

A Smart Workforce Analytics System for Early Driver Attrition Detection and Retention Planning

P. Santhi¹, N. Siva Nagamani², D. Ramesh¹, S. Naresh¹

¹Assistant Professor, ²Associate Professor, ^{1,2}Department of Computer Science and Engineering (AI & ML)

^{1,2}Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

ABSTRACT

Driver attrition has become a major operational issue in ride-hailing platforms, as it directly impacts service availability, platform efficiency, and user satisfaction. Accurately identifying drivers who are likely to leave at an early stage enables organizations to implement proactive retention strategies and minimize workforce instability. Traditional approaches to attrition prediction mainly rely on survey-based analysis, basic statistical models, and rule-driven systems. However, these methods struggle to handle large-scale, high-dimensional data and fail to capture complex relationships among multiple influencing factors, resulting in limited predictive capability. To address these limitations, this work proposes a hybrid predictive framework that combines deep learning and machine learning techniques for improved accuracy and interpretability. A Long Short-Term Memory (LSTM) network is employed to learn hidden patterns and extract high-level feature representations from structured driver data. These extracted features are then passed to a Greedy Tree (GT) classifier, which performs the final classification by leveraging its efficient decision-making structure. Alongside the proposed model, baseline algorithms including Random Forest (RF), Gradient Boosting (GB), and Support Vector Classifier (SVC) are implemented for comparative performance evaluation. The system is designed as a complete web-based application using the Flask framework, enabling seamless integration between machine learning models and user interaction layers. It incorporates a secure login and authentication mechanism with role-based access control for different users such as HR and employees. User data and credentials are efficiently managed using a Lightning Memory-Mapped Database (LMDB), ensuring fast and lightweight storage operations. The backend architecture supports data preprocessing, model training, prediction, and performance evaluation, while the frontend interface allows users to input data and visualize results effectively. Experimental results demonstrate that the proposed GLSTMT model achieves superior performance compared to traditional methods, reaching an accuracy of highest accuracy on the dataset used in this study.

Keywords: Employee Attrition Prediction, Workforce Analytics, Predictive Modeling, Hybrid Learning, Role-Based Access Control.

1. INTRODUCTION

Driver attrition describes the process through which drivers exit an organization due to factors such as resignation, career transitions, or dissatisfaction with working conditions. It has emerged as a critical issue, particularly in ride-hailing and transportation industries, where workforce continuity plays a vital role in maintaining operational efficiency and service quality. Attrition becomes problematic when the rate of drivers leaving exceeds the rate of hiring, resulting in workforce shortages, increased workload on existing employees, and reduced productivity. Elevated attrition levels often lead to higher recruitment and training costs, loss of experienced drivers, and interruptions in service delivery, which significantly affect overall organizational performance as illustrated in Fig. 1. Consequently, effective monitoring and control of attrition are essential for ensuring long-term organizational stability and performance. Driver attrition can be broadly classified based on its nature and cause. Voluntary attrition occurs when drivers independently decide to leave the organization, whereas involuntary attrition arises

when the organization terminates employment. External attrition refers to drivers moving to other companies or opportunities outside the organization, while internal attrition involves transitions such as role changes or promotions within the same organization. Analyzing attrition through these categories helps organizations identify root causes and implement targeted interventions. The attrition rate is a key metric used to evaluate workforce stability and is generally computed as the ratio of the number of drivers leaving to the average number of drivers during a defined time period.

Impact of components in driving employee retention	Average economic context	Economic downturn context	Rank difference between contexts
1 (Most Impactful Factor)	Opportunity	Work	+1 More Important
2	Work	People	+1 More Important
3	People	Opportunity	-2 Less Important
4	Rewards	Rewards	No Change in Importance
5 (Least Impactful Factor)	Organization	Organization	No Change in Importance

Fig. 1: Driving Employee Attrition Strategy

Recent data emphasizes the magnitude of this challenge. Research findings suggest that nearly one-third of newly hired drivers leave within their first six months. Data from the Job Openings and Labor Turnover Survey (JOLTS) indicates that approximately 3 to 4.5 million employees resign from their jobs every month in the United States. According to the Bureau of Labor Statistics, attrition rates can reach up to 57.3% in certain sectors, while the overall industry average remains close to 19%. Furthermore, the Society for Human Resource Management (SHRM) reports that the average cost per hire is around USD 4,129. Maintaining a retention rate of approximately 90% is generally considered optimal, suggesting that organizations should aim to keep attrition levels below 10% to achieve sustainable growth and operational efficiency. The research objectives are as follows

