E-ISSN:2250-0758

P-ISSN:2394-6962

Research Article

Logistic Regression

International Journal of Engineering and Management Research

2025 Volume 15 Number 5 October

Publisher

Vandana Publications

www.vandanapublications.com

Predictive Modelling for Customer Purchase Behaviour: A Logistic Regression Approach Based on Age and Estimated Salary

Selvakumar S[1*], Yogeshwaramoorthi K[2], Jegathambal PMG[3]

DOI:10.5281/zenodo.17301302

[1*] Selvakumar S, UG Student, Department of CSE AI & DS, VELS University, Chennai, Tamil Nadu, India.

[2] Yogeshwaramoorthi K, UG Student, Department of CSE AI & DS, VELS University, Chennai, Tamil Nadu, India.

[3] P.M.G. Jegathambal, Assistant Professor, Department of CSE, VELS University, Chennai, Tamil Nadu, India.

Customer purchase prediction has become a critical requirement in the insurance industry, where businesses strive to maximize customer acquisition while minimizing marketing costs. Accurate forecasting of whether a potential customer will purchase an insurance policy allows companies to focus on high potential leads and optimize their strategies. In this study, we propose a predictive modelling approach using logistic regression to classify customers based on two key demographic features: Age and Estimated Salary. A dataset of over 1,000 customer records was pre-processed, visualized, and divided into training and testing subsets using an 80:20 ratio. The logistic regression model was trained to identify significant patterns influencing purchase decisions and to estimate the probability of policy adoption. To enhance usability, the trained model was deployed in a Streamlit based web application that includes secure user authentication, interactive input fields, decision boundary visualization, and a leaderboard to track predictive outcomes. Experimental results demonstrate that the logistic regression model achieves an accuracy of approximately 90%, with strong interpretability through coefficient analysis and decision boundary visualization. This work highlights the potential of combining machine learning models with lightweight, interactive applications to support business analysts and decision-makers. The proposed framework offers a scalable, interpretable, and cost-effective solution for insurance companies seeking to strengthen customer targeting. Future work will focus on incorporating additional demographic and behavioral features, applying advanced ensemble models, and integrating large-scale realworld datasets to further enhance prediction performance.

Keywords: Logistic Regression, Customer Purchase Prediction, Insurance Analytics, Streamlit Application, Decision Boundary Visualization, Predictive Modeling

Corresponding Author	How to Cite this Article	To Browse
Selvakumar S, UG Student, Department of CSE AI & DS, VELS University, Chennai, Tamil Nadu, India. Email: sk.selvakumar379@gmail.com	Selvakumar S, Yogeshwaramoorthi K, Jegathambal PMG, Predictive Modelling for Customer Purchase Behaviour: A Logistic Regression Approach Based on Age and Estimated Salary. Int J Engg Mgmt Res. 2025;15(5):34-43. Available From https://ijemr.vandanapublications.com/index.php/j/article/view/1793	

Manuscript Received	Review Round 1	Review Round 2	Review Round 3	Accepted
2025-09-01	2025-09-15			2025-10-01
Conflict of Interest	**Funding**	**Ethical Approval**	**Plagiarism X-checker**	**Note**
None	Nil	Yes	4.71	

1. Introduction

The insurance industry is one of the fastest-growing sectors worldwide, driven by increasing awareness of financial security and risk management. However, the industry is also highly competitive, with companies facing challenges in identifying and converting potential customers. In this context, understanding customer behaviour has become an essential factor in designing effective marketing strategies, improving operational efficiency, and optimizing resource allocation. The ability to predict whether a customer will purchase an insurance policy enables organizations to prioritize high-potential leads, reduce acquisition costs, and improve overall customer satisfaction.

Traditional approaches to customer segmentation often rely on demographic analysis, surveys, and manual classification techniques. While these methods provide basic insights, they tend to oversimplify customer behaviour and fail to capture the complex interactions between demographic and financial variables. As a result, businesses risk targeting the wrong customers, leading to wasted resources and lower profitability.

