A Project Report on

EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements

Submitted in partial fulfillment of the requirements for the award
of the degree of

Bachelor of Engineering

in

Information Technology

by

Soham Dalvi (21104010)
Sankalp Gunjal (21104087)
Siddharth Devare (21104136)
Sumit Mesta (21104069)

Under the Guidance of

Ms. Sonal Jain

Department of Information Technology
NBA Accredited
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W)-400615
UNIVERSITY OF MUMBAI

Academic Year 2024-2025

Approval Sheet

This Project Report entitled ***"EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements"*** Submitted by ***"Soham Dalvi"(21104010),"Sankalp Gunjal"(21104087),"Siddharth Devare" (21104136), "Sumit Mesta"(21104069)***is approved for the partial fulfillment of the requirement for the award of the degree of ***Bachelor of Engineering*** in ***Information Technology*** from ***University of Mumbai***.

Ms. Sonal Jain
Guide

Dr. Kiran Deshpande
HOD, Information Technology

Place: A. P. Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements"*** submitted by ***"Soham Dalvi" (21104010),"Sankalp Gunjal" (21104087),"Siddharth Devare" (21104136),"Sumit Mesta" (21104069)*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Information Technology***,to the University of Mumbai,is a bonafide work carried out during academic year 2024-2025.

Ms. Sonal Jain
Guide

Dr. Kiran Deshpande
HOD, Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Internal Examiner(s)

1.

2.

Place: A. P. Shah Institute of Technology, Thane
Date:

Acknowledgement

We have great pleasure in presenting the report on **EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements.** We take this opportunity to express our sincere thanks towards our guide **Ms. Sonal Jain** for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards her constant encouragement, support and guidance through the development of project.

We thank **Dr. Kiran B. Deshpande** Head of Department for his encouragement during the progress meeting and for providing guidelines to write this report.

We express our gratitude towards BE project co-ordinators, for being encouraging throughout the course and for their guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Soham Dalvi
(21104010)

Sankalp Gunjal
(21104087)

Siddharth Devare
(21104136)

Sumit Mesta
(21104069)

Declaration

We declare that this written submission represents our ideas in our own words and where others ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Soham Dalvi (21104010)

Sankalp Gunjal (21104087)

Siddharth Devare (21104136)

Sumit Mesta (21104069)

Date:

Abstract

In the current educational landscape, traditional assessment systems remain primarily focused on academic performance, often neglecting vital aspects such as co-curricular achievements, soft skills, and individual growth metrics. This narrow approach fails to offer a holistic understanding of student development, leading to incomplete evaluations and missed opportunities for skill recognition. To address this gap, this project introduces a machine learning-driven, cross-platform web application designed to acknowledge and analyze student achievements across academic and non-academic domains. Leveraging Machine Learning (ML) and Generative AI the platform constructs dynamic student profiles, identifies learning gaps, and recommends content and feedback in real time. A digital credentialing integration using Credly ensures secure validation of accomplishments, while an ML-based certificate verification module enhances data authenticity. This project features an intuitive dashboard for stakeholders to visualize performance trends using Graphical representations, and incorporates gamification elements to increase motivation and retention. Built using scalable technologies such as Node.js and Flask, this solution promotes inclusive, automated, and data-informed assessment practices, redefining how institutions evaluate student success in a digitally evolving educational ecosystem.

Contents

List of Figures

List of Tables

List of Abbreviations

AI: Artificial Intelligence
ML: Machine Learning
API: Application Programming Interface
GPT: Generative Pre-trained Transformer
LLaMa: Large Language Model Meta AI
NLP: Natural Language Processing
UI: User Interface
JWT: JSON Web Token
URL: Uniform Resource Locator
IP: Internet Protocol
CSRF: Cross-Site Request Forgery
JSON: JavaScript Object Notation
MBTI: Mayers-Briggs Type Indicator
PDF: Portable Document Format
CNN: Convolutional Neural Network

Chapter 1

Introduction

In modern education systems, the drawbacks of traditional evaluation approaches are inevitable. The mere focus on academic performance overshadows the brilliant co-curricular engagement and the development of soft skills within a student. This definitely creates an imbalance between the comprehensive assessment of the true outcome or growth of a student. The key to overcoming this challenge lies in using graphs and charts to convert complex student achievement data into clear visualizations, providing a comprehensive view beyond traditional assessment metrics.

The latest breakthroughs in AI and ML have opened avenues for transformative innovations in education. Research has proved that AI-driven adaptive learning will boost student participation and performance levels. Additionally, adaptive AI-based evaluations will provide an appropriate assessment in response to students learning abilities and will produce actionable insights both for the learners and the instructor. The above-mentioned concepts are foundational to EduFlex, an AI-based platform set to transform assessment processes in schools.

EduFlex has borrowed methodologies from leading research. The generative AI module, as based on Pesovski et al. [1], creates personalized learning paths with dynamic crafting to address the gaps in knowledge. Similarly, Yang et al.'s [2] research on AI-driven profiling inspired EduFlex's analytics engine to identify trends in student performance and give actionable feedback. Furthermore, the study of Wolz et al. [3] about digital credentialing inspired the development of systems like Credly to authenticate and share various accomplishments by students in this platform. These elements altogether redefine how education success is assessed.

Real-time feedback, and validation of digital certificates are some of the features offered through the platform. There is robust functionality ensured across devices by using scalable technologies such as Node.js, Flask(Python), and MongoDB. ML-based certificate validation along with interactive dashboards in current implementations empowers educators and students with actionable insights, creating a culture of continuous growth with holistic learning.

The ultimate purpose of EduFlex is to transform educational assessments into a holistic system that accounts for both academic and co-curricular achievements. This paper discusses the design, methodology, implementation, and potential of EduFlex in establishing new standards for inclusive and dynamic learning environments that emphasize innovation in modern education.

1.1 Motivation

In the modern digital era, students actively curate and showcase their social media profiles, yet the significance of maintaining a well-structured academic profile is often overlooked. An academic portfolio that accurately reflects achievements, skills, and certifications is essential for shaping future educational and career opportunities. However, traditional assessment methods lack a unified approach to tracking and evaluating a student's holistic development, including academic, co-curricular, and extracurricular accomplishments. This gap highlights the need for a comprehensive system that allows students to effectively manage and present their academic credentials.

Additionally, certificate forgery and credential fraud have become significant concerns, particularly in colleges. The absence of a reliable verification system undermines the credibility of student achievements. To address this challenge, a robust certificate verification mechanism is essential to ensure the authenticity and integrity of student records. By leveraging artificial intelligence, machine learning, and secure digital credentialing, this project aims to develop a transparent, data-driven platform that enables students to track their progress, receive personalized insights, and present verifiable credentials within academic and professional domains.

1.2 Problem Statement

Traditional student evaluation systems focus mainly on academic grades, often ignoring co-curricular activities, soft skills, and overall personal growth. This limited approach makes it difficult for educators and employers to truly understand a student's full potential. Additionally, students do not have a structured way to showcase their achievements in a manner similar to social media profiles, which limits their opportunities for academic and career advancement.

Furthermore, there is no single platform to manage and verify student credentials, leading to inefficiencies in academic validation and hiring processes. The increasing risk of fake certificates adds to the challenge, making it difficult for institutions and companies to trust the authenticity of student records. Without a standardized and verifiable system, recruiters must rely on time-consuming background checks, slowing down the hiring process. A verified student profile, including authenticated certifications, would help companies easily assess candidates for jobs, ensuring transparency and making recruitment faster and more efficient.

1.3 Objectives

- To develop a web-based, cross-platform framework that provides a holistic view of student achievements by integrating academic, co-curricular, and extracurricular records.
- To implement AI-powered psychometric assessments using llama 3.1:8b model to identify students learning gaps and evaluate cognitive abilities, personality traits, and career readiness.
- To build a data visualization platform using Google-Charts that presents clear, actionable insights through interactive dashboards, helping educators and students make informed decisions.
- To ensure ML-based certificate verification based on CNN model to detect tampered certificates and maintain the authenticity of student achievements.
- To develop a scalable cloud-based infrastructure for real-time data processing, seamless accessibility, and reliable system performance.
- To enable digital credentialing via Credly API to validate and showcase verified student accomplishments.

1.4 Scope

- Can provide holistic student assessment by collecting and analyzing academic records, co-curricular participation, and extracurricular achievements for a well-rounded evaluation.
- Can validate certificates using machine learning to detect tampered or forged documents, ensuring authenticity.
- Can conduct psychometric assessments to evaluate cognitive abilities, personality traits, and career readiness, offering personalized insights.
- Can offer interactive data visualization with real-time dashboards, graphs, and reports for better progress tracking and decision-making.
- Can enhance student engagement through gamification elements like badges, points, and rewards for academic and extracurricular achievements.
- Can ensure scalability and security with a cloud-based, cross-platform system that provides seamless accessibility and data protection.
- Can enable digital credentialing through Credly API integration, allowing students to receive standardized, verifiable digital credentials for their achievements.

Chapter 2

Literature Review

- **Review Paper 1** : "Generative AI for customizable learning experiences."

  According to Ivica Pesovski, Ricardo Santos, Roberto Henriques, and Vladimir Trajkovik, generative AI technologies are paving the way for highly personalized learning environments by dynamically adjusting learning materials to suit individual student needs. Their 2024 study published in Sustainability presents a framework where AI is used to generate quizzes and adaptive content based on real-time assessment of student progress. By integrating with existing Learning Management Systems (LMS), the platform ensures smooth delivery of learner-specific content. The AI system observes how students interact with material and identifies knowledge gaps, enabling the creation of targeted educational resources that enhance engagement and learning efficiency. This research supports the growing movement toward individualized education where AI systems play a central role in delivering customized, real-time instruction [1].

- **Review Paper 2** : "AI-Powered Smart and Personalized Education Platform"

  According to Dunusinghe A. V., Ranasinghe T. K. S. A., Gamage J. G. A. C. H., Perera K. G. D. T., Samantha Thelijjagoda, and Poojani Gunatilake, an AI assistant addresses the shortcomings of online education, focusing on personalized learning and career guidance. Their study emphasizes how AI provides tailored recommendations, assesses performance, suggests subjects, and creates learning plans based on individual needs. The primary objectives include accurate grade prediction, advising on subject streams, offering customized assessments, and developing personalized learning strategies. The paper reveals how data analytics and machine learning are used to predict grades, while a multi-tenancy architecture enhances the management of educational data. Personalized assessments take into account both academic performance and soft skills, while machine learning techniques help predict the most suitable subject streams for students. Additionally, engagement detection is used to optimize the learning process. This AI system has the potential to reshape online education, although the authors highlight the need for ongoing improvements to further enhance learning outcomes [2].

- **Review Paper 3** : "Digital credentials in higher education institutions: A literature review."

  According to Elena Wolz, Matthias Gottlieb, and Hans Pongratz, digital credentials are emerging as a reliable and transparent method of validating educational achievements in higher education institutions. Their 2021 literature review published in Innovation Through Information Systems explores how digital badges and blockchain-based verification systems can be used to document both academic and non-academic accomplishments. These credentials are tamper-proof, portable, and verifiable across platforms, offering increased trust in the educational ecosystem. The authors highlight the shift toward micro-credentialing and stackable certificates, which can represent a wide variety of competencies beyond traditional degrees. Their review underscores how digital credentialing can enhance employability and lifelong learning by making skills more visible and verifiable [3].

- **Review Paper 4** : "Enhancing Personalised Learning and Student Engagement Using Generative AI"

  According to Adebowale Owoseni, Oluwaseun Kolade, and Abiodun Egbetokun, generative AI technologies like ChatGPT can significantly enhance both student engagement and the personalization of learning experiences. In their 2024 chapter in Generative AI in Higher Education: Innovation Strategies for Teaching and Learning, the authors analyze how generative models enable dynamic content creation and real-time learner interaction. By tailoring course material to individual student profiles, these models support greater comprehension and retention. The system also provides in-the-moment adjustments based on learner responses, offering flexibility in pacing and complexity. This approach shifts traditional static learning paradigms into more responsive and engaging experiences, demonstrating the potential of AI to create more inclusive and effective digital classrooms [4].