- To design and implement a hybrid predictive model that integrates Long Short-Term Memory LSTM and GT techniques for accurate driver attrition prediction.
- To perform a comparative analysis of the proposed model with baseline algorithms such as RF, GB, SVC using standard evaluation metrics.
- To develop a Flask-based web application with secure login and role-based access control, supported by LMDB for efficient user data management.
- To provide interpretable insights through visualization and performance evaluation modules, enabling organizations to make data-driven decisions for reducing attrition rates.

2. LITERATURE SURVEY

Recent advancements in ride-hailing platforms (RHPs) have focused on optimizing operational efficiency through pricing strategies, commission rate modeling, driver behavior analysis, and platform competition. Researchers are increasingly exploring mathematical modeling, system dynamics, and data-driven approaches to improve platform profitability, service quality, and sustainability.

2.1 Pricing and Commission Rate Optimization Models

Several studies have focused on optimizing commission rates and pricing strategies to improve RHP operations. Deng et al. [1] developed queue-theory-based mathematical models to analyze the

relationship between commission rates, driver switching behavior, and platform performance under different market conditions such as duopoly and competition.

Cohen and Zhang [2] proposed an endogenous model for two-sided ride-hailing platforms to determine optimal service pricing and commission rates analytically. Similarly, Sun and Ertz [3] introduced a system dynamics framework, identifying commission rates, order pricing, and investment levels as critical factors influencing platform efficiency.

Zhong et al. [4] examined competition between ride-hailing platforms and traditional taxi services, proposing pricing strategies under both regulated and unregulated environments to enhance operational outcomes.

2.2 Operational Strategies and Platform Competition

Research has also explored operational strategies such as order assignment and platform competition. Sun et al. [5] identified that different order assignment strategies significantly impact platform performance and efficiency.

Xu et al. [6] proposed a generalized fluid framework incorporating commission rate mechanisms to model ride-hailing systems more effectively. Ke and Qian [7] developed an optimal pricing scheme to improve operational efficiency.

Bandiera et al. [8] introduced a mathematical model to simulate interactions between service providers and customers, while Xi et al. [9] proposed a multi-leader–multi-follower model to analyze competitive behavior among platforms.

Yao and Zhang [10] addressed pricing challenges in multi-modal transport systems using a many-to-many matching framework, improving allocation efficiency across different transport modes.

2.3 Regulatory Policies and Market Dynamics

Regulatory policies play a crucial role in shaping ride-hailing operations. Sun et al. [11] investigated the impact of different regulatory frameworks on platform growth and operational performance. Their study highlights how regulations influence pricing, driver participation, and overall market dynamics. Despite these contributions, there remains a lack of research addressing the trade-off between platform profitability and customer service quality. Additionally, the interaction between commission rates, driver behavior, and system parameters is not fully understood.

2.4 Emerging Trends and Research Gaps

Although existing studies provide valuable insights, several limitations remain. Many models assume a stable number of drivers and simplify driver behavior, ignoring long-term variations in driver supply [12]. Future research should consider dynamic changes in driver availability and their impact on pricing strategies.

Furthermore, most studies focus on competition between platforms, while real-world scenarios often involve both competition and cooperation (coopetition) among platforms [13], which requires further exploration.

Zhang et al. [14] highlighted the emergence of electric ride-hailing vehicles (ERVs) as a sustainable mobility solution. Their study shows that electric vehicles significantly reduce carbon emissions compared to traditional fuel-powered vehicles, emphasizing the role of green technologies in future ride-hailing systems.

2.5 Key Contributions

- Development of queue-theory-based and analytical models to optimize commission rates and understand driver switching behavior in dynamic ride-hailing markets [1], [2].
- Identification of critical operational factors such as pricing, commission rates, and investment levels using system dynamics modeling to improve platform performance [3].

