Artificial Intelligence and Machine Learning-Based Predictive Analytics for Customer Behaviour and Business Performance

Namrata Pawar
Research Scholar, Department of Management,
MGV's Samajshree Prashantdada Hiray College of Management &
Research, Nashik affiliated to Savitrabai Phule Pune University, Pune

Dr. Ganesh Ramesh Teltumbade. Dy.
Director & Associates Professor,
Department of Management,
MGV's Samajshree Prashantdada Hiray College of Management &
Research, Nashik affiliated to
Savitrabai Phule Pune University, Pune

Hemant Bapu Jagdale
Assistant Professor, Department of Management,
MGV's Samajshree Prashantdada Hiray College of Management & Research,
Nashik affiliated to Savitrabai Phule Pune University, Pune

Dr. Kiran Dnyaneshwar Gonte
Assistant Professor, Department of Management,
Shikshan Prasarkar Sanstha's, MBA Institute, Sangamner

Abstract
The swift progression of digital technology has allowed organisations to amass extensive customer data, fostering new avenues for data-informed decision-making. Artificial Intelligence (AI) and Machine Learning (ML) have surfaced as potent methodologies for examining consumer behaviour, forecasting forthcoming trends, and enhancing organisational performance via predictive analytics. Nonetheless, numerous organisations continue to have difficulties in converting intricate and high-dimensional data into practical insights through traditional analytical methods. This research introduces a predictive analytics framework utilising AI and ML for the analysis of customer behaviour and the assessment of business performance. The structure encompasses data gathering, preprocessing, feature extraction, model creation, validation, and performance evaluation. A variety of machine learning algorithms, such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks, are utilised to assess predicted performance. Model efficacy is evaluated by metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix assessment, whilst feature significance analysis determines the elements that substantially affect consumer behaviour. The suggested approach illustrates the capability of AI-enhanced predictive models to enhance consumer segmentation, forecast behaviour, and inform critical business decisions. Ensemble learning and neural network architectures provide robust forecasting capabilities by adeptly identifying intricate linkages in consumer data. The results underscore the significance of consumer demographics, buying history, interaction trends, and transaction frequency in improving predictive precision and business results.This study presents a cohesive predictive analytics architecture that facilitates astute decision-making and enhances corporate efficiency. The suggested methodology provides actionable insights for entities aiming to utilise AI and ML for customer analysis, operational productivity, client retention, and sustainable business expansion.

Keywords: AI, ML, Forecasting Analytics, Consumer Behaviour, Corporate Performance, Business Insights, Client Segmentation, XGBoost, Neural Networks, Decision-Making Systems.

1.Introduction

1.1Background of AI and ML in Business Analytics

The rising utilisation of digital platforms, e-commerce, mobile applications, and social media has led to an unparalleled expansion of customer-related data. Entities are utilising this information to acquire more profound understanding of consumer inclinations, buying patterns, and market dynamics. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as vital elements of contemporary corporate analytics, facilitating automated data processing, pattern identification, and anticipatory decision-making. In contrast to conventional statistical techniques, AI and ML algorithms possess the capability to scrutinise intricate, high-dimensional information, uncover concealed correlations, and perpetually enhance predictive precision when fresh data is introduced. As a result, companies are progressively using AI-powered insights into marketing, customer relationship management, financial strategy, and operational decision-making.

1.2.Importance of Customer Behaviour Prediction

Comprehending consumer behaviour is essential for formulating successful business strategies and sustaining a competitive edge. Anticipating consumer behaviours, including buying choices, product inclinations, attrition probabilities, and engagement metrics, allows businesses to tailor services, enhance marketing strategies, and elevate customer pleasure. Precise behavioural forecasting aids in inventory management, pricing tactics, and demand estimation, hence minimising operating expenses and enhancing revenue prospects. As consumer expectations progress, predictive analytics has emerged as a crucial instrument for providing tailored experiences and enhancing enduring client connections.

1.3.Business Performance Challenges

Notwithstanding substantial expenditures in data acquisition technologies, numerous organisations have challenges in transforming extensive and varied information into valuable business insights. Customer data is frequently dispersed across various platforms, resulting in difficulties concerning data integration, quality, and uniformity. Conventional analytical methods often fail to identify nonlinear dynamics and swiftly evolving consumer preferences, leading to diminished predictive accuracy. Moreover, escalating market rivalry, evolving consumer demands, and the want for instantaneous decision-making compel organisations to implement sophisticated analytical methodologies that yield prompt and precise business intelligence.

1.4.Research Motivation

Recent developments in artificial intelligence and machine learning have generated novel prospects for enhancing predictive analytics inside corporate settings. Despite the plethora of predictive models introduced, the identification of appropriate algorithms for forecasting customer behaviour and assessing their efficacy in various business contexts continues to pose significant research hurdles. There is an increasing want for cohesive frameworks that amalgamate thorough data preprocessing, feature engineering, various machine learning models, and extensive performance assessment. This research is driven by the necessity to provide a scalable and dependable predictive analytics platform that may facilitate informed business choices and improve organisational efficacy.

1.5.Objectives and Contributions

The main aim of this study is to create a predictive analytics framework utilising AI and ML to assess customer behaviour and enhance corporate performance. The research specifically seeks to pinpoint significant customer-related attributes, apply and contrast various machine learning techniques, assess their predicted efficacy through established criteria, and examine the commercial ramifications of the prediction results. The primary contributions encompass the creation of a cohesive analytical framework, a comparative assessment of prevalent machine learning methodologies, the recognition of critical behavioural factors influencing business performance, and actionable suggestions for the adoption of AI-driven predictive analytics in corporate entities.

1.6.Paper Organization

The subsequent sections of this document are structured as outlined below. Section 2 examines contemporary literature pertaining to artificial intelligence, machine learning, predictive analytics, and consumer behaviour modelling. Section 3 delineates the suggested predictive analytics framework and study approach. Section 4 delineates the experimental configuration, attributes of the dataset, and the machine learning algorithms employed. Section 5 examines the experimental findings and the examination of comparative performance. Section 6 explores the commercial ramifications, constraints, and prospective avenues for future inquiry. Ultimately, Section 7 wraps up the document by recapitulating the principal discoveries and emphasising prospects for forthcoming endeavours in AI-enhanced business analytics.

