Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

InsureAI: Predictive Customer Conversion And Lead Prioritization System

P. Malleswari
Computer Science and Engineering
Gayatri Vidya Parishad College of Engineering(Autonomous)
Visakhapatnam, India.

Metta Rupasri
Computer Science and Engineering
Gayatri Vidya Parishad College of Engineering(Autonomous)
Visakhapatnam, India.

N. Vivek Rakesh Ratan
Computer Science and Engineering
Gayatri Vidya Parishad College of Engineering(Autonomous)
Visakhapatnam, India

Parsha Uday
Computer Science and Engineering
Gayatri Vidya Parishad College of Engineering(Autonomous)
Visakhapatnam, India

Mediboyina Chaitanya
Computer Science and Engineering
Gayatri Vidya Parishad College of Engineering(Autonomous)
Visakhapatnam, India

Abstract - Customer conversion prediction is a critical challenge in the insurance industry, where identifying potential customers can significantly improve sales efficiency and resource utilization. Traditional rule-based and manual lead evaluation techniques often lack scalability and fail to provide accurate predictions. This paper proposes an intelligent machine learning–based system for predicting the likelihood of customer conversion using historical insurance data. The proposed framework analyses demographic, behavioural, and campaign-related attributes to generate conversion probability scores and classify customers into different priority levels. In addition, the system integrates predictive modelling with interactive dashboards that support customer management, lead prioritization, and real-time status updates. Experimental results demonstrate that machine learning models can effectively support data-driven decision-making, reduce manual effort, and improve customer conversion rates in insurance sales operations.

Keywords

Customer Conversion Prediction, Machine Learning, Insurance Analytics, Lead Scoring, Predictive Modelling, Data-Driven Decision Making

1.INTRODUCTION

The insurance industry has experienced rapid digital transformation with the increasing adoption of data analytics and artificial intelligence technologies. Insurance companies collect vast amounts of customer data through marketing campaigns, customer registrations, policy inquiries, and service interactions. However, effectively utilizing this data to identify potential customers and improve policy sales remains a significant challenge. Sales teams often rely on manual analysis and traditional marketing strategies to evaluate leads, which can be inefficient and may not accurately identify customers with high purchase intent.

Customer conversion prediction plays a crucial role in improving the efficiency of insurance marketing and sales processes. By analysing historical customer information and behavioural patterns, organizations can predict which customers are more likely to purchase insurance policies. This allows insurance agents to focus their efforts on high-potential leads, thereby optimizing time, resources, and marketing investments. However, traditional lead evaluation methods often fail to capture complex relationships within large datasets, limiting their effectiveness in predicting customer behaviour.

With the advancement of machine learning and predictive analytics, intelligent models can be developed to analyse customer attributes and generate accurate conversion predictions. Machine learning algorithms can learn patterns from historical data and identify factors that influence customer decisions. These predictive insights enable insurance companies to adopt data-driven strategies for lead prioritization, customer segmentation, and sales optimization.

In addition to predictive modelling, modern web technologies allow the development of interactive platforms that simplify customer data management and visualization. Integrating machine learning models with web-based systems can provide real-time insights, allowing insurance agents to track leads, manage customer information, and monitor prediction outcomes through intuitive dashboards.

2. LITERATURE SURVEY

A. Literature Review

The adoption of Artificial Intelligence (AI) and Machine Learning (ML) in the insurance industry has significantly improved customer analytics, risk prediction, and marketing decision-making processes. Insurance organizations

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

generate large volumes of data related to customer demographics, policy purchases, claims history, and marketing campaigns. By analysing this data using predictive analytics, companies can identify potential customers, optimize sales strategies, and improve operational efficiency. Several research studies have explored machine learning models for insurance analytics and customer behaviour prediction.

N. Shriram, V. S. Tomar, and S. Kokila (2023) [1] proposed a lead scoring analysis system using machine learning techniques. Their research utilized historical customer data to classify leads based on their likelihood of conversion. Machine learning algorithms were applied to identify patterns in customer behaviour and marketing responses. The results demonstrated that predictive lead scoring can significantly improve marketing effectiveness. However, the study mainly focused on prediction models and did not provide a complete system for real-time customer management or visualization.