The rapid growth of data-driven decision-making has transformed how industries address this challenge. With the availability of large-scale customer datasets, predictive analytics has emerged as a powerful tool for forecasting purchasing behaviour. Among various machine learning techniques, logistic regression is particularly well suited for binary classification tasks such as predicting whether a customer will purchase insurance or not. It is widely adopted due to its simplicity, interpretability, computational efficiency, and robustness. Unlike complex models such as deep neural networks, logistic regression provides clear insights into the relative importance of features, making it easier for business analysts to interpret results and take informed actions.

In this paper, we present a logistic regression-based predictive model for forecasting customer purchase behaviour using two key features: Age and Estimated Salary. These variables were chosen due to their strong correlation with purchasing power and decision-making tendencies in the insurance market. To make the system accessible and practical for real-world use, the model is integrated into a Streamlit web application, enabling business users to interactively input customer data, generate predictions in real time, visualize decision boundaries, and track outcomes through a leaderboard. The application also incorporates a secure login system, ensuring restricted access and controlled usage.

The main contributions of this work are summarized as follows:

- Development of a predictive model using logistic regression to forecast customer purchase behaviour with high accuracy.
- Integration into a user-friendly Streamlit application, complete with authentication, interactive inputs, and visualization features.
- Evaluation of model performance using accuracy, confusion matrix, precision, recall, and decision boundary analysis.
- Practical business insights, showing how age and salary—two simple features—can significantly improve customer targeting and segmentation.

The remainder of this paper is structured as follows: Section II reviews related work in predictive modeming and logistic regression applications. Section III describes the methodology, including dataset preprocessing, model formulation, and evaluation metrics. Section IV discusses the system implementation within the Streamlit framework. Section V presents the experimental results and key findings. Finally, Section VI concludes the paper and suggests future research directions.

2. Literature Review

The prediction of customer purchase behaviour has been widely explored across domains such as finance, retail, healthcare, and insurance. Early research in the insurance sector relied heavily on traditional demographic segmentation and survey-based analysis, which provided only limited insights due to their inability to handle complex, multidimensional data. With the rise of machine learning and data mining, predictive modelling has gained increasing attention as a way to uncover hidden patterns in customer data and improve marketing efficiency.

Among various predictive models, logistic regression has emerged as one of the most widely adopted methods for binary classification tasks. Its popularity is attributed to several advantages:

- Interpretability: Logistic regression provides direct insights into the influence of each feature through its coefficients, allowing businesses to easily understand customer behaviour.
- Efficiency: It is computationally inexpensive and works well even with small or medium-sized datasets.
- Robustness: It performs reliably across a wide range of applications, including insurance adoption, credit risk modelling, and medical diagnosis.

Numerous studies have validated the effectiveness of logistic regression in purchase prediction. For example, Kumar and Singh (2018) demonstrated that logistic regression outperformed decision trees in predicting whether customers would purchase retail products, achieving an accuracy above 85%. Similarly, Sharma and Bansal (2019) applied logistic regression to insurance datasets and reported competitive accuracy compared to support vector machines, while also highlighting its interpretability as a major advantage for business decision making.

While logistic regression remains a powerful tool, researchers have also explored more advanced models. Random forests and gradient boosting methods often achieve higher predictive accuracy due to their ability to capture non-linear patterns, but they sacrifice interpretability. Support vector machines (SVMs) and deep learning models have also been applied in customer analytics, offering strong performance but requiring greater computational resources and expertise. In practice, the choice of model depends on the trade-off between predictive performance and ease of interpretation. For industries like insurance—where decision transparency is critical—logistic regression remains highly valuable.

Another important research direction is the integration of predictive models into interactive web applications. Traditional predictive analytics often required specialized software or technical expertise, limiting accessibility for business users. Recent studies have highlighted the importance of embedding machine learning models into lightweight, user-friendly platforms to bridge the gap between technical analysis and practical decision-making. Tools such as Streamlit provide an effective solution, enabling rapid deployment of predictive systems with interactive features, real-time visualization, and secure authentication. However, literature in this area is still limited, and only a few works have emphasized the role of interactive deployment in enhancing adoption by non-technical users.

Building on these findings, our study applies logistic regression to predict customer insurance purchase behaviour using a dataset of over 1,000 records. Unlike many prior works that focus only on predictive performance, we extend the contribution by integrating the model into a Streamlit web application with login authentication, decision boundary visualization, and a leaderboard for tracking outcomes. This combination of interpretability, accuracy, and practical usability positions our work as a meaningful contribution to both academic research and real-world industry applications.