- **Review Paper 5** : "The impact of gamification in educational settings on student learning outcomes: a meta-analysis"

  According to Rui Huang, Albert D. Ritzhaupt, Max Sommer, Jiawen Zhu, Anita Stephen, Natercia Valle, John Hampton, and Jingwei Li, gamification techniques significantly boost student motivation and learning outcomes. In their 2020 meta-analysis published in Educational Technology Research and Development, they examine the effects of integrating points, badges, leaderboards, and game-like progress tracking into educational platforms. The analysis reveals that gamified environments lead to improved academic performance, increased participation, and enhanced learner satisfaction. By promoting goal-oriented behavior and providing immediate feedback, gamification fosters a deeper sense of engagement and encourages students to take ownership of their learning process. The paper establishes gamification as an effective strategy for active learning across multiple disciplines and educational settings [5].

- **Review Paper 6 [6]** : "Certificate fraud verification model using clustered-based classification approach."

  According to Shamsudeen Mohammed, Lois Nwobodo, and Njoku Ekene, certificate fraud remains a major challenge in digital education systems, and their 2024 study proposes a cluster-based classification model to verify the authenticity of certificates. By analyzing document metadata, font consistency, and layout patterns, the system applies unsupervised machine learning techniques to detect anomalies associated with fraudulent certificates. This classification model enhances verification efficiency, reduces human error, and strengthens trust in digital credentialing. Their approach addresses the increasing need for secure, automated systems capable of ensuring the credibility of academic and professional certifications in both educational and employment contexts [6].

- **Review Paper 7** : "Machine Learning Techniques for Identity Document Verification in Uncontrolled Environments: A Case Study"

  According to Alejandra Castelblanco, Jesus Solano, Christian Lopez, Esteban Rivera, Lizzy Tengana, and Martín Ochoa, verifying identity documents in uncontrolled environments poses significant technical challenges. In their 2020 case study published in Pattern Recognition, the authors apply machine learning techniques such as convolutional neural networks (CNNs) to assess and validate identity documents captured under varying environmental conditions. By incorporating image preprocessing, segmentation, and feature extraction, the model is able to recognize and verify document authenticity even in suboptimal settings. This work contributes to the broader field of secure document authentication, with applications in both educational credentialing and national identification systems [7].

- **Review Paper 8** : "Guided Student Profile System: Mentor Feedback Integration"

  According to V. R. Sadasivam, V. Sanjaykumar, A. Sheik Imam, and R. Vasanthan, student profiling systems can be significantly improved by integrating mentor feedback into performance evaluation. Presented at the 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS), their model enables guided profiling by combining academic metrics with qualitative mentor inputs. This hybrid profiling technique helps provide a more accurate reflection of student abilities, progress, and potential. It also facilitates real-time interventions by mentors, thereby improving the educational experience through personalized academic counseling and development [8].

- **Review Paper 9** : "Academic profile development: An investigation of differentiation processes based on students' level of achievement and grade level."

  According to Moritz Breit, Martin Brunner, Antoine Fischbach, Rachel Wollschläger, Ulrich Keller, Sonja Ugen, and Franzis Preckel, academic profiles evolve based on students' achievements and grade levels. Their 2024 study in the Journal of Educational Psychology investigates how differentiation in academic performance influences the development of student profiles over time. By analyzing longitudinal data, the study identifies distinct developmental patterns among high- and low-achieving students. The research provides insights into academic self-concept formation and underlines the importance of targeted support systems for various learner segments, thereby informing policy-making in curriculum design and student evaluation [9].

- **Review Paper 10** : "A Study of the Effects of an Individual's Personality and Characteristics on Job Behavior Using the Myers-Briggs Type Indicator"

  According to Manyata Yadav, personality traits play a significant role in shaping job behavior, and the Myers-Briggs Type Indicator (MBTI) can be a useful tool in understanding these dynamics. Her 2023 study, published in the International Journal of Innovative Research in Engineering and Management, explores how personality dimensions affect work habits, team collaboration, and leadership potential. The paper categorizes participants by MBTI type and correlates these with behavioral tendencies in organizational contexts. This research offers a foundation for integrating personality-aware frameworks in educational and professional development programs, enhancing individual performance through better alignment of roles and traits [10].

Chapter 3

Project Design

3.1 Existing System

The traditional student evaluation and certification processes suffer from several limitations that hinder efficiency, accuracy, and engagement. Institutions and employers rely heavily on manual verification methods, leading to delays, inconsistencies, and errors in validating student credentials. This process is often cumbersome and time-consuming, reducing the reliability of certifications.

One of the most significant drawbacks is the high risk of forgery. Certificates can be easily manipulated using image editing software, making it difficult to ensure the authenticity of a students achievements. Without an automated detection mechanism, institutions and employers must rely on visual inspection, which is prone to oversight and fraudulent activities.

Furthermore, there is no real-time student performance tracking in conventional systems. Educators and mentors lack dynamic monitoring tools, preventing timely interventions to assist students who may need additional support. This delayed feedback loop can negatively impact a students learning journey and overall academic performance.

Additionally, traditional platforms do not integrate mental well-being assessments or psychometric testing, which are crucial for understanding a students cognitive abilities, personality traits, and emotional stability. Without such assessments, students may struggle to receive proper career guidance or personalized learning recommendations tailored to their strengths and interests.

Another major limitation is the lack of activity-based recommendations. The absence of an intelligent system that analyzes a students participation in academic and co-curricular activities prevents tailored suggestions that could enhance their learning experience. This leads to a one-size-fits-all approach, which does not cater to individual student needs.

3.2 Proposed System

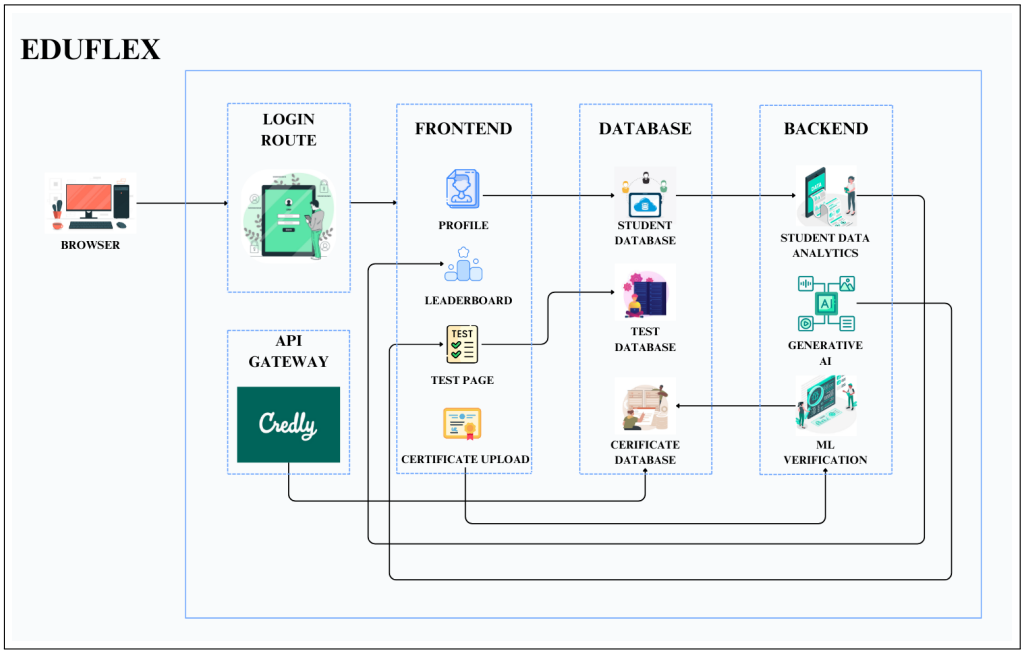


Figure 3.1: System Architecture

To overcome the challenges of manual verification, lack of real-time student performance tracking, and the absence of learning recommendations, EduFlex introduces an automated, intelligent, and scalable solution. By leveraging machine learning, psychometric analysis, and data-driven insights, the system ensures secure verification, dynamic progress tracking (Fig. 3.1) . The proposed system enhances engagement, improves accuracy, and provides educators and students with real-time analytics to make informed decisions.

- Automated Certificate Verification: The system utilizes ML models to detect forged certificates by analyzing metadata, fonts, and image inconsistencies, ensuring tamper-proof validation.
- Real-Time Student Performance Tracking: Educators and mentors receive dynamic insights into student progress, enabling timely interventions to support learning and career growth.
- Psychometric & Activity-Based Recommendations: EduFlex conducts cognitive and personality assessments to provide students with personalized learning paths and career guidance based on strengths and interests.
- Enhanced User Engagement: By incorporating interactive leaderboards and gamified testing, the system encourages active participation and motivation among students.
- Secure & Scalable Architecture: The modular design ensures scalability, high availability, and security, making it adaptable to various educational institutions and corporate environments.

3.2.1 Critical Components of System Architecture

EduFlex is designed to provide a secure, intelligent, and efficient education validation system by integrating multiple components that work seamlessly together. Each module serves a crucial function, from user authentication and data processing to real-time analytics and AI-driven verification. The following sections outline the six core components that define the system's architecture:

1. **Browser Interface**
   - Acts as the **primary access point** for users.
   - Supports **user authentication, certificate uploads, and test-taking functionality**.
2. **Login Route**
   - Ensures **secure authentication** and **access control**.
   - Verifies **user identity** before granting system access.
3. **API Gateway**
   - Facilitates **external integrations** (e.g., Credly for **certificate verification**).
   - Ensures **secure communication** between EduFlex and **third-party platforms**.
4. **Frontend Module**
   - Provides an **intuitive UI** for students, educators, and institutions.
   - Includes the following key features:
     - **Profile Management** – Stores user details and tracks learning progress.
     - **Leaderboard** – Encourages **healthy competition** among students.
     - **Test Page** – Hosts **psychometric and skill-based assessments**.
     - **Certificate Upload** – Enables **secure submission for verification**.
5. **Database Layer**
   - Centralized storage system consisting of:
     - **Student Database** – Stores user credentials, **academic records**, and progress.
     - **Test Database** – Records results of **psychometric and performance-based tests**.
     - **Certificate Database** – Maintains **uploaded certificates for validation**.
6. **Backend Module**
   - The **core processing unit** that drives the system's intelligence:
     - **Student Data Analytics** – Analyzes **student performance trends** for insights.
     - **Generative AI** – Provides **personalized learning suggestions**.
     - **ML Verification** – Detects **fake certificates** using machine learning algorithms.

3.3 System Diagrams

3.3.1 UML Diagram

UML diagrams help in visualizing system behavior, interactions, and workflows. This section covers three key types: Activity Diagrams (showing step-by-step and parallel flows), Use Case Diagrams (mapping user roles and functionalities), and Sequence Diagrams (depicting message flow over time). Together, they offer a clear, structured view of system operations for effective design.

3.3.1.1 Activity Diagram

Figure 3.2: Activity Diagram

This Activity Diagram (Fig. 3.2) showcases the dynamic and conditional flow of the EduFlex platform. After users log in and build their profiles, they can link their Credly accounts to fetch badges or proceed without them. The system allows students to upload achievements, which are then verified using a Machine Learning model. If a certificate is flagged or cannot be verified by ML, it is passed on to the admin for final review ensuring authenticity and integrity. In addition, the platform includes an AI-based Soft Skills Assessment, allowing students to evaluate and reflect on their interpersonal abilities. These diverse insights are then visualized in a personalized dashboard, promoting holistic student development.