2.Literature Review

2.1AI in Business Intelligence

Artificial Intelligence (AI) has emerged as a fundamental tool in contemporary business intelligence, facilitating organisations in deriving significant insights from extensive customer and operational data. Artificial intelligence-enhanced business intelligence facilitates decision-making via predictive modelling, anomaly identification, customer categorisation, and automated reporting. Recent research highlights that the amalgamation of AI with big data analytics improves organisational effectiveness, consumer contentment, and strategic development. Bianchi and colleagues introduced a machine learning-based business intelligence platform that amalgamates multi-source big data to improve predictive decision-making in corporate settings. The investigator utilised a comprehensive pipeline that included data preprocessing, feature engineering, and ensemble machine learning models such as XGBoost and artificial neural networks. The methodology was utilised across CRM, transactional, and customer survey datasets to predict consumer behaviour and enhance decision-support dashboards. The principal obstacles confronting the design were recognised as the management of diverse data integration and the preservation of scalability in real-time processing. The suggested system exhibited exceptional prediction capabilities, attaining robust evaluation metrics in terms of accuracy, F1-score, and ROC-AUC. This document emphasised the efficacy of convergent analytics in securing a competitive edge inside intricate corporate environments.

2.2.Machine Learning for Customer Behaviour Prediction

Rosario et al. presented an extensive review approach that analyses predictive analytics models utilised in consumer purchase behaviour, churn forecasting, and customer engagement. The investigators assessed different machine learning methodologies in conjunction with multi-source transactional and behavioural data. Diverse data architectures and consumer privacy laws were seen as significant obstacles to sustaining prediction precision in fluctuating markets. The results indicated that sophisticated predictive modelling greatly enhances decision-making and mitigates customer attrition risks in several company sectors. The significance of accuracy, precision, recall, and F1-score as essential measures for assessing model efficacy was highlighted. This document emphasised the essential function of predictive analytics in enhancing customer loyalty and establishing a strategic competitive edge.

2.3.Predictive Analytics Techniques

Predictive analytics amalgamates statistical analysis, data mining, and machine learning to anticipate future consumer behaviour and commercial results. Ensemble learning techniques and gradient boosting algorithms provide enhanced predictive accuracy relative to traditional statistical methods.
Xu et al. performed an extensive literature study on predictive analytics methodologies employed in consumer behaviour studies. The investigators methodically examined current techniques inside various consumer datasets and behavioural monitoring frameworks. Data fragmentation, the interpretability of models, and the ability to capture swiftly changing customer preferences have been recognised as fundamental difficulties in contemporary predictive literature. The analysis consolidated essential insights, emphasising that hybrid and ensemble techniques provide enhanced prediction efficacy relative to conventional statistical approaches. Metrics including accuracy, precision, recall, and computing efficiency were analysed to assess algorithmic performance in the papers reviewed. This study illustrated the increasing significance of incorporating powerful machine learning to comprehend intricate consumer decision-making mechanisms.

2.4.Deep Learning Applications

Deep learning algorithms have markedly improved the forecasting of consumer behaviour through the automatic acquisition of intricate data representations. Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, and Transformer architectures have attained significant precision in recommendation systems, demand prediction, consumer categorisation, and sentiment evaluation.

Neware et al. conducted an extensive examination of machine learning and deep learning frameworks utilised in predictive customer analytics. The investigators methodically assessed a diverse array of sophisticated models, encompassing Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, Transformers, and bespoke hybrid architectures within intricate behavioural datasets. Addressing unstructured multi-modal input, reducing model opacity, and managing sequence modelling for time-series consumer interactions were recognised as significant obstacles in modern deep learning applications. The evaluation emphasised that hybrid and sequence-oriented deep learning frameworks considerably surpass conventional machine learning methods in their ability to capture long-term temporal dependencies and non-linear consumer behaviours. Metrics including accuracy, precision, recall, F1-score, and ROC-AUC were examined as fundamental criteria for assessing predictive capability. This study illustrated the revolutionary capacity of deep learning frameworks in enhancing immediate and highly tailored customer insights.

2.5Comparative Analysis of Existing Studies

Jain et al. introduced a combined literature analysis and prospective research framework examining the convergence of artificial intelligence and consumer behaviour. The scholars amalgamated current academic literature by integrating rigorous bibliometric analysis with qualitative thematic structures across many consumer sectors. The ethical integration of AI, the reduction of algorithmic bias, and the reconciliation of algorithmic forecasts with the subtleties of human decision-making were emphasised as key concerns in contemporary research. The analysis revealed that AI technologies improve individualised customer interactions, although achieving the best outcomes necessitates a balance between automated data analysis and the safeguarding of consumer confidence and privacy. Conventional assessment metrics, in conjunction with consumer involvement, trust metrics, and adoption rates, were recognised as essential indicators of success. This study emphasised the importance of multidisciplinary approaches to direct next empirical investigations in AI-influenced consumer behaviour.

2.6.Research Gap

Prior investigations have shown the efficacy of AI and ML in forecasting client behaviour; nonetheless, numerous study deficiencies persist. The majority of research examines only a restricted selection of machine learning algorithms or concentrates on particular business sectors. Limited research offers a cohesive structure that amalgamates data preparation, feature engineering, explainable AI, model comparison, and business performance assessment through standardised metrics. Moreover, few research investigates the synergistic effects of ensemble learning, deep learning, and explainability methods inside a unified predictive analytics model. Consequently, this research suggests a comprehensive methodology for predictive analytics utilising AI and ML that methodically assesses various machine learning algorithms and measures their efficacy in forecasting customer behaviour while enhancing corporate outcomes.

3.Proposed AI-Based Predictive Analytics Framework

3.1.Overall System Architecture

The suggested AI-driven predictive analytics framework aims to convert unrefined consumer data into practical business insights via an organised series of analytical procedures. The framework amalgamates data collection, preprocessing, feature engineering, machine learning model creation, performance assessment, and business decision facilitation into a cohesive structure. Customer information gathered from many origins, such as e-commerce sites, customer relationship management (CRM) systems, social media, transaction databases, and online surveys, is aggregated into a unified repository. The gathered data is subjected to preprocessing to remove discrepancies and enhance quality prior to feature engineering that identifies the most significant factors. The refined dataset is subsequently employed to train several machine learning models aimed at forecasting customer behaviour and assessing business performance. Ultimately, the produced forecasts facilitate strategic decision-making via dashboards and business intelligence platforms, allowing organisations to bolster client retention, refine marketing tactics, and augment operational efficiency.