Nolan Bishop (2022) [2] introduced a B2B lead scoring model based on machine learning for improving lead prioritization in marketing environments. The research highlighted how predictive models can evaluate customer attributes and assign conversion probability scores. This approach enables sales teams to allocate resources more efficiently and focus on high-value prospects. Despite its effectiveness, the model primarily addressed conceptual lead scoring frameworks and lacked integration with a practical platform for managing customer data and tracking lead status.

Mohamed Hanafy and Ruixing Ming (2020) [3] explored machine learning approaches for analysing large-scale auto insurance datasets. Their research demonstrated how big data analytics and predictive modelling techniques can identify hidden patterns within insurance data. The study emphasized the importance of data preprocessing, feature selection, and algorithm selection for improving predictive performance. However, the research focused primarily on insurance data analytics and did not specifically address customer conversion prediction or lead prioritization systems.

A. Kumar and S. Sharma (2019) [4] investigated the application of machine learning algorithms for insurance claim prediction. Their system analysed historical policy and claim data to predict the likelihood of insurance claims. The research showed that classification models can support risk assessment and help insurance companies make informed decisions. Nevertheless, the study mainly addressed claim prediction and risk analysis rather than customer acquisition and marketing optimization.

Data-driven approaches for insurance risk prediction and pricing models (MDPI Journal) [5] highlight the importance of predictive analytics in insurance pricing and risk evaluation. These studies demonstrate how statistical models and machine learning techniques can improve risk assessment and policy pricing strategies. While such approaches contribute significantly to actuarial science, they generally focus on financial risk analysis rather than predicting customer conversion and improving sales efficiency.

The reviewed studies demonstrate that machine learning methods significantly contribute to insurance analytics by enabling better customer behaviour analysis and more accurate predictive modelling. .However, most existing studies primarily focus on predictive algorithms or theoretical frameworks without integrating them into practical systems for customer management and real-time decision support.

B. Summary of Literature

The literature review indicates that machine learning algorithms such as Decision Trees, Random Forest, Logistic Regression, and Gradient Boosting are widely used for predictive analytics in insurance and marketing domains. These models are capable of analysing historical data to identify patterns related to customer behaviour, risk assessment, and marketing responses. Predictive analytics can significantly improve decision-making processes and support data-driven strategies for customer targeting and resource allocation.

However, several limitations are observed in existing research studies:

- Lack of integrated platforms that combine predictive management systems
- Limited support for real-time lead prioritization and monitoring
- Heavy focus on theoretical predictive models rather than practical implementation systems
- Limited integration of machine learning models with modern web-based applications

These limitations highlight the need for an intelligent system that integrates predictive analytics with a practical platform for managing customer data and prioritizing insurance leads. Therefore, the proposed InsureAI system aims to address these challenges by combining machine learning–based customer conversion prediction with a web-based dashboard that supports lead scoring, customer management, and data-driven decision-making.

3. PROPOSED SYSTEM

InsureAI is an intelligent insurance analytics platform designed to assist insurance agents and organizations in identifying potential customers and prioritizing high-value leads. The system leverages machine learning techniques to analyse historical customer data and predict the likelihood of policy conversion. By integrating predictive analytics with an

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

interactive web-based dashboard, the platform enables agents to make informed decisions and improve sales efficiency.

Users interact with the system through a web-based interface developed using React.js, where customer information and marketing data can be entered and managed. These inputs are processed by a backend service built with Node.js and Express.js, which coordinates communication between system components and ensures efficient data processing.

The system performs customer data analysis, conversion prediction, lead scoring, and visualization through an integrated pipeline. A machine learning model analyses customer attributes such as demographic information, marketing responses, and interaction patterns to estimate the probability of policy conversion. Based on these predictions, customers are categorized into different priority levels, allowing agents to focus on leads with higher conversion potential. Customer records and prediction results are securely stored in a MongoDB database, enabling scalable data

management and future analysis. Overall, InsureAI provides a smart and efficient platform for insurance lead management, helping organizations improve marketing strategies, reduce manual effort, and increase customer conversion rates.

A. System Architecture

The architecture of InsureAI consists of several interconnected components including the user interface, backend services, machine learning prediction module, analytics dashboard, and database storage.

The frontend interface allows agents to add, view, and manage customer information through an intuitive dashboard. Customer data such as demographic attributes, interaction history, and campaign responses are entered through the interface.

