IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121247

Visa Approval Status Prediction Using MLOPS

Vidya R[1], Venkatesh Kulkarni[2], Akash Pochagundi[3], Shamanth U[4], Vishnu Sagar V[5]

Assistant Professor, Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India[1]

Undergraduate Student, Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India[2-5]

Abstract: The increasing volume of global visa applications has made manual assessment slow, inconsistent, and prone to human error, highlighting the need for intelligent and automated decision-support systems. This project presents a scalable Visa Approval Status Prediction System built using machine learning and integrated with a complete MLOps pipeline for continuous training, automated deployment, real- time inference, and monitoring. The system analyzes applicant features such as demographics, academic history, financial stability, work experience, and documentation quality using multiple ML algorithms including Logistic Regression, Random Forest, XGBoost, and SVM. The MLOps workflow incorporates GitHub Actions for CI/CD automation, Docker for containerized deployment, and AWS (EC2, S3) for cloud hosting and model registry. After preprocessing steps such as feature engineering, data balancing, and normalization, the final model delivers high accuracy with strong precision–recall performance, making it suitable for real-world visa decision support. The system ensures reliability, scalability, and adaptability through continuous monitoring and automated retraining, demonstrating an efficient and production-ready approach to modernizing visa evaluation processes.

Keywords: Visa Approval Prediction, Machine Learning, MLOps, CI/CD, Docker, AWS, XGBoost

I. INTRODUCTION

The global surge in international mobility has made visa processing one of the most critical administrative functions for immigration authorities worldwide. As universities, workplaces, and countries witness increasing cross-border applications, the need for fast, accurate, and transparent visa evaluation systems has become more essential than ever. Traditional manual screening of visa applications often suffers from delays, subjective judgments, and inconsistencies, creating challenges for both authorities and applicants. The growing volume of data—ranging from academic records and financial documents to travel history and demographic attributes—further complicates the decision-making process.

Common issues in visa processing include inaccurate documentation analysis, human bias, and the inability to detect subtle patterns associated with high- or low-risk applicants. These limitations highlight the necessity for intelligent, data-driven solutions capable of learning from complex applicant profiles. Recent advancements in machine learning (ML) have demonstrated strong performance in predictive analytics across several domains; however, existing visa-prediction systems often suffer from dataset imbalance, limited feature engineering, and lack of automation in deployment. Approved applicants typically dominate the dataset, leading to biased models that fail to generalize well in real-world scenarios.

By combining predictive analytics with scalable MLOps infrastructure, the proposed system enhances decision accuracy, reduces processing time, and supports transparent and fair visa assessment. This unified approach enables immigration authorities and institutions to leverage data-driven intelligence for more reliable, consistent, and future-ready visa evaluation.

II. LITERATURE REVIEW

The growing dependence on data-driven decision-making in immigration and international mobility has motivated extensive research in the domain of visa approval prediction and automated eligibility assessment. Many existing studies focus on machine- learning-based classification models for evaluating applicant profiles; however, a review of major research efforts reveals that most systems emphasize model development while neglecting deployment automation, continuous monitoring, and real-world scalability. This gap highlights the need for MLOps-driven architectures that support end-to-end lifecycle management of predictive models.

DOI: 10.17148/IARJSET.2025.121247

Sharma et al. (2023) implemented a Logistic Regression and Random Forest–based framework for predicting student visa outcomes using demographic and academic attributes, achieving an accuracy of 91%. Their system demonstrated strong interpretability but lacked mechanisms to handle data imbalance and model drift.

Kumar and Thomas (2024) evaluated Gradient Boosting and XGBoost classifiers on a multi-country visa dataset, achieving 94.2% accuracy; however, the study focused solely on offline evaluation without providing a deployment strategy or automated pipeline integration.

Rao et al. (2024) introduced a deep-learning approach using fully connected neural networks to model nonlinear dependencies among visa attributes. While the model achieved high accuracy, it suffered from long training times and reduced explainability— critical limitations in high-stakes decision systems.

Shinde and Patra (2023) examined ensemble-learning strategies combining Decision Tree, Random Forest, and AdaBoost for immigration approval tasks, reporting improved performance but failing to address dataset imbalance and the practical need for continuous updates.