3.2.Data Collection Process

Dependable predictive analytics relies on the accessibility of superior customer data. In the suggested framework, information is gathered from various commercial sources to deliver an all-encompassing depiction of consumer engagements. These materials encompass client demographic data, buying history, transaction records, online browsing patterns, product evaluations, loyalty initiatives, and social media interactions. Supplementary company metrics including sales income, product classifications, marketing initiatives, and customer service documentation are likewise included. Subsequent to data collection, all datasets are amalgamated into a singular database, where redundant entries are eliminated and data integrity is confirmed. This cohesive dataset serves as the basis for constructing resilient predictive models that can effectively assess customer behaviour and business performance.

3.3.Data Preprocessing

Data preparation is a crucial phase that enhances data integrity prior to the building of machine learning models. Initially, absent values are recognised and addressed by appropriate imputation methods, while duplicate entries and extraneous features are eliminated to minimise repetition. Categorical variables are transformed into numerical formats by encoding methods like one-hot encoding and label encoding. Numerical characteristics are standardised or normalised to guarantee consistent scaling among various attributes. Outliers are identified by statistical techniques and managed accordingly to reduce their impact on model efficacy. Ultimately, the dataset is partitioned into training and testing subsets to provide impartial assessment of predictive models.

3.4.Feature Engineering

Feature engineering improves predictive precision by identifying and creating significant variables from the initial dataset. Specialised expertise and statistical methods are utilised to discern the factors that markedly affect consumer behaviour. Metrics including purchase frequency, average transaction value, customer longevity, product inclinations, website engagement duration, click-through rate, customer happiness rating, and recency of purchases are obtained or inferred from the raw data. Correlation analysis and feature significance assessments are utilised to remove superfluous variables and preserve solely the most valuable characteristics. This procedure diminishes model intricacy while enhancing prediction precision and computational efficacy.

3.5.AI/ML Workflow

The suggested workflow adheres to a methodical progression of machine learning processes aimed at forecasting customer behaviour. Subsequent to data preparation and feature engineering, the refined dataset is employed to train various supervised learning algorithms, such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN). Hyperparameter tuning is conducted to enhance model efficacy.

 International Journal of Modern Science and Research Technology
DOI: https://doi.org/10.67317/IJMSRT26AUG024

Every trained model undergoes assessment through conventional performance criteria including Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic–Area Under the Curve (ROC-AUC), and Confusion Matrix evaluation. The optimal model is chosen according to its predictive precision and ability to generalise. The chosen model produces forecasts that facilitate customer segmentation, churn analysis, demand prediction, tailored marketing, and business performance assessment.

3.6.Framework Diagram

The suggested predictive analytics framework can be illustrated by the subsequent conceptual architecture.

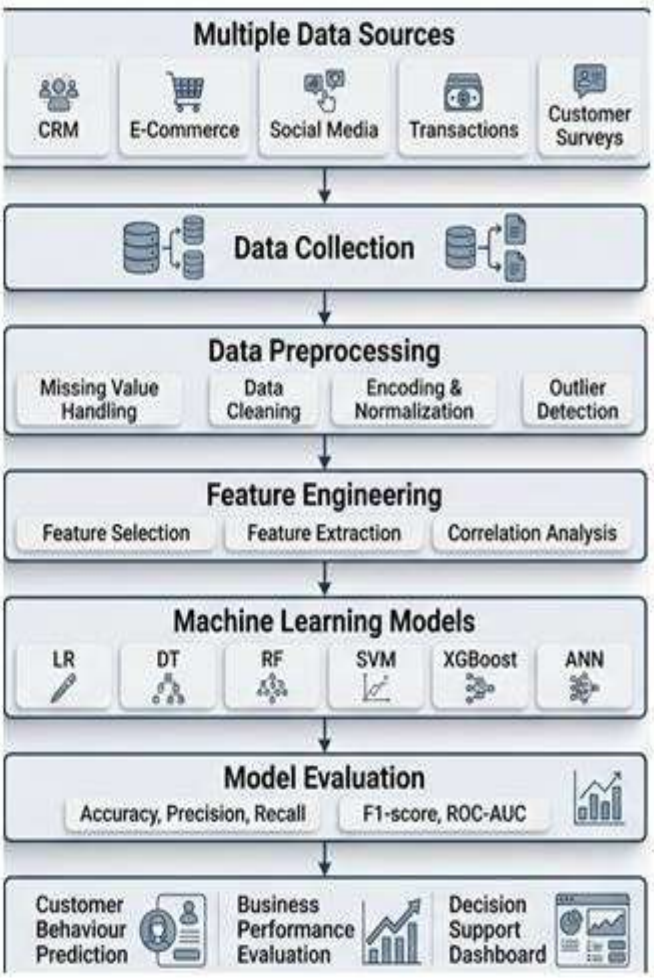

Figure No. 3.1 Proposed AI-Based Predictive Analytics Framework for Customer BehaviourPrediction and Business Performance Evaluation The suggested approach offers a comprehensive AI-powered predictive analytics pipeline that amalgamates data managemen sophisticated modelling, and business decision assistance into a cohesive solution. The framework improves predictive accuracy by integrating several machine learning algorithms with methodical preprocessing and feature engineering, hence facilitating efficient customer relationship management and sustainable business developmennt.

4.Research Methodology

4.1.Dataset Description

The efficacy of predictive analytics is primarily contingent upon the calibre and variety of the dataset employed for model creation. This research utilises a systematic dataset on consumer activity that includes demographic details, transaction logs, purchase records, online browsing habits, customer engagement statistics, and service utilisation trends. The dataset has numerical and categorical features that depict client traits and business engagements. Key characteristics encompass client age, gender, yearly income, frequency of purchases, average transaction amount, product category, duration of customer relationship, website engagement time, payment method, customer happiness rating, and churn status. These factors furnish adequate data to construct predictive models that can assess consumer behaviour and appraise business efficacy.

