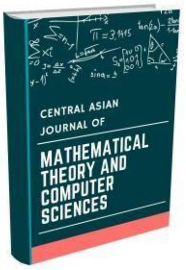

CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS
Volume: 06 Issue: 03 | July 2025 ISSN: 2660-5309

CENTRAL ASIAN STUDIES

Article

Predicting User Engagement in Mobile Applications Using Machine Learning: Insights for Optimizing App Design and Retention

R. Sivakani*[1], G. Rajasekaran[2], S. Manimaran[3], M. Mohamed Sameer Ali[4], S. Suman Rajest[5], R. Regin[6]

1,2,3,4,5. Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India
6. SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India
* Correspondence: sivakani@dhaanishcollege.in

Abstract: The use of machine learning methods to foretell how users will interact with mobile applications is the focus of this research. Predictive models are created to foretell engagement levels by assessing a variety of features, including user demographics, app usage habits, and user feedback. In order to help app developers optimise user experiences and retention methods, our research seeks to shed light on the elements that influence user engagement and to construct accurate prediction models. Machine learning algorithms may be able to predict how users will interact with a mobile app, according to the results. This might lead to better app designs and more downloads. User engagement dynamics are illuminated by this ground-breaking investigation on the complex relationship between user behaviour and app functionality. Results from extensive testing and analysis show that machine learning has great promise as a method for understanding user actions and preferences. These findings demonstrate the impact of predictive analytics on app development and herald a new age of personalised and interactive user experiences. The findings from this study could significantly impact how mobile app optimisation strategies are developed in the future.

Keywords: Forecast User, Revolutionary Power, Optimization Tactics As The Digital World Develops, Machine Learning Algorithms, Mobile App Design, Potential of Machine Learning

Citation: Sivakani, R., Rajasekaran, G., Manimaran, S., Ali, M. M. S., Rajest, S. S., Regin, R. Predicting User Engagement in Mobile Applications Using Machine Learning: Insights for Optimizing App Design and Retention. Central Asian Journal of Mathematical Theory and Computer Sciences 2025, 6(3), 472-489.

Received: 20th Apr 2025
Revised: 28th Apr 2025
Accepted: 7th May 2025
Published: 17th May 2025

1. Introduction

As mobile applications become an essential part of our daily lives in the digital world, app developers and marketers must now be able to anticipate and interpret user interaction. Getting users to engage at a high level is essential to an app's success and longevity in the face of the abundance of apps available to them [44]. The complex behaviors of contemporary app users are frequently difficult to capture by traditional techniques of user engagement analysis [45]. As a result, the use of machine learning techniques to offer more accurate predictions and in-depth insights is increasing [46]. We utilize Jupyter Notebook, which is perfect for writing and debugging code interactively. Using machine learning techniques, this paper explores the field of user engagement prediction in mobile applications, with a particular emphasis on the Random Forest algorithm and K-Means Clustering [47]. These methods provide reliable means of analyzing intricate user data and accurately predicting engagement trends.

This study is crucial even though it has the power to drastically alter app development and marketing strategies. By using machine learning, app developers and marketers can use statistical data to create proactive plans in addition to reactive ones [48]. This modification increases the app's commercial success while simultaneously enhancing user experiences, retention rates, and performance [49]. This study aims to perform two things: first, it will analyze user engagement data by applying the Random Forest and K-Means Clustering algorithms; second, it will create prediction models that can reliably anticipate the amount of user interaction in mobile applications [50]. By fulfilling these goals, the study hopes to offer marketers and app developers practical advice and insights that will help them refine their approaches and increase user engagement [51]. The purpose of the study is to use the Random Forest and K-Means Clustering algorithms to analyze user interaction data in mobile applications, create predictive models that can anticipate user involvement levels in mobile applications with accuracy, offer practical perspectives and suggestions to app developers and marketers to enhance their tactics and optimize user interaction, and promote the development of mobile app analytics and give stakeholders the insight they need to design mobile experiences that are captivating, interesting, and focused on the needs of users [52].

Through this research, we hope to contribute to the growing body of knowledge in mobile app analytics and provide stakeholders with practical tools to aid in their better navigating the cutthroat market [53]. By utilizing machine learning, we hope to give marketers and app developers the knowledge they need to create engaging, intriguing, and user-focused mobile experiences [54]. User engagement in mobile apps is a critical factor for the success and sustainability of app ecosystems [55]. However, predicting user engagement poses a significant challenge for app developers and marketers [56]. The problem statement revolves around the need to leverage machine learning techniques to forecast user engagement accurately [57]. This entails understanding and predicting how users interact with mobile apps over time, including their frequency of use, session duration, in-app actions, and retention rates [58]. Developers may improve user engagement and eventually app success by tackling this obstacle and optimizing app design, functionality, and marketing tactics [59].

The project's goal is to predict users' engagement with mobile applications by utilizing machine learning [60]. The field of machine learning is fundamental to the idea of utilizing it to predict consumers' engagement with mobile applications [61]. The primary objective of this area of computer science is the development of models and algorithms that enable computers to learn from data and form opinions or predictions [62]. Rather than directly writing instructions, machine learning algorithms use statistical approaches to find patterns and relationships in data [63]. The algorithms can then, without explicit programming, generate predictions or judgments based on this knowledge [64]. In order to obtain detailed information on user interactions with mobile apps, the project will create a reliable data gathering system [65]. In order to forecast user involvement patterns, this data will be machine learning-analyzed and safely kept in a single database. In addition to real-time monitoring capabilities enabling proactive management, interactive visualisation tools will aid stakeholders in understanding the data [66]. The ultimate objective is to improve user happiness by transforming the use of mobile apps into a proactive, transparent system [67]. This includes helping app developers better understand what aspects of their apps are most engaging or which segments of their user base are at risk of churning [69].

As more individuals rely on mobile devices for a variety of daily tasks, applications must be both appealing and responsive to user expectations [70]. Machine learning helps address this need by allowing developers to look beyond surface-level metrics and uncover deep insights that guide app refinement [71]. Whether it is adjusting notification timing, redesigning user interfaces, or customizing content delivery, the insights gained from machine learning can direct efforts toward user-focused innovations [72]. Moreover, it allows for predictive planning where app developers can act before engagement metrics decline [73]. The study employs Random Forest and K-Means Clustering to fulfill its dual purpose of classifying user behaviors and predicting future engagement trends [74]. Random Forest, an ensemble learning method, is effective in handling large datasets and capturing non-linear relationships among variables, making it suitable for classifying user interactions based on features such as session length, frequency of use, and demographic information [75]. K-Means Clustering, on the other hand, segments users into distinct groups based on similarities in behavior, helping marketers tailor their engagement strategies to each user segment [76].

By combining the strengths of both algorithms, this research can provide a comprehensive view of user engagement. For instance, clustering helps in identifying patterns among groups—like users who engage heavily during specific times or those who frequently abandon sessions after a few seconds—while classification predicts the engagement level of new or existing users [77]. This dual approach ensures a robust analysis, bridging exploratory and predictive analytics. Data collection will involve multiple stages: initial exploration through user interviews and app usage logs, followed by data cleaning and preprocessing to ensure model accuracy [78]. Features such as click-through rates, session intervals, screen time per visit, and feedback scores will form the core variables in our dataset [79]. These inputs will be normalized and fed into the machine learning models, which will be evaluated using metrics like accuracy, precision, recall, and F1-score [80]. Security and privacy are key concerns in the data collection phase, especially when dealing with user data [81]. The project adheres to standard data protection practices by anonymizing personal data, obtaining informed consent, and ensuring compliance with regulations such as GDPR [82]. These precautions ensure ethical research practices and build trust with end users [83].