- Design of pricing strategies under competitive and regulatory environments, including competition with traditional taxi systems [4].
- Exploration of order assignment strategies and their impact on platform efficiency and service quality [5].
- Proposal of advanced mathematical frameworks such as fluid models, multi-leader–multi-follower models, and many-to-many matching approaches for optimizing platform operations and pricing [6]-[10].
- Analysis of the impact of regulatory policies on platform growth, driver participation, and operational efficiency [11].
- Identification of research gaps, including limited understanding of trade-offs between platform profitability and customer service quality, and insufficient modeling of dynamic driver behavior [12].
- Recognition of co-competition (competition + cooperation) among platforms as an emerging research direction [13].
- Introduction of sustainable mobility concepts, particularly the adoption of electric ride-hailing vehicles (ERVs), highlighting their potential to reduce carbon emissions and improve environmental impact [14].

3. PROPOSED METHODOLOGY

The system architecture integrates data ingestion, preprocessing, machine learning, visualization, and user interaction into a unified Flask-based framework designed for employee analytics and predictive intelligence. Initially, the Driver-Employee-Attrition dataset is processed to extract statistical insights and structured features required for downstream tasks. The processed data is simultaneously utilized for both analytical visualization and multi-model training pipelines to ensure parallel insight generation. A central Flask application coordinates all modules, including routing, authentication, and role-based dashboards for HR and employee users. Secure data handling is achieved using an LMDB key-value store combined with hashing and serialization mechanisms. The MLModelManager incorporates multiple algorithms such as Random Forest, Gradient Boosting, Support Vector Classifier, and the proposed Greedy LSTM Tree model to perform multi-target prediction tasks. Visualization modules generate interpretable graphical outputs, while performance evaluation modules compare all models using standard metrics as illustrated in Fig. 2.

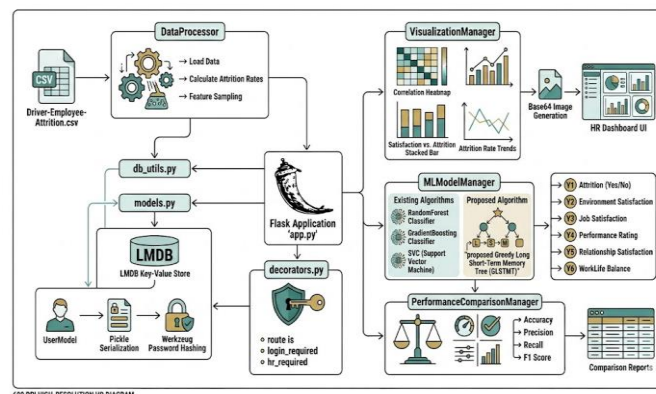


Fig. 2: Proposed system architecture.

Step 1: Data Acquisition and Preprocessing

- The system begins by loading the Driver-Employee-Attrition dataset using the DataProcessor module, ensuring structured ingestion of CSV data.

- It performs data cleaning, removes irrelevant attributes, and computes key statistics such as attrition rates and demographic distributions.
- Feature sampling and preparation are carried out to generate structured inputs for both visualization and machine learning modules.

Step 2: Data Storage and User Management

- User data is managed using an LMDB key-value database integrated through db_utils and models modules.
- The system applies password hashing and serialization techniques to ensure secure storage and retrieval of credentials.
- It supports user registration, authentication, and role assignment (HR or Employee) for controlled access.

Step 3: Application Control and Access Management

- The Flask application acts as the central controller that connects all modules and manages request routing.
- Role-based access control is enforced using decorators such as login_required, hr_required, and employee_required.
- Based on authentication, users are redirected to appropriate dashboards, ensuring secure and role-specific interaction.

Step 4: Machine Learning Model Training

- The MLModelManager handles training using multiple algorithms including RF, SVC, GBC and the proposed GLSTM Tree model.
- It performs preprocessing steps such as label encoding and feature scaling to convert raw data into model-ready format.
- The system supports multi-target prediction including Attrition, Environment Satisfaction, Job Satisfaction, Performance Rating, Relationship Satisfaction, and Work-Life Balance.

Step 5: Performance Evaluation and Comparison

- The PerformanceComparisonManager evaluates all trained models across all targets using metrics such as accuracy, precision, recall, and F1 score.
- It performs systematic comparison between Random Forest, Gradient Boosting, Support Vector Classifier, and Greedy LSTM Tree models.
- The results are structured into comparison reports to identify the most effective model for each prediction task.

Step 6: Visualization and Output Generation

- The VisualizationManager generates analytical visualizations including correlation heatmaps, satisfaction analysis, attrition distribution, and trend patterns.
- All visual outputs are converted into Base64 format for seamless integration into the web-based dashboard.
- The final system presents predictions, insights, and performance metrics through an interactive HR dashboard, supporting data-driven decision-making.