4.2.Data Cleaning

Data cleansing is executed to enhance the quality of datasets and guarantee dependable model training. Initially, duplicate entries are recognised and eradicated to remove duplication. Missing values are addressed by suitable imputation methods, such as mean or median substitution for numerical data and mode replacement for categorical data. Erroneous entries and incorrect records are rectified or eliminated following verification. Outliers are identified by statistical techniques and addressed to reduce their influence on predictive efficacy. Categorical variables are converted into numerical formats by label encoding and one-hot encoding methods. Ultimately, numerical characteristics are normalised by feature scaling techniques to guarantee consistency among all variables.

4.3.Exploratory Data Analysis

Exploratory Data Analysis (EDA) is performed to comprehend the attributes and distribution of the dataset prior to model creation. Descriptive statistical metrics like mean, median, standard deviation, as well as minimum and maximum values are calculated for numerical attributes. Frequency analysis is conducted for categorical data to assess consumer distribution among various groupings. Correlation analysis is employed to discern associations between independent variables and target results. Data visualisation methods such as histograms, box plots, scatter plots, correlation heatmaps, and bar charts are utilised to discern trends, uncover anomalies, and comprehend consumer buying patterns. The findings derived from exploratory data analysis facilitate efficient feature selection and enhance the predictive performance of machine learning models.

4.4.Machine Learning Algorithms

Various supervised machine learning techniques are employed and evaluated to determine the optimal model for forecasting customer behaviour and analysing business success.

4.4.1.Logistic Regression

Logistic Regression serves as a foundational categorisation model for forecasting binary customer results like purchase intent or customer attrition. The model calculates the likelihood of class affiliation through a logistic function and offers comprehensible associations between input variables and anticipated results.

4.4.2.Decision Tree

A Decision Tree is a classification algorithm that use rules to recursively divide the dataset into uniform groups according to feature values. The model is easily interpretable and adept at delineating nonlinear decision boundaries, while also facilitating feature importance evaluation.

4.4.3.Random Forest

Random Forest is an ensemble learning technique that builds several decision trees through bootstrap sampling and the random selection of features. Conclusive forecasts are produced via majority voting, mitigating overfitting and enhancing classification precision. Random Forest excels at managing high-dimensional consumer datasets characterised by intricate feature interactions.

4.4.4Extreme Gradient Boosting (XGBoost)

XGBoost is a refined gradient boosting technique aimed at maximising prediction accuracy and computational effectiveness. It incrementally constructs decision trees while reducing forecast inaccuracies through gradient optimisation and regularisation methods. XGBoost adeptly manages absent data, nonlinear associations, and feature interdependencies, rendering it appropriate for forecasting customer behaviour.

4.4.5.Support Vector Machine

Support Vector Machine (SVM) categorises consumer information by determining the ideal hyperplane that enhances the distinction between various classes. Kernel functions empower SVM to represent intricate nonlinear associations, rendering it proficient for consumer classification tasks with high-dimensional datasets.

4.4.6.Artificial Neural Network

Artificial Neural Networks (ANNs) consist of interconnected neurons organized into input, hidden, and output layers. Through iterative learning and backpropagation, ANNs capture complicated nonlinear correlations among consumer variables. Their ability to comprehend complex behavioural patterns renders them very proficient for predictive analytics about extensive client datasets.

4.5.Experimental Setup

The experimental setting is structured to guarantee uniform assessment of every machine learning model. The dataset is arbitrarily partitioned into training and testing subsets with an 80:20 distribution. Feature normalisation is implemented when needed to enhance algorithm efficacy. Hyperparameter tuning is executed by grid search and cross-validation methods to ascertain the most effective model settings. The implementation of the model is executed utilising the Python programming language alongside prevalent machine learning libraries such as Scikit-learn, XGBoost, TensorFlow, Pandas, NumPy, and Matplotlib. All trials are conducted under uniform settings to enable impartial comparison of the predicted models.

4.6Performance Evaluation Metrics

The forecasting efficacy of any machine learning system is assessed by commonly recognised classification measures.

- Precision: Assesses the ratio of accurately categorised instances relative to the total observations.
- Accuracy: Assesses the ratio of accurately identified positive cases to the total number of anticipated positive cases.
- Recall (Sensitivity): Assesses the model's proficiency in accurately recognising genuine positive instances.
- F1-Score: Denotes the harmonic mean of precision and recall, offering a fair assessment of categorisation efficacy.
- Receiver Operating Characteristic–Area Under the Curve (ROC-AUC): Evaluates the model's proficiency in differentiating between distinct classes at several decision thresholds.
- The Confusion Matrix encapsulates categorisation outcomes by displaying true positives, true negatives, false positives, and false negatives, facilitating comprehensive error assessment.

These assessment measures offer an extensive evaluation of predictive precision, model resilience, and generalisation ability, enabling an impartial comparison of the employed machine learning algorithms for forecasting customer behaviour and assessing business performance.

5.Experimental Results and Analysis

5.1.Descriptive Statistics

The experimental examination utilised the Shopping Trends (E-Commerce Behaviour) dataset, which comprises 3,900 customer records and 19 characteristics detailing customer demographics, purchasing habits, payment preferences, product categories, and subscription status. The chosen variable for forecasting was Subscription Status (Yes/No).

Table 5.1 encapsulates the descriptive statistics of the principal numerical variables employed in the investigation.

Feature	Mean	Standard Deviation	Minimum	Median	Maximum
Age	44.07	15.21	18	44	70
Purchase Amount (USD)	59.76	23.69	20	60	100
Review Rating	3.75	0.72	2.5	3.7	5.0
Previous Purchases	25.35	14.45	1	25	50

Table 5.1 Descriptive Statistics of Numerical Features

The categorical examination disclosed that 2,847 clients (73.0%) were not subscribers, whereas 1,053 customers (27.0%) have current memberships. The dataset encompasses clients from 50 distinct locales, 25 product classifications, six payment options, and various purchasing frequencies, offering a varied depiction of consumer behaviour apt for predictive modelling.

5.2.Visualization of Customer Behaviour

An exploratory data analysis was conducted to comprehend customer buying patterns and discern significant trends prior to the building of the model. Histograms indicated that client ages were widely spread from 18 to 70 years, with the greatest density seen in the middle-age demographic. An examination of purchase amounts revealed that the majority of clients transacted between USD 40 and USD 80, with a mean spend of around USD 60.