3. Methodology

This section outlines the systematic approach adopted for predicting customer insurance purchase behaviour using logistic regression. The methodology includes dataset description, preprocessing steps, feature selection, model development, and evaluation metrics.

A. Dataset Description:

The dataset used in this study contains 1,000+ customer records, each consisting of demographic and financial information relevant to insurance purchase decisions. The key attributes include:

1. Age – The age of the customer (in years).
2. Estimated Salary – Annual income of the customer (in USD or equivalent local currency).
3. Purchased – Target variable indicating whether the customer purchased insurance (1) or not (0).

The dataset was divided into training (80%, 800 records) and testing (20%, 200 records). A larger dataset ensured improved model generalization and reduced the risk of overfitting.

A preliminary exploratory data analysis (EDA) was conducted to examine the distribution of Age and Estimated Salary. Visualization revealed that customers with higher salaries and middle-to-older age groups showed a higher tendency to purchase insurance, highlighting the relevance of these features.

B. Data Preprocessing:

Data preprocessing ensures that the input is clean, consistent, and suitable for model training. The following steps were applied:

1. Handling Missing Values – Missing entries, if any, were handled by removing incomplete rows or imputing them using the mean value of the corresponding feature.
2. Outlier Detection – Outliers in salary distribution were checked to avoid skewed predictions.
3. Feature Scaling – Since Age and Estimated Salary have different scales, we applied standardization (z-score normalization) to transform the features into comparable ranges, improving model convergence.
4. Train-Test Split – The dataset was split into training and test sets with random shuffling to avoid bias.

C. Feature Selection:

This study focuses exclusively on two features—Age and Estimated Salary—as predictors. These variables were chosen because they represent key demographic and financial indicators that significantly influence customer purchase decisions in insurance markets. Although additional features (e.g., occupation, marital status, prior history) could improve predictive power, the current scope emphasizes simplicity and interpretability.

D. Logistic Regression Model:

Logistic regression is a supervised learning algorithm designed for binary classification problems. It predicts the probability that a given input belongs to one of two classes—in this case, whether a customer purchases insurance (1) or not (0).

- The logistic regression model is defined as:

$$P(Y \;=1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}$$

This threshold can be adjusted in business contexts to balance false positives and false negatives depending on risk preferences.

E. Model Evaluation Metrics:

To assess the performance of the logistic regression model, we employed multiple evaluation metrics:

- Accuracy – The percentage of correct predictions.
- Confusion Matrix – Provides detailed insight into correct and incorrect classifications (true positives, true negatives, false positives, false negatives).
- Precision – The proportion of correctly predicted positive cases among all predicted positives.
- Recall (Sensitivity) – The proportion of correctly predicted positive cases among all actual positives.
- F1-Score – Harmonic mean of precision and recall, useful in imbalanced datasets.
- ROC Curve and AUC (Area Under Curve) – Visual and numerical measures of the model's ability to discriminate between the two classes.
- Decision Boundary Visualization – A 2D plot showing how the model separates purchasing vs. non-purchasing customers based on Age and Salary.

F. Implementation Overview:

The model was implemented using Python and the scikit-learn library. After training, the model was serialized with Pickle and integrated into a Streamlit-based web application.

The app allows users to:

- Input Age and Estimated Salary interactively.
- Generate real-time predictions.
- Visualize decision boundaries.
- Track predictions on a leaderboard with user authentication.

4. Implementation

This section presents the practical implementation of the predictive system, focusing on the integration of the logistic regression model within a Streamlit web application. The goal was to create a user-friendly, secure, and interactive platform that combines predictive analytics with real-time visualization.

A. System Architecture:

The architecture of the application is organized into three primary components:

1. Frontend (Streamlit Interface) – Provides the interactive user interface for login, prediction inputs, and visualization. It includes custom backgrounds, sliders, buttons, and dynamic animations to enhance usability.
2. Backend (Model Engine) – Responsible for executing the logistic regression model trained using scikit-learn. The backend processes user inputs (Age and Estimated Salary) and returns probability predictions.
3. Database/Leaderboard – Maintains user authentication details and tracks predictive outcomes, allowing results to be ranked and compared in a leaderboard.