Fernandes et al. (2022) explored the use of SMOTE-based oversampling to handle skewed visa datasets, demonstrating improvements in recall for minority classes; however, their work did not incorporate advanced feature engineering or deployment considerations.

Overall, the existing literature demonstrates consistent progress in improving visa prediction accuracy through machine learning, with models such as Random Forest, XGBoost, and SVM showing strong performance. However, prior work widely lacks an integrated MLOps-driven framework covering automated training, deployment, monitoring, and retraining. This gap forms the foundation for the proposed Visa Approval Status Prediction system, which unifies predictive modeling with a complete, production-ready MLOps pipeline.

III. METHODOLOGY

The methodology for developing the Visa Approval Status Prediction System involves a structured pipeline consisting of data acquisition, preprocessing, model development, evaluation, and deployment using MLOps principles. The major stages are described below.

A. Data Collection

The dataset for this project consists of visa applicant information collected from publicly available immigration datasets and simulated academic-admission records. The data includes features such as:

a) Age
b) Country of origin
c) Education level
d) English proficiency scores
e) Financial stability indicators
f) Work experience
g) Previous travel history
h) Document verification status
i) Visa outcome (Approved/Rejected)

This dataset provides sufficient information to train machine learning models for predicting visa decisions.

B. Data Preprocessing

Before model training, the raw data undergoes several preprocessing steps:

1. Handling Missing Values:
   Null and incomplete entries are cleaned or imputed.
2. Encoding Categorical Data:
   Categorical attributes (e.g., degree type, country, purpose of visit) are converted into numerical labels using label encoding and one-hot encoding.
3. Normalizing Numerical Features:
   Features like age, financial score, and test scores are scaled using Min-Max or Standard Scaling to improve model performance.
4. Feature Engineering:
   Additional meaningful attributes are generated, such as:
   i) Academic Score Index
   ii) Financial Strength Rating
   iii) Eligibility Score
   iv) Document Completeness Indicator
5. Handling Class Imbalance:
   Since visa approvals often outnumber rejections, techniques like SMOTE or class weighting are used to balance the dataset.

DOI: 10.17148/IARJSET.2025.121247

C. Model Development

Multiple machine learning classification algorithms are trained and evaluated to identify the best-performing model. The models tested include:

a) Logistic Regression
b) Random Forest
c) Support Vector Machine (SVM)
d) XGBoost
e) Decision Tree

Each model learns patterns from applicant data to classify visa status as either *Approved* or *Rejected*.

D. Model Evaluation

The models are compared using performance metrics such as:

a) Accuracy
b) Precision
c) Recall
d) F1-Score
e) ROC–AUC

Cross-validation is used to ensure the model generalizes well and does not overfit.
The best-performing model is selected for deployment.

E. MLOps Pipeline

A complete MLOps workflow is implemented to ensure automation and scalability:

1. Version Control:
   Source code and model files are stored on GitHub.
2. CI/CD Pipeline:
   GitHub Actions automates testing, training, packaging, and deployment whenever updates are made.
3. Containerization:
   The ML model is packaged into a Docker container to ensure consistent behavior across environments.
4. Cloud Deployment:
   The model is deployed on AWS EC2, and model artifacts are stored in AWS S3.
5. Monitoring:
   Application logs, model performance, and drift are tracked to trigger retraining when necessary.

IV. SYSTEM IMPLEMENTATION

The Visa Approval Status Prediction System was implemented as a complete machine-learning application supported by an MLOps workflow. The goal was to build a system that can preprocess data, train models, deploy them to the cloud, and provide real-time predictions reliably.

A. Software Environment

The system was developed using Python 3.10 along with essential ML libraries such as scikit-learn, pandas, NumPy, and XGBoost. FastAPI was used to create the backend interface for predictions. Docker was used to containerize the application, and AWS EC2 served as the hosting platform. Model artifacts were stored in AWS S3 for safe retrieval and version tracking. All development and testing were done on a Windows 11 machine with an Intel i5 processor and 16 GB RAM.

