International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 3, Issue. 3 (July – September) , 2026 , Pages: 28 - 33
Received: 24 September 2025 ; *Accepted:* 30 June 2026 ; *Published:* 18 July 2026
https://doi.org/10.63328/IJCSER-V3RI3P4
https://www.youtube.com/watch?v=GHBdbMsjA2o
https://archive.org/details/jsp-ijcser-v3ri3p4
Original Paper : Applied Research – Experimental Research Paper

Customer Churn Prediction using Behavioural and usage Pattern Analysis System

M Swetha [1] , Kandukuri Gowtham [2]

[1] Department of MCA, PSCMR College of Engineering and Technology, Vijayawada
[2] Department of Computer Applications , PSCMR College of Engineering and Technology, Vijayawada, Andhra Pradesh, India; midasalaswetha12@gmail.com, , kandukurigowtham48@gmail.com

** Corresponding Author : Kandukuri Gowtham ; kandukurigowtham48@gmail.com*

Abstract: In today's kind of competitive business world, keeping existing customers is often more cost-effective than going out and grabbing totally new ones. If an organization can notice churn earlier, it can act first, rather than sitting and waiting then reacting a bit late, and in general customer satisfaction gets a stronger push. In this project, we build a machine learning system for customer churn prediction, it flags customers who are more likely to leave a service , by looking at their behaviour signals, how they actually use the service, demographic details, and also account related information. The whole workflow starts with data prepossessing, where we clean up and reorganize the dataset , then we do exploratory data analysis, just to get a better feel for customer behaviour and uncover patterns that might otherwise remain kind of unnoticed. After that, several machine learning classification algorithms are trained and tested, so we can choose the model that produces the most accurate and dependable churn forecasts. For performance checking we rely on familiar evaluation metrics like accuracy, precision, recall, F1-score, and ROC-AUC, which makes comparing results simpler and lowers the "it just seems better" bias. To avoid that "black box" feeling during prediction, the system also uses SHAP (SHapley Additive exPlanations). SHAP kind of helps interpret what each variable is contributing, by showing how different features shift the churn outcome for each individual customer.

Keywords: Customer Churn Prediction, ML, Behavior Analysis, XAI, SHAP, customer retention , Predictive Analytics.

1. Introduction

In today's digital economy, companies work in markets that are highly competitive, where customer retention has become kind of a key factor in driving long-term results. Of course, getting new customers matters too, but keeping the customers you already have is often more cost-effective, and it keeps feeding stable revenue growth. Because of this, organizations are more and more interested in understanding customer behaviour and figuring out what actually steers customer loyalty. One big problem that service providers face is customer churn , meaning when customers decide to stop using a company's products or services. This churn can happen for lots of reasons like dissatisfaction with the service quality [1], concerns about prices, better deals from rivals, or even shifts in what customers prefer over time. If an organization can spot customers who are likely to drift away before it really happens, then it can act with retention strategies that are personalized , like sending targeted offers, improving customer support, or building customized service plans.

That earlier , proactive action reduces the revenue spill , but also helps build trust, and it tends to improve business performance overall. More recent progress in machine learning now makes it practical to inspect huge customer data sets and find subtle patterns that are hard to reveal using classic statistical tools [2]. When machine learning models train on past customer records, they can estimate the chance of churn using demographic details, service usage, billing history, account activity, and different behavioural traits. With these predictive abilities, organizations are able to plan smarter choices and refine customer relationship management , not just react after the problem shows up.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P4

2. Methodology

The proposed customer churn prediction system follows sort of a systematic workflow to identify customers who are likely to stop using a service. The whole thing happens in a kind of sequence of stages, data collection, data preprocessing, feature extraction [3], model training, prediction, and then result storage. Each step helps the prediction model become more accurate and more trustworthy , in a practical way.

2.1. Data Collection

To start, we collect customer data from dependable sources. The dataset usually covers customer demographic information, account details, service usage patterns, billing records, contract information, and also historical churn status. In other words, these inputs are the basis for training and for evaluating the machine learning model.