Results from the study will be visualized using dashboards that show real-time user engagement levels, predicted churn rates, and suggested interventions [84]. Stakeholders can use these visualizations to make quick decisions, such as initiating re-engagement campaigns or modifying app functionalities based on predicted drops in usage [85]. These insights not only inform internal strategy but can also be used in external communications, demonstrating the app's responsiveness to user needs [86]. In the project aims to provide a scalable, data-driven solution to a universal challenge in mobile app development—sustaining and increasing user engagement. By integrating Random Forest and K-Means Clustering into a unified analytical framework, the research combines the power of predictive modeling and behavioral segmentation [87]. This integration allows for a deeper understanding of user dynamics and equips developers and marketers with actionable insights to create engaging, personalized, and user-centric mobile applications [88]. The outcome of this project is expected to contribute to more intuitive app designs, improved user retention strategies, and an overall enhancement of the mobile user experience.

Literature Review

The leveraged power of both randomized trials and machine learning analysis allowed a deep dive into user engagement dynamics. Through a meticulously controlled experiment, user engagement levels were assessed, while machine learning algorithms were applied to uncover intricate behavioral patterns [4]. Advanced statistical techniques supported the modeling and forecasting of engagement outcomes with precision. Although machine learning proved effective, the study's narrow focus limited broader insight into engagement dynamics [5]. Potential biases in the randomized trial may have skewed engagement predictions, and the absence of traditional method comparisons reduced the context for findings [6]. Expanding the scope to include diverse analytical approaches and mitigating experimental biases could improve the robustness of conclusions [7]. Future work should integrate traditional analysis and broader engagement metrics for a more holistic understanding of user interaction dynamics [3].

A combination of supervised and unsupervised learning methodologies was applied to analyze user engagement data. Unsupervised learning revealed hidden trends and user behavior patterns without needing labeled data, offering a deeper understanding of interaction dynamics [8]. Feature engineering techniques extracted essential insights, ensuring accurate modeling [9]. This dual strategy boosted both the interpretability and performance of the models. However, the challenges of unsupervised learning—such as opaque interpretability, unclear feature selection, and potential training data bias—pose risks to result reliability [10]. The absence of labeled outcomes makes result validation difficult. Transparency and clarity in feature extraction are vital to ensure trustworthy outcomes [11]. Mitigation strategies such as bias detection tools, rigorous evaluation methods, and clearly defined features are necessary to strengthen the reliability of unsupervised learning in user engagement prediction studies [2].

User engagement across various platforms like social media and news websites was analyzed using machine learning. Sentiment analysis assessed users' emotional responses to news content, while topic modeling identified trending subjects and their effects on engagement [12]. Algorithms were trained to predict engagement using features like platform type, sentiment polarity, and topical relevance. However, clarity was lacking regarding how sentiment and topic data were extracted and processed [13]. Furthermore, the integration of user behavior across platforms was underexplored, limiting insights into multi-platform engagement patterns. Temporal patterns in engagement dynamics were only lightly addressed, which may affect the validity of predictions [14]. Comprehensive analysis would benefit from detailed methodological explanations and better handling of cross-platform data integration, enabling more accurate engagement predictions and richer behavioral insights over time [1].

Social media engagement data was combined with polling data to predict election outcomes using machine learning. Sentiment analysis and trend tracking helped gauge public opinion from social platforms, while polling data provided complementary statistical grounding [15]. Algorithms analyzed combined datasets to increase the accuracy of electoral predictions. Feature selection was applied to highlight influential predictors [16]. Despite promising results, limitations include potential biases in social media data that could distort predictions. Real-world election dynamics influenced by non-social media factors were inadequately addressed, and there was limited comparison with actual election results [17]. Validation against final outcomes remains crucial to verify prediction reliability. Expanding the dataset to include socio-political indicators and improving bias mitigation in social media data can make these models more robust for election forecasting [18].

An adaptive ad sequencing algorithm was created to enhance user engagement through personalized delivery. Machine learning techniques were employed to customize ad orderings based on user preferences, while real-time feedback loops allowed the system to continuously improve its performance. Statistical evaluation confirmed the effectiveness of adaptive sequencing [21]. However, the adaptability of the algorithm was not deeply examined. The absence of comparative analysis with traditional ad sequencing methods limited the assessment of its relative benefits. Additionally, potential bias within user response data may impact the optimization process and lead to misrepresentations in ad targeting [22]. To ensure accurate personalization, it is crucial to enhance transparency in model behavior and incorporate robust validation measures. Future enhancements could focus on fairness, scalability, and long-term engagement outcomes [19].

Machine learning algorithms were applied to predict students' academic performance using various educational and demographic data points [23]. Data preprocessing techniques managed missing values and normalized features to ensure consistency. Feature selection strategies were employed to identify key factors influencing performance, such as attendance, prior grades, and participation [24]. Predictive models were trained using methods like decision trees and logistic regression and evaluated through cross-validation to ensure generalization. However, limited detail was provided on the criteria for feature selection, and the model training process lacked transparency [25]. Additionally, the impact of potential bias in educational data—such as grading inconsistencies or socio-economic influences—was not fully explored [26]. Comprehensive model validation, including bias analysis and ethical considerations, is essential to ensure reliable and equitable performance predictions in educational systems [20].

A three-tiered recommendation system was built for e-commerce platforms, incorporating collaborative filtering, content-based filtering, and hybrid approaches [27]. Collaborative filtering enabled personalized recommendations based on user-item interaction histories, while content-based filtering used item attributes to refine suggestions. The hybrid model combined both techniques to improve accuracy and user satisfaction [28]. However, the system offered limited insights into individual user preferences beyond basic interactions [29]. Issues related to scalability and efficiency, especially when handling large datasets, were not adequately addressed. Furthermore, no comparative analysis was conducted with existing recommendation algorithms to assess performance benefits [30]. Improving this system would involve enhancing personalization, increasing computational efficiency, and incorporating explainable AI techniques to provide users with transparency in recommendations, thereby building trust and optimizing the e-commerce user experience [31].

Recent advances in artificial intelligence and machine learning applications in business intelligence were explored by reviewing case studies and industry reports [33]. These sources highlighted innovative use cases and emerging challenges associated with AI integration into strategic business processes. Interviews or surveys with industry experts provided insights into real-world adoption strategies [34]. Theoretical frameworks and conceptual models were examined to understand AI's transformative impact. Nonetheless, these discussions lacked empirical validation and did not thoroughly address ethical concerns, such as algorithmic bias and data privacy. Moreover, the practical challenges associated with large-scale AI implementation—like system integration, cost, and skill gaps—were underexplored [35]. To advance the field, future studies should incorporate measurable results, real-world deployment data, and rigorous ethical evaluations, ensuring responsible and sustainable adoption of AI in business [32].

The impact of values in pandemic-themed advertising on consumer engagement was studied using qualitative methods, including interviews and focus groups [37]. Consumer responses and perceptions were gathered to assess how value-laden messages influenced behavior [38]. Content analysis was applied to identify common themes in pandemic-related advertisements. Statistical analysis then measured the correlation between conveyed values and consumer engagement metrics, such as click-through rates or brand loyalty. However, the study explored cultural value differences only minimally, reducing generalizability across diverse markets. Additionally, emotional factors driving engagement were not thoroughly discussed, and the strategic application of findings to future marketing efforts was insufficiently developed [39]. More inclusive research involving multicultural perspectives and emotion-driven analytics would enrich understanding and provide advertisers with actionable insights for future value-based campaigns [36].

Machine learning algorithms were employed to predict the satisfaction of Chinese psychotherapy clients using structured data extracted from counseling sessions [41]. Variables such as session length, frequency, therapist interaction quality, and emotional tone were analyzed using supervised learning models like logistic regression and decision trees [42]. These models provided valuable insights into client satisfaction trends and therapeutic effectiveness. However, the study lacked empirical validation using external datasets and did not adequately address ethical issues related to mental health data usage.

Additionally, real-world application challenges, such as patient privacy, therapist variability, and model deployment in clinical settings, were not sufficiently explored [43]. Future research should prioritize ethical compliance, include diverse patient profiles, and validate findings through long-term clinical implementation to ensure responsible and effective use of machine learning in psychotherapy [40].