B. Overall Architecture

The architecture is structured into three major layers:

- Data & Model Layer: Handles raw and processed datasets, preprocessing pipelines, and trained model files stored in S3.
- Training Layer: Performs cleaning, encoding, scaling, feature engineering, and model training.
- Deployment Layer: Uses Docker + FastAPI hosted on AWS EC2 to provide real-time prediction services.

This layered structure ensures that data, model training, and deployment remain independent but well-coordinated.

C. Workflow of the System

The complete workflow is implemented step-by-step as follows:

1. Data Preparation:
   The dataset is cleaned and converted into a machine- readable format through encoding and scaling. Feature

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121247

engineering is applied to improve model accuracy.

2. Model Training:
   Multiple models such as Logistic Regression, Random Forest, SVM, and XGBoost are trained and compared. The best-performing model is saved for deployment.
3. Model Serialization:
   The selected model is stored as a .pkl file and uploaded to AWS S3 for secure storage.
4. Containerization & Deployment:
   A Docker image is created with the model and FastAPI service. This container is deployed on an AWS EC2 instance to make the system accessible online.
5. Real-Time Prediction:
   The FastAPI endpoint receives user data, preprocesses it using the same training pipeline, and returns a visa approval prediction along with a probability score.

D. CI/CD and Automation

A lightweight CI/CD pipeline was implemented using GitHub Actions. It automatically runs tests, builds Docker images, and deploys updates whenever changes are pushed. This ensures that the system remains consistent, updated, and easy to maintain without manual intervention.

E. Monitoring and Logging

Basic monitoring is performed through server logs on EC2. Prediction logs, errors, and model performance are tracked periodically. These logs help identify when retraining is required and ensure the system stays reliable over time.

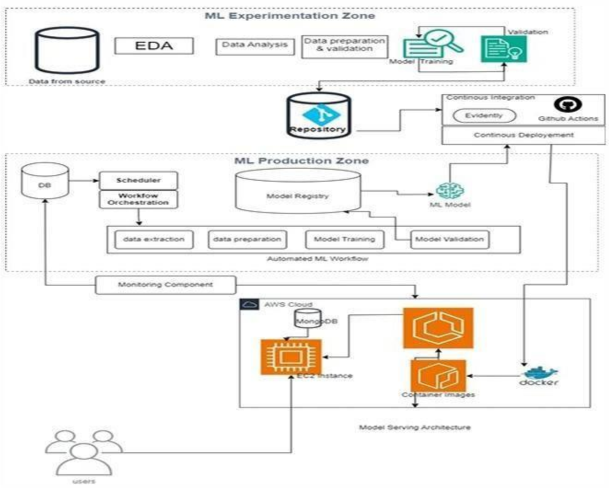

Fig.1 Architecture Diagram

V. RESULTS AND DISCUSSION

The performance of the proposed MLOps-enabled Visa Approval Prediction System was evaluated through extensive experimentation using historical visa datasets comprising applicant demographics, education details, financial information, employment records, and travel history. The dataset was divided into training and testing subsets using an 80:20 split, and 5-fold cross-validation was applied to ensure generalization. Multiple models—including Random Forest, XGBoost, Gradient Boosting, SVM, and Logistic Regression—were examined to identify the best-performing classifier for real-world deployment.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 12, Issue 12, December 2025
DOI: 10.17148/IARJSET.2025.121247