2.2. Data Preprocessing

The gathered data can have missing entries , duplicate records, or formats that don't quite match, and that can weaken model performance. So preprocessing is done to make data quality better. The primary tasks include: - Dealing with missing values and removing duplicate records [4]. Turning categorical attributes into numerical values using encoding approaches. - Normalizing, or scaling numerical features when it's needed. - Arranging the data into a steady format that most machine learning algorithms can use smoothly.

2.3. Feature Extraction

After preprocessing, the relevant customer attributes are picked out for model development. Feature extraction changes the processed data into a usable form, sort of like a translation, so machine learning models can benefit from it. Methods like feature selection, encoding, and scaling [5], keep improving prediction accuracy while cutting away extra complexity that otherwise gets in the way.

2.4. Model Training

The prepared dataset is split into a training chunk and a testing chunk, sometimes a bit uneven, depends on what ends up working best in real practice. The machine learning algorithms we picked get trained on the training data so they can slowly learn the inner signals that relate to customer churn [6]. In this phase the models begin to catch faint links between customer traits and the likelihood of leaving the service, I mean, it starts subtle at first, honestly, like barely there but still somehow noticeable.

2.5. Prediction and Classification

Once the training is done , the model gets looked at using fresh records it hasn't seen before. Before any real prediction, the exact preprocessing routine is run again , and the feature extraction part is redone for the data that just came in. After that, it assigns every customer into either Likely To Churn or Not Likely To Churn, and it also provides a confidence score, so you can tell how sturdy or dependable the outcome is kind of, basically.

2.6. Result Storage

The final prediction output, along with the customer details and those confidence values, is saved into the database. These stored entries can be used for dashboards and reporting, ongoing supervision of performance, and even for future planning [6]. They also help an organization observe customer behaviour more closely and then run retention actions that chip away at churn little by little, not all at once.

3. Related Work and System Architecture

The proposed Customer Churn Prediction System is meant to, in a sort of automatic way, flag customers who might stop using a service. It does this by looking at their behavioural patterns, service usage, and account related information. In general, the architecture has a few linked modules that kind of work in tandem so the system can process customer data, produce workable predictions, and also help with business decisions, even if the decision makers do not want to dig too deep into the raw data. It all starts at the User Interface, where an admin or business analyst either uploads the customer dataset or types in customer details for analysis [7]. From there the input data moves to the Data Preprocessing Module. In this step missing values, duplicate records, and small inconsistencies are dealt with. Categorical variables are turned into numerical values, and numerical features are normalized , so everything is aligned with what machine learning algorithms usually expect. After prepossessing, the cleaned dataset is passed to the Feature Engineering Module. Here the system picks the most useful customer attributes and discards unnecessary fields.

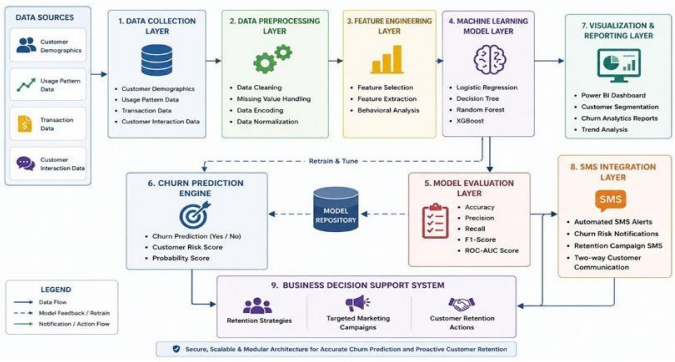


Figure. 1 System Architecture

Through feature selection and transformation, the dataset quality improves and it also supports better model prediction performance, in a more measurable way. This part is often the difference between "it runs" and "it actually predicts well". Next the

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P4

processed data gets fed into a Machine Learning Model, ya know, and from there Classification algorithms are trained using historical customer records, so the model learns the customer behaviour "leanings", the service usage patterns, the billing history and other connected relationships. The intent is basically to pull out customers who are likely to churn versus customers who are expected to remain with the service for longer [8]. When the model finishes, the Prediction Module then puts every customer into one of these two groups. Likely to Churn Not Likely to Churn [9]. Finally, the predicted label along with a confidence score is shown through the workflow continues [10] based on whatever reporting or follow up steps the organization uses.