3.3.1.2 Use Case Diagram

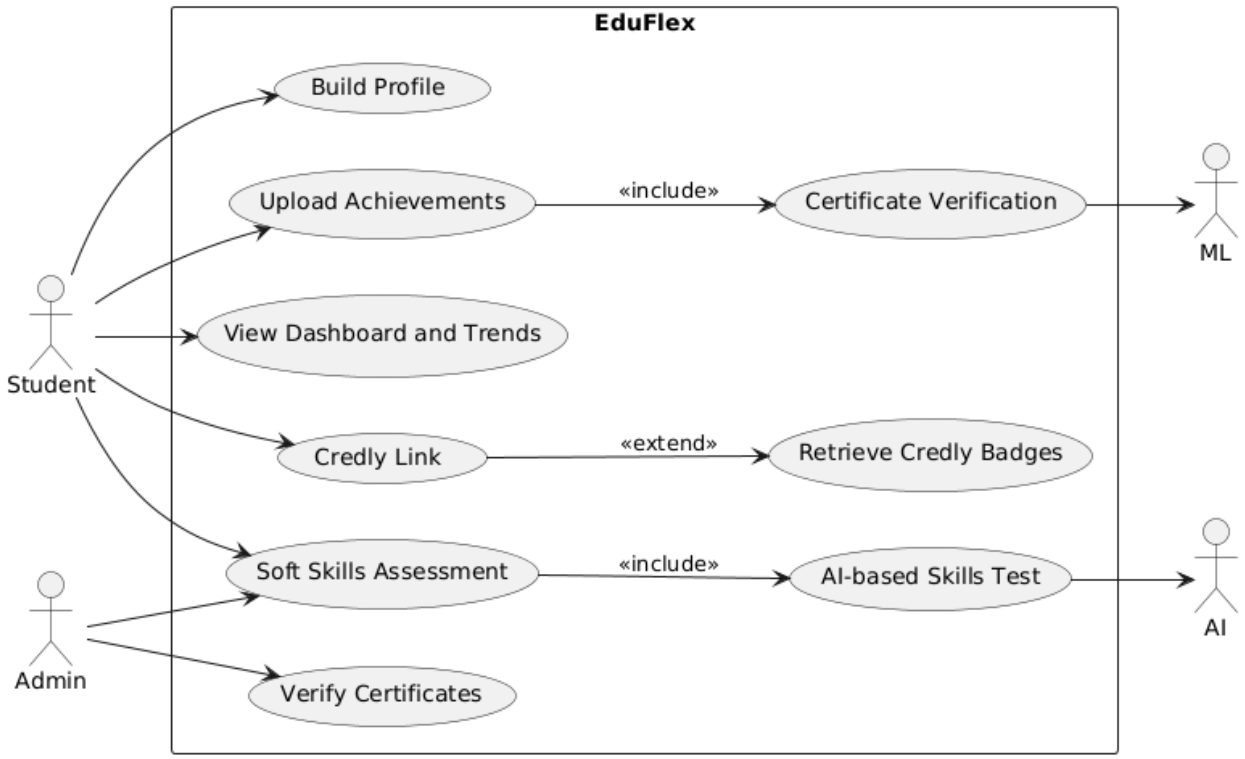


Figure 3.3: Use Case Diagram

The Use Case Diagram (Fig. 3.3) represents the core interactions within the EduFlex platform, highlighting the roles of the Student, Admin, and external components like ML and AI. Students can build profiles, link their Credly accounts, upload achievements, take soft skills assessments, and view dashboards. The system includes automated interactions such as certificate verification using Machine Learning and an AI-based soft skills test. Admins play a crucial role in verifying certificates flagged by the ML model.

3.3.1.3 Sequence Diagram

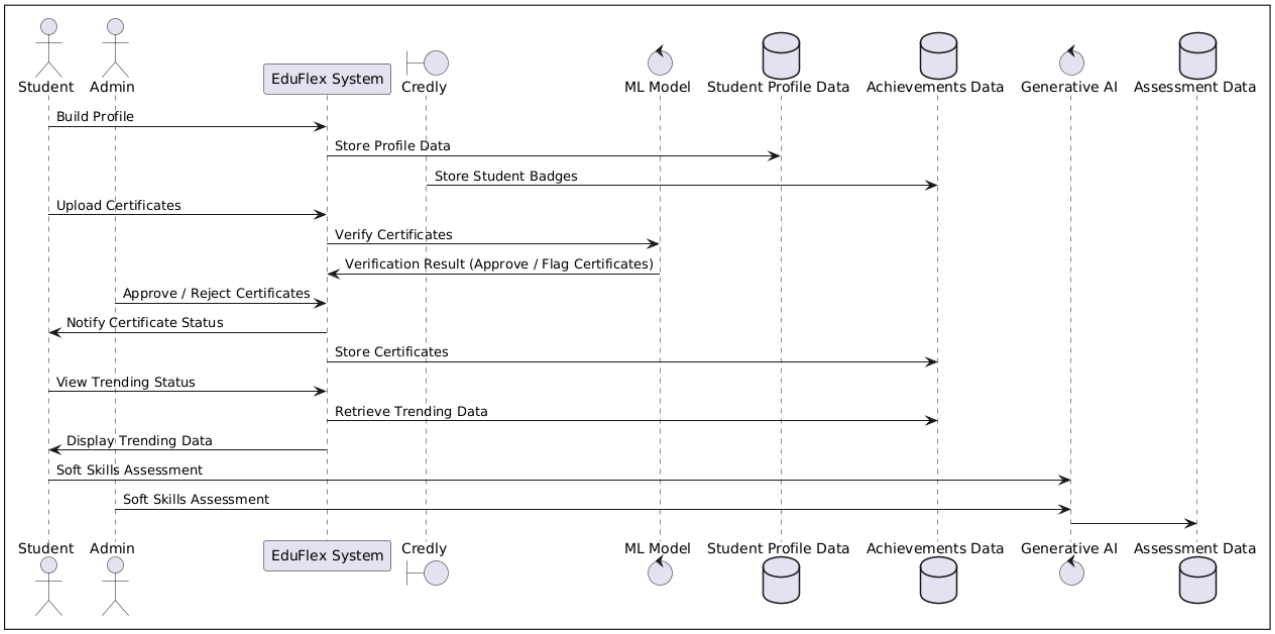


Figure 3.4: Sequence Diagram

The Sequence Diagram (Fig. 3.4) outlines the chronological interaction between key components of the EduFlex system - Student, Admin, ML Model, Credly, Generative AI, and multiple data repositories. It begins with the student building a profile and uploading certificates. The system interacts with Credly to fetch badges and with the ML Model to verify certificates. If flagged, the admin reviews and makes the final decision. Upon approval, certificates are stored, and students are notified of the status.

Additionally, the student can view trending status data, which is fetched and displayed from the backend. The process also includes a dedicated flow for AI-based Soft Skills Assessment, where responses are analyzed and stored, contributing to a well-rounded profile. This diagram (Fig. 3.4) emphasizes seamless automation with strategic human intervention, ensuring accuracy, personalization, and transparency.

3.4 Formula and Methodology

3.4.1 Methodology: Certificate Verification Model

To detect tampered certificates effectively, a machine learning-based approach is implemented, focusing on key aspects such as dataset preparation, feature extraction, model selection, training, validation, and deployment.

3.4.1.1 Dataset Collection

A diverse dataset of authentic and tampered certificates is compiled. Genuine certificates are collected from verified sources, while tampered versions are created using document editing tools like Adobe Photoshop, Microsoft Word, and various online PDF editors. The modifications include altering names, dates, course details, and institutional seals, ensuring the dataset represents real-world forgery techniques.

3.4.1.2 Feature Extraction

Relevant features are extracted from PDF certificates using `pdfminer` and `PyMuPDF`. These features include:

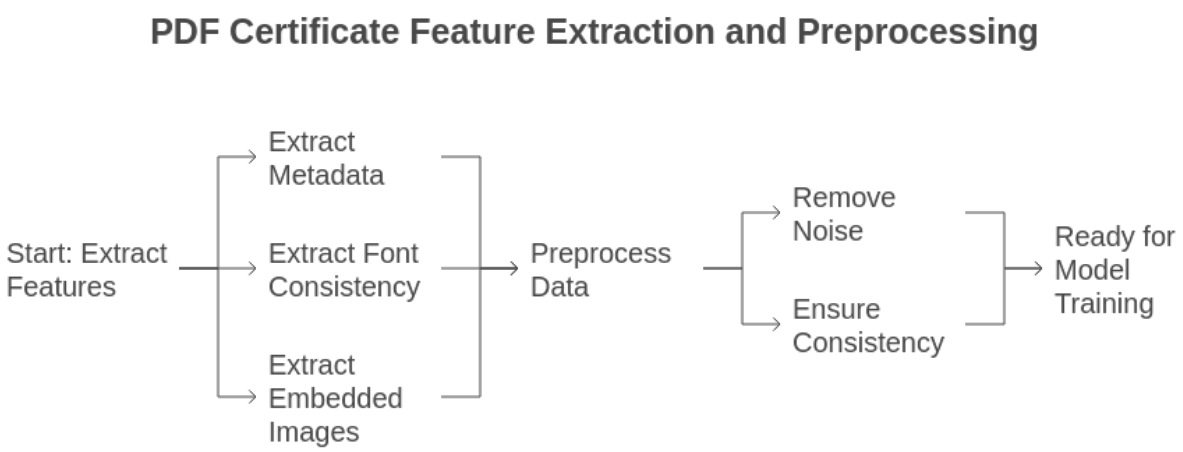


Figure 3.5: Feature Extraction Process Flow

- Metadata attributes (e.g., PDF producer, creation date, modification history)
- Font consistency (e.g., font style, size, alignment)
- Embedded images and color distributions

The extracted data is preprocessed to remove noise and ensure consistency before model training(Fig. 3.5).

3.4.1.3 Model Selection and Training

A Random Forest classifier is used for certificate verification due to its efficiency in handling structured data and its ability to detect subtle inconsistencies in tampered documents. The model is trained with the extracted features and validated using an 80-20 train-test split, with an additional 20% validation set from the training data.

Random Forest Parameters:

- Number of Trees: 100
- Criterion: Gini Impurity
- Maximum Depth: None (full tree expansion allowed)
- Minimum Samples Split: 2
- Minimum Samples per Leaf: 1

The model is evaluated using accuracy, precision, and recall to ensure high detection reliability.

3.4.1.4 Deployment Strategy

The trained model is deployed as a Flask-based REST API with the core validation logic implemented in the `/validate-certificate-two` endpoint. The processing pipeline ensures:

- Secure certificate uploads and validation
- Automated feature extraction and classification
- JSON-based response with results (Real/Fake) and potential tampering indicators

Additionally, an auxiliary endpoint `/checkcertlevel` is included to categorize certificates by type (academic/sports) and level (local/global), enhancing usability in various verification scenarios.

This approach provides an efficient, scalable, and automated solution for certificate fraud detection, ensuring document authenticity in educational and professional settings.

3.4.2 Working of Model:

A. CNN In our model, a convolutional neural network is used where the first nodes take the PDF features and process them by passing the data to the next array of nodes. At each layer, the data is filtered and passed forward to the next nodes. Initially, the input layer consists of 7 nodes. In the next layer, the number of nodes increases to 64, followed by a dropout to 32 nodes to refine the features and prevent overfitting. Finally, the output layer consists of 2 nodes, producing an output of either 1 or 0, representing "Real" or "Fake," respectively, as illustrated in Fig. 3.6.

