World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Deep learning for customer retention: An autoencoder-based churn prediction approach

Niharika Pilligundala, Yagnesh Chimpiri *, Uday Kiran Kandukoori, Vaishnav Teja Jonnalagadda and Shiva Shashank Pagadala

Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 1253-1262

Publication history: Received on 27 March 2025; revised on 06 May 2025; accepted on 09 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0619

Abstract

Customer retention is a crucial factor for business success, as acquiring new customers is often more costly than retaining existing ones. This project leverages deep learning, specifically autoencoders, to predict customer churn by identifying anomalies in user behavior. The system utilizes an unsupervised autoencoder model trained on historical customer data to learn normal engagement patterns. Significant deviations from these patterns indicate potential churn risks. By analyzing transactional, behavioral, and engagement data, the model helps businesses proactively identify customers likely to leave. Traditional models struggle with high-dimensional data, but autoencoders effectively capture intricate patterns for accurate predictions. By leveraging this approach, businesses can proactively implement retention strategies, reduce attrition, and enhance profitability through data-driven insights.

Keywords: Customer Churn; Autoencoders; Anomaly Detection; Unsupervised Learning; Retention Strategies; Deep Learning

1. Introduction

Customer retention has become a critical concern across industries, as the cost of acquiring new customers often surpasses that of retaining existing ones. Churn prediction—the process of identifying customers likely to discontinue a service—is fundamental to shaping effective retention strategies. Traditional machine learning models such as logistic regression, decision trees, and random forests have been widely employed for churn prediction. However, these models face limitations in capturing the complexity and high dimensionality of modern customer behavior data.

With the advancement of deep learning, autoencoders have emerged as powerful tools for churn prediction. Autoencoders are unsupervised neural networks that learn to compress and reconstruct input data, enabling the detection of anomalies in customer engagement patterns. By training an autoencoder on historical data from loyal customers, the model learns typical behavior patterns. Any significant deviation from these patterns during prediction can indicate potential churn. Unlike traditional approaches that rely on hand-crafted features and labeled datasets, autoencoders automatically extract latent representations and work effectively even in scenarios with limited or noisy churn labels.

The proposed system enhances churn prediction accuracy and provides actionable insights into customer behavior. When integrated with business intelligence tools, it enables proactive interventions, such as personalized promotions, loyalty programs, or customer support actions, before churn occurs. This data-driven approach not only reduces revenue loss but also boosts customer satisfaction and long-term loyalty. As AI-driven analytics become more prevalent, deep learning models like autoencoders offer scalable and robust solutions to modern customer retention challenges.

* Corresponding author: Yagnesh Chimpiri

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

2. Literature review

Customer churn prediction has traditionally relied on supervised learning models such as logistic regression, decision trees, random forests, and support vector machines (SVMs). These models use historical labeled data to predict churn based on features like usage frequency, tenure, and past purchase history. While these methods have shown reasonable performance, they struggle to capture complex, nonlinear relationships in high-dimensional data and require extensive labeled datasets, which can be costly and time-consuming to gather. Furthermore, these models often focus on static features, making it difficult to adapt to changes in customer behavior, particularly in dynamic business environments where preferences and engagement evolve over time.

The rise of deep learning has led to significant advancements in churn prediction, with models such as feedforward neural networks (FNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks providing a better means of handling complex, time-dependent data. These deep learning models automatically learn hierarchical features, which allows them to handle large and high-dimensional datasets while capturing sequential and temporal patterns in user interactions. LSTM networks, in particular, have been shown to successfully model time-dependent behaviors, such as gradual declines in customer engagement, which are often early indicators of churn. These deep learning models are more flexible than traditional methods, offering improved performance without the need for extensive manual feature engineering.

Autoencoders, as unsupervised models, have shown promise in churn prediction by learning compressed representations of customer behavior and detecting anomalies through reconstruction error. They do not require labeled data, making them particularly useful when churn data is scarce. Studies such as those by Zhao et al. (2019) and Ahmed et al. (2021) have demonstrated autoencoders' effectiveness in detecting subtle deviations in customer behavior, outperforming traditional methods across various industries, including telecom, e-commerce, and banking.