4. Results and Discussions

This chapter presents the results obtained from the Customer Churn Prediction Using Behavioural & Usage Pattern Analysis System. The results cover model performance evaluation, behavioural pattern analysis findings, customer risk classification outputs, visualization outputs, and system execution screenshots.

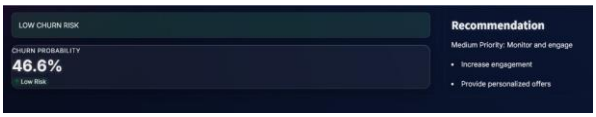
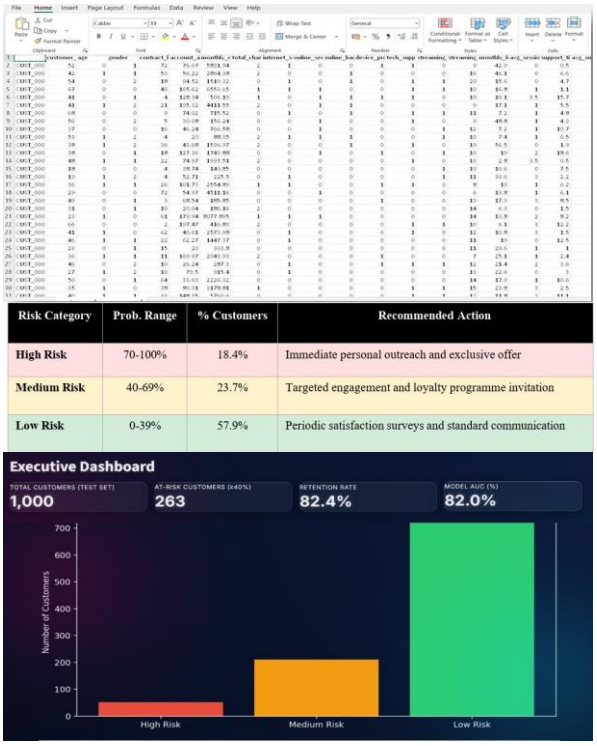


Figure. 2 Customer risk classification

The system successfully classifies customers into three distinct risk categories based on their individual churn



Risk Category	Prob. Range	% Customers	Recommended Action
High Risk	70-100%	18.4%	Immediate personal outreach and exclusive offer
Medium Risk	40-69%	23.7%	Targeted engagement and loyalty programme invitation
Low Risk	0-39%	57.9%	Periodic satisfaction surveys and standard communication

Figure. 3 Data visualizing and report

The system's visualization and reporting module generates a comprehensive set of interactive charts and analytical displays within the Streamlit web dashboard. These visualizations provide business stakeholders with clear, accessible, and actionable insights derived from the churn prediction analysis.

4.1. Challenges and Limitations

Even though the proposed customer churn prediction system shows reliable performance, there are a few challenges and limitations that can, make it just a bit less effective when used in everyday real world setups [11], especially outside the neat conditions of testing

4.2. Data quality and availability

The prediction model performance relies heavily on how good the dataset is. Missing values, duplicate records, inconsistent information, and simply not enough customer data can lower prediction accuracy quite a lot. Also, collecting a large [12], balanced and current dataset is often hard, sometimes it takes too much time, or the data is not fresh enough.

4.3. Class imbalance

In most churn datasets, non-churning customers are usually much more numerous than churners. This kind of skew can push the model to prefer the majority group, which in turn reduces its ability to correctly flag the customers who are likely to leave.

4.4. Shifting customer behaviour

Customer tastes, market dynamics, and company rules keep changing over time. A model trained on older patterns may lose strength, unless it is refreshed regularly using newer customer information. Otherwise, the whole thing can drift, and the predictions get less aligned with what is happening now, not yesterday.

4.5. Feature selection

Picking the most relevant customer attributes is important, but also difficult. If we add irrelevant or redundant features, model quality may fall. But if we remove useful variables, predictions can become inaccurate, or strangely optimistic , depending on what was left out.