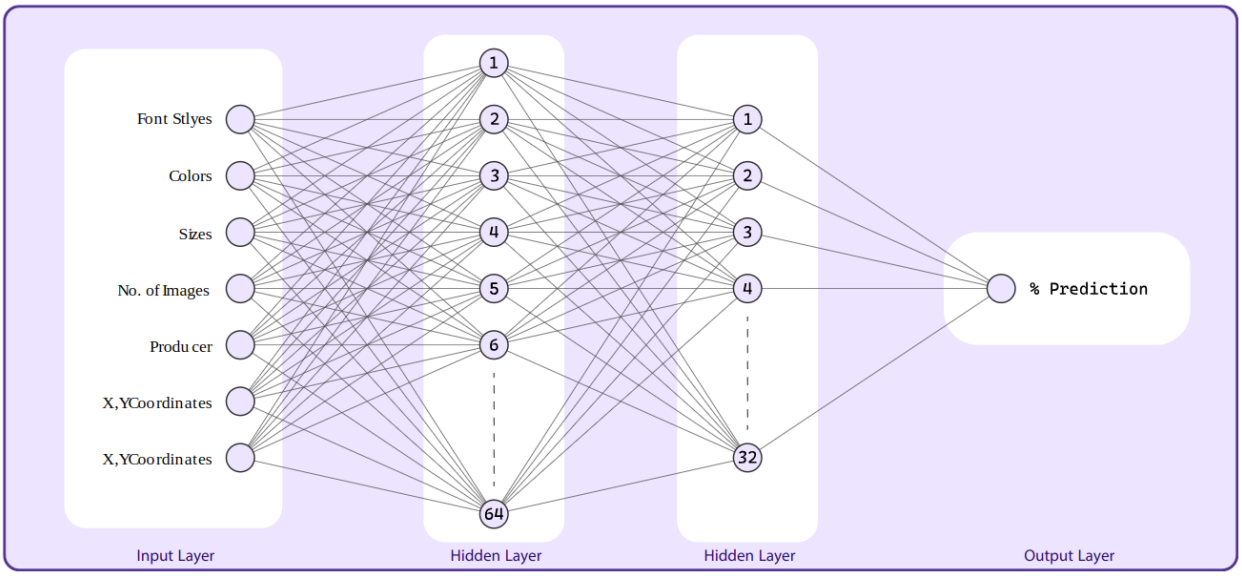


Figure 3.6: Convolutional Neural Network

B. ReLU Activation Function in Feature Extraction In our model, CNN is used for feature extraction from PDF documents. The extracted features include font size, style, color, text positioning (bounding box coordinates), number of images, and the PDF producer name. ReLU plays a critical role in this process by ensuring that the features captured by the CNN are processed effectively before being passed to the Random Forest classifier. ReLU is a widely used activation function in CNN, defined as:

$$f(x) = \max(0, x)$$

It outputs the input directly if it is positive and outputs zero otherwise. This activation function is computationally efficient and enables faster learning in deep neural networks. The specific roles of ReLU in our model are as follows:

- **Introducing Non-Linearity:** Real-world data patterns, such as the relationships between font size, style, and possible tampering, are non-linear. ReLU helps CNN capture these complex patterns, allowing the model to make better decisions regarding document integrity.
- **Preventing Vanishing Gradients:** Unlike other activation functions like sigmoid or tanh, which squash values into a limited range, ReLU does not restrict the gradient. This ensures that stronger gradients are preserved during back propagation, leading to faster and more stable training.
- **Efficient Feature Extraction for Classification:** After feature extraction by the CNN, ReLU ensures that only significant activations (positive values) contribute to the final classification. Irrelevant activations are suppressed, helping the Random Forest classifier make more informed predictions.
- **Sparse Activation and Computational Efficiency:** ReLU outputs zero for negative inputs, which introduces sparsity in the activations. This sparsity reduces unnecessary computations, making the feature extraction process more efficient.

Impact on the CNN-Random Forest Pipeline :

Features extracted by the CNN are refined using ReLU, then flattened and fed into a Random Forest classifier. ReLU enhances feature relevance and prevents vanishing gradients, ensuring meaningful input to the classifier. The Random Forest aggregates decisions from multiple trees to classify the PDF as *tampered* or *not tampered*, improving the hybrid model's accuracy and stability.

3.4.3 Formula

3.4.3.1 Application of Random Forest in Certificate Verification

Random Forest is an ensemble learning method that enhances classification accuracy by training multiple decision trees on different subsets of data and combining their outputs. For certificate verification, it effectively analyzes structured features extracted from PDF documents, such as font properties, metadata attributes, image count, and text positioning. How Random Forest Works in This Model:

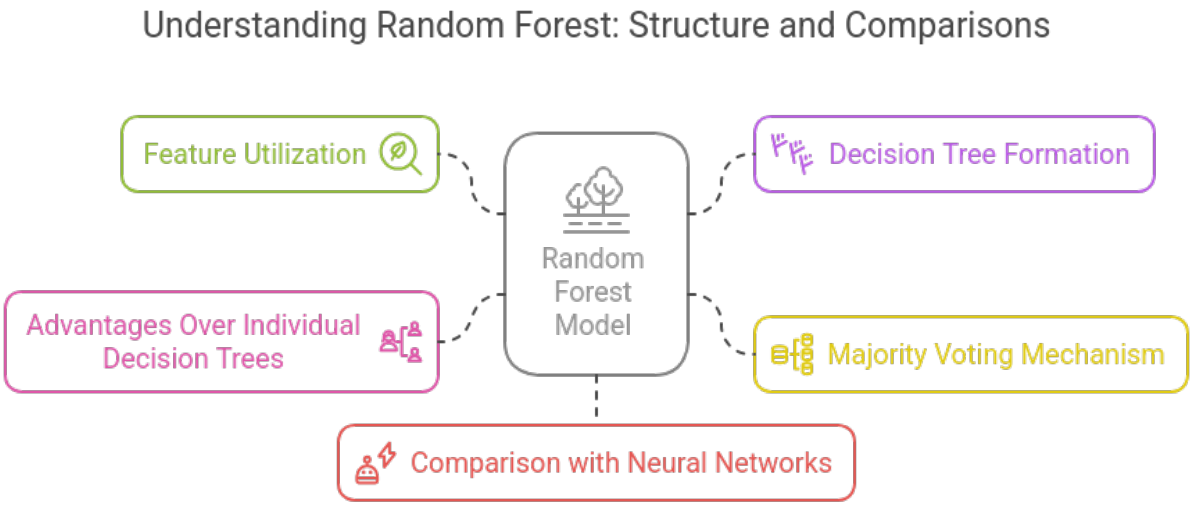


Figure 3.7: Random Forest Structure

- Feature Utilization:

  The model takes preprocessed numerical and categorical features, including font size, color consistency, character positioning, metadata attributes, and embedded image count.

- Decision Tree Formation:

  Trees analyze individual features to detect anomalies, such as mismatched font styles or inconsistent metadata timestamps.

- Majority Voting Mechanism:

  The final classification is determined by majority voting, making the model robust against noise and overfitting.

- Advantages Over Individual Decision Trees:

  Handles Missing or Noisy Data: Since each tree trains on different data subsets, it mitigates the impact of outliers.

For structured certificate verification (Fig. 3.7), Random Forest provides a strong baseline model, offering interpretability and efficiency. However, deep learning methods like neural networks may be preferable for analyzing intricate document manipulations beyond structural inconsistencies.

How Random Forest makes decisions in our model:

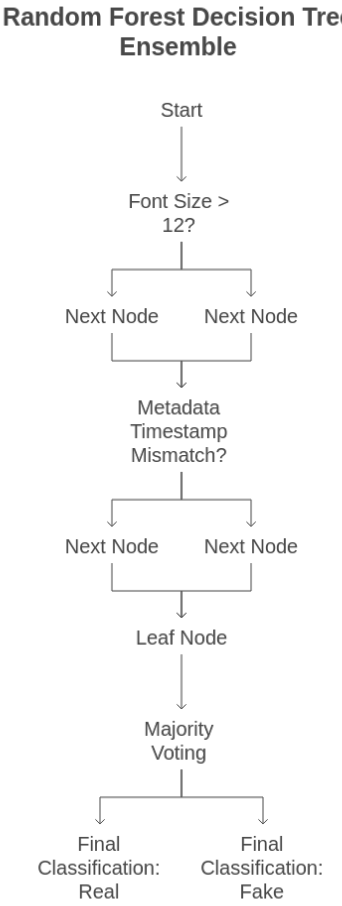


Figure 3.8: Working of Random Forest

The flowchart illustrates the working of a Random Forest Decision Tree Ensemble used for certificate tampering detection. Each tree evaluates features such as font size and metadata timestamp to reach a leaf node (Fig. 3.8), where the outcome is determined. The final classification—real or fake—is decided through majority voting across all trees in the ensemble.

Formula:

$$P(\text{Tampered}) = \frac{\text{Number of Trees that Predict "Tampered"}}{\text{Total Number of Trees}}$$

Chapter 4

Project Implementation

4.1 Code Snippets

```python
@app.route('/validate-certificate-two', methods=['POST'])
def validate_certificate_two():
    try:
        # Get the JSON data from the request
        data = request.json
        username = data.get('username')
        filename = data.get('filename')

        if not username or not filename:
            return jsonify({'error': 'Username and filename are required'}), 400

        base_dir = 'uploads'  # Update this path to the actual base directory
        file_path = os.path.join(base_dir, username, secure_filename(filename))
        print("The original file path:", file_path)

        # Ensure the file exists
        if not os.path.isfile(file_path):
            print("File not found at validate certificate")
            return jsonify({'error': 'File not found'}), 404

        # Process the PDF file to extract features
        features = extract_font_information_with_metadata_and_images(file_path)
        if not features:
            return jsonify({'error': 'No features extracted from PDF'}), 500

        # Prepare data for prediction
        df = prepare_data_for_prediction(features)

        # Apply preprocessing and make predictions
        X_processed = preprocessor.transform(df)
        predictions = model.predict(X_processed)

        # Determine the final result
        all_real = True
        fake_producer = None

        for i, pred in enumerate(predictions):
            if pred <= 0.5:
                all_real = False
                fake_producer = features[i]['Producer']
                break

        # Return the result based on prediction
        if all_real:
            return jsonify({"result": "Real"})
        else:
            return jsonify({"result": "Fake", "Edited_By": f"{fake_producer}"})
```

Figure 4.1: Verification Model

The 'validate-certificate-two' function operates as a Flask endpoint that implements a sophisticated PDF certificate authentication system. Upon receiving a POST request with username and filename parameters, it locates the PDF in the uploads directory and extracts seven critical features: font style, color, size, text positioning coordinates (x0, y0, x1, y1),

and PDF producer metadata. These features are preprocessed through a standardization pipeline before being analyzed by a CNN model, which evaluates each text element to generate confidence scores – values below 0.5 (Fig. 4.1), indicate potential tampering. The system responds with either a "Real" status for genuine certificates or a "Fake" status along with the name of the suspected editing software for compromised documents. The implementation includes comprehensive error handling for missing files, extraction failures, and processing errors, ensuring robust operation while maintaining clear communication through HTTP status codes and descriptive error messages.

```python
@app.route("/fetch-badges",methods =['GET'])
def fetch_badges_2():
    url = request.args.get('url') + '/badges'
    if not url :
        return jsonify({'error': "Url not provided"}), 400
    try:
        responce = requests.get(url,headers={'User-Agent':'Mozilla/5.0'})

        if responce.status_code ==200:
            soup = BeautifulSoup(responce.text,'html.parser')

            badges = soup.find_all('div', class_ ='cr-standard-grid-item-content')
            certificates = []
            for badge in badges:
                # Extract details for each badge
                title = badge.find('div', class_='cr-standard-grid-item-content__title')
                subtitle = badge.find('div', class_='cr-standard-grid-item-content__subtitle')
                certificates.append({
                    'certificate_name': title.text.strip() if title else 'N/A',
                    'issuer_name': subtitle.text.strip() if subtitle else 'N/A',
                })

            return jsonify(certificates),200
        elif responce.status_code != 200:
            return jsonify({"error": f"Scrapper status code ${responce.status_code}"}), 500
    except ExceptionGroup as e:
        return({"error" : f"Internal server error , ${e.message}"}),500

################################################################################        Validat
```