2. Materials and Methods

The suggested system functions through a structured, multi-stage pipeline that begins with data collection and validation. Unprocessed data is compiled from a range of sources, including user interactions and app logs. To ensure reliable and efficient data retrieval, techniques such as web scraping and API queries are utilized. Maintaining the integrity and quality of the data is crucial; therefore, stringent validation and cleansing processes are applied to remove inconsistencies, errors, and noise that might compromise the reliability of subsequent analysis. Following data collection, exploratory data analysis (EDA) is conducted to gain insights into data distributions, correlations, and trends [89]. This phase uses tools like correlation matrices, scatter plots, and histograms to visually interpret patterns and uncover potential predictive features [90]. EDA helps in understanding the nature of the dataset and informs the identification of key variables that significantly influence user engagement [91].

Once potential predictors are identified, the system progresses to feature engineering and selection. This involves developing new features and refining existing ones to improve the model's predictive capacity [92]. Statistical tests and feature importance scores are used to assess which attributes contribute most to prediction accuracy. Techniques like recursive feature elimination are employed to reduce dimensionality and simplify the feature set, enhancing model performance and interpretability [93]. Model development and training form the core of the system. Machine learning models are constructed to forecast user interactions using ensemble methods that provide robust and diverse predictions [94]. Optimization algorithms are implemented during training to fine-tune parameters and minimize prediction errors [95]. The process is designed to maximize model generalizability and predictive power. Evaluation of model performance follows, using a variety of metrics such as accuracy and F1-score to assess effectiveness on unseen data [96]. Different models are compared rigorously, utilizing statistical tests and model selection criteria to identify the most effective one [97]. These comparisons help determine which models generalize best to new data and are most suitable for deployment.

Prior to deployment, data preprocessing and normalization are carried out to standardize input formats and prepare the data for analysis. This includes encoding categorical variables, scaling numerical features, and dividing the dataset into training and testing sets. Techniques like feature scaling and standardization are employed to improve model stability and convergence rates during training. Dimensionality reduction and clustering techniques are then used to distill high-dimensional data into manageable forms without losing essential information. Methods such as Principal Component Analysis (PCA) and t-SNE reduce feature dimensions, while clustering algorithms like K-means and DBSCAN group similar data points [98]. These methods help in understanding user segmentation and interaction patterns, offering deeper insights into engagement behaviors. Cross-validation and model fine-tuning are integral to ensuring model reliability and avoiding overfitting. Techniques such as k-fold or leave-one-out cross-validation are applied to test the model's ability to generalize. Optimization strategies, including gradient descent, are used to adjust hyperparameters based on validation data, and iterative training cycles refine the model's accuracy.

Ensemble learning is employed to further enhance prediction accuracy. By combining outputs from multiple models using methods like bagging and boosting, the system aims to leverage the strengths of individual models. These ensemble approaches are then compared with standalone models, and the best-performing model is selected based on evaluation metrics and alignment with business objectives. Sensitivity analysis is conducted to verify the chosen model's robustness and reliability under various scenarios. Deployment of the final model is executed using practical methods such as API integration or containerization, facilitating its integration into operational systems. Continuous monitoring mechanisms are established to track real-time performance and gather feedback, enabling iterative improvements. This ensures that the model remains effective as user behavior and app features evolve. Finally, the report is organized in a structured manner. The upcoming chapter will provide an in-depth literature review, examining previous work and foundational theories relevant to the study. Chapter 3 will then describe the project in comprehensive detail, outlining the methodology, tools, data, and processes involved in achieving the research objectives.

Though not without limitations, there are several techniques and approaches available for forecasting user engagement with mobile apps. Many app developers use basic descriptive analytics, such tracking session duration, the number of active users, and retention rates, to gain insight into user behavior. These metrics have limited predictive value and may not fully capture the underlying factors that drive user engagement, even while they provide important information about app usage patterns. A/B testing is a popular technique that compares two versions of an app or feature set to evaluate which one performs better in terms of user engagement metrics. Even though A/B testing can be useful in identifying effective design adjustments or enhancements, it is typically limited to evaluating predetermined hypotheses and may not capture advanced design revisions or feature additions. Cohort analysis examines users' interaction habits over a prolonged period of time by grouping them based on shared characteristics or behaviors. Understanding user retention and churn rates is one thing that cohort analysis might be good at, but it might not be able to predict future user behavior or identify the root causes of engagement. To predict user engagement, some app developers utilize heuristics or rule-based techniques; examples include setting thresholds for idleness or criteria for churning users. Although rule-based systems can be easily configured, their flexibility may be insufficient to accommodate the complicated behavior of users in changing settings.

By utilizing cutting-edge machine learning techniques and resolving important issues, the suggested system for forecasting user involvement in mobile apps seeks to overcome the drawbacks of current approaches. The suggested system is made up of multiple parts. The system will gather and incorporate data from a variety of sources about user interactions, app usage trends, demographics, and contextual factors. Data from user profiles, in-app surveys, usage logs from apps, and information from other sources like social media or location-based data may all be included in this. Selection and Feature Engineering: In order to identify pertinent predictors of user involvement and extract valuable features from the data gathered, the system will engage in feature engineering. This could entail choosing features with the highest predictive power for user engagement, finding pertinent patterns and trends, and converting raw data into actionable insights. Machine Learning Models: To create predictive models for user interaction, the system will make use of a range of machine learning algorithms, such as clustering, regression, and classification methods. These models will forecast user behavior and optimize app performance by utilizing the features that have been chosen. The predictive models will be evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score once the system has trained them on historical data. The cross-validation techniques will be employed to assess the robustness and generalization ability of the models. Real-Time Prediction and Recommendation: Based on user participation with mobile apps, real-time predictions and recommendations will be produced after the predictive models have been trained. Stakeholders and app developers will be able to optimize app features, designs, and marketing strategies based on expected user behavior with the system's assistance. Continuous Improvement and Feedback Loop: The system will have a feedback loop mechanism to collect user input and update the predictive models iteratively. Because of this, the system will be able to adapt.

Module Description

The process depicted in the image showcases the predictive power of machine learning in understanding user behavior. It begins with the preparation of a dataset, which undergoes preprocessing steps such as transformation and cleaning to ensure its quality. Next, machine learning algorithms are employed to build a predictive model using the training data. This model is then evaluated using separate test data to assess its accuracy and reliability. Once validated, the trained model can predict user engagement levels based on new input from users. These predictions empower businesses to make informed decisions and take appropriate actions to enhance user experiences and optimize outcomes. . It begins with the preparation of a dataset, which undergoes preprocessing steps such as transformation and cleaning to ensure its quality by leveraging machine learning in this manner, organizations can gain valuable insights into user behavior, enabling them to tailor their strategies effectively, see Figure 1.

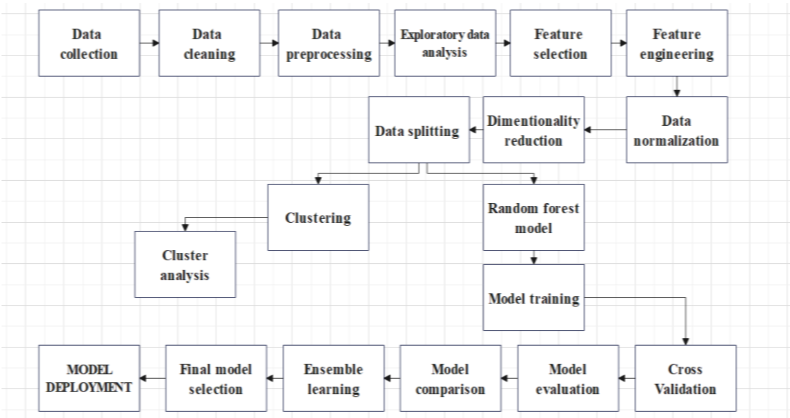

Figure 1. Data Flow Diagram

The pipeline for data analysis depicted in the image encompasses a comprehensive approach to ensure the quality and efficacy of the data-driven process. Beginning with data collection, it emphasizes the importance of obtaining reliable and relevant data. Subsequent steps focus on cleaning and preparing the data, essential for removing inconsistencies and ensuring consistency across the dataset. Investigative data analysis delves into the properties of the dataset, uncovering valuable insights that inform subsequent stages. Cluster analysis identifies underlying patterns within the data, while feature engineering and selection enhance the performance of models by refining the input variables. Techniques such as normalization and dimensionality reduction optimize the data for modeling, maximizing its utility and effectiveness. Data separation facilitates the training of models, predominantly employing random forest techniques to build robust predictive models. These models undergo rigorous comparison, assessment, and group selection to identify the most suitable candidate. Validation through cross-validation ensures the robustness and reliability of the chosen model. Investigative data analysis delves into the properties of the dataset. Ultimately, the validated model is implemented for practical use, empowering stakeholders to make well-informed decisions and take strategic actions based on reliable data insights, see Figure 2.