4.6. Model interpretability

Even if machine learning models produce accurate results, some of them still behave like black box models, meaning , they do not naturally show their reasoning. Without the right explanation methods, it can be tricky for business users to grasp why a specific customer has been predicted to churn. Explainable AI approaches, like SHAP, improve clarity.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P4

5. Conclusion and Future Work

The fast expansion of digital communication channels has greatly contributed to the acceleration of misinformation proliferation, thereby posing a serious risk to societies. Human detection of fake news is not only very challenging and almost impossible, but also slow, and it totally depends on human perception, which makes it unreliable. This paper suggests a misinformation detection system that utilize deep learning methods. The system we have developed undergoes Natural Language Processing (NLP) techniques for text data pre-processing, and it uses Long Short-Term Memory (LSTM) networks for performing the classification task. The model does a very good job at extracting textual features, semantic connections, and contextual senses to be able to judge whether the input text is true or false. Tests done at various stages were able to show that the system has the potential to produce results with very low errors in accuracy, sensitivity, and specificity giving an overall good score of 0.9 in F-measure, thus establishing its efficiency in detection of misinformation.

This system will work automatically, save a lot of human resources, and still manage to give us efficient solutions to the problem of live identification of such information that is leading people astray. It will make human resources spent on fact-verification check very little and make people have more confidence in platforms of digital information. This work shows that deep learning methods are capable of making a great impact in the fight against fake news and in sustaining the dependability of online data. The next version of the system is expected to have the ability of recognizing the words through visual and other forms of possibly encoded data such as video or audio and vice versa, and also deliberately lumber to the call of different languages so as to have a wider and more accurate detection of misinformation. The proposed customer churn prediction system gives an effective way to spot customers who might decide to discontinue a service. Still, there is room to make it a bit stronger by pushing its prediction quality, scalability, and overall usefulness in real situations, even if the current version is already on a good level.

References

[1]. M. Z. Alotaibi and M. A. Haq, "Customer Churn Prediction for Telecommunication Companies Using Machine Learning and Ensemble Methods," *IEEE Access*, vol. 12, pp. 45678–45695, 2024.

[2]. L. A. Parulekar, A. Patil, S. Kulkarni, and P. Deshmukh, "Customer Churn Prediction," in *Proceedings of the International Conference on Computing, Communication and Intelligent Systems (ICCCIS)*, 2024, pp. 215–221.

[3]. P. S. Ramya, K. Sreenivasulu, and P. Ravi Kiran, "Customer Churn Prediction Using Machine Learning and Deep Learning Techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 13, no. 2, pp. 120– 132, 2025.

[4]. A. Goyal, R. Sharma, P. Gupta, and S. Verma, "Developing an Effective Churn Prediction Model for Telecommunications: Enhancing Customer Retention through Advanced Machine Learning Techniques," *Expert Systems with Applications*, vol. 245, Art. no. 123456, 2026.

[5]. J. Hopfield et al., "Rigorous bounds on the storage capacity of the dilute hopfield model," Proceedings of the National Academy of Sciences, vol. 79, pp. 2554–2558, 1982.

[6]. Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.

[7]. R. K. S, T. Y, and S. C. U, "Implementation of Pulmonary Disease Detection Model using Deep Learning," International Journal of Research and Development in Engineering Sciences, vol. 7, no. 2, p. 19, Mar. 2025, doi: 10.63328/ijrdes-v7ri2p4.

[8]. Ravi Samikannu, Retrieval-Augmented Multi-Agent Architecture for Hallucination Reduction in Large Language Models, International Journal of Computer Science , Engineering and Artificial Intelligence , Vol. 2, Issue. 4 (October – December) , 2025 , Pages: 7 – 13. DOI : https://doi.org/10.63328/IJCSEAI-V2RI4P2

[9]. Ujval Mutyala ," Embodied Agentic AI for Autonomous Task Planning and Execution in Dynamic Environments ", International Journal of Computer Science , Engineering and Artificial Intelligence , Vol. 2, Issue. 4 (October – December) , 2025 , Pages: 22 – 30. DOI : https://doi.org/10.63328/IJCSEAI-V2RI4P4

[10]. Dereje Weyessa ," Scalable Data Deduplication for Privacy-Preserving Media Sharing in Mobile Cloud Environments ", International Journal of Computer Science , Engineering and Artificial Intelligence , Vol. 2, Issue. 2 (April – June) , 2025 , Pages: 30 – 34. DOI : https://doi.org/10.63328/IJCSEAI-V2RI2P5