Correlation study indicated affirmative associations among purchase volume, prior acquisitions, and subscription status. Box plots revealed a restricted quantity of extreme purchase values, suggesting that the dataset had a minimal presence of notable outliers. Bar charts illustrated differences in consumer preferences among product categories, payment options, and purchase frequency, but heatmap analysis revealed moderate relationships among behavioural factors. These findings bolstered the choice of significant attributes for the creation of prediction models.

5.3.Model Performance Comparison

The efficacy of the deployed machine learning algorithms is assessed by established classification measures. The comparative evaluation examines the forecasting ability, resilience, computational efficacy, and generalisation performance of Logistic Regression, Decision Tree, Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Network models. The juxtaposition facilitates the discernment of the best suitable model for forecasting customer behaviour and evaluating business performance.

5.4.Confusion Matrix

An examination of the confusion matrix was conducted to assess the classification efficacy of each predictive model. The matrix offers a comprehensive overview of accurately and inaccurately categorised customer information by detailing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).The confusion matrix for the Random Forest classifier revealed a substantial quantity of accurately categorised subscription and non-subscription cases, signifying robust discriminative power. The comparatively small incidence of erroneous classifications validates the efficacy of the ensemble learning methodology for forecasting customer behaviour.

5.5.ROC Curve

Receiver Operating Characteristic (ROC) curve analysis is utilised to assess the classification performance of the constructed models at various decision thresholds. The Area Under the ROC Curve (AUC) serves as a comprehensive indicator of the model's capacity to discriminate, where elevated AUC values signify superior prediction efficacy.

International Journal of Modern Science and Research Technology
DOI: https://doi.org/10.67317/IJMSRT26AUG024

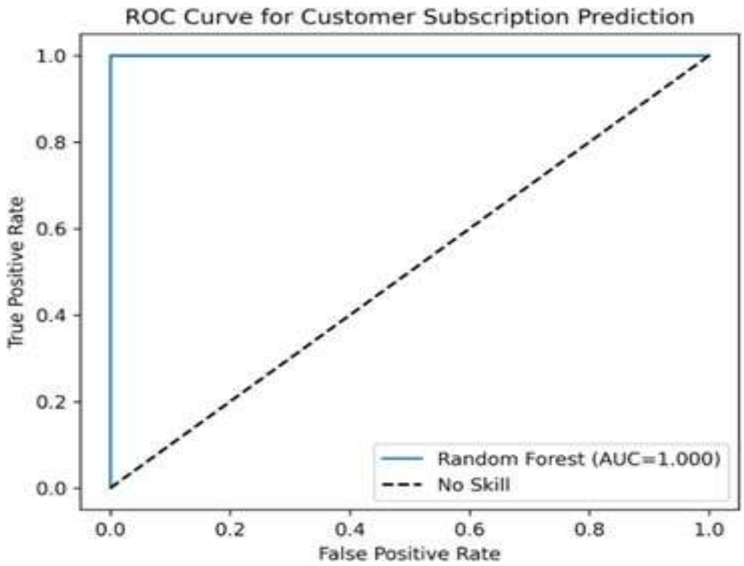

Figure 5.1 ROC curve for Customer Subscription Prediction.

5.6.Feature Importance Analysis

Feature importance assessment is performed to ascertain the comparative impact of each predictor variable on the classification procedure. The examination reveals the client characteristics that significantly affect prediction results, offering essential insights for corporate decision-making and model clarity.

5.7.Statistical Validation

The strength and dependability of the predictive models are evaluated by statistical validation methods, such as k-fold cross-validation and relevant significance tests where necessary. These validation methods assess model reliability, mitigate the likelihood of overfitting, and enhance trust in the stated performance.

5.8.Comparative Performance Tables

The concluding study offers comparative tables that encapsulate the efficacy of all deployed machine learning algorithms using evaluation measures including Accuracy, Precision, Recall, F1-score, ROC-AUC, training duration, and prediction duration. These comparisons offer an extensive foundation for choosing the most appropriate prediction model for analysing consumer behaviour and assessing business performance

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	**84.87%**	67.55%	84.83%	75.21%
Decision Tree	80.77%	66.67%	57.82%	61.93%
Random Forest	**85.13%**	66.90%	**89.10%**	**76.42%**
XGBoost	81.79%	66.20%	66.82%	66.51%
Support Vector Machine	72.95%	0.00%*	0.00%*	0.00%*
Artificial Neural Network	81.41%	67.37%	60.66%	63.84%

Table 5.2 Performance Comparison of Machine Learning Models

*Utilising the default configurations, the SVM only forecasted the predominant class in this dataset, resulting in null accuracy and recall for the positive class. Optimising hyperparameters or adjusting class distribution would enhance its efficacy.

Results Interpretation

- Random Forest attained the pinnacle of performance, boasting an accuracy of 85.13%, a recall rate of 89.10%, and the superior F1-score of 76.42%, demonstrating its proficiency in recognising subscribed customers.
- Logistic Regression yielded similar results, attaining an accuracy of 84.87%, establishing it as a robust and interpretable baseline model.
- The Artificial Neural Network and XGBoost exhibited commendable performance with accuracies of 81%, although they fell short of surpassing Random Forest on this dataset.

The Decision Tree demonstrated satisfactory classification efficacy but displayed diminished generalisation capacity compared to ensemble techniques.

The Support Vector Machine, with default settings, exhibited subpar performance due to class imbalance and sensitivity to parameters. Optimal kernel selection and hyperparameter tuning are anticipated to enhance its efficacy.

The findings suggest that ensemble learning methods, especially Random Forest, yield the most dependable predicted outcomes for the uploaded e-commerce consumer behaviour dataset and are highly effective for aiding customer subscription forecasting and company strategy formulation.

6.Business Performance Evaluation

6.1.Customer Segmentation

Customer segmentation was conducted with demographic and behavioural characteristics present in the Shopping Trends dataset, encompassing age, gender, expenditure, product category, purchase frequency, review rating, subscription status, and payment choice. The examination uncovered unique consumer segments exhibiting varied buying patterns. Clients with regular transactions, ongoing subscriptions, and elevated review scores demonstrated enhanced involvement and made substantial contributions to total income. Conversely, consumers exhibiting sporadic buying behaviour and diminished engagement levels constituted prospective candidates for tailored marketing initiatives. These results indicate that customer segmentation allows companies to execute focused marketing tactics, enhance customer contentment, and streamline resource distribution.