4. Result Description

The results of the research demonstrate the successful implementation of an AI-driven HR attrition prediction system integrated with a Flask-based web interface. The research provides a seamless workflow from user registration to HR analytics and real-time attrition prediction. Each component, including data visualization, model training, and predictions, is presented through interactive screens and dashboards. The system allows HR managers to analyse employee trends, understand key attrition

factors, and make informed decisions based on predictive analytics. Performance evaluation of different models validates the effectiveness of the chosen algorithms in predicting employee turnover accurately. Fig. 2 shows the registration page where new users, including employees or HR personnel, input their personal and professional details. The form captures essential data fields required for subsequent analysis and prediction. User input is validated to ensure data consistency and integrity. Fig. 3 this dashboard presents comprehensive insights into employee attrition statistics. Various charts, graphs, and tables summarize employee demographics, satisfaction levels, and other key features. The dashboard integrates outputs from the preprocessing and EDA steps to provide actionable analytics for decision-making

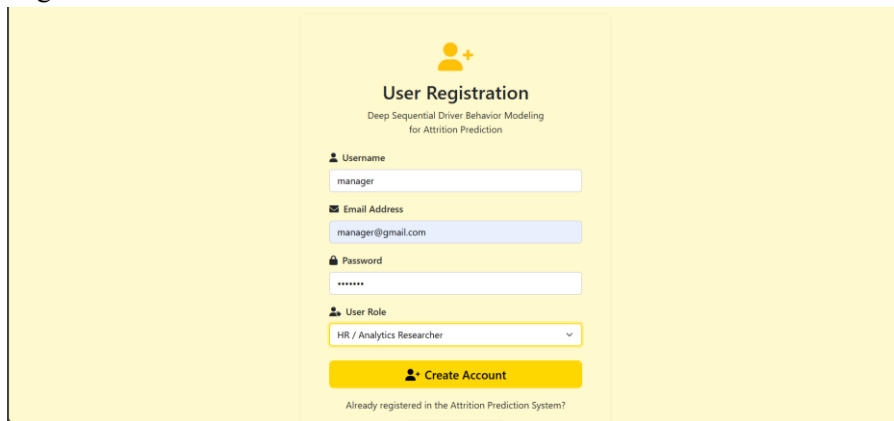


Fig. 2: HR register page

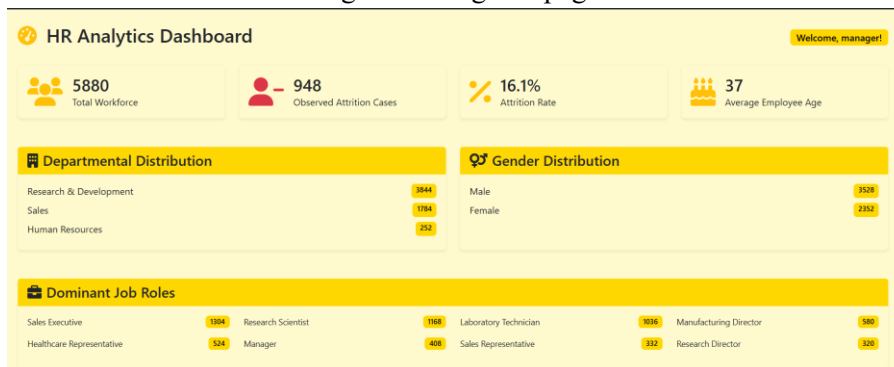
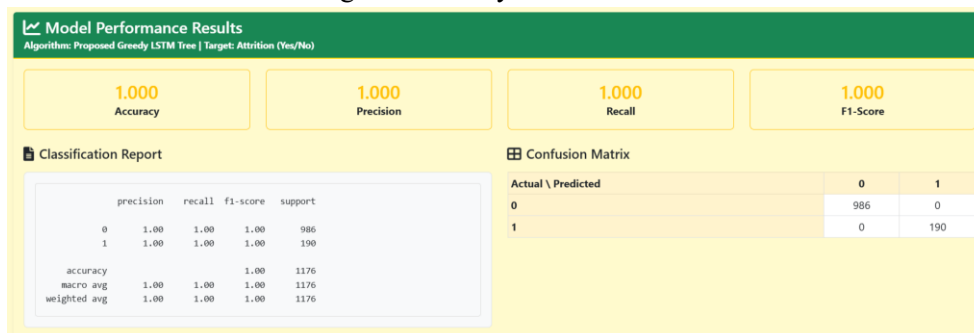
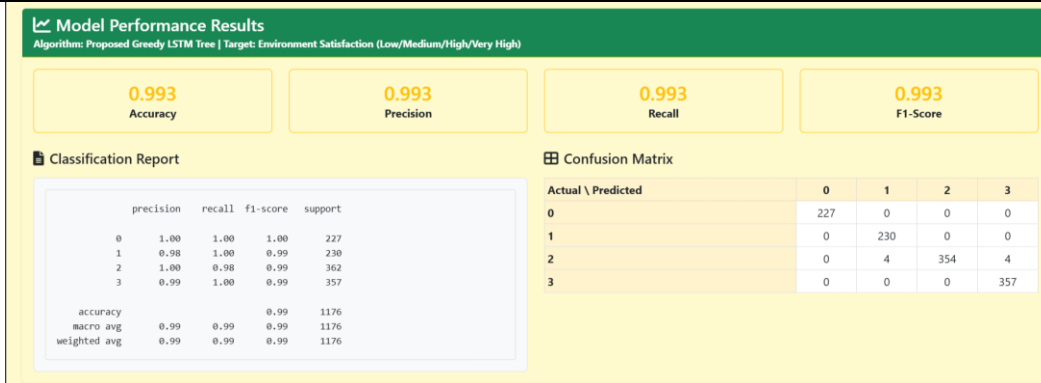