The backend server processes the data, performs validation and preprocessing, and sends structured information to the machine learning prediction module.

The machine learning module analyses customer attributes and generates a conversion probability score indicating the likelihood that a customer will purchase an insurance policy. Based on this probability, customers are classified into different priority levels such as high-priority, medium-priority, and low-priority leads.

The prediction results are then displayed on the analytics dashboard, which provides visual insights into customer distribution, conversion probabilities, and lead prioritization.

All customer records, prediction outputs, and interaction data are securely stored in a MongoDB database, ensuring scalability and efficient data retrieval.

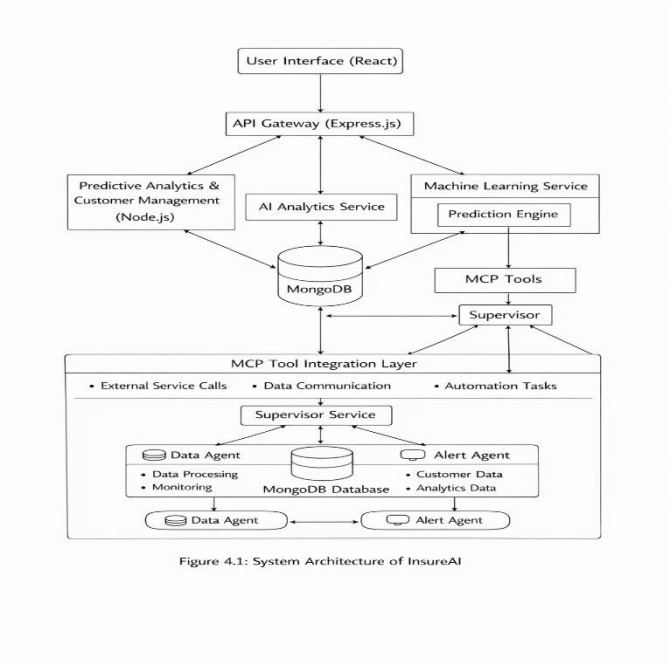


Figure 4.1: System Architecture of InsureAI

B. System Modules

The InsureAI platform consists of multiple functional modules that collectively perform customer data management, predictive analytics, and lead prioritization.

1) User Authentication and Session Management Module

This module manages user registration, login, and session

handling. User credentials are securely stored in MongoDB, and password encryption is implemented using bcrypt hashing mechanisms.

Authentication tokens are generated using JSON Web Tokens (JWT) to maintain secure user sessions. This module ensures that only authorized agents can access customer data and prediction results.

2) Customer Data Input and Management Module

This module allows agents to add, update, and manage customer information within the system.

Customer attributes may include:

- Age
- Gender
- Occupation
- Income level
- Previous insurance interactions
- Marketing campaign responses
- Contact history

The module performs data preprocessing including normalization, removal of duplicate entries, and feature formatting. This ensures that customer information is converted into a structured format suitable for machine learning analysis.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

3) Customer Conversion Prediction Module

The conversion prediction module uses machine learning techniques to estimate the likelihood that a customer will purchase an insurance policy.

The prediction model analyses multiple customer attributes and identifies patterns in historical data. Based on these patterns, the system calculates a conversion probability score for each customer.

The predicted probability values are then used to categorize customers into different lead priority levels:

- **High Priority Leads** – Customers with high probability of conversion
- **Medium Priority Leads** – Customers with moderate conversion likelihood
- **Low Priority Leads** – Customers with low conversion probability

This classification helps insurance agents focus their efforts on the most promising prospects.

4) Lead Scoring and Prioritization Module

The lead scoring module assigns a priority score to each customer based on the prediction results.

This score helps agents:

- Identify high-value prospects quickly
- Allocate marketing resources efficiently
- Reduce time spent on low-potential leads

The system continuously updates lead scores as new customer data or interactions are recorded.

5) Analytics Dashboard and Visualization Module

The dashboard module provides interactive visualizations that help agents understand prediction outcomes and customer insights.

The dashboard may display:

- Total number of customers
- High-priority leads
- Conversion probability distribution
- Customer statistics and analytics

These visual insights allow agents to monitor sales performance and make data-driven decisions.