The prediction table is designed to be interactive and

- The user logs in using credentials
- Age and Salary inputs are entered.
- The backend model processes inputs and returns a prediction.
- The result is displayed with visualizations and optionally recorded on the leaderboard.

This architecture ensures modularity, making the system easy to extend in the future (e.g., with more features or advanced models).

B. Login System

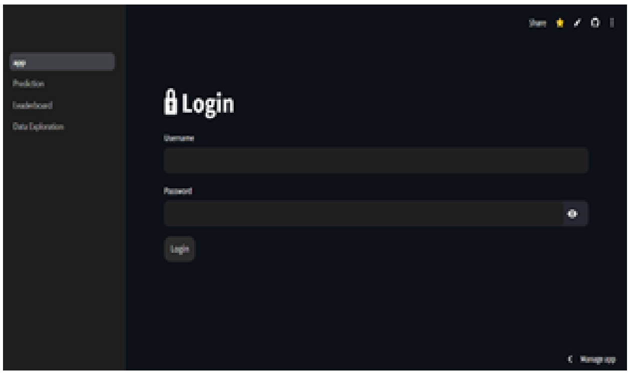


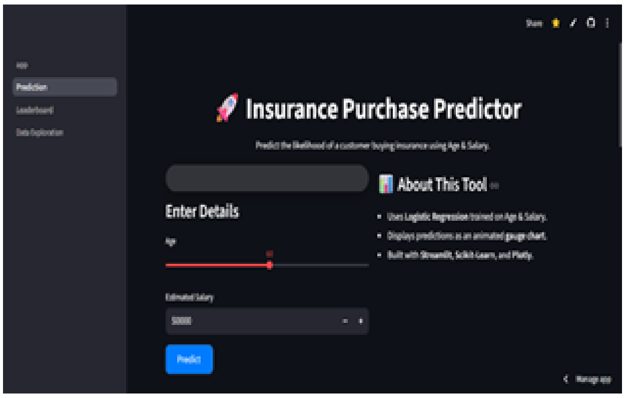
Figure 1: Login Page.

If the login attempt is successful, users are redirected to the prediction page; otherwise, an error message is displayed. In a production environment, login details can be stored in a secure database with hashed passwords to prevent unauthorized access. Multi-User support can be added to enable wider deployment.

C. Prediction Interface:

The prediction page is designed by using Streamlit, Scikit-Learn & Plotly.

- Age Input – A slider widget allows users to select values between 18 and 100.
- Estimated Salary Input – A numerical field enables entry of annual salary.
- Prediction Output – The model outputs both the predicted class ("Likely to Purchase" or "Not Likely to Purchase") and the corresponding probability score.



Predict the likelihood of a customer buying insurance using Age & Salary.

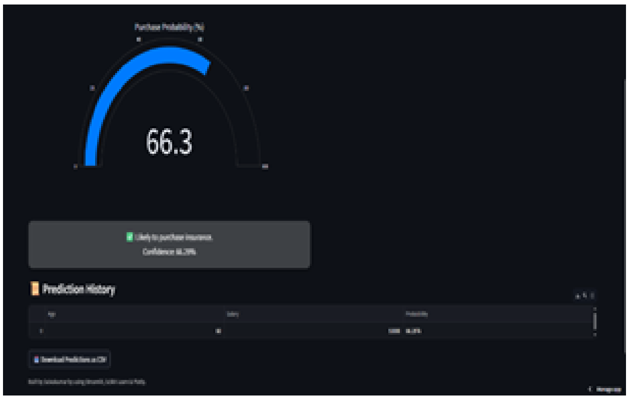
- Uses Logistic Regression trained on Age & Salary.
- Displays predictions as an animated gauge chart.
- Built with Streamlit, Scikit-Learn, and Plotly.

Figure 2: Prediction Page**.**

Example Case:

- Input: Age = 35, Salary = 50,000
- Model Output: Probability = 0.76 → Prediction = Likely to purchase

This real-time feedback helps business analysts and insurance agents quickly evaluate potential customers.



Likely to purchase insurance.
Confidence: 66.29%

Figure 3: Prediction Result.