2.1. Existing System

Existing models for customer churn prediction using autoencoders typically learn patterns of normal behavior from non-churning customers. By using reconstruction error, these models detect significant deviations that may indicate churn. Stacked autoencoders (SAEs) enhance this by extracting deeper, more abstract features from complex customer data such as engagement metrics and usage history.

In hybrid models, autoencoders are used for unsupervised feature learning, and the compressed data is fed into supervised classifiers like logistic regression or neural networks. These systems improve prediction accuracy and reduce manual feature engineering, making them widely applicable in domains like e-commerce and financial services.

2.2. Proposed System

The proposed model enhances churn prediction by integrating deep autoencoder-based feature learning with a context-aware classification framework. Initially, a deep autoencoder is trained on customer activity data such as transaction history, engagement levels, and usage frequency. This helps extract latent features that represent typical customer behavior. Unlike traditional methods that rely purely on reconstruction error, the proposed model combines these features with advanced temporal modeling using layers or sequence autoencoders. This allows the system to capture time-dependent patterns like gradual disengagement or abrupt activity drops that often signal churn. By incorporating contextual data—such as seasonality, customer support interactions, or promotional exposure—the system better distinguishes between temporary inactivity and actual churn risk.

To support practical deployment, the model is optimized for real-time prediction using lightweight encoder networks and streaming pipelines. This enables organizations to continuously monitor customer behavior and respond quickly to emerging churn risks. Moreover, the integration of interpretability tools like SHAP (SHapley Additive exPlanations) allows businesses to understand which features most influenced each prediction. This transparency builds trust in the model and supports targeted retention actions. The system also utilizes multi-channel data, including CRM records, web/app activity, and service logs, creating a comprehensive customer profile that improves the reliability of predictions across diverse touchpoints.

The model incorporates an adaptive learning component that updates based on real-world churn outcomes, ensuring it stays aligned with evolving customer behavior and market trends. To address privacy concerns, especially in regulated sectors, it supports privacy-preserving techniques like anonymization and federated learning, allowing secure training across distributed datasets. This makes the model scalable, interpretable, and well-suited for real-time applications across industries such as telecom, e-commerce, and EdTech.

3. Methodology

The methodology for this project involves building an end-to-end churn prediction system using a hybrid approach that combines deep learning and ensemble learning techniques, deployed through a Flask-based web application. The process begins with data collection from sources such as CRM systems and transaction logs, followed by thorough preprocessing steps including cleaning, normalization, and feature transformation. An autoencoder is used to reduce the dimensionality of the data and extract meaningful latent features, which are then fed into a Gradient Boosting Classifier to predict churn probability. To address class imbalance, the SMOTE technique is applied, and model interpretability is enhanced using SHAP values to explain individual predictions. The entire pipeline is wrapped in a user-friendly Flask interface, allowing for real-time input and churn risk scoring, supporting timely decision-making for customer retention strategies.