Figure 4.2: Credly Badges

The fetch-badges function is a Flask endpoint (Fig. 4.2) that scrapes and processes badge information from Credly's digital credential platform. When called with a GET request containing a URL parameter, it appends '/badges' to the URL and makes an HTTP request using the requests library with a Mozilla user agent header. Upon receiving a successful response (status code 200), it uses BeautifulSoup to parse the HTML content and locate all div elements with the class 'cr-standard-grid-item-content'. For each badge found, it extracts the title and subtitle information from specific div classes ('cr-standard-grid-item-contenttitle' and 'cr-standard-grid-item-contentsubtitle'), storing them in a list of dictionaries with 'certificatename' and 'issuername' keys, using 'N/A' as a fallback if either element is missing. The function handles errors by returning appropriate HTTP status codes and error messages, particularly for cases where the URL is missing (400), scraping fails (500), or when other exceptions occur, ultimately returning the processed badge information as a JSON response with status code 200 when successful.

```python
@app.route('/upload', methods=['POST'])
def upload_file():
    if 'file' not in request.files:
        return jsonify({'message': 'No file part'}), 400

    file = request.files['file']
    selection = request.form.get('selection', 'id')

    if file.filename == '':
        return jsonify({'message': 'No selected file'}), 400

    if file:
        filename = file.filename
        file_path = os.path.join(app.config['UPLOAD_FOLDER'], filename)
        file.save(file_path)

        if filename.endswith('.pdf'):
            if selection == 'id':
                data = extract_ids_from_pdf(file_path)
            elif selection == 'name':
                data = extract_names_from_pdf(file_path)
            else:
                return jsonify({'message': 'Invalid selection type'}), 400
        elif filename.endswith('.xlsx') or filename.endswith('.xls'):
            if selection == 'id':
                data = extract_ids_from_excel(file_path)
            elif selection == 'name':
                data = extract_names_from_excel(file_path)
            else:
                return jsonify({'message': 'Invalid selection type'}), 400

        else:
            return jsonify({'message': 'Unsupported file type'}), 400
        cleaned_data = []
        for item in data:
            # Remove unwanted characters, e.g., parentheses, special symbols
            cleaned_item = re.sub(r'[(){}\[\],;]', '', item).strip()  # Adjust regex as needed
            if cleaned_item:  # Only add non-empty items
                cleaned_data.append(cleaned_item)

        print("MENTOR DATA:", cleaned_data)
        return jsonify({'data': cleaned_data})

    return jsonify({'message': 'Failed to upload file'}), 500
```

Figure 4.3: Mentees Batch creation

The upload_file Flask endpoint (Fig. 4.3) facilitates mentor batch creation by processing both PDF and Excel files (.pdf, .xlsx, .xls) through a POST request. When a file is uploaded along with a selection parameter ('id' or 'name'), the function saves it to the 'apiuploads' directory and processes it accordingly. For PDFs, it uses pdfplumber to extract tables, while Excel files are processed using pandas. The extraction method varies based on the selection parameter: 'id' mode extracts numeric identifiers while 'name' mode extracts text entries, ignoring numbers and short strings. The extracted data undergoes cleaning to remove special characters and empty entries using regex (re.sub()). This processed data is then returned as a JSON response, enabling mentors to efficiently import student IDs or names from their existing documents, streamlining the batch creation process while handling potential errors with appropriate HTTP status codes (400 for invalid requests, 500 for processing failures).

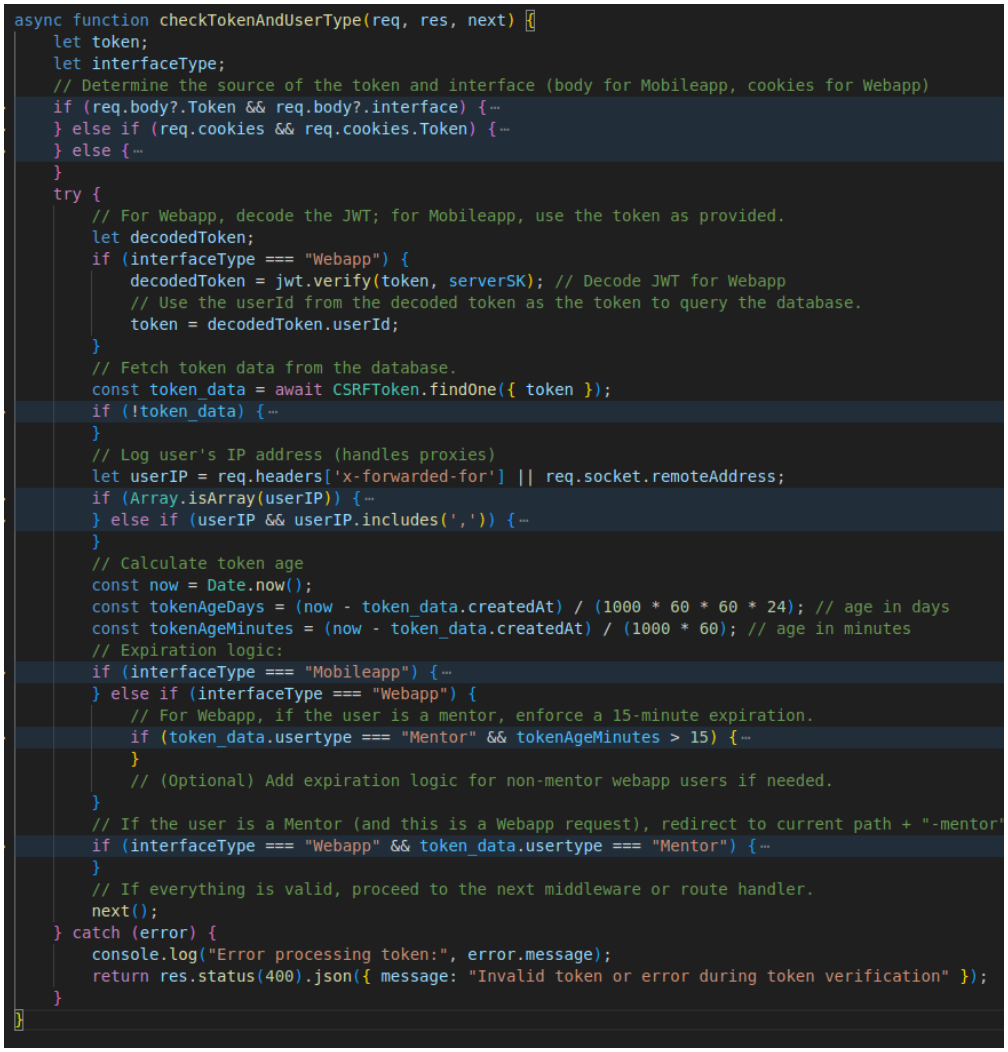


Figure 4.4: Check Session

The checkTokenAndUserType middleware function (Fig. 4.4) implements a comprehensive authentication and authorization system that handles both web and mobile application requests. It first determines the interface type and extracts the authentication token either from the request body (mobile) or cookies (web), with JWT decoding specifically for web tokens. The middleware performs several security checks: verifying token existence in the database, logging user IP addresses (including proxy handling), and enforcing different expiration policies (30 days for mobile tokens, 15 minutes for web mentor sessions). For web requests from mentors, it automatically redirects to mentor-specific routes by appending "-mentor" to the URL. The function implements robust error handling for invalid tokens, expired sessions, and verification failures, returning appropriate HTTP status codes and messages. The middleware uses the CSRFToken model for token storage and verification, integrates with a logging system for security events, and incorporates environment variables for JWT secret management, making it a secure and flexible authentication layer that accommodates different user types and interfaces while maintaining strict security protocols.

4.2 Steps to Access the System

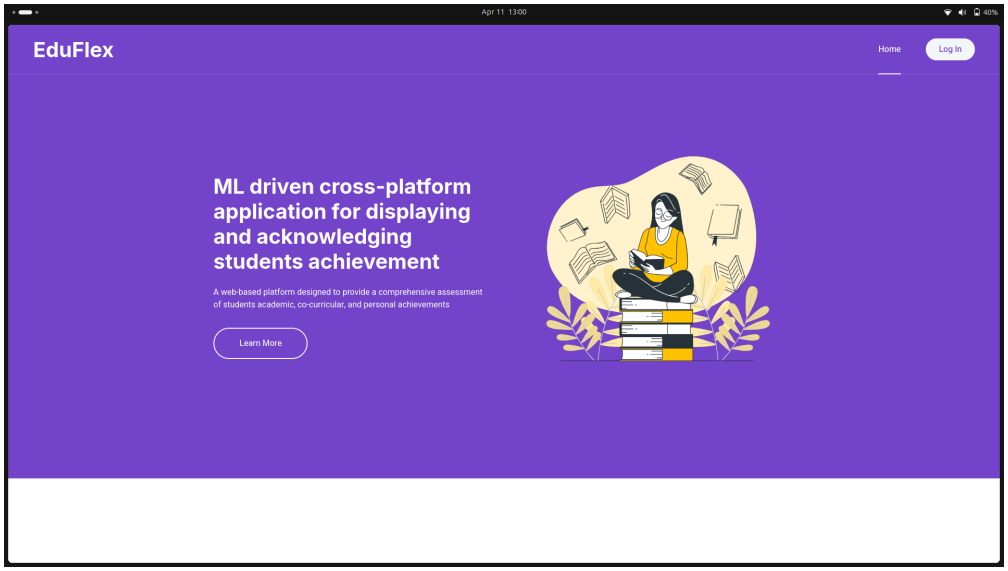


Figure 4.5: Landing Page

The EduFlex landing page (Fig. 4.5) offers an interactive overview of core features like certificate verification, skill assessment, and personalized insights, enabling users to quickly grasp how the platform supports and enhances their academic journey.

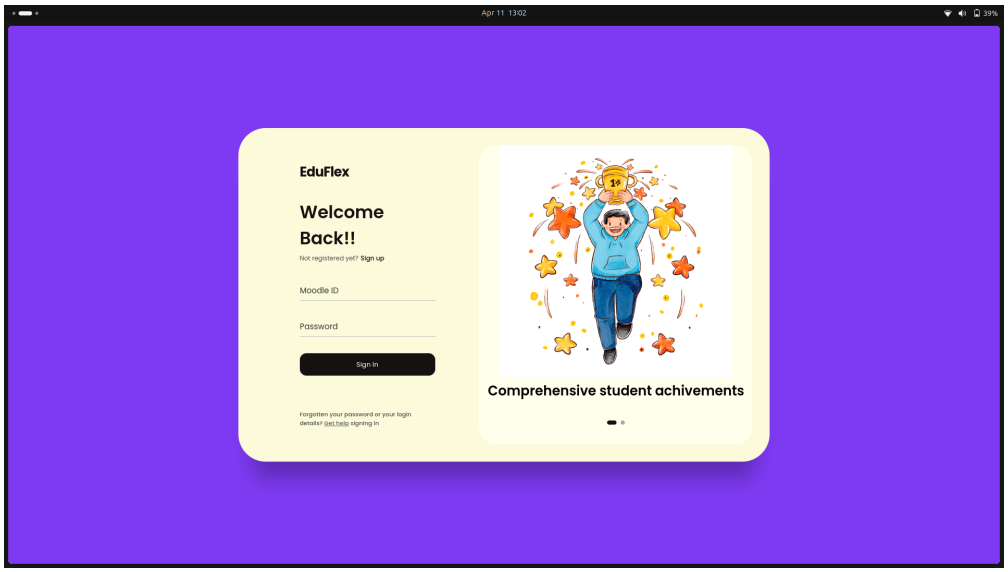


Figure 4.6: Login and Register

EduFlex provides a dedicated login and registration (Fig. 4.6) system where students can create their accounts to access personalized dashboards, track their progress, and manage certificates. The same login page is also used by mentors and faculty, allowing them to verify student data, view insights, and provide necessary approvals, ensuring a unified and secure access experience for all users.