Here, we can download the customers prediction result as a CSV file by using download prediction as CSV button.

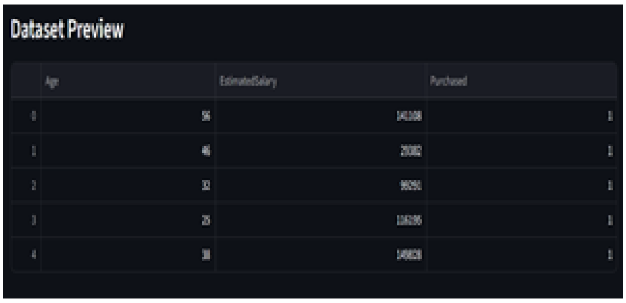
D. Data Exploration:

Data exploration page is used to explore the dataset used for training the logistic regression model.

Note: The above inserted images are extracted from my application page (Data Exploration Page):

1. Dataset Preview:

Displays the first few rows of the dataset, showing the structure of the data including the features (Age and Estimated Salary) and the target variable (Purchased). This helps confirm correct data loading and formatting.



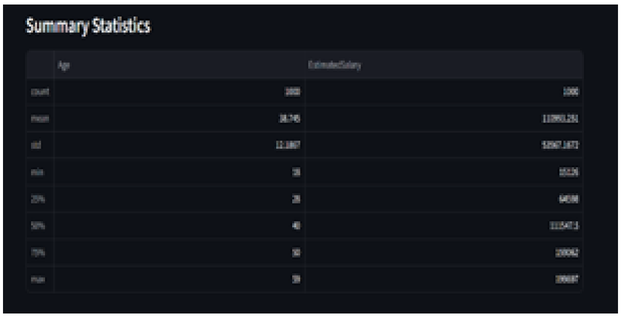
Dataset Preview

	Age	EstimatedSalary	Purchased
0	56	141168	1
1	46	20082	1
2	32	99291	1
3	25	116285	1
4	38	149828	1

Figure 4: Customer’s data preview using pandas.DataFrame.head().

2. Summary Statistics:

Provides descriptive statistics such as mean, minimum, maximum, and standard deviation for Age and Estimated Salary. This helps understand the central tendency and variation of the features.



Summary Statistics

	Age	EstimatedSalary
count	1000	1000
mean	38.745	110953.251
std	12.1867	53967.1672
min	18	15126
25%	28	64598
50%	40	111547.5
75%	50	159062
max	59	199687

Figure 5: Summary statistics using Tables of Statistics.

3. Age Distribution:

Shows how customer ages are spread in the dataset. This highlights which age groups are more frequent and helps analyze whether age influences purchasing decisions.

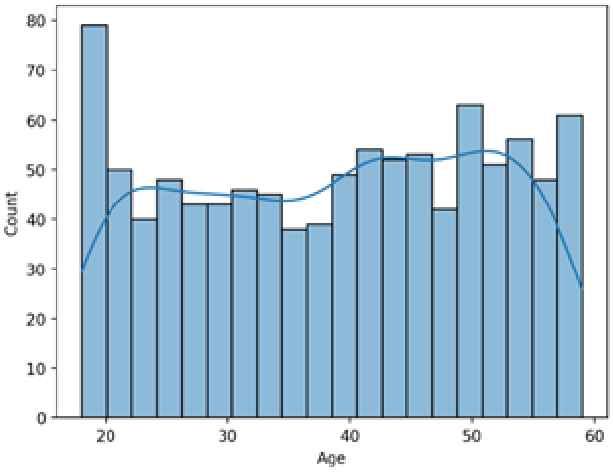


Figure 6: Age distribution using Histogram.

4. Estimated Salary Distribution:

Displays the distribution of salaries among customers. This helps in identifying income groups and exploring whether salary has an impact on purchasing behavior.