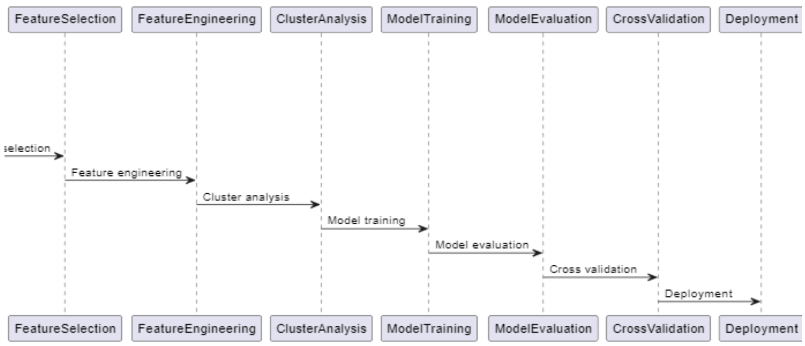

Figure 2. UML Diagram

The user engagement system process for a mobile application is shown in the UML diagram. Users engage with the mobile application, starting the flow of data through different phases. In order to guarantee data quality and comprehend its properties, the procedure starts with data cleaning, preprocessing, and exploratory data analysis. After feature engineering and selection, pertinent characteristics are refined for analysis, and cluster analysis is used to determine user subgroups. The model is then cross-validated after being trained on the chosen features and performance assessed. When the verified model is finally implemented, the application can interact with users in an efficient manner according to their interactions and attributes.

An illustration of the process of creating and assessing a machine learning model is provided by the activity diagram. The dataset must first be gathered, and then preprocessing must be done to prepare the data for analysis. Random selection is used to divide the dataset into training and testing sections. The training dataset is then used to apply machine learning techniques after features have been classified. Model accuracy is evaluated using performance metrics such as HHH. Feature extraction helps to further refine relevant portions of the data. Finally, the testing dataset assesses the model's effectiveness and capacity for generalization , see Figure 3.

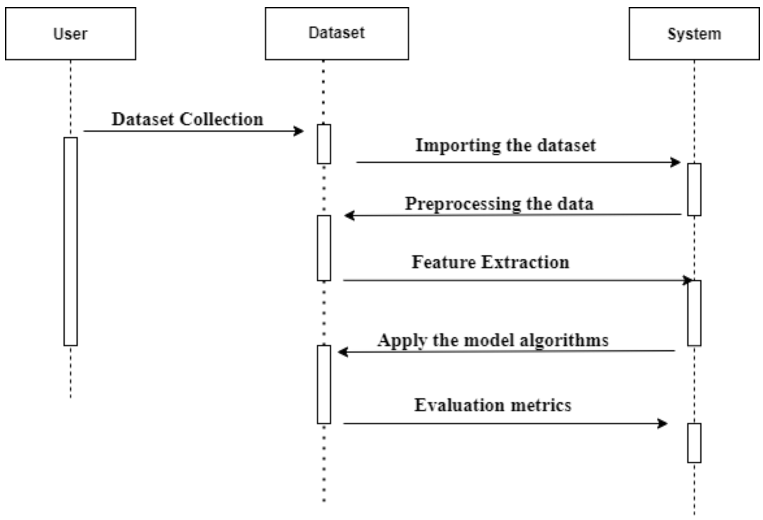

Figure 3. Sequence Diagram.

The sequence diagram shows the user's interaction with a system for data analysis and model application. The user initially compiles and imports the dataset into the system. To ensure the data's quality and suitability for analysis, preprocessing is applied. Using feature extraction, pertinent data components are refined for modeling. The user then processes the data and applies several model algorithms. Metrics for evaluation are used to assess how well applied models perform. In the end, the system offers feedback or results to the user based on the chosen algorithms and evaluation outcomes.

Module Description

Our entire project is divided into three modules. The first module, data preparation, begins with the collection and retrieval of relevant datasets using data mining algorithms. This process involves gathering app usage logs, user engagement metrics, demographic information, and feature engineering data. Once collected, the datasets undergo a thorough cleaning process to address missing values, outliers, and anomalies. This is accomplished using a combination of statistical techniques and domain expertise to ensure data quality and integrity. Following data cleaning, the next phase involves data transformation and feature engineering. During this stage, the raw data is converted into a structured format suitable for analysis. Meaningful features are extracted and engineered to enhance the performance of predictive models. Mathematical techniques, including feature scaling and normalization, are applied to standardize the data. Transformations such as logarithmic scaling, one-hot encoding, and polynomial feature generation may also be used to enrich the dataset and improve model interpretability and accuracy.

(x - min(x)) / (max(x) - min(x)) is the value of scaled.

Where:

a. The initial feature value was X.
b. Min_X denotes the feature’s minimal value.
c. Max_X denotes the feature’s maximum value

Scaled_X = (X - mu) / sigma

Where

a. The initial feature value was X.
b. The feature's mean is called mu.
c. sigma represents the feature's standard deviation

X_transformed = log(X + c)

Where

a. X. denotes the first feature value.
b. To prevent taking the logarithm of zero or negative values, particularly for variables with zero values, c is a constant, usually a modest positive number, that is added.

If lambda! = 0, Transformed = (X^lambda - 1) / lambda.

Log(X) = X_transformed, if lambda = 0.

Naive Bayes is a probabilistic machine learning technique primarily used for classification tasks. It is based on Bayes' theorem and assumes independence between features given the class label, which simplifies computation. Despite its simplicity, Naive Bayes often performs remarkably well, particularly in high-dimensional domains such as text classification. The first step in applying the Naive Bayes algorithm is data preparation, where the dataset is cleaned and preprocessed to ensure consistency and eliminate irrelevant information. This involves handling missing values, encoding categorical variables, and scaling numerical features to create a standardized dataset suitable for modeling. Following this, the feature selection process identifies and selects the most relevant attributes that influence classification outcomes. This step is crucial for reducing dimensionality and computational complexity, and techniques such as information gain or chi-square testing may be employed.

Once the features are selected, the next step is probability calculation. Here, the algorithm determines the prior probability of each class based on their frequency in the dataset. These prior probabilities represent the likelihood of each class occurring independently of the input features. The subsequent step involves conditional probability estimation using Bayes' theorem. Naive Bayes assumes that features are conditionally independent given the class, which enables the calculation of the probability of each feature value occurring within each class. These conditional probabilities are computed for every possible feature-class combination.

In the classification phase, the model uses Bayes' theorem to calculate the posterior probability of each class based on the feature values of a new instance. The class with the highest posterior probability is then assigned to that instance. Mathematically, the process begins with initialization, where the prior probabilities P(Ck)P(C_k)P(Ck) for each class CkC_kCk are established, along with the conditional probabilities P(Xi∣Ck)P(X_i∣C_k)P(Xi∣Ck) for each feature XiX_iXi. In the assignment step, these prior and conditional probabilities are computed for a new data instance with features x=(x1,x2,...,xn)x = (x_1, x_2, ..., x_n)x=(x1,x2,...,xn). The update step refines the probabilities using maximum likelihood estimation based on the observed frequencies of class labels and feature-value pairs in the training dataset.