[11]. N P Patnaik M , " Low-Latency Sensor Fusion Architecture for Real-Time Motorsport Diagnostics ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 1, no. 1, p. 1-7, December 2024, DOI: https://doi.org/10.63328/IJCSEAI-V1RI1P1

[12]. D Aruna Kumari , " Low-Channel Wearable EEG Seizure Detection Using Geodesic Features and Riemannian Manifold Learning ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 2, p. 15-22, April 2026, doi: https://doi.org/10.63328/IJCSEAI-V3RI2P3

How To Cite:

M Swetha, Kandukuri Gowtham , " Customer Churn Prediction using behavioural and usage Pattern Analysis System ", International Journal of Computational Sciences and Engineering Research (IJCSER) , Volume 3, Issue 3 , pp 28 - 33 , July 2026. CrossRef DOI: https://doi.org/10.63328/ijcser-v3ri3p4

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

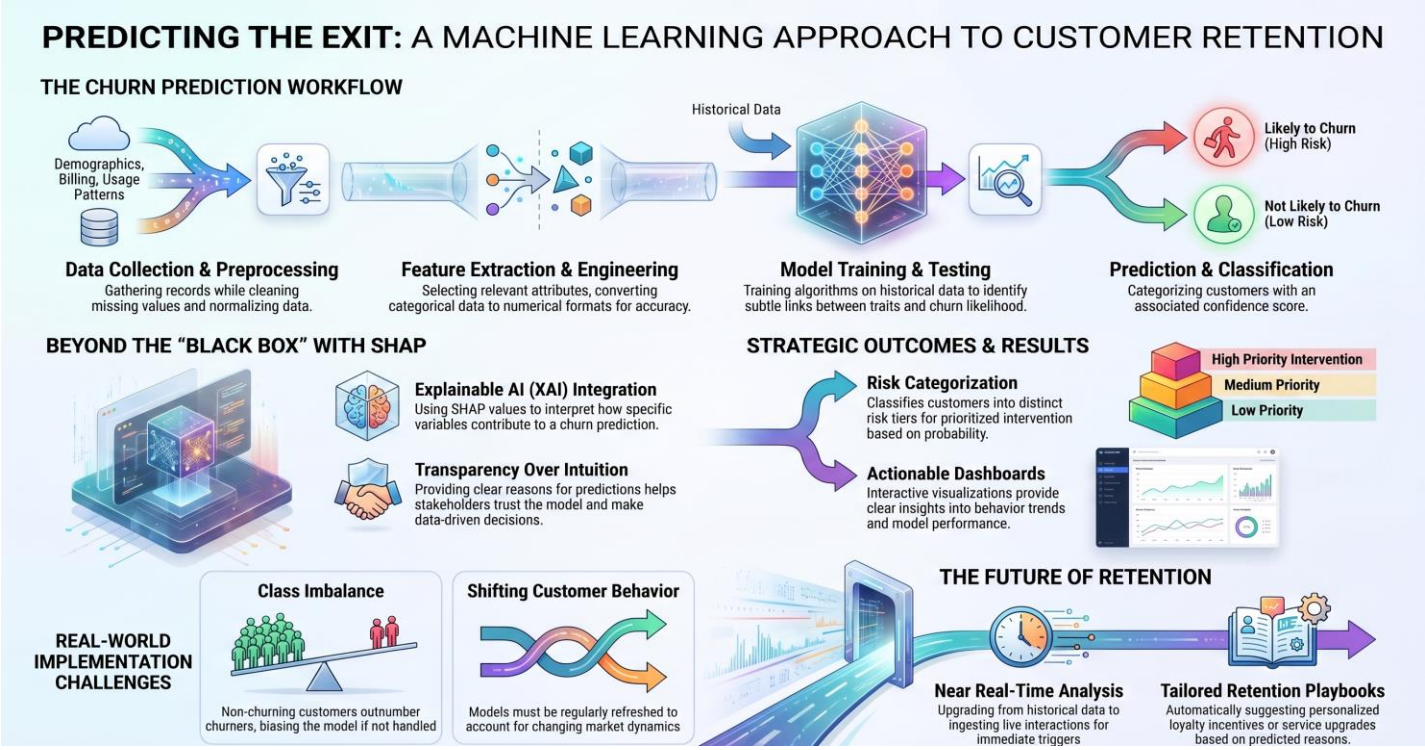
Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