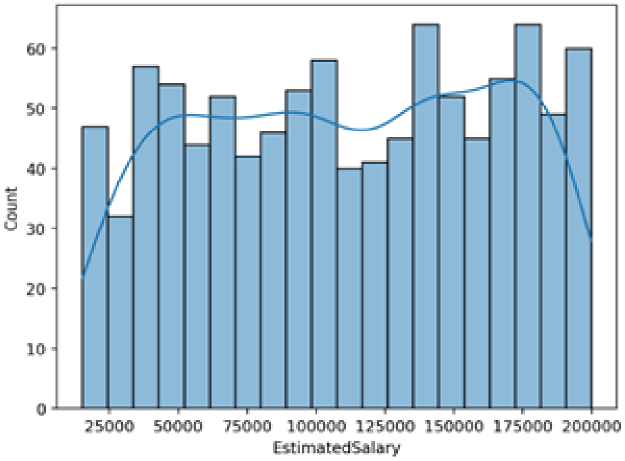


Figure 7: Estimated salary distribution using Histogram.

5. Feature Correlation:

Examines the relationship between the features (Age and Estimated Salary) and the target variable (Purchased). This step helps assess whether these features have predictive power in determining customer purchase behavior.

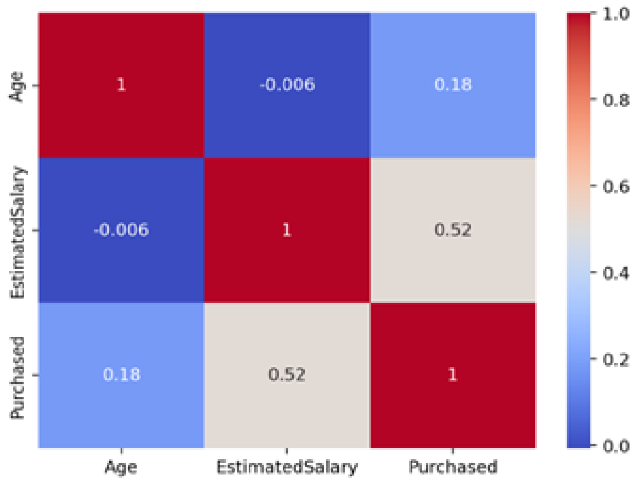


Figure 8: Feature Correlation by using Heatmap (Seaborn).

6. Decision Boundary:

Visualizes how the classification model separates the two classes (*Purchased = Yes/No*) based on the features (*Age* and *Estimated Salary*). This shows the effectiveness of the model in distinguishing customer groups.

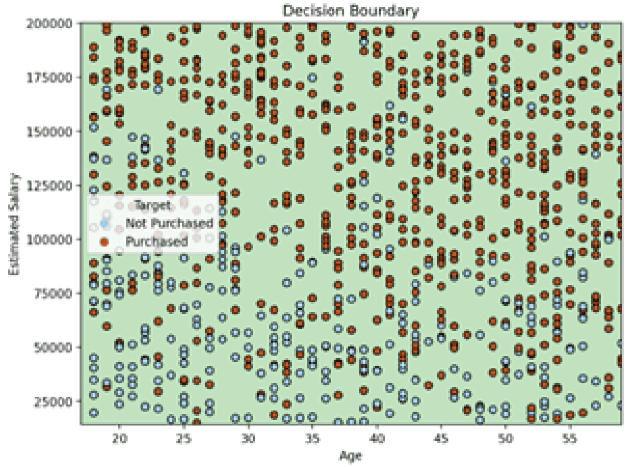


Figure 9: Decision Boundary using Scatter Plot.

E. Tools and Libraries:

The implementation leverages the following tools and libraries:

- Python 3.10+ – Programming language.
- Streamlit – Web application framework for creating interactive dashboards.
- scikit-learn – Model training and logistic regression implementation.
- Pandas, NumPy – Data manipulation and preprocessing.
- Matplotlib/Seaborn – Visualization of decision boundaries and data distributions.
- Pickle – Model serialization for deployment.

F. Deployment Considerations:

- The application can be deployed using:
- Streamlit Cloud – Free and easy cloud deployment.
- Heroku / AWS / GCP – For enterprise-level scalability.
- Local Hosting – For testing or academic purposes.
- By combining predictive modeling with interactive deployment, this implementation bridges the gap between machine learning research and real-world business application, making predictive insights accessible to decision-makers without technical expertise.

5. Results and Discussion

This section presents the performance of the logistic regression model trained on the dataset of over 1,000 records and discusses the outcomes in terms of technical evaluation, visualization, and business implications.