This iterative process continues across the training data, updating the probabilities until convergence or a predefined stopping criterion is met. Finally, for any new instance, the model outputs the class with the highest posterior probability as the predicted classification. After training, the model applies Bayes' theorem to calculate these probabilities and make predictions, providing a robust and interpretable method for classification tasks, see Figure 4.

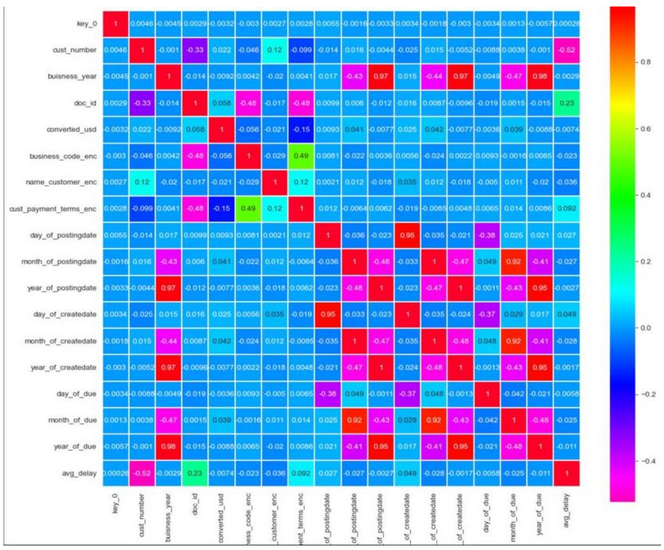

Figure 4. Heat map.

Testing procedures for sports data visualization are meticulously crafted to ensure that the software not only meets requirements, is error-free, and performs as intended but also delivers a superior product that is dependable, user-friendly, and friendly. Part of the plan involves selecting appropriate testing methods, including unit, functional, integration, performance, and user acceptability testing. The testing strategy also includes developing test cases and scenarios, carrying out regression testing, and collaborating with stakeholders. Delivering a superior product that meets end users' expectations and is dependable, user-friendly, and friendly is the ultimate goal of the testing approach. A well-planned testing strategy is required to guarantee that the sports data visualization software performs as planned.

3. Results

Consider Real-time processing enables the system to handle user data and generate predictions instantly, providing timely insights that are crucial for app optimization. By analyzing data as it is received, developers can make immediate adjustments to enhance user experience and app performance. The model demonstrates high predictive accuracy and reliability, which significantly supports informed decision-making for app developers and stakeholders. This strong performance ensures that the insights derived from the model are dependable and actionable. Adaptability is a core strength of the system, allowing it to respond effectively to changes in user preferences and app dynamics. This ensures that the model remains relevant and continues to deliver value as usage patterns and expectations evolve. Resource optimization is achieved through intelligent data analysis and predictive modeling, which allows for better allocation of resources. This reduces operational inefficiencies, maximizes utilization, and minimizes wastage, contributing to overall system efficiency. The system integrates user feedback seamlessly into its learning process, enabling iterative refinement of the predictive models. This feedback loop enhances both the accuracy and relevance of predictions, leading to improved engagement outcomes over time. The approach is cost-effective, providing a practical solution for predicting user engagement without incurring high development or operational costs. This ensures a greater return on investment for both app development and marketing strategies, see Table 1.

Table 1. Performance Metrics.

Metrics / Algorithms	Random forest & K-Means Clustering	Logistic Regression	SVM
Accuracy	0.90	0.83	0.85
Precision	0.88	0.77	0.84
Recall	0.93	0.84	0.82
F1-score	0.92	0.81	0.85

The proposed method combining Random Forest and K-Means clustering offers several advantages. Random Forest is effective in handling non-linear relationships, making it suitable for complex classification tasks. K-Means clustering complements this by providing insights into user segments, allowing for better understanding and targeting of different user groups. However, this approach also comes with limitations. Random Forest models typically require more computational resources during training when compared to simpler models like logistic regression. Additionally, K-Means clustering necessitates specifying the number of clusters (K) beforehand, which can be a non-trivial decision, especially in the absence of clear guidance from the data.

Logistic Regression, an existing method, has its own strengths. It performs well with high-dimensional data and is favored for its simplicity, ease of interpretation, and efficiency when dealing with small to medium-sized datasets. Despite these advantages, logistic regression has notable drawbacks. It is highly sensitive to outliers, and it relies on the assumption of a linear relationship between the target variable and predictor features. As a result, its performance may degrade when applied to data characterized by complex, non-linear patterns. Support Vector Machines (SVM), another existing technique, are particularly effective in high-dimensional spaces and can manage both linear and non-linear data. SVMs are also capable of defining non-linear decision boundaries using kernel functions, enhancing their predictive power in varied contexts. Nonetheless, SVMs are not without disadvantages. The models can be difficult to interpret, which limits their transparency and usability for decision-making. Moreover, training SVMs on large datasets can be computationally intensive, posing challenges in terms of scalability and resource consumption.

4. Discussion

User engagement prediction in mobile applications is vital for developers and marketers aiming to optimize strategies and enhance user experiences. The metrics from the Random Forest model highlight its accuracy and reliability in predicting user engagement, crucial for informed decision-making in app development and marketing. Additionally, metrics derived from K-Means clustering further validate the efficacy of this approach in segmenting users based on their engagement characteristics. This segmentation insight is invaluable for tailoring app features and content to specific user segments, thereby enhancing overall user satisfaction and retention. Combining insights from both Random Forest and K-Means clustering yields a comprehensive understanding of user engagement dynamics, leading to a refined model that enhances predictive accuracy and interpretability. Comparative analysis underscores the complementary strengths of Random Forest and K-Means clustering, with Random Forest capturing nonlinear relationships and complex patterns, while K-Means efficiently identifies distinct user groups based on their similarity in engagement behavior.

The synergistic use of these algorithms not only improves predictive accuracy but also enhances the interpretability of the model, enabling developers and marketers to gain deeper insights into user behavior and tailor app strategies accordingly. In conclusion, the integration of Random Forest and K-Means clustering represents a potent approach for app optimization, offering compelling results in predicting user engagement within mobile applications. Opportunities for further research and improvement include exploring advanced machine learning algorithms, leveraging expansive datasets, and investigating the impact of external factors such as app updates, marketing campaigns, and user demographics on user engagement. Overall, the amalgamation of Random Forest and K-Means clustering models has yielded compelling results in predicting user engagement within mobile applications. This integrated approach not only improves predictive accuracy but also enhances the understanding of user behavior, paving the way for more effective app optimization strategies.

5. Conclusion

In the proposed strategy for forecasting mobile app user engagement advances app development and optimisation. The system provides app developers with a complete solution to improve user experience, app success, and app performance using cutting-edge machine learning algorithms, data-driven decision making, and ethics. App developers may adapt to changing user needs by accurately projecting engagement metrics like session duration, app usage frequency, and retention rates. A multi-core CPU (Inter Core i5), 4GB GPU (Nvidia GTX 1650), 1TB HDD, and Python for data analysis, Jupyter Notebook, Tensorflow, and Scikitlearn are utilised. The technology also helps app developers choose features, advertising, and app design, improving app performance and giving them a competitive edge in the mobile app market. Flexible developers that pay attention to user input and market changes may keep their programs competitive and relevant. The system prioritises ethics and privacy to comply with data protection laws and maintain user trust and loyalty. App developers who follow ethical guidelines and handle data properly can generate long-term customer relationships and a good industry reputation. Overall, the suggested methodology addresses the challenges of forecasting user engagement while supporting moral and user-centred design principles to create and improve apps. Data-driven decision making and advanced analytics can help developers succeed and innovate in the fast-paced, competitive mobile app development industry.

REFERENCES

[1] A. Srivastava, “Use of Python in Data Science, Data Integration and Data Engineer,” Int. J. Sci. Res. Eng. Manag., vol. 8, no. 7, 2024.

[2] G. Kashyap, "Explainable AI (XAI): Methods and techniques to make deep learning models more interpretable and their real-world implications," Int. J. Innov. Res. Eng. Multidisciplinary Phys. Sci., vol. 13, no. 1, pp. 1–8, Jan. 2025.