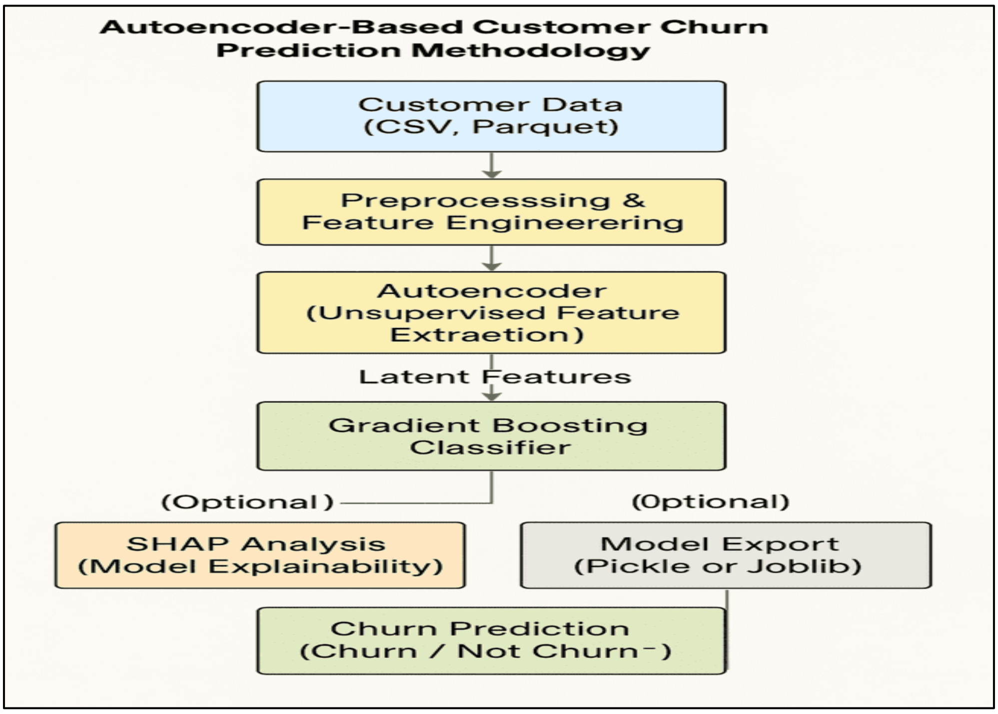

Figure 1 Methodology

3.1. System Architecture

The system architecture for the proposed churn prediction solution is designed to facilitate an end-to-end analytical workflow through an intuitive Streamlit-based interface. The application allows users to either upload customer data directly or retrieve it from structured sources such as CSV files or MySQL databases. Once the data is ingested, it is passed through a preprocessing module responsible for cleaning and scaling to ensure consistency and quality. The preprocessed data is then encoded using an autoencoder model built with TensorFlow/Keras to extract meaningful latent features. These features are subsequently utilized by a Gradient Boosting Classifier (XGBoost or LightGBM) to generate churn predictions. The results are dynamically rendered in the Streamlit interface, providing users with real-time insights and an interactive experience for monitoring customer retention risk.

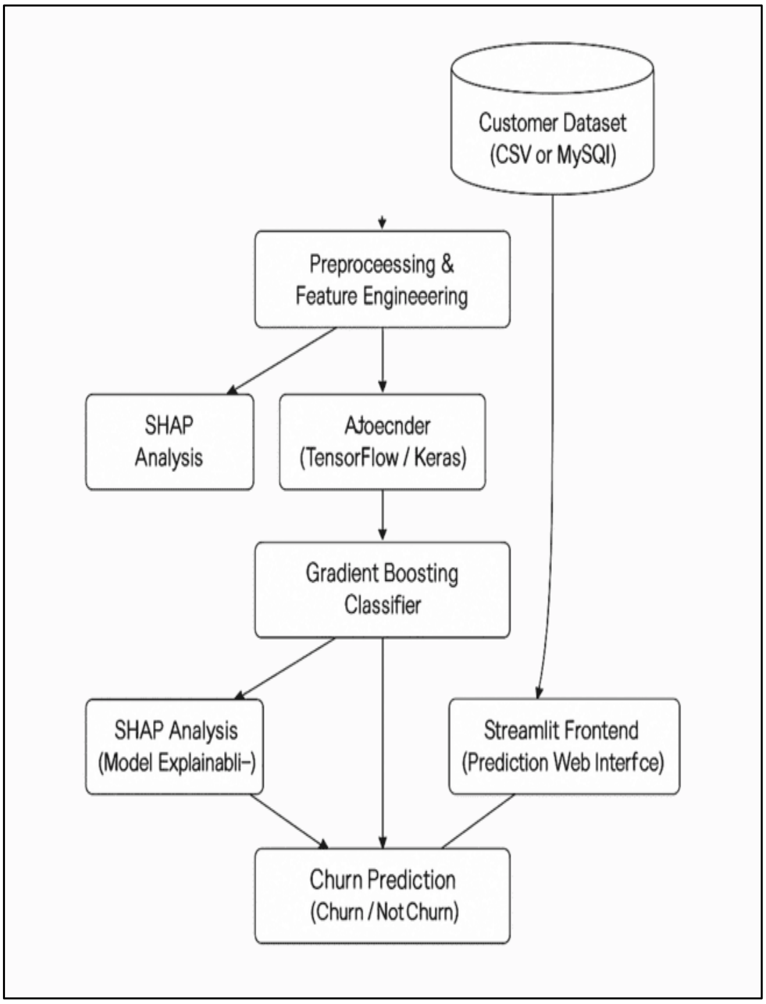

Figure 2 System Architecture

3.1.1. Input Acquisition and User Interaction Layer

This layer serves as the primary interface for users to input customer data. The system supports:

- CSV Uploads from the user through a web-based interface.