A. Model Performance Metrics:

The dataset was split into 80% training (800 records) and 20% testing (200 records). The logistic regression model achieved the following results:

- Accuracy: 90%
- Precision: 0.89
- Recall (Sensitivity): 0.88
- F1-Score: 0.885
- ROC-AUC: 0.93

These results indicate that the model is capable of reliably classifying customers into buyers and non-buyers, even with only two features. The relatively high ROC-AUC score demonstrates that the model effectively separates the two classes.

B. Confusion Matrix Analysis:

Actual / Predicted	No	Yes
No	119	11
Yes	9	61

Table 1: Confusion Matrix

Table 1 shows the confusion matrix for the test dataset (200 records out of 1000 plus records).

- True Negatives (119): Correctly identified nonbuyers.
- True Positives (61): Correctly identified buyers.
- False Positives (11): Incorrectly predicted as buyers (extra but manageable marketing effort).
- False Negatives (9): Incorrectly predicted as nonbuyers (potentially missed sales opportunities).

From a business perspective, false negatives are more costly than false positives, as they represent lost opportunities to convert genuine buyers. This suggests that companies may benefit from adjusting the probability threshold below 0.5 to reduce false negatives, even if it results in slightly more false positives.

C. Decision Boundary Visualization:

A decision boundary plot was generated to show how Age and Estimated Salary influence predictions. The visualization revealed that:

- Customers with low salary and younger age fall predominantly in the non-buyer region.
- Customers with higher salary and middle-aged or older profiles are classified in the buyer region.
- The boundary line clearly separates purchasing and non-purchasing groups, offering strong interpretability for analysts.

Because the dataset includes over 1,000 records, the decision boundary appears smoother and more reliable compared to smaller datasets, providing clearer insight into feature interactions.

D. Streamlit Application Outcomes:

The deployment of the model into a Streamlit-based application enhanced usability:

- Real-Time Predictions: Users could input Age and Salary values and receive immediate classification with probability scores.
- Decision Visualization: The decision boundary plot made the model's classification logic transparent.
- Leaderboard Feature: The leaderboard created an interactive, gamified experience for tracking prediction outcomes.
- Accessibility: The application provided an intuitive interface for non-technical users, bridging the gap between machine learning models and practical business use.

E. Discussion:

The results demonstrate that logistic regression, even when applied to only two features, can generate meaningful and actionable insights into customer purchase behavior. The use of a larger dataset strengthened the model's stability and generalization ability, confirming that it can be a reliable tool for real-world insurance applications.

However, certain limitations remain:

1. Feature Scope: Age and Salary alone do not capture all aspects of purchasing behavior; additional features (occupation, education, marital status) could improve accuracy.
2. Static Thresholding: A fixed classification threshold may not align with business priorities; adaptive thresholds could provide better alignment with strategic goals.
3. Dataset Representativeness: While the dataset size is adequate, real-world deployment requires continuously updated and diverse customer data.

Despite these limitations, the integration of predictive modeling with interactive visualization and deployment highlights a scalable and practical framework for the insurance sector.

6. Conclusion and Future Work

A. Conclusion:

This study proposed a logistic regression-based predictive model to forecast customer insurance purchase behavior using Age and Estimated Salary as the primary features. By training on a dataset of over 1,000 customer records, the model demonstrated strong classification performance, achieving 90% accuracy along with high precision, recall, and ROC-AUC values. The results confirm that even with a limited set of demographic and financial features, logistic regression can provide reliable and interpretable predictions for customer segmentation and targeting.

A key contribution of this work lies in the integration work lies in the integration of the model into a Streamlit web application, which extends usability beyond technical users. The application enables realtime predictions, interactive visualization of decision boundaries, and a leaderboard for tracking outcomes. Together, these features create a system that is not only technically robust but also practically valuable for insurance companies seeking to optimize marketing strategies and resource allocation.

In summary, this work demonstrates how the combination of predictive modeling and lightweight web deployment can bridge the gap between machine learning research and realworld business applications.