[3] G. Kashyap, "Knowledge transferring techniques: AI in learning transfer," Science Data & Learning Machine Intelligence Artificial Journal, vol. 4, no. 1, pp. 1898–1902, Dec. 2023.

[4] B. Chandrashekar, S. Boggavarapu, S. Pundir, C. Yosepu, S. Chepyala and G. Manikandan, "Machine Learning Prediction Approach For Financial Forecast System In Stock Exchange Marketing Management," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1433–1438.

[5] S. Boggavarapu, G. Ramkumar, P. R. Gedamkar, A. Kaneria, S. Pundir and R. Selvameena, "Research on Unmanned Artificial Intelligence Based Financial Volatility Prediction in International Stock Market," 2024 5th International Conference on Recent Trends in Computer Science and Technology, Jamshedpur, India, 2024, pp. 16–20.

[6] M. Chugh and S. Vyas, “Digital Twin: A Promising New Era for Elderly Healthcare,” in Digital Twins for Sustainable Healthcare in the Metaverse, IGI Global Scientific Publishing, 2025, pp. 55–76.

[7] S. Vyas, M. Chugh, and S. Gupta, “Comprehensive study of digital forensics in contemporary domain,” in AIP Conference Proceedings, vol. 2916, no. 1, AIP Publishing, 2023.

[8] S. Gupta and S. Vyas, “Contemporary role of edge-AI in IoT and IoE in healthcare and digital marketing,” in Edge-AI in Healthcare, CRC Press, 2023, pp. 75–84.

[9] S. Gupta, M. Chugh, and S. Vyas, “Understanding immersive technologies for autism detection: A study,” Automation and Computation, pp. 364–370, 2023.

[10] Gokulakrishnan S., P. Chakrabarti, B. T. Hung, et al., “An optimized facial recognition model for identifying criminal activities using deep learning strategy,” Int. J. Inf. Technol., vol. 15, pp. 3907–3921, 2023.

[11] S. N. V. J. D. Kosuru, R. Praveen Kumar, R. Kumar, S. Gokulakrishnan, A. Sethy, and V. Sreenivas, “A comparative assessment for examining the performance of reconfigurable multiband MIMO antennas for communication systems,” J. Inf. Syst. Eng. Manag., vol. 10, no. 10s, 2025.

[12] R. Kumar, S. Gokulakrishnan, S. N. V. J. D. Kosuru, R. Praveen Kumar, and R. T. Radha, “An efficient fuzzy logic and artificial intelligence based optimization strategy for big data healthcare system,” Edelweiss Appl. Sci. Technol., vol. 9, no. 3, 2025.

[13] S. Gokulakrishnan and J. M. Gnanasekar, “Efficient and privacy for data integrity and data replication in cloud computing,” Int. J. Innov. Technol. Explor. Eng., 2019.

[14] S. Gokulakrishnan and J. M. Gnanasekar, “Data integrity and recovery management in cloud systems,” in Proc. 4th Int. Conf. Inventive Syst. Control (ICISC), Coimbatore, India, 2020, pp. 645–648.

[15] A. K. Tyagi and S. R. Addula, “Artificial intelligence for malware analysis: A systematic study,” Artificial Intelligence-Enabled Digital Twin for Smart Manufacturing. Wiley, pp. 359–390, 15-Oct-2024.

[16] G. S. Sajja and S. Reddy Addula, "Automation Using Robots, Machine Learning, and Artificial Intelligence to Enhance Production and Quality," 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), Lucknow, India, 2024, pp. 1-4.

[17] N. Nasib et al., “Systematic Analysis based on Conflux of Machine Learning and Internet of Things using Bibliometric analysis,” JISIoT, vol. 13, no. 1, pp. 196–224, 2024.

[18] S. Almotairi et al., “Personal data protection model in IOMT-blockchain on secured bit-Count Transmutation data encryption approach,” Fusion: Practice and Applications, vol. 16, no. 1, pp. 152–170, 2024.

[19] S. Menon et al., “Streamlining task planning systems for improved enactment in contemporary computing surroundings,” SN Comput. Sci., vol. 5, no. 8, 2024.

[20] S. S. Nair, G. Lakshmikanthan, N. Belagalla, S. Belagalla, S. K. Ahmad and S. A. Farooqi, "Leveraging AI and Machine Learning for Enhanced Fraud Detection in Digital Banking System: A Comparative Study," 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT), Bhimtal, Nainital, India, 2025, pp. 1278-1282.

[21] S. S. Nair, G. Lakshmikanthan, J.ParthaSarathy, D. P. S, K. Shanmugakani and B.Jegajothi, "Enhancing Cloud Security with Machine Learning: Tackling Data Breaches and Insider Threats," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 912-917.

[22] G. Lakshmikanthan, S. S. Nair, J. Partha Sarathy, S. Singh, S. Santiago and B. Jegajothi, "Mitigating IoT Botnet Attacks: Machine Learning Techniques for Securing Connected Devices," 2024 International Conference on Emerging Research in Computational Science, Coimbatore, India, 2024, pp. 1-6.

[23] G. Lakshmikanthan and S. S. Nair, "Zero trust architecture: Redefining security parameters for remote-first organizations," International Research Journal of Modernization in Engineering Technology and Science, vol. 2, no. 3, pp. 1003–1013, 2020.

[24] S. Sreekandan Nair and G. Lakshmikanthan, "Open Source Security: Managing Risk in the Wake of Log4j Vulnerability", International Journal of Emerging Trends in Computer Science and Information Technology, vol. 2, no. 4, pp. 33–45, Nov. 2021.

[25] T. S. Chu, S. S. Nair, and G. Lakshmikanthan, "Network intrusion detection using advanced AI models: A comparative study of machine learning and deep learning approaches," Int. J. Commun. Netw. Inf. Secur., vol. 14, no. 2, pp. 359–365, Aug. 2022.

[26] R. Aravindhan, R. Shanmugalakshmi, K. Ramya and Selvan C., "Certain investigation on web application security: Phishing detection and phishing target discovery," 2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2016, pp. 1-10.

[27] R. Aravindhan and R. Shanmugalakshmi, "Comparative analysis of Web 3.0 search engines: A survey report," 2013 International Conference on Advanced Computing and Communication Systems, Coimbatore, India, 2013, pp. 1-6.

[28] R. Vadisetty, "Efficient Large-Scale Data based on Cloud Framework using Critical Influences on Financial Landscape," in 2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC), 2024, pp. 1–6.

[29] R. Vadisetty and A. Polamarasetti, "Generative AI for Cyber Threat Simulation and Defense," in 2024 12th International Conference on Control, Mechatronics and Automation (ICCMA), 2024, pp. 272–279.

[30] R. Vadisetty and A. Polamarasetti, "Gen AI for Real- Time Traffic Prediction and Autoscaling in Cloud Computing Education 4.0," in 2024 13th International Conference on System Modeling & Advancement in Research Trends (SMART), 2024, pp. 735–741.

[31] R. Vadisetty and A. Polamarasetti, "Quantum Computing For Cryptographic Security With Artificial Intelligence," in 2024 12th International Conference on Control, Mechatronics and Automation (ICCMA), 2024, pp. 252–260.

[32] R. Vadisetty and A. Polamarasetti, "AI-Augmented Skill Development Roadmaps: Tailoring 12-Month Learning Paths for Future-Ready Careers in Education 4.0 and Industry 4.0," in 2024 13th International Conference on System Modeling & Advancement in Research Trends (SMART), 2024, pp. 655–661.

[33] U. K. Lilhore, V. Dutt, T. A. Kumar, M. Margala, and K. Raahemifar, Math Optimization for Artificial Intelligence: Heuristic and Metaheuristic Methods for Robotics and Machine Learning. De Gruyter, 2025.

[34] M. S. Dr. Dilipkumar A. Ode, J. D. K. Dr. Krishnendu Roy, M. M. C. Dr. Birajlakshmi Ghosh, and R. S. Dr. Amar Baliram Abhrange, AI & Chatgpt Tools For Teaching Learning Process. Redshine Publication, 2024.