6) Database and Data Management Module

All system data is stored in MongoDB, which provides flexible and scalable data storage.

The database stores:

- Customer information
- Prediction results
- Lead priority scores
- User authentication data

This ensures efficient data retrieval and supports future analytics and system scalability.

C. Advantages of the Proposed System

The proposed InsureAI platform offers several advantages compared to traditional manual lead evaluation methods.

First, the integration of machine learning-based predictive analytics enables accurate customer conversion prediction and supports data-driven decision-making.

Second, the lead scoring mechanism allows insurance agents to prioritize high-value prospects, improving marketing efficiency and reducing manual effort.

Third, the interactive analytics dashboard provides clear visualization of customer insights, enabling agents to quickly interpret prediction results and optimize their strategies.

Fourth, the system integrates predictive models with a modern web-based platform, allowing real-time customer management and monitoring of lead status.

Finally, the use of scalable technologies such as MongoDB, Express.js, React.js, and Node.js (MERN stack) ensures system flexibility, performance, and ease of future expansion.

4. METHODOLOGY

The methodology of InsureAI describes the systematic process through which customer data is analysed to predict the likelihood of insurance policy conversion. The system follows a structured workflow that integrates data preprocessing, machine learning prediction, lead scoring, and analytics visualization. This approach enables insurance agents to identify potential customers and prioritize high-value leads efficiently.

A. Customer Data Acquisition and Preprocessing

The first stage of the methodology involves collecting customer data through the web-based interface developed using **React.js**. Insurance agents enter customer information such as demographic details, contact history, and marketing responses into the system.

The collected data may include attributes such as:

- Age
- Gender
- Occupation
- Income level
- Previous insurance interactions
- Marketing campaign responses
- Customer engagement history

Before being used for prediction, the data undergoes a **preprocessing stage** to ensure consistency and reliability. The preprocessing operations include:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

- Removal of duplicate entries
- Handling missing values
- Feature normalization and encoding
- Data formatting and structuring

These preprocessing steps transform raw customer information into a structured dataset suitable for machine learning analysis.

B. Customer Conversion Prediction

After preprocessing, the structured customer dataset is passed to the machine learning prediction model, which estimates the probability of customer conversion.

The prediction model is trained using a historical insurance dataset containing customer attributes and their corresponding policy purchase outcomes. The dataset is divided into training and testing sets using an 80–20 split, where 80% of the data is used to train the model and 20% is used for evaluation.

The machine learning algorithm analyses patterns within the dataset and identifies relationships between customer attributes and policy purchase behaviour. Based on these patterns, the model generates a conversion probability score that represents the likelihood of a customer purchasing an insurance policy.

This predictive analysis enables insurance agents to make informed decisions and focus on customers with higher purchase potential.

C. Lead Scoring and Prioritization

After predicting the conversion probability, the system assigns a lead priority score to each customer. The lead scoring mechanism categorizes customers into different priority levels based on their predicted conversion probability.

The lead categories include:

- High Priority Leads: Customers with a high probability of policy conversion
- **Medium Priority Leads:** Customers with moderate likelihood of conversion
- **Low Priority Leads:** Customers with low probability of purchasing a policy

This classification helps insurance agents prioritize their follow-up activities and allocate marketing resources more effectively.

D. Backend Processing and API Communication

The backend of the system is implemented using Node.js and Express.js, which handle communication between the frontend interface and the machine learning prediction service.

When customer data is submitted through the frontend interface, it is sent to the backend API for processing. The backend performs the following operations:

- Data validation and formatting
- Communication with the machine learning prediction module
- Storing customer information and prediction results in the database

The backend ensures efficient coordination between system components and maintains data integrity throughout the prediction process.

E. Data Storage and Management

All customer data, prediction results, and system information are securely stored in a MongoDB database. MongoDB provides a flexible and scalable storage system capable of handling large volumes of customer data.

The database stores the following information:

- Customer demographic details
- Marketing interaction history
- Prediction results and conversion probabilities
- Lead priority classifications
- User authentication data

This centralized data storage allows the system to efficiently retrieve and analyse customer information for future predictions and analytics.