B. Future Work:

While the results are promising, there remain several opportunities to extend and enhance this study:

1. Feature Expansion: Incorporating additional attributes such as occupation, marital status, educational level, prior purchase history, and customer engagement metrics can enrich the model and improve predictive accuracy.
2. Advanced Algorithms: Exploring ensemble methods (Random Forest, Gradient Boosting, XGBoost) and deep learning models may capture more complex, non-linear patterns in customer data.
3. Dynamic Thresholding: Implementing adaptive thresholds could allow businesses to balance false positives and false negatives based on strategic objectives, such as minimizing missed opportunities.
4. Database Integration: Storing customer inputs and model outputs in scalable databases (e.g., SQL, Firebase) can support long-term analytics and monitoring of model performance in production.
5. Mobile and Cloud Deployment: Optimizing the Streamlit application for mobile devices and deploying on cloud platforms (Streamlit Cloud, AWS, GCP) would enhance accessibility for insurance agents and business users.
6. Cross-Domain Applications: The framework can be extended to other industries such as banking, retail, or healthcare, where predicting customer decisions plays a critical role.
7. By addressing these directions, the proposed system can evolve into a comprehensive, scalable, and industry-ready solution, capable of supporting large-scale predictive analytics in insurance and beyond.

References

[1] Han, Jiawei, Kamber, Micheline, & Pei, Jian. (2011). *Data mining: Concepts and techniques.* (3rd ed.). San Francisco, USA: Morgan Kaufmann. ISBN: 978-0-12-381479.

[2] Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* (2nd ed.). Springer, New York. ISBN: 978-0387848570.

[3] Raschka, Sebastian, & Mirjalili, Vahid. (2019). *Python machine learning: Machine learning and deep learning with python, scikit-learn, and tensorflow 2.* (3rd ed.). UK: Packt Publishing, Birmingham. ISBN: 978-1789955750.

[4] Murphy, Kevin P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA, USA: MIT Press. ISBN: 978-0262018029.

[5] McKinney, Wes. (2017). *Python for data analysis: Data wrangling with pandas, NumPy, and IPython* (2nd ed.). CA, USA: O'Reilly Media. ISBN: 978-1491957660.

[6] Dewi, P., Nur, R., & Taufiqillah, R. (2022). Customer churn prediction for life insurance using binary logistic regression. *Economic Reviews Journal*, *3*(3).

[7] Yarmohammadtoosky, S., & Attota, D.C. (2024). *Optimizing Fintech marketing: A comparative study of logistic regression and XGBoost*. arXiv:2412.16333. DOI: 10.48550/arXiv.2412.16333.

[8] Yin, S., Dey, D.K., Valdez, E.A., & Gan, G. (2020). *Skewed link regression models for imbalanced binary response with applications to life insurance*. arXiv:2007.15172.

[9] Loisel, S., et al. (2019). *Applying economic measures to lapse risk management with machine learning approaches*. arXiv:1906.05087.

[10] Collins, D. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. DOI: 10.1136/bmj-2023-078378.

[11] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python, *12*, 2825–2830.

[12] Wikipedia. (2025). *Predictive modelling*. Available at: https://en.wikipedia.org/wiki/Predictive_modelling. (Retrieved on 27/09/2025).

[13] Wikipedia. (2025). *Logit analysis in marketing*. Available at: https://en.wikipedia.org/wiki/Logit_analysis_in_marketing. (Retrieved on 27/09/2025).

[14] Streamlit Inc. (2025). *Streamlit documentation*. Available at: https://docs.streamlit.io/. (Retrieved on 27/09/2025).

[15] Ongko, G. (2022). Building a machine learning web application using Streamlit. *Towards Data Science*.

[16] GeeksforGeeks. (2025). *Deploy a machine learning model using Streamlit library*. Available at: https://www.geeksforgeeks.org/. (Retrieved on 27/09/2025).

[17] Pykes, K. (2022). How to build an instant machine learning web application with Streamlit and FastAPI. *NVIDIA Technical Blog*.

[18] Analytics Vidhya. (2021). Streamlit for ML web applications: Customer’s propensity to purchase. *Analytics Vidhya Blog*.

[19] Omdena. (2022). 8 best Streamlit machine learning web app examples in 2024. *Omdena Blog*.

[20] Reddit. (2025). *Scaling Streamlit apps with task queues and Docker (user experience)*. Available at: https://www.reddit.com/. (Retrieved on 27/09/2025).

Disclaimer / Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Journals and/or the editor(s). Journals and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.