[35] R. Vadishetty, "Efficient Deep Fake Detection Technique on Video and Audio Dataset Using Deep Learning BT - Proceedings of 5th International Ethical Hacking Conference," 2025, pp. 137–155.

[36] R. Aravindhan, R. Shanmugalakshmi, and K. Ramya, "Circumvention of nascent and potential Wi-Fi phishing threat using association rule mining," Wireless Pers. Commun., vol. 94, no. 3, pp. 2331–2361, 2017.

[37] R. Aravindhan and R. Shanmugalakshmi, "Visual analytics for semantic based image retrieval (SBIR): semantic tool," International Journal of Latest Trends in Engineering and Technology, vol. 7, no. 2, pp. 300–312, 2016.

[38] R. Aravindhan and R. Shanmugalakshmi, "Multistage fuzzy classifier based phishing detection using LDA and CRF features followed by impersonated entity discovery," International Journal of Control Theory and Applications, vol. 10, no. 29, pp. 33–42, 2017.

[39] C. Selvan, A. Ragunathan, and U. M. Ashwinkumar, "Mitigating phishing threats in unmanned aircraft systems (UAS) through multi-stage defense strategies," in Analyzing and Mitigating Security Risks in Cloud Computing, IGI Global, 2024, pp. 125–162.

[40] S. R. Addula and A. K. Tyagi, “Future of computer vision and industrial robotics in smart manufacturing,” Artificial Intelligence-Enabled Digital Twin for Smart Manufacturing. Wiley, pp. 505–539, 15-Oct-2024.
[41] S. R. Addula and G. Sekhar Sajja, "Automated Machine Learning to Streamline Data-Driven Industrial Application Development," 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), Lucknow, India, 2024, pp. 1-4.
[42] S. R. Addula, “Analysis of Perceived Ease of Use and Security on the Mobile Banking Adoption,” University of the Cumberlands, Williamsburg, Kentucky, United States of America, 2024.
[43] S. R. Addula, A. K. Tyagi, K. Naithani, and S. Kumari, “Blockchain-empowered internet of things (IoTs) platforms for automation in various sectors,” Artificial Intelligence-Enabled Digital Twin for Smart Manufacturing. Wiley, pp. 443–477, 15-Oct-2024.
[44] S. R. Addula, K. Meduri, G. S. Nadella, and H. Gonaygunta, “AI and blockchain in finance: Opportunities and challenges for the banking sector,” Nternational J. Adv. Res. Comput. Commun. Eng., vol. 13, no. 2, 2024.
[45] Z. H. Jaber, M. Ihsan, S. Gokulakrishnan, H. A. Alshaibani, F. H. Alsalamy, and H. Al-Aboudy, “Distributed self-localization with improved optimization with machine learning in IoT applications,” in Proc. Asian Conf. Commun. Netw. (ASIANComNet), 2024.
[46] G. Parasa, D. K. Nayak, A. Gangopadhyay, M. Stanlywit, S. Gokulakrishnan, and P. Sharma, “Exploring the role of artificial intelligence in enhancing chatbot functionality,” in Proc. Int. Conf. Adv. Comput., Commun. Netw. (ICAC2N), 2024.
[47] M. A. Mookambal and S. Gokulakrishnan, “Potential subscriber detection using machine learning,” in Proc. ICIPCN 2020, J. I. Z. Chen, J. M. R. S. Tavares, S. Shakya, and A. M. Iliyasu, Eds., Adv. Intell. Syst. Comput., vol. 1200, Springer, Cham, 2021, pp. 425–435.
[48] M. Chugh and S. Vyas, “Exploring the incredible potential and opportunity of the metaverse world,” in Strategies and Opportunities for Technology in the Metaverse World, IGI Global, 2023, pp. 34–47.
[49] S. Gupta, M. Chugh, and S. Vyas, “A Study on AI-Empowered Smart Healthcare: Key Challenges and Opportunities,” in Proc. Int. Conf. Smart Systems and Advanced Computing, Cham: Springer Nature Switzerland, 2022, pp. 255–266.
[50] S. Vyas, S. Gupta, and M. Chugh, “Metaverse Technologies and Applications: A Study,” in Proc. Int. Conf. Cybersecurity in Emerging Digital Era, Singapore: Springer Nature Singapore, 2022, pp. 287–300.
[51] M. Chugh, A. Pandey, and S. Vyas, “A Comprehensive Study on the Association Between Personality Traits and Software Development,” in Proc. 4th Int. Conf. Information Management & Machine Intelligence, 2022, pp. 1–6.
[52] M. Chugh, S. Gupta, and S. Vyas, “Leveraging the Potentiality of Blockchain Technology for Waste Management in Smart City Development,” in Proc. Int. Conf. Information and Management Engineering, Singapore: Springer Nature Singapore, 2022, pp. 377–387.
[53] S. Gupta, M. Chugh, and S. Vyas, “A Study on AI-Empowered Smart Healthcare: Key Challenges and Opportunities,” in Proc. Int. Conf. Smart Systems and Advanced Computing, Cham: Springer Nature Switzerland, 2022, pp. 255–266.
[54] S. Boggavarapu, G. S. Navale, G. Manikandan, N. Senthamilarasi, K. Lakshminarayana and H. R. Goyal, "A Novel Intelligent AI with Automated Defense Attack Data Privacy System Design," 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), Jamshedpur, India, 2024, pp. 99–104.
[55] S. Boggavarapu, S. S. Ali, G. Manikandan, R. Mohanraj, D. P. Singh and R. R, "Flying Neural Network-Based Optimistic Financial Early Alert System in AI Model," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1367–1373.
[56] F. Ahamed, S. Biswal, S. Sekhar Nanda, S. Pundir, T. Soubhari and S. Boggavarapu, "Intelligent Unmanned AI Detection Model for Financial Volatility in Stock Exchange," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1422–1426.
[57] G. Kashyap, "AI in blockchain-enabled healthcare systems: Transforming healthcare delivery and data security," ESP Journal of Engineering & Technology Advancements, vol. 3, no. 8, pp. 109, 2023.
[58] G. Kashyap, "Knowledge graphs and NLP: Integrating structured knowledge into NLP systems for better reasoning and contextual understanding," Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci., vol. 13, no. 2, pp. 1–7, Mar.–Apr. 2025.