System Workflow

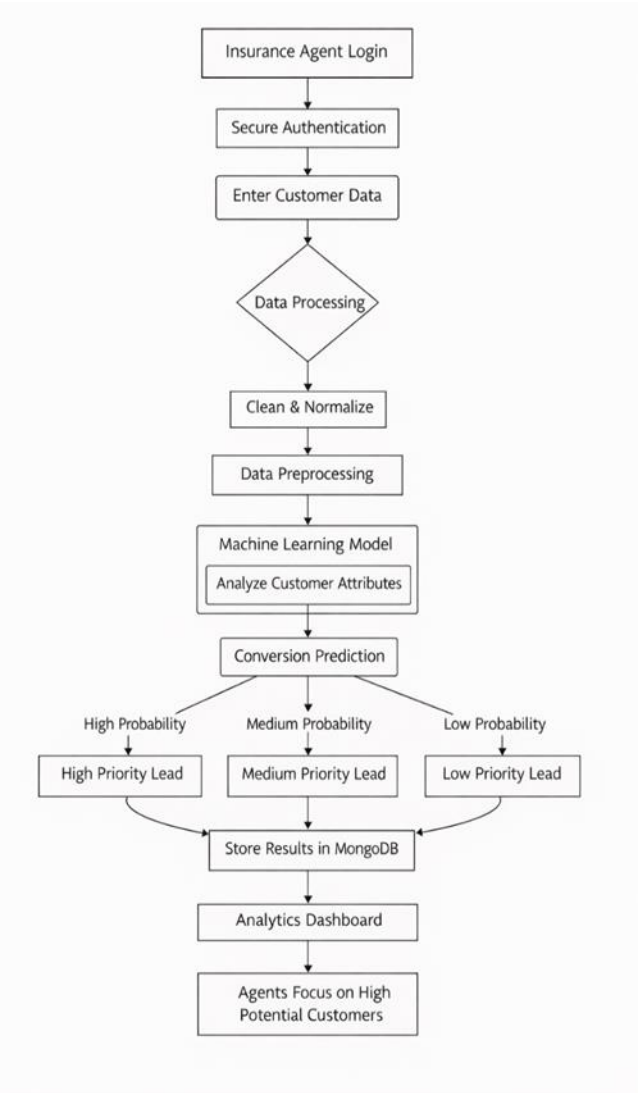
The InsureAI platform operates through a structured workflow that transforms customer data into predictive insights and lead prioritization while ensuring secure role-based access control for different users of the system.

The workflow consists of the following steps:

1. Users log into the system through secure authentication, where role-based access control ensures that users (such as administrators and permitted for their role.
2. Insurance agents enter customer information through the web-based interface developed using React.js.
3. The system preprocesses the customer data by cleaning, normalizing, and converting it into a structured format suitable for machine learning analysis.
4. The machine learning model analyses the processed data and predicts the probability of insurance policy conversion for each customer.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

5. The system assigns a lead priority score based on the predicted conversion probability, categorizing leads into high, medium, or low priority.
6. Prediction results and customer data are stored in the MongoDB database for future analysis and tracking.
7. The analytics dashboard displays customer insights, lead prioritization, and overall system statistics to authorized users.
8. Insurance agents utilize these insights to focus on high-potential customers and improve their sales performance.
9. Administrators can monitor system activities and analytics, manage users, and oversee customer data to ensure efficient system operation.



5. RESULTS AND DISCUSSION

This section presents the experimental results and performance evaluation of the proposed **InsureAI platform**. The evaluation focuses on the effectiveness of the machine learning model for customer conversion prediction, the functionality of the lead prioritization system, and the usability of the web-based analytics dashboard.

The system was tested using multiple customer records containing demographic attributes, marketing responses, and interaction history. The complete workflow—from customer data entry to conversion prediction and lead prioritization—was validated through the implemented MERN-based platform.

The generated outputs include **conversion probability prediction, lead priority classification, and interactive dashboard analytics**, which assist insurance agents in identifying potential customers and improving marketing strategies.

A. Machine Learning Model Performance

The InsureAI system uses a machine learning model to analyse historical customer data and predict the likelihood of insurance policy conversion.

The model was trained using a dataset containing customer attributes such as:

- Age
- Gender
- Occupation
- Income level
- Previous insurance interactions
- Campaign responses
- Customer engagement history

The dataset was divided using an **80–20 train–test split**, where 80% of the data was used for training and 20% for testing and evaluation.