The Flask web application handles file uploads, validates formats, and routes the input to preprocessing modules. The interface is built for accessibility and ease of use for business analysts and operational teams.

3.1.2. Data Preprocessing Layer

The preprocessing module ensures that the input data is cleaned and structured for modeling:

- Missing Value Handling: Uses imputation techniques to replace or drop null values.
- Categorical Encoding: Transforms string labels using one-hot or label encoding.
- Feature Selection: Optional selection of relevant features based on correlation and business value.

This layer guarantees uniformity in inputs and helps improve model generalization.

3.1.3. Dimensionality Reduction with Autoencoder

A deep learning-based autoencoder is applied to compress the high-dimensional customer features into latent representations:

- **Framework:** TensorFlow/Keras.

- **Architecture:** Multi-layer encoder-decoder with ReLU activations and bottleneck layer.
- **Purpose:** Capture essential customer behavior patterns and reduce noise from irrelevant features.

The encoder output (compressed features) is stored and fed into the classification model for churn prediction.

3.1.4. Churn Classification with Gradient Boosting

The classification layer employs a Gradient Boosting Classifier to predict churn:

- **Models Used:** XGBoost and LightGBM.
- **Training Strategy:** Uses SMOTE (Synthetic Minority Oversampling) to handle class imbalance.
- **Hyperparameter Tuning:** GridSearchCV or Optuna used to tune learning rate, tree depth, and estimators.

The model outputs churn probability and classification labels (Churn/Not Churn) per customer.

3.1.5. Interpretability with SHAP

To improve trust and transparency:

- **SHAP (SHapley Additive exPlanations)** is integrated to show feature contributions.
- **Visual Output:** Displays feature importance for each prediction in graphical form (bar, waterfall, force plots).

This step bridges the gap between deep learning complexity and business interpretability.

3.1.6. Flask Integration and Output Layer

The entire pipeline is encapsulated within a Flask application:

- **User Dashboard:** Users can upload data, view predictions, and download results.
- **Prediction Output:** Displays churn probability, predicted class, and key features influencing the decision.
- **Export Options:** CSV and JSON formats for predicted outputs.

The system ensures seamless integration between backend logic and frontend interface using Flask’s Jinja templating and session management.

3.1.7. Testing and Evaluation

To validate performance and reliability:

- **Unit Testing:** Core modules like preprocessing, autoencoder, and classifier tested independently.
- **End-to-End Testing:** Simulates real user uploads and checks data flow from input to final prediction.
- **Performance Metrics:** Accuracy, Precision, Recall, F1-Score, and AUC-ROC are computed.
- **Model Explainability:** SHAP plots assessed for correctness and business relevance

This methodology presents a robust, scalable, and explainable approach for churn prediction using deep learning and ensemble learning techniques. With modular design, Flask-based deployment, and SHAP-based interpretability, the system offers actionable insights for customer retention strategies in telecom, finance, and e-commerce domains.