[59] G. Kashyap, "Self-supervised learning: How self-supervised learning approaches can reduce dependence on labeled data," Int. J. Innov. Res. Creat. Technol., vol. 10, no. 4, pp. 1–10, Jul. 2024.
[60] A. Srivastava, “AI in Healthcare and its Future,” J. Artif. Intell. Cloud Comput., vol. 1, no. 1, pp. 1–2, Mar. 2022.
[61] A. Srivastava, “Cloud Replacing Traditional Database,” Int. J. Multidiscip. Res., vol. 7, no. 2, pp. 1–2, Mar.–Apr. 2025.
[62] B. Qian, N. Al Said, and B. Dong, “New technologies for UAV navigation with real time pattern recognition,” Ain Shams Engineering Journal, vol. 15, no. 3, p. 102480, 2024.
[63] B. Zakarneh, N. Annamalai, N. Al Said, and F. Aljabr, “Revolutionizing language learning through ChatGPT: An analysis of English language learners,” International Journal of English Language and Literature Studies, vol. 14, no. 1, pp. 1–16, 2025.
[64] D. Gura, B. Dong, D. Mehiar, and N. Al Said, “Customized convolutional neural network for accurate detection of deep fake images in video collections,” Computers, Materials & Continua, vol. 79, no. 2, 2024.
[65] I. A. Mohammed, R. Sofia, G. V. Radhakrishnan, S. Jha, and N. Al Said, “The Role of Artificial Intelligence in Enhancing Business Efficiency and Supply Chain Management,” 2024.
[66] K. M. Al Said, N. Al Said, and E. Hattab, “Impact of Social Networks in Educational Media,” Journal of Information and Communication Convergence Engineering, vol. 18, no. 4, pp. 230–238, 2020.
[67] N. Al Said and B. Z. Al Rawashdeh, “Information and computer technologies in media specialist preparation,” Information Development, vol. 38, no. 3, pp. 380–390, 2022.
[68] N. Al Said and K. Al Said, “Assessment of acceptance and user experience of human computer interaction with a computer interface,” International Association of Online Engineering, 2020.
[69] N. Al Said and K. M. Al Said, “The effect of visual and informational complexity of news website designs on comprehension and memorization among undergraduate students,” AI & Society, pp. 1–9, 2022.
[70] N. Al Said and Y. Gorbachev, “An Unmanned Aerial Vehicles Navigation System on the Basis of Pattern Recognition Applications,” Journal of Southwest Jiaotong University, vol. 55, no. 3, 2020.
[71] N. Al Said, “Artificial Intelligence in Banking: Enhancing Customer Experience and Operational Efficiency,” in Contemporary Challenges in Multidisciplinary Research: A Collaborative Approach, Volume 1, Jan. 21, 2025.
[72] N. Al Said, “Data Mining for Managing and Using Online Information on Facebook,” Journal of Advances in Information Technology, vol. 14, no. 4, pp. 769–776, 2023.
[73] N. Al Said, “Human-AI Collaboration in Business: Opportunities and Challenges,” in Contemporary Challenges in Multidisciplinary Research: A Collaborative Approach, Volume 1, Jan. 21, 2025.
[74] N. Al Said, “Mobile application development for technology enhanced learning: An applied study on the students of the college of mass communication at Ajman University,” International Journal of Emerging Technologies in Learning (IJET), vol. 15, no. 8, pp. 57–70, 2020.
[75] N. Al Said, “The Ethical Implications of AI in Workplace Automation,” in Contemporary Challenges in Multidisciplinary Research: A Collaborative Approach, Volume 1, Jan. 21, 2025.
[76] N. Al Said, “The Future of Work: Economic Implications of Automation and AI,” in Contemporary Challenges in Multidisciplinary Research: A Collaborative Approach, Volume 1, Jan. 21, 2025.
[77] N. Al Said, “The Impact of Computer Information Systems on the Adequacy of the Institutional Work (Case Study),” International Journal of Computer Science Engineering and Information, 2019.
[78] N. Al Said, D. Gura, and D. Karlov, “Efficiency of smart ai based voice apps and virtual services operating with chatbots,” Mendel, vol. 28, no. 2, pp. 9–16, 2022.
[79] N. Al Said, L. Vorona Slivinskaya, and E. Gorozhanina, “Data mining in education: Managing digital content with social media analytics in medical education,” Interactive Learning Environments, vol. 32, no. 8, pp. 3983–3995, 2024.
[80] O. Akylbekov, N. Al Said, R. Martínez García, and D. Gura, “ML models and neural networks for analyzing 3D data spatial planning tasks: Example of Khasansky urban district of the Russian Federation,” Advances in Engineering Software, vol. 173, p. 103251, 2022.
[81] R. Ivanova, N. Gaifullina, and N. Al Said, “The role of social networks in the development of skills of professional communication: An empirical study,” International Journal of Web Based Learning and Teaching Technologies, 2022.

[82] S. R. ElDadah and N. Al Said, “A process based framework for automatic categorization of web documents,” in 13th Workshop for PhD Students in Object Oriented Programming: Summary and …, 2003.

[83] S. Zhang, N. Gavrilovskaya, N. Al Said, and W. S. Afandi, “A new approach to snow avalanche rescue using UAV pictures based on convolutional neural networks,” Journal of Real Time Image Processing, vol. 20, no. 4, p. 65, 2023.

[84] T. Argyros, C. Ermopoulos, V. Pavlaki, and N. Al Said, “Extracting cyber communities through patterns,” in Proc. 2003 SIAM International Conference on Data Mining, 2003, pp. 259–263.

[85] Z. Wang and N. Al Said, “Analog Computing and a Hybrid Approach to the Element Base of Artificial Intelligence Applications,” International Review of Automatic Control (IREACO), vol. 13, no. 5, pp. 206–213.

[86] A. R. Yeruva, P. Choudhari, A. Shrivastava, D. Verma, S. Shaw, and A. Rana, "Covid-19 Disease Detection using Chest X-Ray Images by Means of CNN," in Proceedings of the 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022.

[87] A. R. Yeruva, R. S. Jadhav, R. Roopa, S. Preetha, R. Priya, K. N. Mishra, and K. Loganathan, "Advancing Health Monitoring with Cognitive IoT, Rapid Machine Learning, and Mechanical Systems," in Advances in Mechatronics and Mechanical Engineering, IGI Global, Pennsylvania, United States of America, pp. 301–312, 2024.

[88] S. M. Abdulrahman, R. R. Asaad, H. B. Ahmad, A. A. Hani, S. R. Zeebaree, and A. B. Sallow, “Machine learning in nonlinear material physics,” J. Soft Comput. Data Min., vol. 5, no. 1, pp. 122–131, 2024.

[89] S. M. Abdulrahman, A. A. Hani, S. R. Zeebaree, R. R. Asaad, D. A. Majeed, A. B. Sallow, and H. B. Ahmad, “Intelligent home IoT devices: An exploration of machine learning-based networked traffic investigation,” J. Ilm. Ilmu Terapan Univ. Jambi, vol. 8, no. 1, pp. 1–10, 2024.

[90] H. B. Ahmad, R. R. Asaad, S. M. Almufti, A. A. Hani, A. B. Sallow, and S. R. Zeebaree, “Smart home energy saving with big data and machine learning,” J. Ilm. Ilmu Terapan Univ. Jambi, vol. 8, no. 1, pp. 11–20, 2024.

[91] S. M. Almufti, H. B. Ahmad, R. B. Marqas, and R. R. Asaad, “Grey wolf optimizer: Overview, modifications and applications,” Int. Res. J. Sci. Technol. Educ. Manag., vol. 1, no. 1, p. 1, 2021.

[92] S. Almufti, R. Asaad, and B. Salim, “Review on elephant herding optimization algorithm performance in solving optimization problems,” Int. J. Eng. Technol., vol. 7, no. 1, pp. 6109–6114, 2018.

[93] S. Almufti, R. Marqas, and R. Asaad, “Comparative study between elephant herding optimization (EHO) and U-turning ant colony optimization (U-TACO) in solving symmetric traveling salesman problem (STSP),” J. Adv. Comput. Sci. Technol., vol. 8, no. 2, p. 32, 2019.

[94] B. Sallow, R. R. Asaad, H. B. Ahmad, S. M. Abdulrahman, A. A. Hani, and S. R. M. Zeebaree, “Machine learning skills to K–12,” J. Soft Comput. Data Min., vol. 5, no. 1, pp. 132–141, 2024.

[95] R. R. Asaad, S. M. Abdulrahman, and A. A. Hani, “Advanced encryption standard enhancement with output feedback block mode operation,” Acad. J. Nawroz Univ., vol. 6, no. 3, pp. 1–10, 2017.

[96] R. R. Asaad, S. M. Abdurahman, and A. A. Hani, “Partial image encryption using RC4 stream cipher approach and embedded in an image,” Acad. J. Nawroz Univ., vol. 6, no. 3, pp. 40–45, 2017.

[97] A. Hani, A. B. Sallow, H. B. Ahmad, S. M. Abdulrahman, R. R. Asaad, S. R. M. Zeebaree, and D. A. Majeed, “Comparative analysis of state-of-the-art classifiers for Parkinson’s disease diagnosis,” J. Ilm. Ilmu Terapan Univ. Jambi, vol. 8, no. 2, pp. 409–423, 2024.

[98] R. R. Ihsan, S. M. Almufti, B. M. Ormani, R. R. Asaad, and R. B. Marqas, “A survey on cat swarm optimization algorithm,” Asian J. Res. Comput. Sci., vol. 10, no. 2, pp. 22–32, 2021.