The machine learning model analyses patterns within the dataset and generates a **conversion probability score** for each customer.

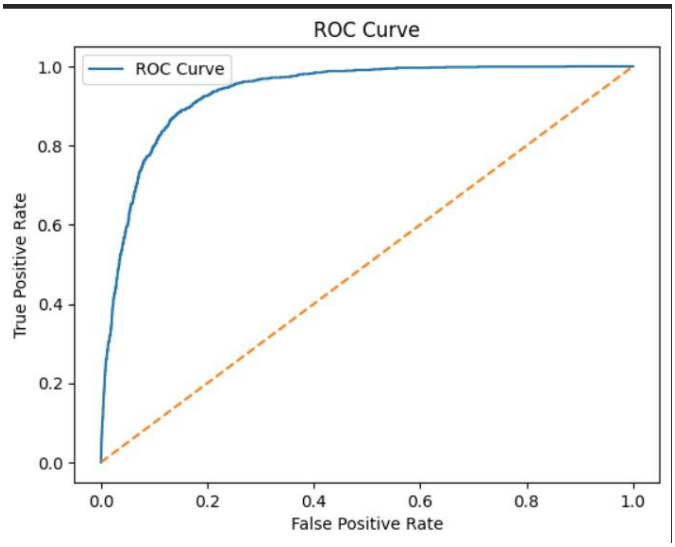
The model performance was evaluated using standard classification metrics such as:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-score**

The experimental results demonstrate that the machine learning model effectively identifies customers with a high likelihood of purchasing an insurance policy.

	Model	Train Accuracy	Test Accuracy
0	Logistic Regression	0.902345	0.898706
1	Decision Tree	0.915340	0.898043
2	Random Forest	1.000000	0.903572
3	SVM	0.890456	0.883999
4	Neural Network	0.909257	0.909572
5	XGBoost	0.936961	0.909322

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

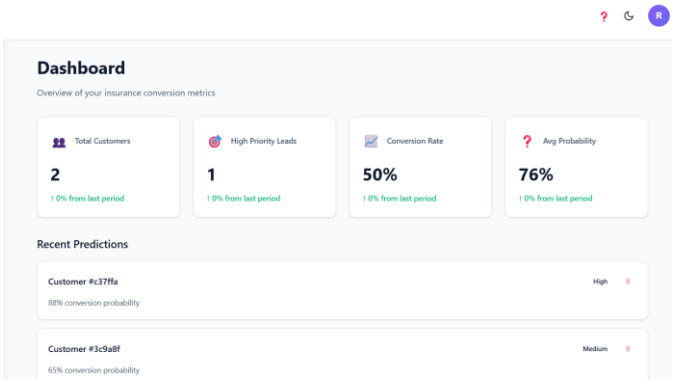


B. Customer Conversion Prediction Results

The trained machine learning model successfully predicts the conversion probability based on customer attributes.
When customer data is submitted through the system interface, the model processes the structured features and generates a **conversion probability score** indicating the likelihood of policy purchase.
Based on the predicted probability, customers are categorized into three lead priority levels:

- **High Priority Leads** – Customers with high conversion probability
- **Medium Priority Leads** – Customers with moderate conversion likelihood
- **Low Priority Leads** – Customers with low conversion probability

This classification enables insurance agents to prioritize high-value prospects and improve overall sales efficiency.

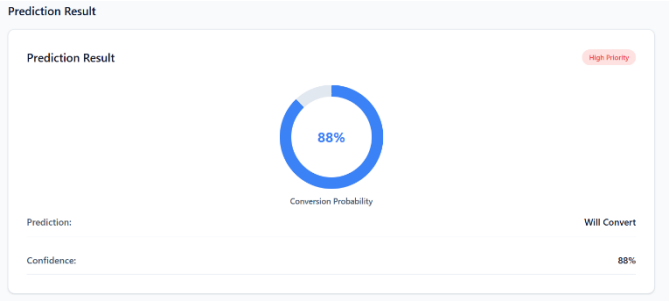


C. Lead Scoring and Prioritization

The lead scoring module evaluates each customer based on the predicted conversion probability.
The system automatically assigns priority levels that help agents:

- Identify high-potential customers
- Focus marketing efforts on valuable leads
- Reduce time spent on low-probability prospects

The lead prioritization mechanism improves resource allocation and increases the efficiency of insurance sales operations.