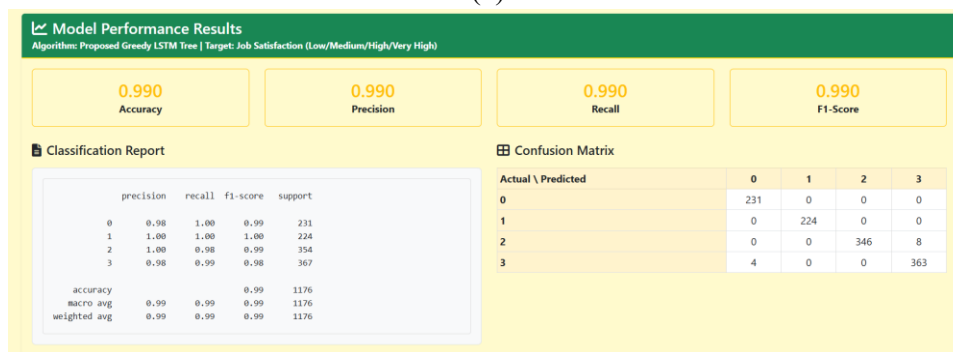
Fig. 3: HR analytics dashboard



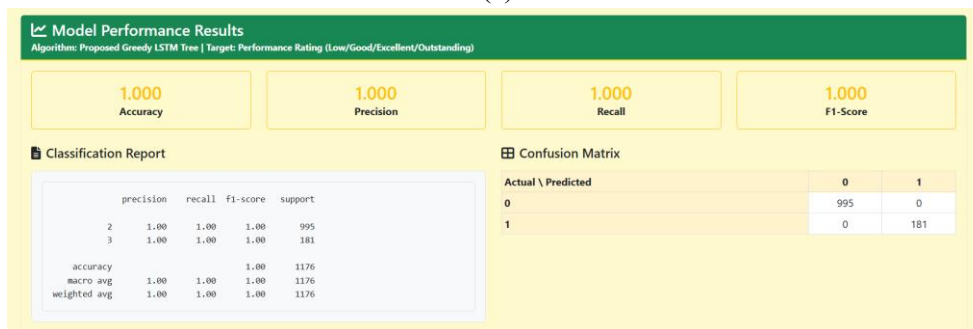
(a)