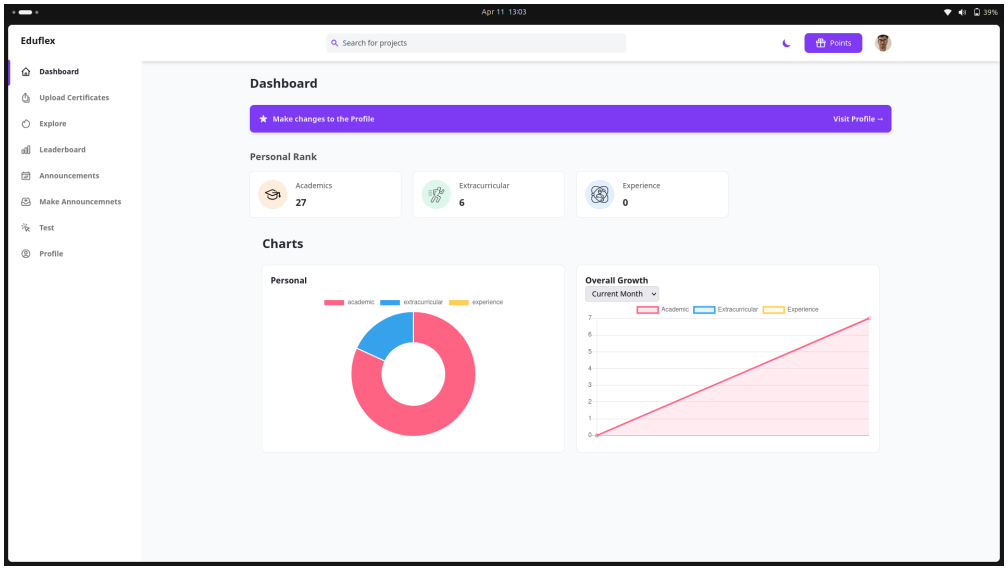


Figure 4.7: Student Dashboard

After logging in, students are directed to a personalized dashboard (Fig. 4.7) where they can view their earned points across different domains like academics, co-curricular, and extracurricular activities. The dashboard also provides interactive graphs and visualizations, allowing students to track their progress and growth over time in a clear and engaging manner.

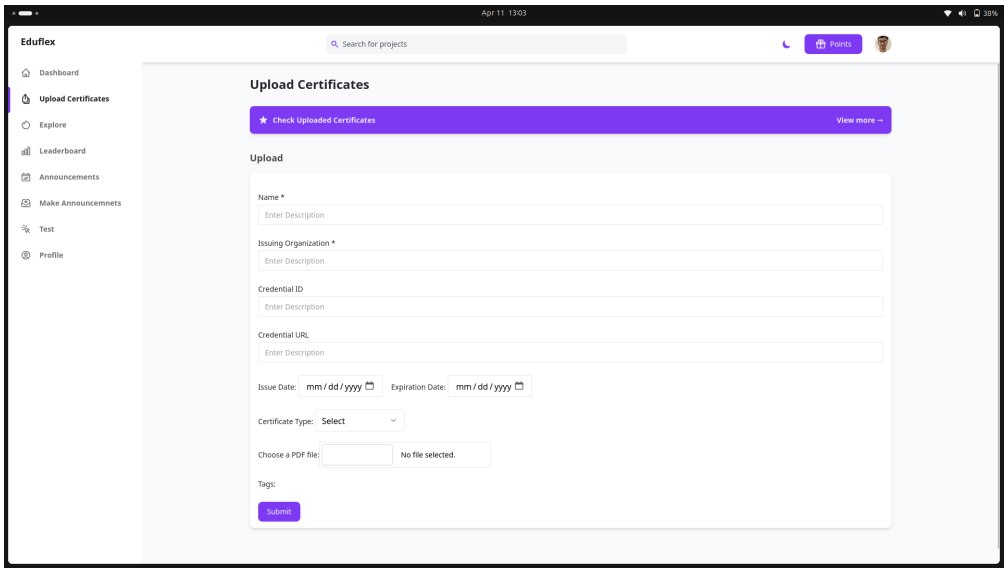


Figure 4.8: Upload Certificate

There is a dedicated "Upload Certificates"(Fig. 4.8) page for students, where they can easily upload their achievement certificates. This page allows students to submit certificates across various domains, which are then verified by the system and reflected in their dashboard, contributing to their overall progress and points.

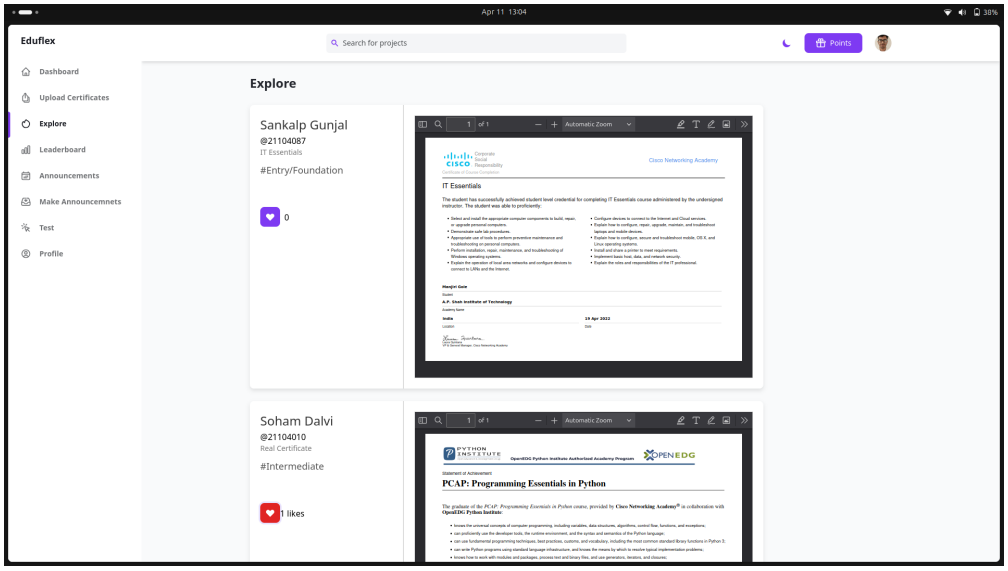


Figure 4.9: Explore Page

There is also an "Explore" (Fig. 4.9) page where all the certificates uploaded by students are displayed. This page acts as a public gallery, allowing students and mentors to view and appreciate the achievements shared within the community, fostering a culture of healthy competition and recognition.

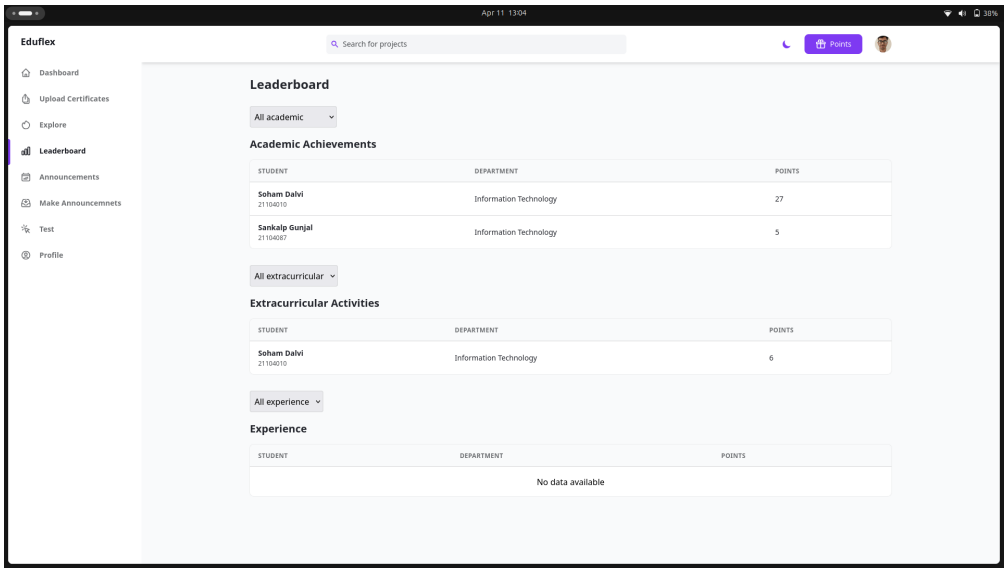


Figure 4.10: Leaderboard

The platform also features a dedicated "Leaderboard" (Fig. 4.10) page, where students are ranked based on their performance across three major domains — Academic, Extracurricular, and Experience. This leaderboard provides a healthy competitive environment, motivating students to actively participate and improve in all areas. Additionally, it offers filter options, allowing users to view rankings based on specific certificate types or domains for a more focused comparison.

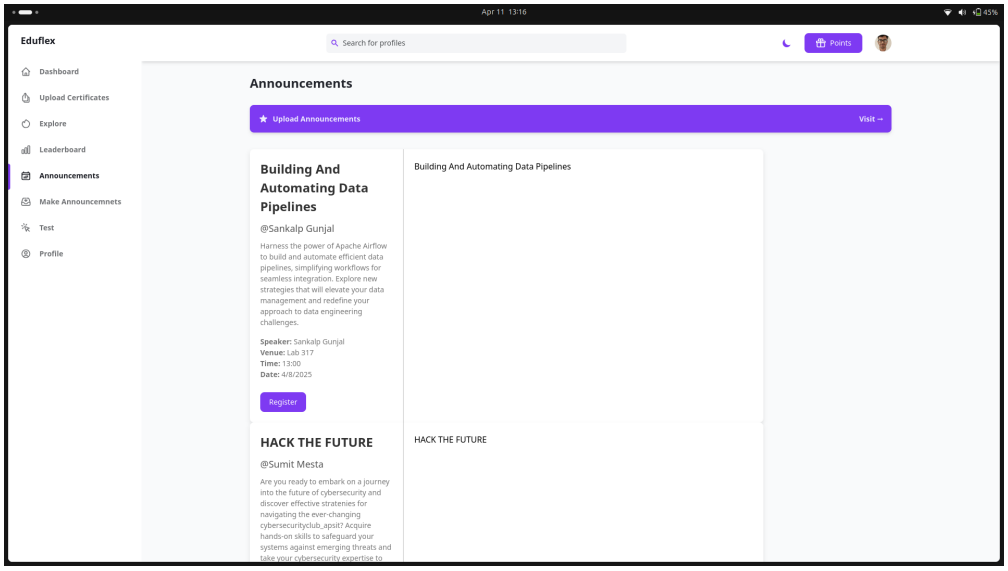


Figure 4.11: Announcements

There is also an "Announcements" (Fig. 4.11) page within the platform where all upcoming events, competitions, and opportunities are posted. Students can directly register for these events through this page, making the process seamless and accessible. Additionally, mentors have the functionality to download the list of registered students, enabling them to efficiently manage and coordinate event participation.

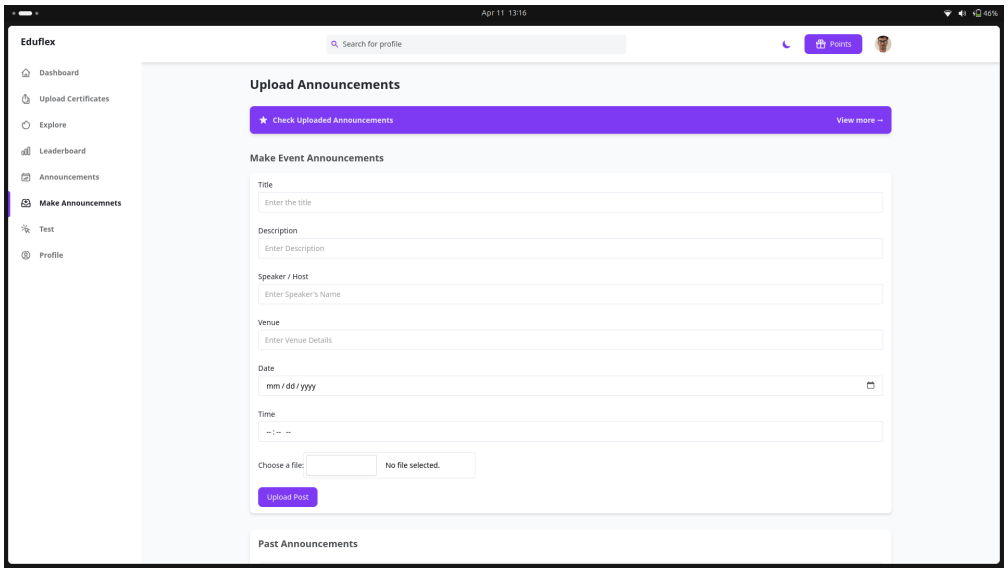


Figure 4.12: Make Announcements

There is also an "Upload Announcement" (Fig. 4.12) page where both students and mentors can create and publish announcements related to events, workshops, or opportunities. This allows them to provide details like event description, date, registration link, and other necessary information. Once uploaded, these announcements become visible on the main Announcements page, enabling other students to view and register directly from the platform.