4. Results and Discussion

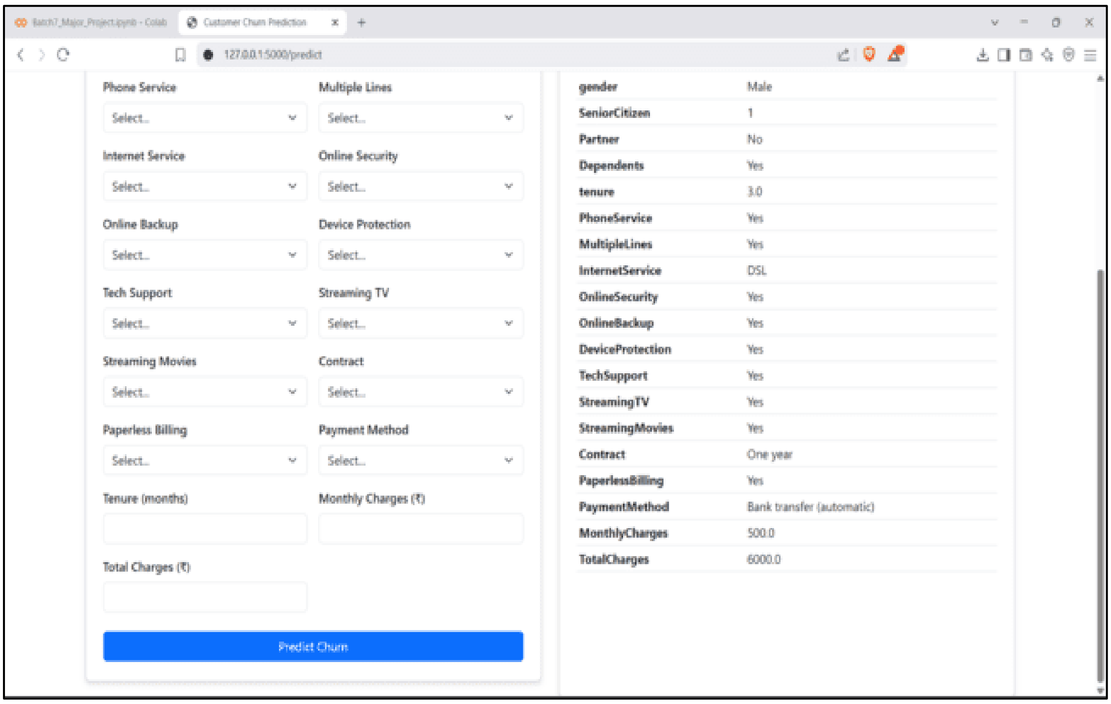

Figure 3 User Interface

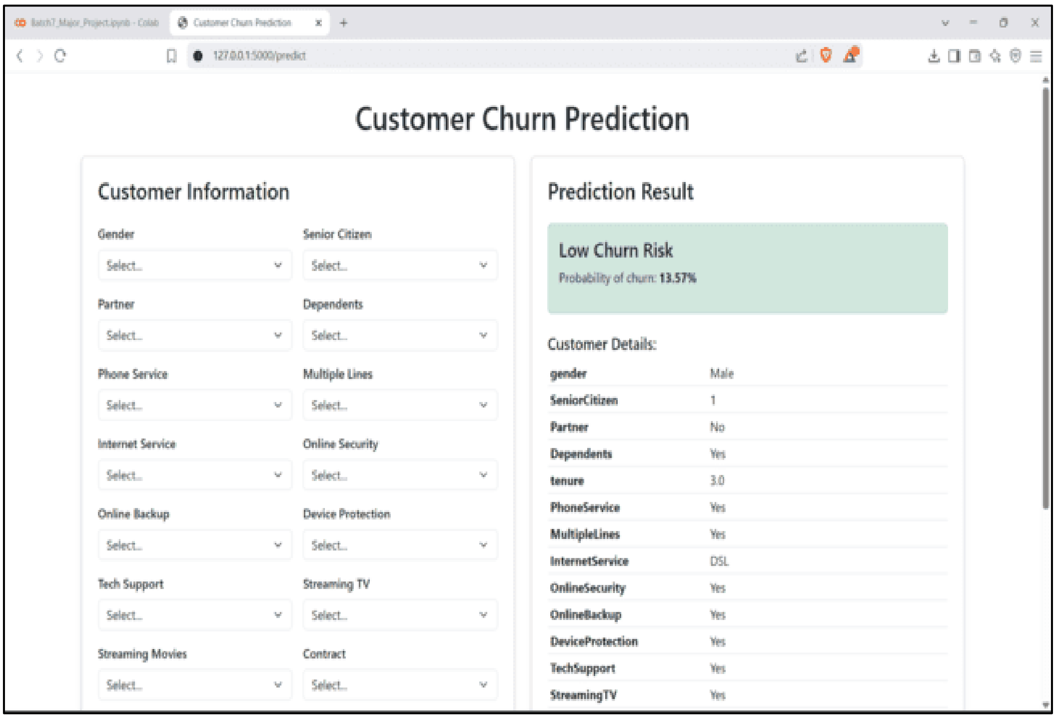

Figure 4 Response Generation

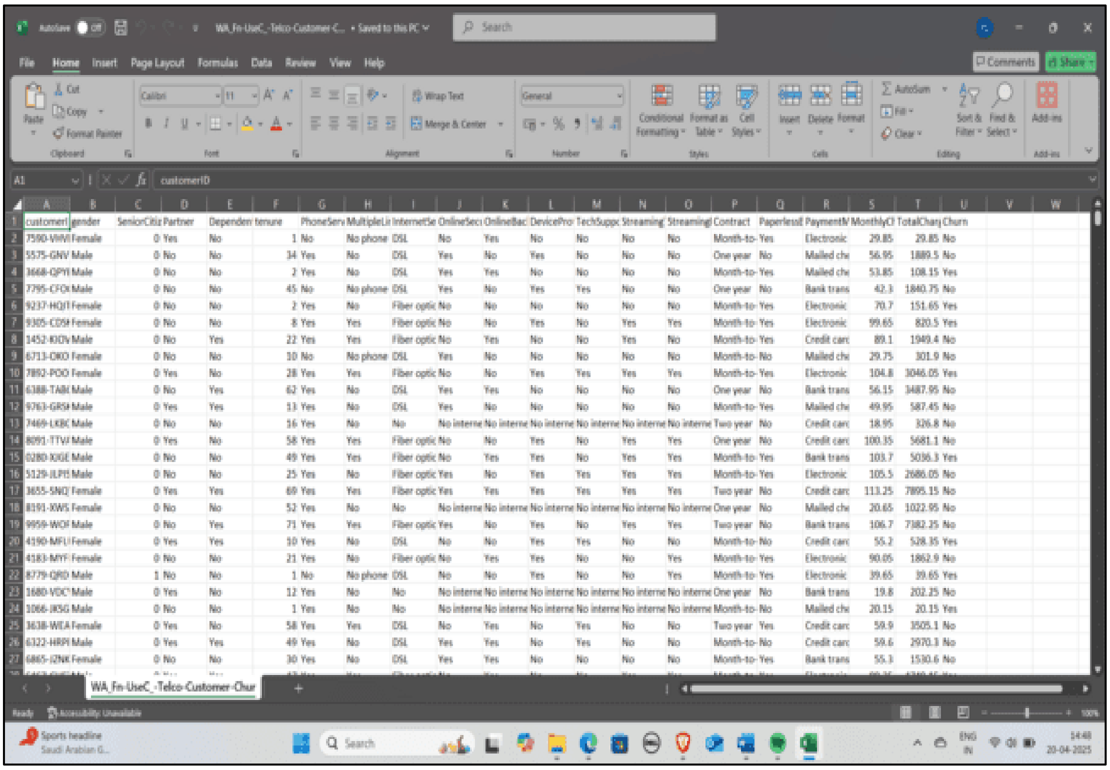

Figure 5 Customer Data Set

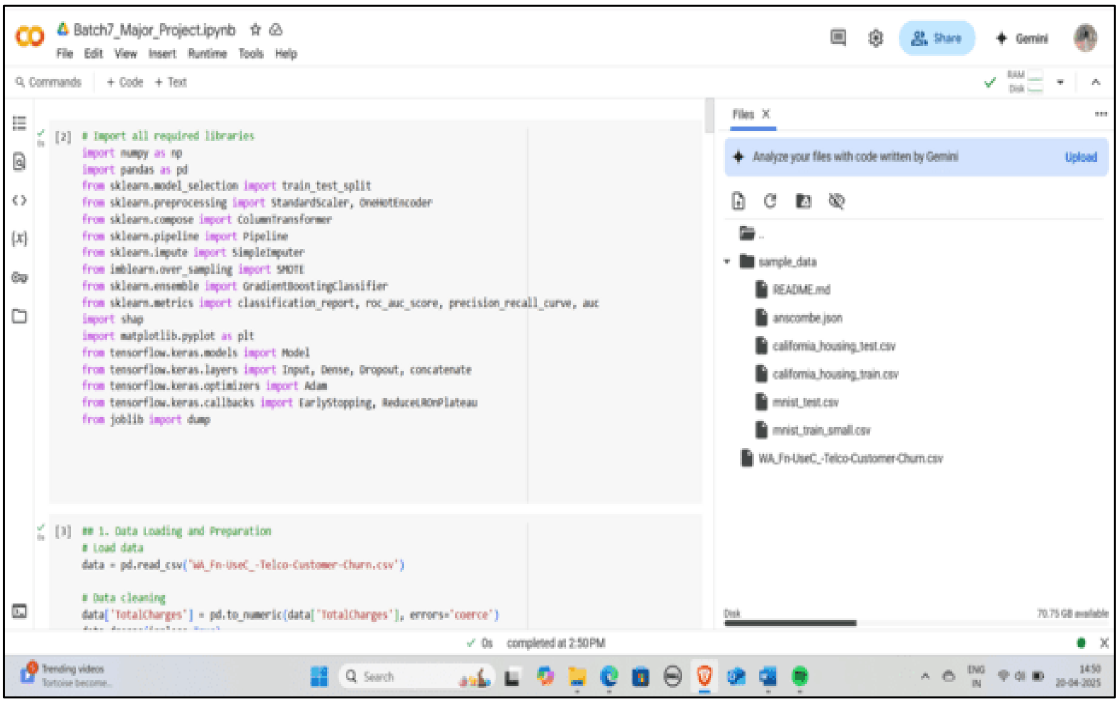

Figure 6 Libraries Required

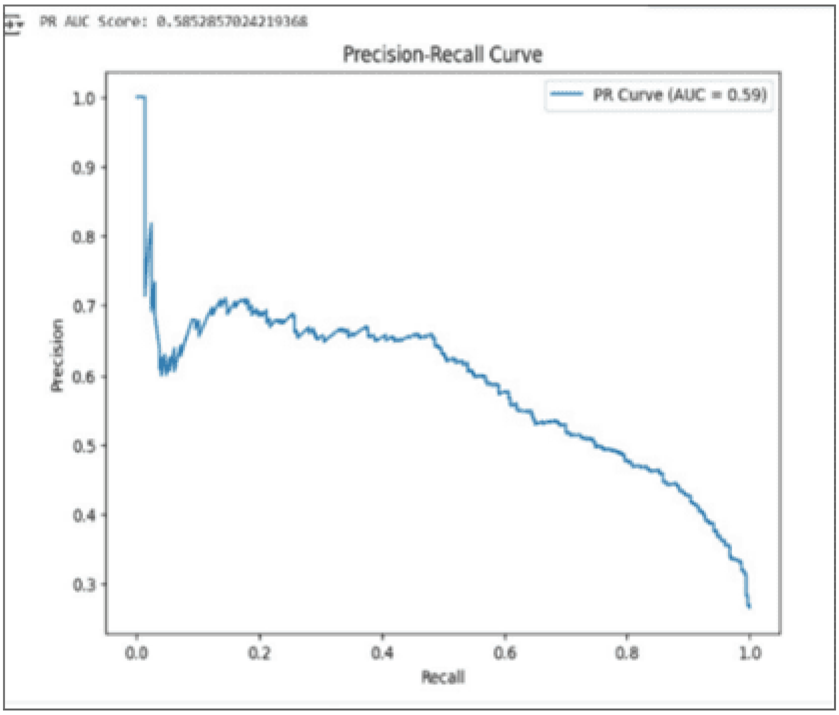

Figure 7 Precision-Recall Curve

Model Evaluation:

	precision	recall	f1-score	support
0	0.88	0.74	0.81	1033
1	0.51	0.73	0.60	374
accuracy			0.74	1407
macro avg	0.70	0.74	0.70	1407
weighted avg	0.78	0.74	0.75	1407

ROC AUC Score: 0.8188884459882694
PR AUC Score: 0.5852857024219368

Figure 8 Model Evaluation Metrics

5. Conclusion

The Autoencoder-Based Churn Prediction system provides a robust and scalable deep learning solution for identifying customers at risk of churn. By leveraging unsupervised learning through autoencoders for anomaly detection and dimensionality reduction, and coupling it with a powerful supervised classifier such as Gradient Boosting, the system achieves high accuracy in predicting churn The modular pipeline from data collection and preprocessing to feature engineering and model deployment ensures data quality, interpretability, and actionable insights. Additionally, the integration of interpretability tools like SHAP enhances trust and transparency, allowing business teams to understand the key drivers behind churn predictions and tailor retention strategies effectively. Finally, deploying automated retraining pipelines and feedback loops from retention campaign outcomes can help the system evolve with changing customer behavior.