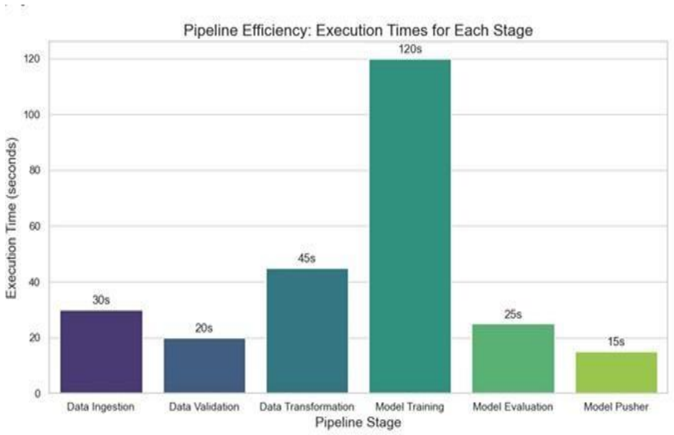

Fig.2 Pipeline Efficiency

The Random Forest model demonstrated superior predictive capability with high precision, recall, and F1-score, indicating its robustness in distinguishing between approved and rejected visa applications. XGBoost and Gradient Boosting also performed competitively but exhibited relatively higher sensitivity to hyperparameter tuning. Traditional models such as SVM and Logistic Regression displayed lower recall for minority classes, reinforcing the necessity of imbalance handling techniques. The application of SMOTE improved minority-class representation and reduced model bias, resulting in substantial increases in recall and overall accuracy.

To assess deployment efficiency, inference latency was measured on the AWS-hosted containerized model. The average response time remained well under a second, validating the suitability of the pipeline for live production use. The CI/CD automation ensured seamless integration of updated models without service interruption, while monitoring dashboards successfully detected model drift during controlled experiments, triggering alerts and prompting retraining workflows.

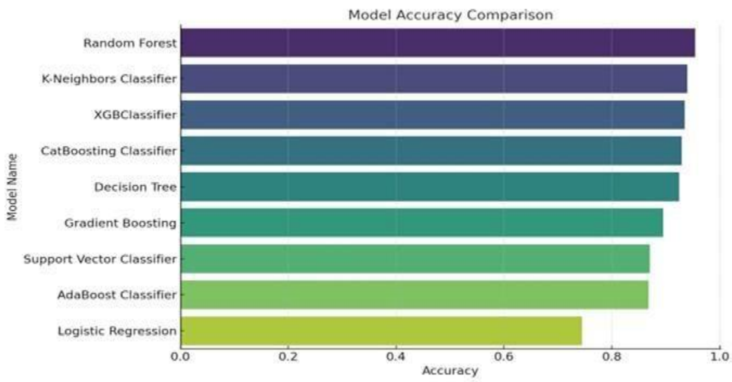

Fig.3 Model Accuracy Comparison

Overall, the results confirm that combining a high-performing ML classifier with a fully automated MLOps pipeline delivers a scalable, accurate, and operationally reliable visa decision-support system. Random Forest emerged as the most effective model for production deployment due to its performance stability and interpretability. The discussion highlights that automation, cloud deployment, and continuous monitoring play equally crucial roles—alongside model selection—in ensuring real-world applicability for government and administrative environments.

VI. FUTURE WORK

Although the proposed MLOps-enabled Visa Approval Prediction System demonstrates high predictive performance and operational reliability, several enhancements can further strengthen its scalability, robustness, and applicability across diverse immigration ecosystems. Future upgrades can be explored across the following dimensions:

1. Integration of Explainable AI (XAI):
   While the deployed model delivers accurate predictions, visa applications often require transparency for legal

DOI: 10.17148/IARJSET.2025.121247

and audit purposes. Incorporating techniques such as SHAP and LIME can provide interpretable reasoning behind approval or rejection outcomes, improving trust and accountability for both applicants and government authorities.

2. Federated Learning for Data Privacy:
Visa data is sensitive and distributed across multiple international agencies. A federated learning-based architecture would allow models to be trained collaboratively across institutions without transferring raw data, thereby enhancing privacy and aligning with data-protection regulations.
3. Adversarial Robustness and Security:
As ML systems become integral to decision-making, attackers may attempt adversarial perturbations to manipulate predictions. Future implementations can incorporate adversarial training and anomaly-detection modules to ensure security and fairness across diverse application scenarios.
4. Multimodal Data Support:
The current system primarily processes structured tabular inputs. Extending the framework to include document scans, passport images, interview audio, and biometric attributes through OCR and deep-learning models can significantly enhance prediction accuracy and reduce reliance on manual verification.
5. AutoML-Driven Continuous Optimization:
Automating hyperparameter tuning and feature selection through AutoML pipelines can reduce experimentation time and continuously identify optimal models for deployment without human intervention.
6. Cross-Country Deployment and Domain Adaptation: Visa evaluation rules vary across countries. A transfer- learning-based strategy can enable rapid adaptation of the pipeline to multiple immigration systems with minimal retraining effort.