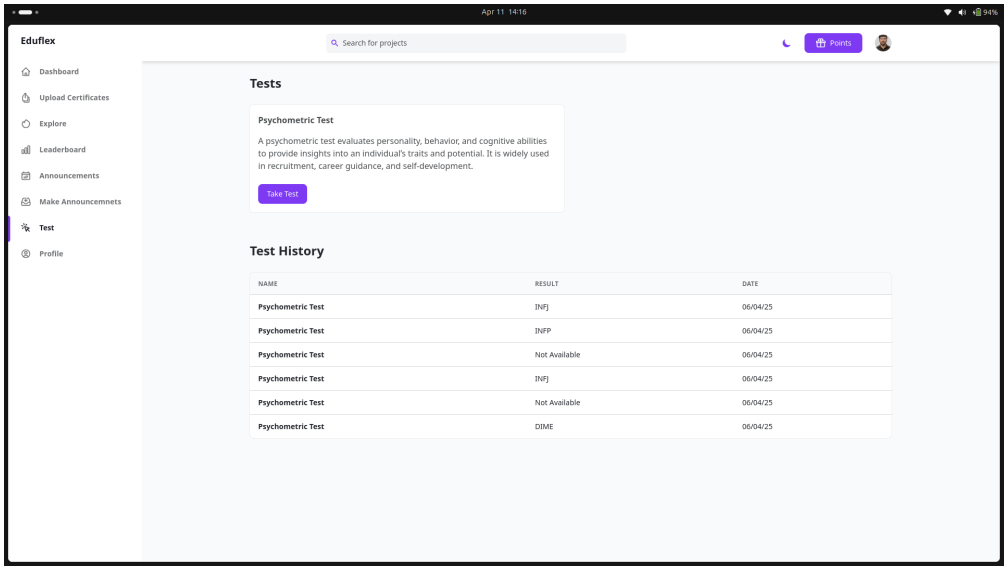


Figure 4.13: Test Results History

There is also a dedicated "Test Dashboard" (Fig. 4.13) where both students and mentors can access information related to tests conducted on the platform. Students can view their previous test results, scores, and performance analysis. They can also take new tests directly from this page.

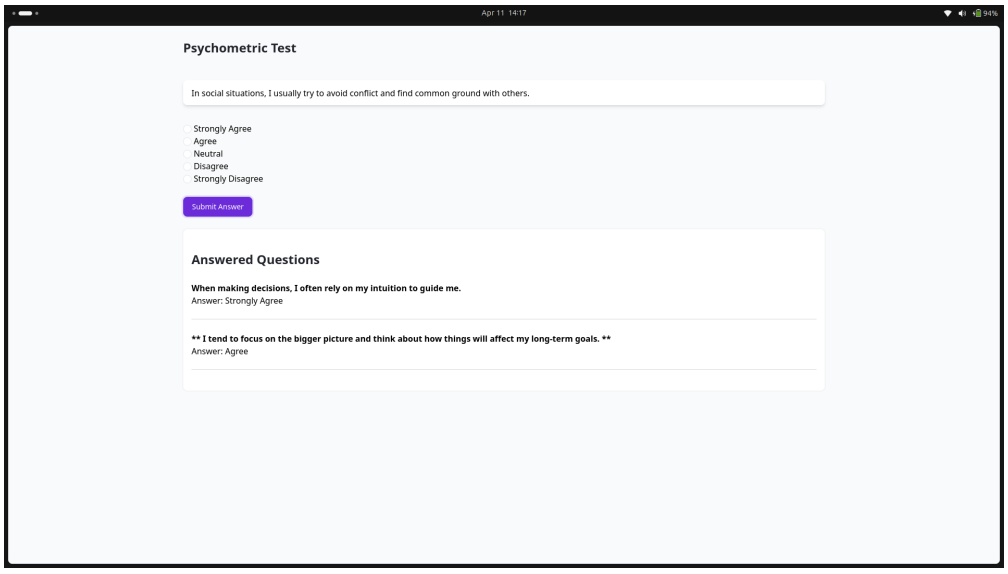


Figure 4.14: Psychometric-test

There is also a dedicated Psychometric Test page (Fig. 4.14) where students can take AI-driven psychometric tests. In this test, the questions are dynamically generated based on the student's previous answers, allowing for a more personalized and adaptive assessment. This helps in analyzing the student's personality traits, strengths, and areas of improvement in a more accurate and interactive manner.

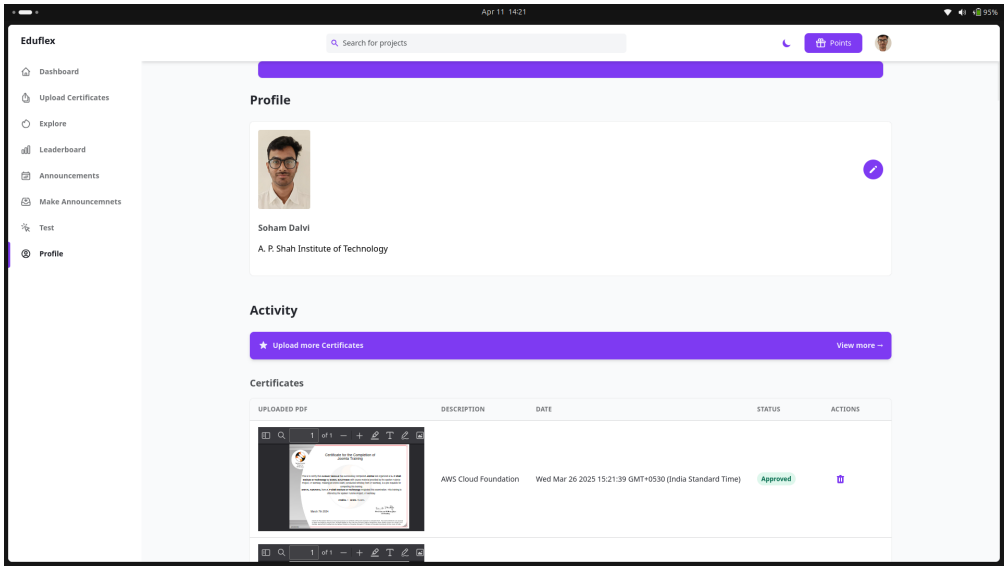


Figure 4.15: Student Profile

There is also a Student Profile page (Fig. 4.15) where students can view all their uploaded posts and certificates. Additionally, they can update their educational details, add new skills, showcase their projects, and view their earned Credly badges, providing a complete overview of their achievements and personal growth within the platform.

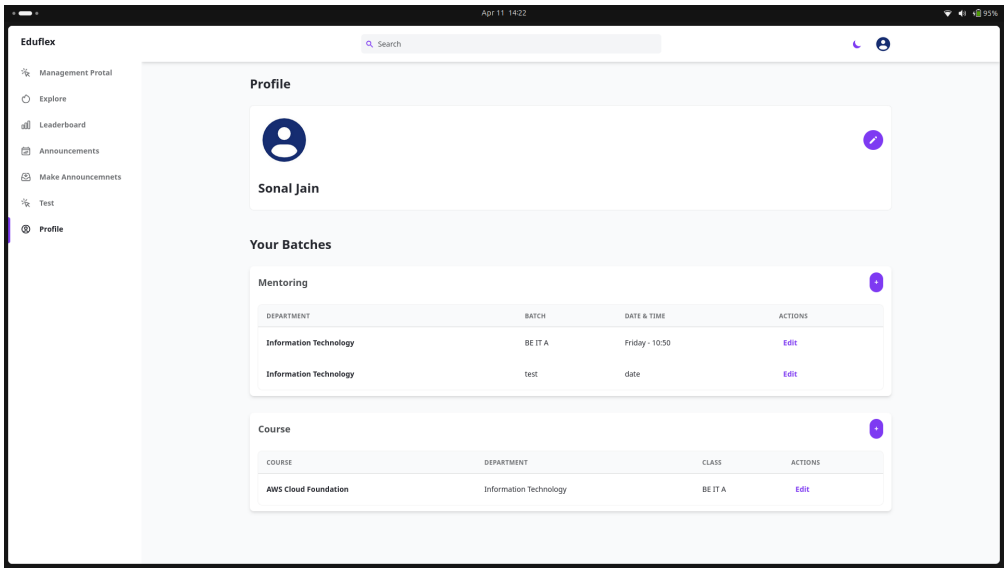


Figure 4.16: Mentor Profile

There is also a Mentor Profile page (Fig. 4.16) where mentors can update their personal information, create batches of students, and add or manage courses. This allows mentors to efficiently organize students and deliver structured learning content within the platform.

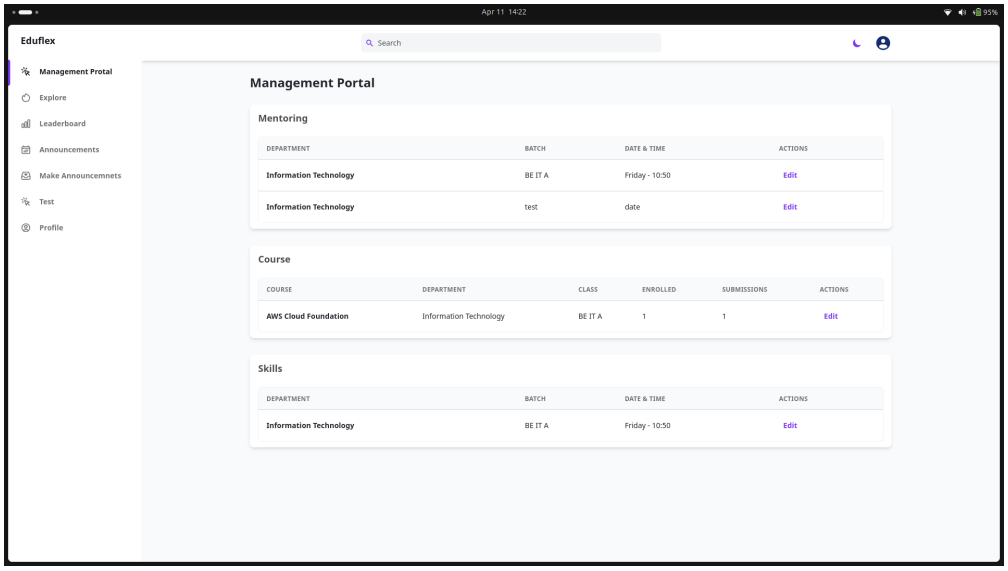
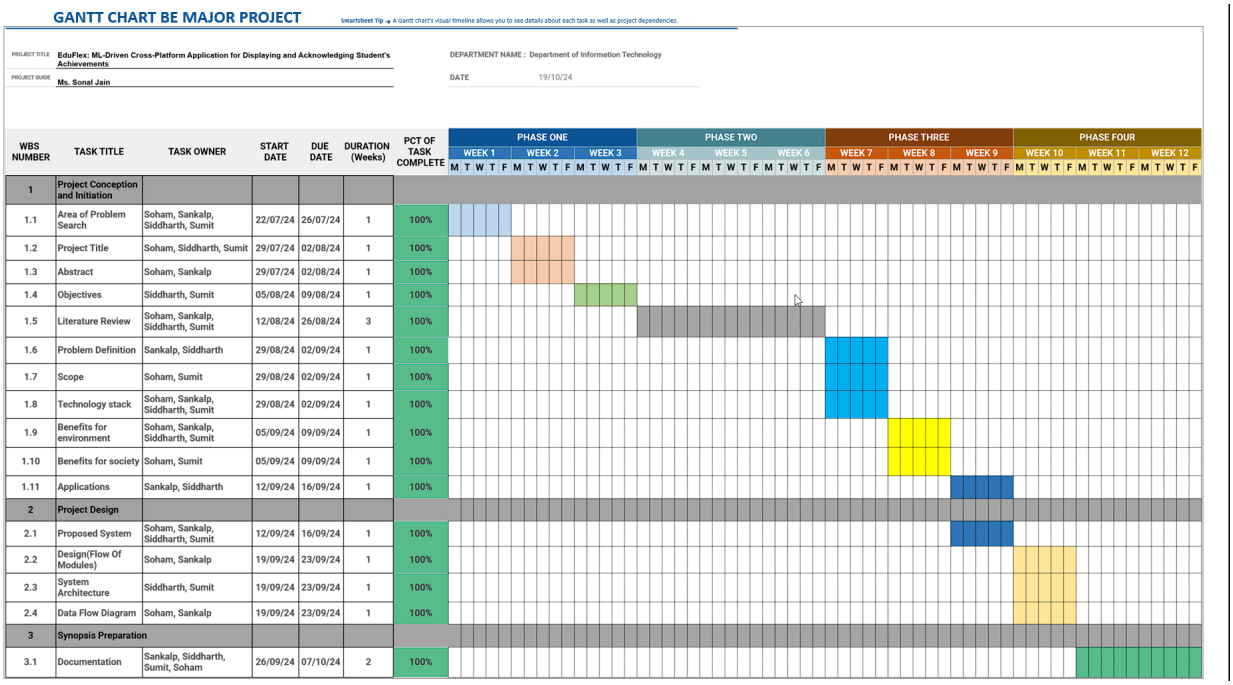


