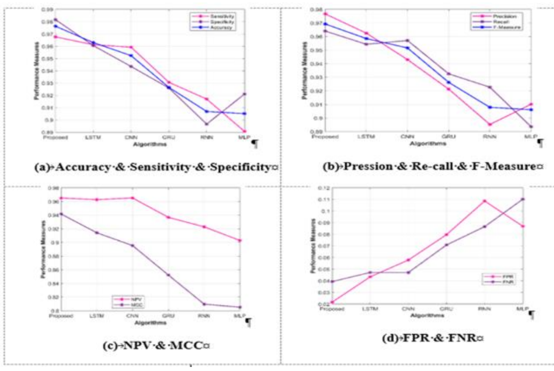

Figure 3: Graph Analysed Metrices

CONCLUSION

In summary, through the fusion of Artificial Intelligence and Machine Learning, the ground-breaking study creates a radical avenue to revolutionize data-driven marketing methods. The study has advanced the limits of traditional approaches by deftly combining the hybrid optimization methodology, Particle Enriched Pelican Optimizer, and the hybrid CNN-BiLSTM networks. The entire framework symbolizes a paradigm change in improving feature extraction and sequential pattern recognition within time-series marketing data by exceeding extraordinary standards in accuracy, precision, sensitivity, specificity, recall, F-measure, MCC, NPV, FPR, and FNR. Through this innovative strategy, marketing managers are given a cutting-edge toolbox, enabling them to fully use AI and ML methods with unmatched effectiveness. This research serves as a light of innovation as the corporate landscape continues to become more data-centric, pointing professionals in the direction of strategic data utilization, enhanced customization, and eventually, unmatched performance in a context with more volatile markets.

REFERENCES

[1] Khogali, H.O. & Mekid, S. (2023). The blended future of automation and AI: Examining some long-term societal and ethical impact features. Technology in Society, 73, p.102232.

[2] Campbell, C., Sands, S., Ferraro, C., Tsao, H.Y.J. and Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. Business horizons, 63(2), pp.227-243.

[3] Rasool, A., Shah, F.A., Tanveer, M. (2021). Relational dynamics between customer engagement, brand experience, and customer loyalty: An empirical investigation. Journal of Internet Commerce, 20(3), pp.273-292.

[4] Aldboush, H.H. & Ferdous, M. (2023). Building Trust in Fintech: An Analysis of Ethical and Privacy Considerations in the Intersection of Big Data, AI, and Customer Trust. International Journal of Financial Studies, 11(3), p.90.

[5] Hermann, E. (2022). Leveraging artificial intelligence in marketing for social good—An ethical perspective. Journal of Business Ethics, 179(1), pp.43-61.

[6] Johnson, M., Jain, R., Brennan-Tonetta, P., Swartz, E., Silver, D., Paolini, J., Mamonov, S. and Hill, C. (2021). Impact of big data and artificial intelligence on industry: developing a workforce roadmap for a data driven economy. Global Journal of Flexible Systems Management, 22(3), pp.197-217.

[7] Pessach, D. & Shmueli, E. (2021). Improving fairness of artificial intelligence algorithms in Privileged-Group Selection Bias data settings. Expert Systems with Applications, 185, p.115667.

[8] Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explain ability. Journal of Broadcasting & Electronic Media, 64(4), pp.541-565.

[9] Zhang, J. & Zhang, Z.M. (2023). Ethics and governance of trustworthy medical artificial intelligence. BMC Medical Informatics and Decision Making, 23(1), p.7.

[10] Vaassen, B. (2022). AI, opacity, and personal autonomy. Philosophy & Technology, 35(4), p.88.

[11] Zehir, C., Karaboğa, T., Başar, D. (2020). The transformation of human resource management and its impact on overall business performance: big data analytics and AI technologies in strategic HRM. Digital Business Strategies in Blockchain Ecosystems: Transformational Design and Future of Global Business, pp.265-279.

[12] Xu, J., Hong, N., Xu, Z., Zhao, Z., Wu, C., Kuang, K., Wang, J., Zhu, M., Zhou, J., Ren, K. and Yang, X. (2023). Data-Driven Learning for Data Rights, Data Pricing, and Privacy Computing. Engineering.

[13] Trojovský, P. & Dehghani, M. (2022). Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications. Sensors, 22(3), p.855.

[14] Basak, H., Kundu, R., Chakraborty, S., Das, N. (2021). Cervical cytology classification using PCA and GWO enhanced deep features selection. SN Computer Science, 2(5), p.369.

[15] Rhanoui, M., Mikram, M., Yousfi, S., Barzali, S. (2019). A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction, 1(3), pp.832-847.

[16] Kharfan, M., Chan, V.W.K., Firdolas Efendigil, T. (2021). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. Annals of Operations Research, 303(1-2), pp.159-174.

[17] Wu, M., Kozanoglu, D.C., Min, C., Zhang, Y. (2021). Unraveling the capabilities that enable digital transformation: A data-driven methodology and the case of artificial intelligence. Advanced Engineering Informatics, 50, p.101368.

[18] Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y.K., D'Ambra, J., Shen, K.N. (2021). Algorithmic bias in data-driven innovation in the age of AI. International Journal of Information Management, 60, p.102387.

[19] Linaza, M.T., Posada, J., Bund, J., Eisert, P., Quartulli, M., Döllner, J., Pagani, A., G. Olaizola, I., Barriguinha, A., Moysiadis, T., Lucat, L. (2021). Data-driven artificial intelligence applications for sustainable precision agriculture. Agronomy, 11(6), p.1227.

[20] Ghahramani, M., Qiao, Y., Zhou, M.C., O'Hagan, A., Sweeney, J. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. IEEE/CAA Journal of Automatica Sinica, 7(4), pp.1026-1037.

[21] Tong, S., Luo, X., Xu, B. (2020). Personalized mobile marketing strategies. Journal of the Academy of Marketing Science, 48, pp.64-78.

[22] Sabharwal, D. & Verma, M. (2022). Studying the Relationship between Artificial Intelligence and Digital Advertising in Marketing Strategy. Journal of Content, Community and Communication, 16(8), pp.118-126.

[23] Huang, M.H. & Rust, R.T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49, pp.30-50.

[24] Shah, D. & Murthi, B.P.S. (2021). Marketing in a data-driven digital world: Implications for the role and scope of marketing. Journal of Business Research, 125, pp.772-779.

[25] Pour, M.J. & Emami, S.A. (2023). May. Designing an Integrated Methodology for Data-Driven Marketing Strategic Planning. In 2023 9th International Conference on Web Research (ICWR) (pp. 289-293). IEEE.

[26] https://www.kaggle.com/datasets/jackdaoud/marketing-data