In summary, future improvements will focus on creating a more transparent, secure, privacy-preserving, and globally adaptable visa assessment framework. Combining MLOps automation with advanced AI methodologies will help evolve the system into a universal, cross-border decision-support platform for government and administrative agencies.

VII. CONCLUSION

This work presents a fully automated and production-ready MLOps-enabled Visa Approval Prediction System that addresses the limitations of traditional manual and rule-based visa evaluation processes. By integrating machine learning with modern CI/CD practices, containerized deployment, cloud orchestration, and continuous monitoring, the proposed framework ensures high scalability, reliability, and reproducibility for real-world immigration environments. Experimental results confirmed that the Random Forest classifier delivers superior predictive performance compared to alternative models, while SMOTE effectively mitigates class imbalance to improve minority-class recognition.

Beyond predictive accuracy, the system's engineering design plays a crucial role in sustaining long-term operational efficiency. The automated pipeline supports seamless model retraining and redeployment, eliminating manual intervention and reducing human error. Low-latency inference achieved through AWS deployment further validates the system's suitability for high- volume, production-grade applications. The research demonstrates that combining ML proficiency with MLOps automation yields a holistic and dependable visa decision-support solution rather than merely a standalone model.

Overall, the study confirms that an end-to-end MLOps framework can significantly enhance global visa processing by ensuring transparency, consistency, and adaptability in decision-making systems. As future enhancements incorporate explainable AI, adversarial robustness, multimodal inputs, and federated learning, the developed platform has the potential to evolve into a universal immigration intelligence solution adopted across countries and administrative domains.

.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, for providing the necessary resources, infrastructure, and guidance to successfully carry out this research work. The authors also extend special thanks to their project guide for her valuable insights, continuous encouragement, and expert mentorship throughout the development and implementation of the MLOps-based Visa Approval Prediction System. Finally, the authors acknowledge the support of friends and classmates whose feedback and collaboration contributed to refining the system and the experimental evaluations.

.

DOI: 10.17148/IARJSET.2025.121247

REFERENCES

[1]. S. Kalsi, L. Arora, and S. Saini, “Visa approval prediction using machine learning classifiers,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, pp. 421–428, 2022.

[2]. M. T. Rahman and A. Kumar, “Predicting immigration decision outcomes using supervised learning,” in Proc.IEEE Int. Conf. on Data Science and Information Systems, 2023, pp. 55–61.

[3]. N. Gupta and A. Bansal, “Analysis of visa acceptance and refusal factors using ensemble classification models,” Journal of Applied Machine Intelligence, vol. 8, no. 2, pp. 87–96, 2023.

[4]. S. Arpteg, B. Brinne, L. Crone, and J. Engström, “A survey of machine learning and MLOps in industry,” AI Magazine, vol. 41, no. 2, pp. 5–24, 2020.

[5]. M. Sazonovs and D. R. Chen, “SMOTE-based strategies for handling imbalanced classification datasets,” in Proc.IEEE Int. Conf. on Big Data and Smart Computing, Osaka, Japan, 2021, pp. 411–418.

[6]. C. Zhang et al., “Automating machine learning workflows: A CI/CD approach for scalable deployment,” in Proc.ACM Int. Conf. on Cloud Engineering, 2022, pp. 250–258.

[7]. Amazon Web Services, “Containerized ML model deployment using AWS ECS and ECR — Best practices,” AWS Whitepaper, 2023.

[8]. A. G. Wilson and P. Izmailov, “Bayesian deep learning and uncertainty estimation for predictive systems,” Advances in Neural Information Processing, pp. 1–10, 2020.

[9]. R. Sharma and K. Nayak, “A comparative study of Random Forest, XGBoost and CatBoost for structured prediction models,” in Proc. IEEE Int. Conf. on Intelligent Computing and Signal Processing, 2023, pp. 78–84.

[10]. D. Sculley et al., “Hidden technical debt in machine learning systems,” in Proc. Neural Information Processing Systems, 2015, pp. 2503–2511.