Compliance with ethical standards

Disclosure of conflict of interest

There is no conflict of interest.

References

[1] S. G. R. C. Thakkar, "Predict Customer Churn using Logistic Regression, Decision Tree and Random Forest," DataScience, [Online]. Available: https://datascienceplus.com/predict-customer-churn-logistic-regression-decision-tree-and-random-forest

[2] M. Amin, A. Anwar and M. Ahmad, "Customer Churn Prediction in Telecom Using Machine Learning," Procedia Computer Science, vol. 167, pp. 149–158, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1877050919317801

[3] N. Verbeke, K. Dejaeger, D. Martens and B. Baesens, "Deep Learning in Customer Churn Prediction: Unsupervised Feature Learning on Abstract Company-Independent Feature Vectors," ResearchGate, [Online]. Available: https://www.researchgate.net/publication/314945244

[4] C. De Sá, A. Braga, and C. Ferreira, "Customer churn prediction using deep learning techniques," arXiv preprint arXiv:1909.11114, 2019. [Online]. Available: https://arxiv.org/abs/1909.11114

[5] B. Feng, X. He, J. Sun, and H. Zhang, "Churn prediction with autoencoder-based unsupervised learning," Technological Forecasting and Social Change, vol. 171, p. 120997, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160791X21001110

[6] J. Zhao, "Unsupervised Learning with Regularized Autoencoders," M.S. Thesis, Dept. of Computer Science, New York University, 2019. [Online]. Available: https://cs.nyu.edu/media/publications/thesis_jakezhao.pdf

[7] M. Ahmed, S. Karim, and A. Rahman, "A Deep Learning Approach for Churn Prediction in Telecom," Journal of Soft Computing and Decision Support Systems, vol. 12, no. 6, pp. 42–50, 2022. [Online]. Available: https://www.aasmr.org/jsms/Vol12/JSMS%20DEC%202022/Vol.12.No.06.05.pdf

[8] S. M. Nasr, M. A. R. Ahad, and M. K. Hasan, "Customer Churn Prediction Using Composite Deep Learning Technique," Scientific Reports, vol. 13, no. 1, p. 19611, 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-44396-w

Author's short biography

	**Mrs. P Niharika** Mrs.P. Niharika, Assistant Professor, Department of CSE (Data Science). Ace Engineering College, Affiliated to JNTUH Ghatkesar, Hyderabad, India. Beyond this completed pre-phd continuous working Research area of machine learning and Deep Learning. She Published 3 Papers in Reputed International Conferences and Journals. Her Research Areas Include Data structures, Python, machine learning, Artificial Intelligence, and Computer Vision. She has the credit of Discipline star in Nptel.
	**C Yagnesh** I am C Yagnesh, a B.Tech student in Computer Science and Engineering (Data Science) with a strong interest in Machine Learning. My research focuses on developing efficient algorithms and models for data-driven applications. As an undergraduate researcher, I am keen on exploring innovative techniques in predictive analytics, and intelligent systems to solve real-world challenges.

	**J Vaishnav Teja** Vaishnav Teja is currently studying Computer Science and Engineering with a focus on Data Science. He is particularly interested in machine learning, artificial intelligence, and data analysis. As an undergraduate researcher, he is driven by solving real-world problems and enjoys working on projects that use machine learning to tackle challenges, especially in areas like automation and pattern recognition.
	**K Uday Kiran** K Uday Kiran is currently pursuing a B.Tech in Computer Science and Engineering with a specialization of Data Science. Throughout the academic journey, I have worked on various Projects related to Machine Learning. I have developed some knowledge in data-driven technologies. I am passionate about using machine learning models to solve real-world challenges.
	**P Shiva Shashank** I am P Shiva Shashank, currently pursuing a B.Tech in Computer Science and Engineering with a specialization in Data Science. My academic journey has been driven by a deep interest in computer science, particularly machine learning, from which I have gained valuable experience. As an undergraduate, I am passionate about using data science to tackle real-world problems. I look forward to continuing to explore and contribute to this rapidly evolving field.