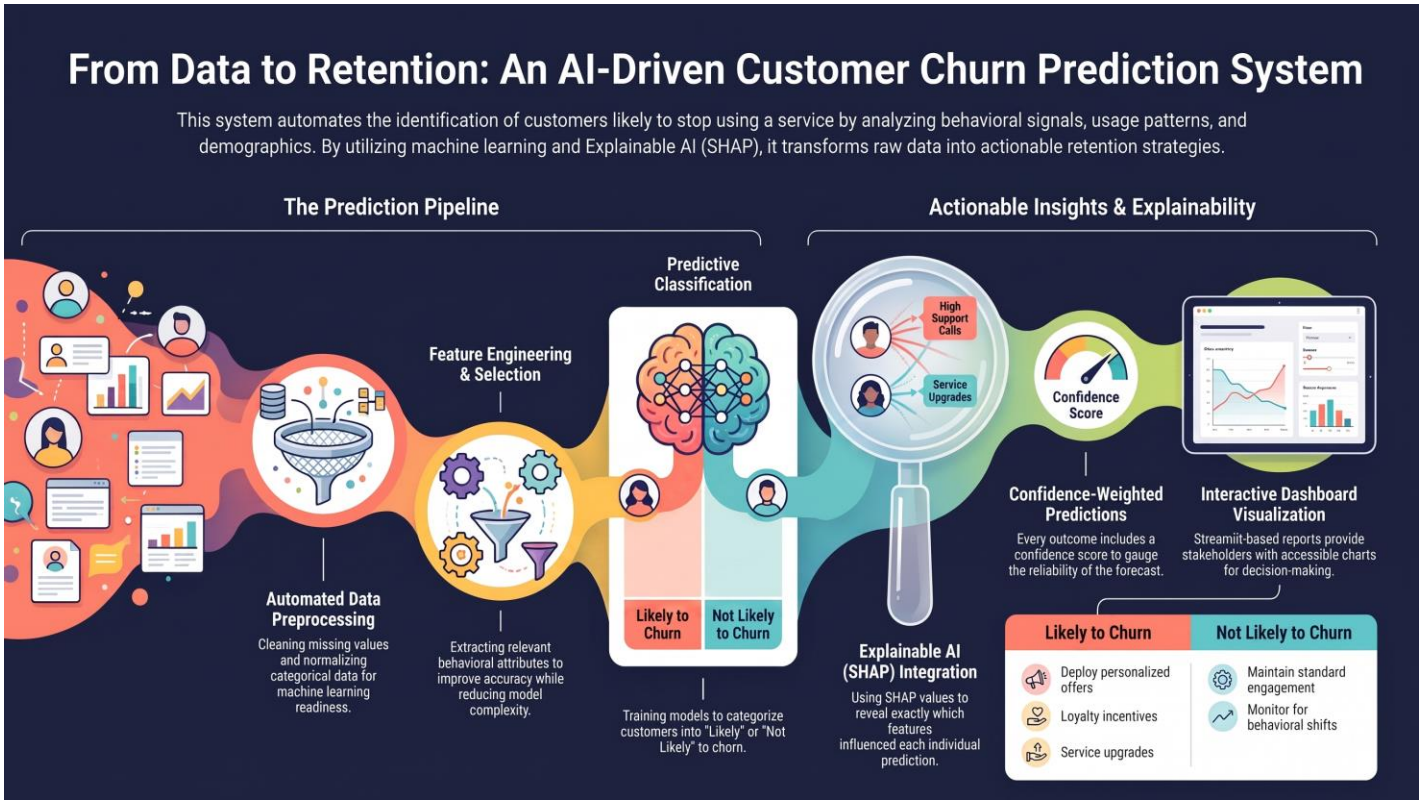
Acknowledgements

We thank everyone who inspired our work.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P4

: : Overview of the Paper for Researchers : :





ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P4