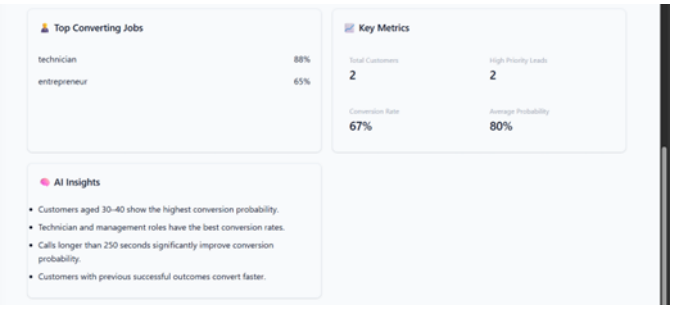


D. Analytics Dashboard Visualization

The InsureAI platform includes an interactive analytics dashboard that provides visual insights into customer data and prediction outcomes.
The dashboard displays:

- Total number of customers
- Number of high-priority leads
- Conversion probability distribution
- Customer analytics statistics

These visualizations allow insurance agents and administrators to monitor system performance and make **data-driven decisions**.



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

E. System Usability and Functional Evaluation

The complete InsureAI system was tested to evaluate the functionality of each module, including:

- User authentication
- Customer data management
- Machine learning prediction
- Lead prioritization
- Dashboard visualization

The MERN stack architecture ensures smooth interaction between the frontend, backend, and machine learning services. The system successfully processes customer data, generates predictions, and displays insights through the dashboard.

6. CONCLUSION

This research presented **InsureAI**, an intelligent insurance analytics platform designed to predict customer conversion probability and assist insurance agents in identifying high-potential leads.

The system integrates **machine learning-based predictive analytics with a web-based dashboard built using the MERN stack**, enabling efficient customer data management and lead prioritization. By analysing historical customer data and marketing interactions, the system predicts the likelihood of policy conversion and categorizes customers into different priority levels.

The proposed platform improves marketing efficiency by helping agents focus on high-value prospects, reducing manual effort, and supporting data-driven decision-making. The use of scalable technologies such as **MongoDB, Express.js, React.js, and Node.js** ensures system flexibility and performance.

Overall, InsureAI demonstrates an effective approach to applying machine learning techniques in the insurance domain, contributing to the development of intelligent customer analytics systems for modern insurance organizations.

7. FUTURE WORK

The InsureAI platform can be further enhanced by incorporating several improvements to increase its scalability and predictive capabilities.

Future work includes expanding the training dataset to improve model accuracy and support more diverse customer profiles. Additional machine learning algorithms and ensemble techniques can also be integrated to enhance prediction performance.

The analytics dashboard can be improved by incorporating advanced data visualization and real-time monitoring features. Integration with external marketing platforms or CRM systems may enable automated customer tracking and campaign analysis.

Furthermore, deploying the platform on **cloud infrastructure** can improve system scalability and allow large-scale real-world deployment. Developing a dedicated **mobile application** for insurance agents may also enhance accessibility and usability.

These improvements will further strengthen the InsureAI platform and enable more intelligent and efficient insurance lead management systems.

8. REFERENCES

[1] N. Shriram, V. S. Tomar, and S. Kokila, “Lead Score Analysis Using Machine Learning,” *International Journal of Research and Technology Innovations (IJRTI)*, Apr. 2023.

[2] N. Bishop, “The Relevance of Lead Prioritization: A B2B Lead Scoring Model Based on Machine Learning,” *ResearchGate Publication*, Mar. 2024.

[3] M. Hanafy and R. Ming, “Machine Learning Approaches for Auto Insurance Big Data,” *ResearchGate Publication*, Feb. 2020.

[4] A. Kumar and S. Sharma, “Insurance Claim Prediction Using Machine Learning Algorithms,” *International Journal of Engineering Technology and Management Sciences (IJETMS)*, vol. 7, no. 1, Jan. 2021.

[5] Various Researchers, “Data-Driven Approaches for Insurance Risk Prediction and Pricing Models,” *MDPI Data Journal*, Jun. 2024.