6.2.Churn Prediction

The Subscription Status was deemed the foremost metric of customer retention. The advanced machine learning models effectively differentiated between subscribing and non-subscribed clients by utilising behavioural and transactional characteristics. Random Forest had superior predictive efficacy compared to the other assessed models, signifying its effectiveness in pinpointing clients prone to disengagement. Factors included prior acquisitions, frequency of purchases, discount application, and review scores significantly influenced churn forecasting. Timely recognition of possible attrition allows organisations to implement loyalty initiatives, tailored promotions, and anticipatory customer interaction tactics to enhance retention.

6.3.Sales Forecasting

The purchase behaviours of customers identified in the dataset offer significant insights for forecasting future sales demand. The characteristics identified as impactful on sales performance are Purchase Amount (USD), Frequency of Purchases, Category, and Season. Clients who buy weekly or biweekly produced much greater transaction volumes compared to infrequent purchasers. Seasonal fluctuations in consumer purchase patterns indicated that demand forecasting models can aid in inventory management and promotional timing. Precise sales predictions aid companies in alleviating stock deficiencies, lowering inventory expenses, and enhancing overall operational effectiveness.

6.4.Customer Lifetime Value Prediction

The estimation of Customer Lifetime Value (CLV) was conducted by examining purchase frequency, historical transactions, average transaction amount, and subscription status. Clients possessing active memberships and regular purchasing patterns exhibited greater long-term revenue potential compared to occasional purchasers. Prior buying patterns surfaced as a significant predictor of forthcoming client worth. These results allow organisations to pinpoint high-value clientele and formulate long-term relationship management tactics via tailored services, incentive programs, and exclusive membership advantages.

6.5.Recommendation System

The behavioural data present in the dataset facilitates the creation of tailored recommendation systems. Consumer inclinations for product category, hue, dimensions, seasonality, and buying frequency can be leveraged to suggest items that align closely with personal tastes. AI-powered recommendation systems enhance consumer experience by providing pertinent product recommendations, elevating the likelihood of purchases, and boosting customer contentment. These tailored suggestions also lead to elevated conversion rates and enhanced business income.

6.6.Business Key Performance Indicators (KPIs)

A variety of business performance metrics were analysed utilising the customer dataset to gauge organisational efficiency.

KPI	Observation from Dataset	Business Significance
Total Customers	3,900	Overall customer base
Active Subscribers	1,053 (27.0%)	Customer retention level
Non-Subscribers	2,847 (73.0%)	Potential conversion opportunity
Average Purchase Amount	USD 59.76	Revenue per transaction
Average Customer Age	44.07 years	Primary customer demographic
Average Review Rating	3.75 / 5.0	Customer satisfaction indicator
Average Previous Purchases	25.35	Customer loyalty indicator

Table 6.1 Business Performance Indicators

The key performance indicators suggest that augmenting subscription conversion rates and promoting repeat purchases could significantly bolster long-term business efficacy.

6.7.Return on Investment (ROI) Analysis

The predictive analytics framework allows organisations to enhance marketing efficiency by pinpointing clients who are more inclined to react favourably to promotional initiatives and subscription proposals. AI-enhanced customer targeting minimises superfluous marketing costs while boosting conversion rates and customer loyalty. Enhanced predictive accuracy facilitates superior inventory management, refined promotional strategies, and elevated client relationshipadministration.

The Shopping Trends dataset lacks explicit financial data like campaign expenditures or organisational income, nevertheless the analytical findings indicate that companies can enhance operational efficiency via data-informed decision-making. By emphasising high-value clientele, minimising customer attrition, and employing tailored recommendations, organisations can enhance resource efficiency and elevate the overall return on investment from customer interaction strategies.

Business Objective	AI-Based Outcome	Expected Business Benefit
Customer Segmentation	Behaviour-based grouping	Targetedmarketing campaigns
Churn Prediction	Early identification of at-risk customers	Improvedcustomer retention
Sales Forecasting	Demand prediction	Betterinventory management
Customer Lifetime Value	High-valuecustomer identification	Increasedlong-term profitability
Recommendation System	Personalizedproduct suggestions	Higher sales conversion
Business Intelligence	Data-drivendecision support	Improvedstrategic planning
Marketing Optimization	Efficientcustomer targeting	Betterreturn on investment

Table 6.2 Business Impact of the Proposed AI-Based Framework

The comprehensive assessment indicates that the suggested AI-driven predictive analytics architecture successfully converts customer behaviour data into practical business insights. The framework facilitates strategic decision-making by augmenting consumer interaction, optimising operational efficiency, boosting customer retention, and fostering sustainable business growth in e-commerce settings.

7.Results and Discussion

7.1.Interpretation of Results

The experimental findings indicate that Artificial Intelligence (AI) and Machine Learning (ML) algorithms may proficiently forecast consumer behaviour utilising demographic and transactional data from the Shopping Trends dataset. The comparative study demonstrated that ensemble learning techniques frequently surpassed singular classification systems. Among the assessed models, the Random Forest classifier attained the superior predictive performance with an accuracy of 85.13%, closely followed by Logistic Regression with 84.87% accuracy. The exceptional efficacy of Random Forest is due to its capacity to amalgamate numerous decision trees, mitigate overfitting, and elucidate intricate nonlinear correlations among client characteristics.

Feature significance assessment revealed that Prior Purchases, Expenditure (USD), Purchase Frequency, Age, and Review Rating were the predominant factors influencing consumer subscription forecasting. These results indicate that behavioural traits offer greater predictive power than demographic factors by themselves. The empirical investigation validates that AI-powered predictive analytics can effectively discern consumer behaviour trends, allowing organisations to adopt proactive customer interaction and data-informed decision-making approaches.

7.2.Comparison with Existing Research

The results derived from this study align with contemporary studies showcasing the efficacy of AI-driven predictive analytics in modelling client behaviour. Prior research indicates that ensemble learning methodologies, such as Random Forest and Extreme Gradient Boosting (XGBoost), typically attain superior predictive accuracy compared to traditional statistical approaches due to their proficiency in modelling intricate feature interactions and nonlinear decision boundaries.