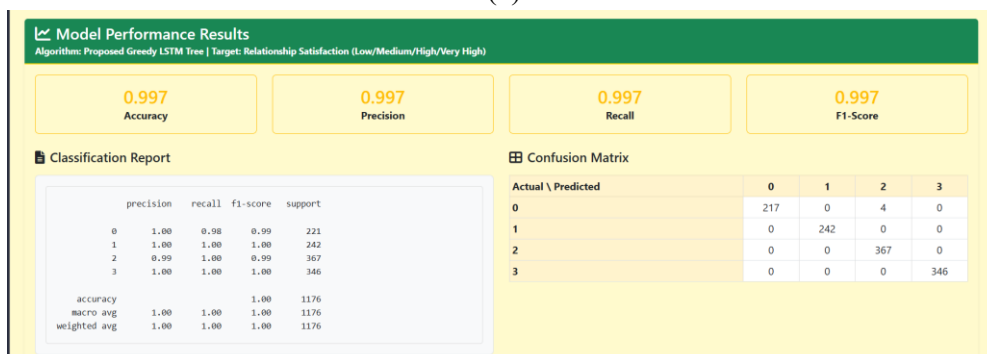
(b)



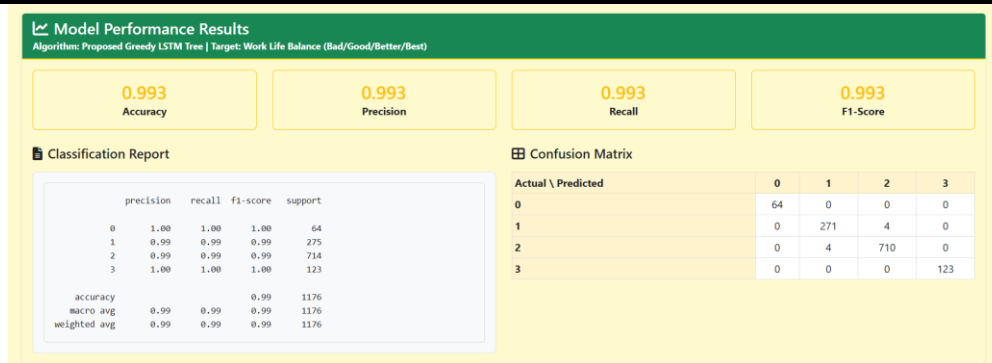
(c)



(d)



(e)



(f)

Fig. 4: Performance and classification reports of various target attributes from GLSTMT Model (a) Attrition, (b) Environment Satisfaction, (c) Job Satisfaction, (d) Performance Rating, (e) Relationship Satisfaction, (f) Work Life Balance.

Fig. 4 presents the performance results and classification reports generated by the proposed GLSTMT model for multiple target attributes related to driver attrition analysis.

Fig. 4 (a) The Attrition classification report illustrates the model’s capability to accurately distinguish between drivers who remain in the organization and those who leave. The evaluation metrics such as precision, recall, and F1-score demonstrate the strong classification performance of the proposed model in attrition prediction.

Fig. 4 (b) The Environment Satisfaction analysis evaluates the effectiveness of the proposed model in classifying satisfaction levels associated with the driver work environment. The classification report indicates accurate identification of satisfaction categories ranging from low to very high.

Fig. 4 (c) The Job Satisfaction report demonstrates the model’s capability to classify driver satisfaction levels related to job roles, responsibilities, and engagement within the organization. The evaluation metrics reflect the strong predictive performance of the proposed hybrid model.

Fig. 4 (d) The Performance Rating report evaluates the classification accuracy of the GLSTMT model in predicting driver performance categories. The results demonstrate precise identification of performance levels based on historical driver attributes.

Fig. 4 (e) The Relationship Satisfaction report reflects the model’s ability to classify variations in satisfaction related to relationships between drivers and management. The classification metrics highlight accurate identification of interpersonal satisfaction levels.


Fig. 4 (f) The Work Life Balance report measures the model’s capability to classify drivers according to their balance between professional responsibilities and personal life. The results demonstrate the strong performance of the proposed model in identifying work-life balance levels within the driver dataset.

DistanceFromHome 1	Education 2	EducationField Life Sciences	Gender Female
HourlyRate 94	JobInvolvement 3	JobLevel 2	JobRole Sales Executive
MaritalStatus Single	MonthlyIncome 5993	MonthlyRate 19479	NumCompaniesWorked 8
OverTime Yes	PercentSalaryHike 11	StockOptionLevel 0	TotalWorkingYears 8

+ Show More Fields (5 remaining)

Generate Prediction

Prediction Results
Target: Attrition (Yes/No) | Algorithm: Proposed Greedy LSTM Tree



Yes

Predicted Outcome

Attrition (Yes/No)

Prediction Details

- ✔ Target Variable: Y1
- ✔ Algorithm Used: Proposed Greedy LSTM Tree
- ✔ Prediction: Yes

Fig. 5: Predictions on test data.