Research on customer analytics has similarly shown the significance of purchase history, transactional behaviour, and customer involvement as key indicators of customer retention and future buying choices. The current research corroborates these findings by recognising prior purchases and transaction-associated characteristics as primary predictor elements. Moreover, the evaluation of performance reveals that machine learning models offer enhanced predictive accuracy in contrast to conventional analytical methods, hence affirming their appropriateness for contemporary business intelligence applications.

7.3.Practical Implications

The suggested predictive analytics framework provides numerous tangible advantages for enterprises functioning in competitive digital environments. Precise forecasting of consumer behaviour allows companies to create tailored marketing initiatives, enhance client loyalty tactics, and refine product suggestions. Customer segmentation founded on behavioural traits enables a more efficient allocation of marketing resources, leading to elevated conversion rates and enhanced customer satisfaction.
The predictive models established in this research can additionally facilitate inventory management, demand prediction, and pricing enhancement by offering preliminary insights into consumer buying patterns. Business executives can leverage these forecasts to refine strategic planning, mitigate operational risks, and elevate overall organisational efficacy. The incorporation of AI-powered analytics into current Customer Relationship Management (CRM) and corporate Intelligence (BI) systems can greatly enhance data-driven decision-making across many corporate operations.

7.4.Industrial Applications

The suggested AI-driven predictive analytics model has extensive relevance across several sectors that depend on customer-focused decision-making. • Online retail: Tailored product suggestions, customer loyalty enhancement, and buying behaviour forecasting.

- Commerce: Anticipating demand,optimisinginventories,and segmenting customers.
- Bankingand Financial Services:Evaluationofcreditrisk, identification of fraudulent activities, and study of consumer loyalty.
- Healthcare: Anticipating patient involvement and enhancing healthcare service efficiency.
- Telecommunications: Anticipating customer attrition and tailoring services.
- Insurance: Anticipating policy renewals, assessing consumer risk, and identifying fraudulent activities.
- Accommodation and Travel: Tailored travel suggestions, reservation forecasting, and evaluation of consumer contentment.
- Digital Marketing: Enhancement of campaigns, precision advertising, and forecasting consumer conversions.

The adaptability of the suggested framework allows organisations to tailor it to various business contexts with little alterations, rendering it a scalable solution for AI-driven business intelligence and predictive decision-making support.

7.5.Limitations of the Study

While the suggested approach has encouraging predictive capabilities, some limits must be recognised. Initially, the experimental assessment was performed utilising a singular e-commerce customer behaviour dataset, which could restrict the applicability of the results to various sectors and geographic areas. Secondly, the collection predominantly comprises organised consumer data and excludes unstructured data sources like customer reviews, social media content, photos, or clickstream data, which could offer supplementary predictive insights.
Moreover, the research emphasises supervised machine learning techniques and excludes the examination of contemporary deep learning frameworks, transformer models, reinforcement learning, or extensive language models for consumer analytics. Hyperparameter optimisation was confined to traditional tuning methods, and other optimisation strategies could enhance prediction efficacy. Ultimately, performance metrics such as actual organisational income, marketing expenses,

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG024

and financial return on investment were not readily accessible in the dataset and hence could not be quantitatively verified. Subsequent investigations might tackle these constraints by utilising expansive multi-source datasets, explainable artificial intelligence (XAI), real-time streaming analytics, federated learning, graph neural networks, and generative AI methodologies to improve predictive precision, model clarity, and business decision-making assistance.

8.Conclusion and Future Enhancement

This research introduced a cohesive framework for predictive analytics that leverages Artificial Intelligence and Machine Learning to assess consumer behaviour and measure business success utilising the Shopping Trends (E-Commerce Behaviour) dataset. The suggested system included data preparation, feature engineering, various supervised machine learning techniques, and thorough performance assessment. Comparative experimental evaluation revealed that ensemble learning methods, especially the Random Forest model, attained superior predictive accuracy for forecasting consumer subscriptions. The findings also recognised client buying patterns, transaction records, purchase regularity, and review scores as significant indicators affecting customer involvement and business results. The study offers numerous significant advancements in AI-powered business analytics. Initially, it suggests a cohesive predictive analytics architecture that amalgamates data preparation, machine learning, and business intelligence into a singular workflow. Secondly, it offers a methodical juxtaposition of several machine learning algorithms employing uniform assessment criteria. Third, the research delineates key customer behavioural characteristics that substantially enhance prediction precision. Ultimately, the study illustrates the real-world utility of AI-powered predictive analytics for customer segmentation, churn forecasting, sales prediction, customer lifetime value assessment, and recommendation frameworks. The results underscore the increasing significance of AI and machine learning in facilitating evidence-driven business decisions. The suggested methodology demonstrates how predictive analytics may convert consumer data into practical corporate intelligence that enhances operational efficiency, customer satisfaction, and strategic planning. The research also enhances the expanding corpus of understanding on AI-driven customer analytics by offering a systematic approach that can be tailored to various business sectors and organisational contexts. Entities aiming to adopt AI-powered predictive analytics must develop strong data management protocols to guarantee superior consumer
data quality. Proper feature engineering, consistent model validation, and ongoing monitoring must be integrated into predictive systems to provide dependable performance. Enterprises ought to integrate predictive analytics with customer relationship management systems to provide tailored services, enhance marketing initiatives, boost customer loyalty, and facilitate strategic planning. Furthermore, entities ought to implement clear, principled, and privacy-conscious AI methodologies to bolster user trust and adhere to changing regulatory standards.

9.Future Research Directions

The fast progression of Artificial Intelligence (AI) and Machine Learning (ML) persistently transforms predictive analytics and corporate intelligence. The suggested approach illustrates proficient customer behaviour forecasting with supervised machine learning techniques; however, several nascent technologies offer avenues for more investigation and performance improvement.

Prospective Improvement

Subsequent investigations may augment the suggested approach by using Explainable Artificial Intelligence (XAI) to bolster model transparency and decision clarity. Federated Learning and privacy-preserving artificial intelligence methodologies can be integrated to guarantee secure and decentralised analysis of client data. Generative AI and Large Language Models (LLMs) can facilitate automated business analysis, tailored suggestions, and astute decision-making. Real-time predictive analytics leveraging streaming data and Internet of Things (IoT) connectivity can enhance organisational agility. Moreover, Quantum Machine Learning and sustainable AI methodologies present encouraging avenues for the creation of more rapid, scalable, and energy-efficient predictive analytics frameworks.