The prediction screen 5 presents the attrition results for employees, indicating the probability of leaving. HR managers receive actionable insights to identify high-risk employees and take preventive measures to reduce turnover.

5. CONCLUSION

This study presents the development of an intelligent employee attrition prediction system using advanced data analytics and learning-based techniques. The framework integrates data preprocessing, exploratory analysis, and multiple predictive models including SVC, RF, GB, and GLSTMT to ensure accurate and reliable outcomes. A Flask-based web application enables secure login, role-based access control, and interactive dashboards for HR analytics. The system also utilizes LMDB for efficient and lightweight user data management. Experimental results show that hybrid and ensemble approaches achieve better predictive performance compared to conventional methods. Visual analysis highlights key factors influencing attrition, such as job satisfaction, work-life balance, and tenure. The system supports early identification of high-risk employees, enabling proactive retention strategies.

REFERENCES

- [1] Chen, X.; Deng, H. Improving Ride-Hailing Platform Operations in Dynamic Markets: A Drivers' Switching Perspective. *Systems* 2025, 13, 80. <https://doi.org/10.3390/systems13020080>
- [2] Cohen, M.C.; Zhang, R. Competition, and coopetition for two-sided platforms. *Prod. Oper. Manag.* 2020, 31, 1997–2014.
- [3] Sun, S.; Ertz, M. Dynamic evolution of ride-hailing platforms from a systemic perspective: Forecasting financial sustainability. *Transp. Res. Part C Emerg. Technol.* 2021, 125, 103003.
- [4] Zhong, Y.; Yang, T.; Cao, B.; Cheng, T.C.E. On-demand ride-hailing platforms in competition with the taxi industry: Pricing strategies and government supervision. *Int. J. Prod. Econ.* 2022, 243, 108301.
- [5] Sun, L.; Teunter, R.H.; Hua, G.; Wu, T. Taxi-hailing platforms: Inform or Assign drivers? *Transp. Res. Part B Methodol.* 2020, 142, 197–212.
- [6] Xu, Z.; Yin, Y.; Chao, X.; Zhu, H.; Ye, J. A generalized fluid model of ride-hailing systems. *Transp. Res. Part B Methodol.* 2021, 150, 587–605.
- [7] Ke, J.; Qian, S. Leveraging ride-hailing services for social good: Fleet optimal routing and system optimal pricing. *Transp. Res. Part C Emerg. Technol.* 2023, 155, 104284.

-
- [8] Bandiera, C.; Connors, R.D.; Viti, F. Mobility service providers' interacting strategies under multi-modal equilibrium. *Transp. Res. Part C Emerg. Technol.* 2024, 168, 104766.
- [9] Xi, H.; Li, M.; Hensher, D.A.; Xie, C.; Gu, Z.; Zheng, Y. Strategizing sustainability and profitability in electric Mobility-as-a-Service (E-MaaS) ecosystems with carbon incentives: A multi-leader multi-follower game. *Transp. Res. Part C Emerg. Technol.* 2024, 166, 104758.
- [10] Yao, R.; Zhang, K. Design an intermediary mobility-as-a-service (MaaS) platform using many-to-many stable matching framework. *Transp. Res. Part B Methodol.* 2024, 189, 102991.
- [11] Sun, S.; Wang, Z.; Wang, W. The impact of regulatory policy on the growth of ride-hailing platform: System dynamics perspective. *Phys. A Stat. Mech. Appl.* 2023, 632, 1–21.
- [12] Chen, X.G.; Deng, H.; Guan, S.Q.; Han, F.X.; Zhu, Z.H. Cooperation-oriented multi-modal shared mobility for sustainable transport: Developments and challenges. *Sustainability* 2024, 16, 11207.
- [13] Ma, L.; Tao, Z.; Wei, Q.; Huo, B. Cooperate with aggregation platform or not? Optimal decision for the on-demand ride service platform. *Res. Transp. Bus. Manag.* 2025, 59, 101251.
- [14] Lyu, W.; Hu, Y.; Liu, J.; Chen, K.; Liu, P.; Deng, J.; Zhang, S. Impact of battery electric vehicle usage on air quality in three Chinese first-tier cities. *Sci. Rep.* 2024, 14, 21.