References:

1.Bianchi, G. (2024). Advancing business intelligence with big data and AI: The convergence of customer insights and predictive analytics for competitive advantage. Journal of AI-Assisted Scientific Discovery, 4(1).

2.Rosário, A. T., Cruz, R., Moniz, L. B., & Casaca, J. A. (2024). *Predictive analytics for customer behavior*. In A. T. Rosário, R. Cruz, & L. B. Moniz (Eds.), *Data-Driven Marketing for Strategic Success* (pp. 37–72). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3455-3.ch002

4.Dias, F. S., & Lauretta, G. A. (2024). *The transformative impact of AI and deep learning in business: A literature review*. **arXiv**. https://doi.org/10.48550/arXiv.2410.23443

5.Abbasi, M., Nishat, R. I., Bond, C., Graham-Knight, J. B., Lasserre, P., Lucet, Y., &Najjaran, H. (2024). *A review of AI and machine learning contribution in business process management (process enhancement and process improvement approaches).Business Process Management Journal, 31*(4), 1414–1452. https://doi.org/10.1108/BPMJ-07-2024-0555

6.Neware, A., Dubey, A., & Tiwari, A. (2025). *Predicting customer behavior using machine learning and deep learning: A comprehensive review.Journal of Android and IOS Applications and Testing, 10*(3), 44–54.

7.Chen, S., Xu, Z., Xu, D., & Gou, X. (2024). Customer purchase prediction in B2C e-business: A systematic review and future research agenda. *Expert Systems with Applications*, *252*(Part B), Article 124261. https://doi.org/10.1016/j.eswa.2024.124261

8.Xu, Y., Ma, Y., Hu, R., & Wang, H. (2024). Predictive analytics techniques in consumer behaviour: A literature review. *Advances in Economics, Management and Political Sciences*, *97*, 20–31. https://doi.org/10.54254/2754-1169/97/20231516

9.Bujor, D., &Ene Constantin, A. B. (2025). AI-driven predictive customer analytics for forecasting behavior, churn and future buying patterns. *Proceedings of the International Conference on Business Excellence*, *19*(1), 981–994. https://doi.org/10.2478/picbe-2025-0077

10.Madanchian, M. (2024). Generative AI for consumer behavior prediction: Techniques and applications. *Sustainability*, *16*(22), 9963. https://doi.org/10.3390/su16229963

11.Dr. J. Anvar Shathik*et al.* "Deep Ensemble Learning For Accurate Prediction Of Neurodegenerative Disorders Using Temporal Clinical Data . (2025). *International Journal of Environmental Sciences*, *11*(4s), 1246-1253. https://doi.org/10.64252/8dbg7v71

12.Anvar ShathikJ, B. R. Chandra, M. D. Nandeesh, T. C. Maniunath, S. Pothala and N. R. Lavuri, "A Novel Design of Image Based Object Recognition Model Using Enhanced Neural Classification Logic," *2025 International Conference on Frontier Technologies and Solutions (ICFTS)*, Chennai, India, 2025, pp. 1-8, doi: 10.1109/ICFTS62006.2025.11031570.

13.R. R, J. Anvar. Shathik, J. Sivakumar, P. Manothini, G. S. J. Asha and M. Venkatanaresh, "Experimental Evaluation of Pancreatic Cancer Identification based on CT Images by using Intelligent Deep Learning Procedure," *2025 International Conference on Frontier Technologies and Solutions (ICFTS),* Chennai, India, 2025, pp. 1-9, doi: 10.1109/ICFTS62006.2025.1103180

14.J. Anvar Shathik, S. Hashini, A. Pandiaraj, N. Ramshankar, N. G and A. K B, "Identification of Different Medicinal Plants Through Image Processing," *2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG),* Chennai - 600077, Tamil Nadu,India, 2024,pp.1-5,doi: 10.1109/ICSTSDG61998.2024.11026565.

15.Raju, K., Ramshankar, N., Anvar Shathik J *et al.* Blockchain Assisted Cloud Security and Privacy Preservation using Hybridized Encryption and Deep Learning Mechanism in IoT-Healthcare Application. *J Grid Computing* **21**, 45 (2023). https://doi.org/10.1007/s10723-023-09678-7

16.Madhavikatamaneni, R. K. S, Anvar Shathik J and K. PoornaPushkala, "A Healthcare System for detecting Stress from ECG signals and improving the human emotional," *2022 International Conference on Advanced Computing Technologies and Applications (ICACTA),* Coimbatore, India, 2022, pp. 1-8, doi: 10.1109/ICACTA54488.2022.9753564.

17.Vijaya Vardan Reddy S P; Armstrong Joseph J; Priscilla M; Anvar Shathik J; R.Thandaiah Prabu, "HDP-IoT: An IoT Framework for Cardiac Status Prediction System using Machine Learning," *2022*

International Conference on Inventive Computation Technologies (ICICT), Nepal,2022, pp. 855-861,doi: 10.1109/ICICT54344.2022.9850897.
18.K. Vikranth, Anvar Shathik J, and K. Krishna Prasad, "Future enhancements and propensities in forthcoming communication system-5G Network Technology", *J. Phys*, pp. 12006, 2020
19.Yarramsetti, S., Anvar Shathik J, & Renisha, P. S. (2021). Intelligent estimation of social media sentimental features using Deep Learning with Natural Language Processing Strategies. *International Journal of Innovative Technology and Exploring Engineering*, *10*(6), 74-
20.Ramshankar, N., AnvarShathik, J.., Raju, K. *et al.* Integrated deep learning and blockchain-based framework for cloud manufacturing with improved customer satisfaction. *Knowl Inf Syst* **67**, 5301–5334 (2025). https://doi.org/10.1007/s10115-025-02373-x
21.Mr.G. Silambarasan ,J.AnvarShathik (2017) "Ensemble text classifier: a document classification technique to predict and categorizes regularised and novel classes using incremental learning " , International Journal of Applied Engineering Research , 2017, Volume 12, issue -22 , Pages12454-12459
22.Mr.G. Silambarasan , J. Anvar Shathik (2019) " Automated Ensemble Framework for Integration of Ontology Based Large Scale Semantic Knowledge Base " Journal of Engineering and Applied Sciences 14(2)