Figure 4.17: Management Portal

There is a Mentor Management Portal (Fig. 4.17) where mentors can review the certificates uploaded by students and have the ability to approve or reject them based on their authenticity. Additionally, mentors can assign points to students skills, allowing for a more personalized and credible evaluation of the students achievements.

4.3 Project Timeline

The graphical representation below outlines the key phases and deliverables of our project.



GANTT CHART BE MAJOR PROJECT

Smartsheet Tip → A Gantt chart's visual timeline allows you to see details about each task as well as project dependencies.

PROJECT TITLE: EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements

PROJECT GUIDE: Ms. Sonal Jain

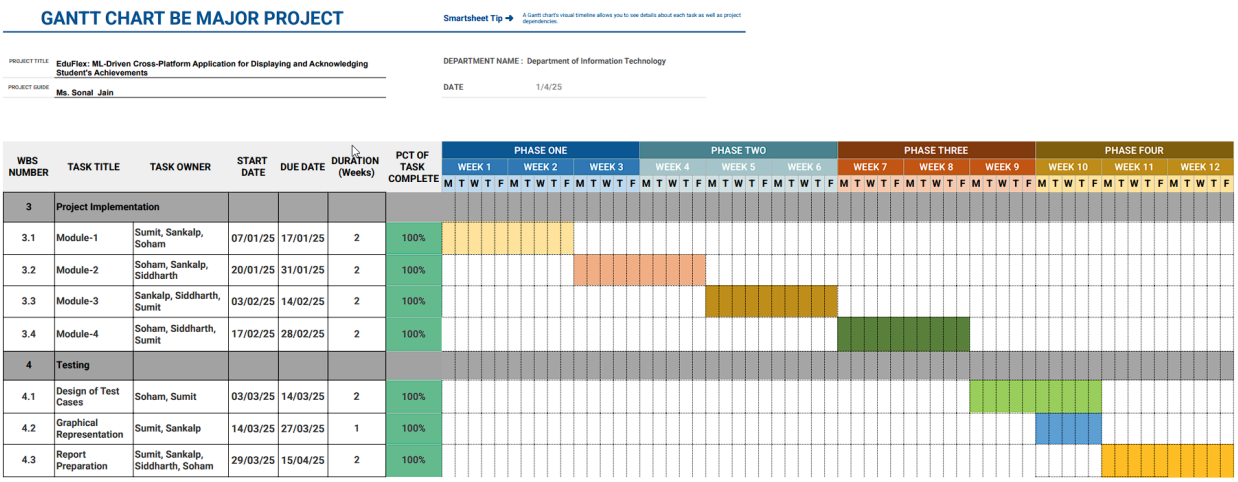
DEPARTMENT NAME : Department of Information Technology

DATE: 19/10/24

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION (Weeks)	PCT OF TASK COMPLETE
1	Project Conception and Initiation					
1.1	Area of Problem Search	Soham, Sankalp, Siddharth, Sumit	22/07/24	26/07/24	1	100%
1.2	Project Title	Soham, Siddharth, Sumit	29/07/24	02/08/24	1	100%
1.3	Abstract	Soham, Sankalp	29/07/24	02/08/24	1	100%
1.4	Objectives	Siddharth, Sumit	05/08/24	09/08/24	1	100%
1.5	Literature Review	Soham, Sankalp, Siddharth, Sumit	12/08/24	26/08/24	3	100%
1.6	Problem Definition	Sankalp, Siddharth	29/08/24	02/09/24	1	100%
1.7	Scope	Soham, Sumit	29/08/24	02/09/24	1	100%
1.8	Technology stack	Soham, Sankalp, Siddharth, Sumit	29/08/24	02/09/24	1	100%
1.9	Benefits for environment	Soham, Sankalp, Siddharth, Sumit	05/09/24	09/09/24	1	100%
1.10	Benefits for society	Soham, Sumit	05/09/24	09/09/24	1	100%
1.11	Applications	Sankalp, Siddharth	12/09/24	16/09/24	1	100%
2	Project Design					
2.1	Proposed System	Soham, Sankalp, Siddharth, Sumit	12/09/24	16/09/24	1	100%
2.2	Design(Flow Of Modules)	Soham, Sankalp	19/09/24	23/09/24	1	100%
2.3	System Architecture	Siddharth, Sumit	19/09/24	23/09/24	1	100%
2.4	Data Flow Diagram	Soham, Sankalp	19/09/24	23/09/24	1	100%
3	Synopsis Preparation					
3.1	Documentation	Sankalp, Siddharth, Sumit, Soham	26/09/24	07/10/24	2	100%

PHASE ONE: WEEK 1, WEEK 2, WEEK 3 | PHASE TWO: WEEK 4, WEEK 5, WEEK 6 | PHASE THREE: WEEK 7, WEEK 8, WEEK 9 | PHASE FOUR: WEEK 10, WEEK 11, WEEK 12

Figure 4.18: Timeline for Sem VII



GANTT CHART BE MAJOR PROJECT

Smartsheet Tip → A Gantt chart's visual timeline allows you to see details about each task as well as project dependencies.

PROJECT TITLE: EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements

PROJECT GUIDE: Ms. Sonal Jain

DEPARTMENT NAME : Department of Information Technology

DATE: 1/4/25

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION (Weeks)	PCT OF TASK COMPLETE
3	Project Implementation					
3.1	Module-1	Sumit, Sankalp, Soham	07/01/25	17/01/25	2	100%
3.2	Module-2	Soham, Sankalp, Siddharth	20/01/25	31/01/25	2	100%
3.3	Module-3	Sankalp, Siddharth, Sumit	03/02/25	14/02/25	2	100%
3.4	Module-4	Soham, Siddharth, Sumit	17/02/25	28/02/25	2	100%
4	Testing					
4.1	Design of Test Cases	Soham, Sumit	03/03/25	14/03/25	2	100%
4.2	Graphical Representation	Sumit, Sankalp	14/03/25	27/03/25	1	100%
4.3	Report Preparation	Sumit, Sankalp, Siddharth, Soham	29/03/25	15/04/25	2	100%

PHASE ONE: WEEK 1, WEEK 2, WEEK 3 | PHASE TWO: WEEK 4, WEEK 5, WEEK 6 | PHASE THREE: WEEK 7, WEEK 8, WEEK 9 | PHASE FOUR: WEEK 10, WEEK 11, WEEK 12

Figure 4.19: Timeline for Sem VIII

Chapter 5

Testing

5.1 Software Testing

Software testing is a critical phase in the software development lifecycle that ensures the accuracy, reliability, and robustness of a system before deployment. It helps identify defects and improve software quality. Testing can be broadly classified into manual and automated testing. Manual testing relies on human effort to verify system behavior, while automated testing uses tools to execute predefined test cases. In this project, testing was conducted manually, including feature testing, unit testing, and security testing, focusing on functional validation and performance evaluation.

5.1.1 Manual Testing

Table 5.1: Manual Testing

Test ID	**Module**	**Scenario**	**Steps**	**Expected Result**	**Status**
MT-01	Login/Sign Up	Login with valid credentials	Enter credentials and click submit	Redirected to dashboard	Pass
MT-02	Certificate Upload	Upload tampered certificate	Upload fake PDF	Marked as *Fake*, editor shown	Pass
MT-03	Announcements	Post and view new event	User posts event → All users see it	Announcement visible to all users	Pass
MT-04	Event Registration	Register for event	Open announcement → Register	Registration confirmation displayed	Pass
MT-05	Mentor Verification	Mentor verifies certificate	Mentor reviews and accepts/rejects	Certificate status updated	Pass
MT-06	Skill Ranking	Mentor ranks student skills	Mentor assigns skill scores	Rankings updated in leaderboard	Pass
MT-07	Batch Creation	Upload student file to create batch	Upload CSV or Excel with student list	Batch created and students grouped	Pass

5.1.2 Feature Testing

Table 5.2: Feature Testing

ID	Feature	Description	Status
FT-01	Certificate Verification	Detect real/fake using CNN-RF model	Pass
FT-02	Psychometric Testing	Generate MBTI questions via LLaMA	Pass
FT-03	Event Posting & Registration	Users can post announcements and register	Pass
FT-04	Filtered Leaderboard	Filter leaderboard by certificate types	Pass
FT-05	Mentor Certificate Review	Mentors verify submitted certificates	Pass
FT-06	Skill Ranking	Mentors rank students by skills	Pass
FT-07	Batch Creation	Mentors upload student files to create batches	Pass

5.1.3 Unit Testing

Table 5.3: Unit Testing

Component	Function	Input	Expected Output	Status
Certificate Validator	validate_certificate()	PDF file	"Real" / "Fake" + editor (if fake)	Pass
Announcement Module	create_announcement()	Title, desc, date	Event created and visible	Pass
Registration Module	register_user_for_event()	User ID, Event ID	Confirmation or error	Pass
Leaderboard Engine	filter_leaderboard_by_type()	Category string	Filtered leaderboard list	Pass
Mentor Verification	verify_certificate()	Certificate ID, status	Status updated	Pass
Skill Evaluator	rank_student_skills()	Map of scores	Scores stored	Pass
Batch Manager	upload_batch_file()	CSV/XLS file	Batch created	Pass

5.1.4 Security Testing

Table 5.4: Security Testing

ID	Scenario	Description	Expected Result	Status
ST-01	SQL Injection in Login	Attempt login using ‘’ OR ’1’=’1‘	Login should fail	Pass
ST-02	Unauthorized Certificate Verification	Non-mentor tries to verify certificate	Access denied	Pass
ST-03	Batch File Tampering	Upload malformed batch file	File rejected with error	Pass
ST-04	Sensitive Info Leak	View leaderboard data	Only public fields shown	Pass
ST-05	Repeated Skill Ranking	Rank same student repeatedly	Overwrite allowed or limited	Pass

Chapter 6

Result and Discussions

This chapter presents a thorough evaluation of the EduFlex project, discussing the outcomes of the investigation and highlighting the contributions made through its development. The discussion logically leads to inferences and identifies potential directions for future improvements and research.

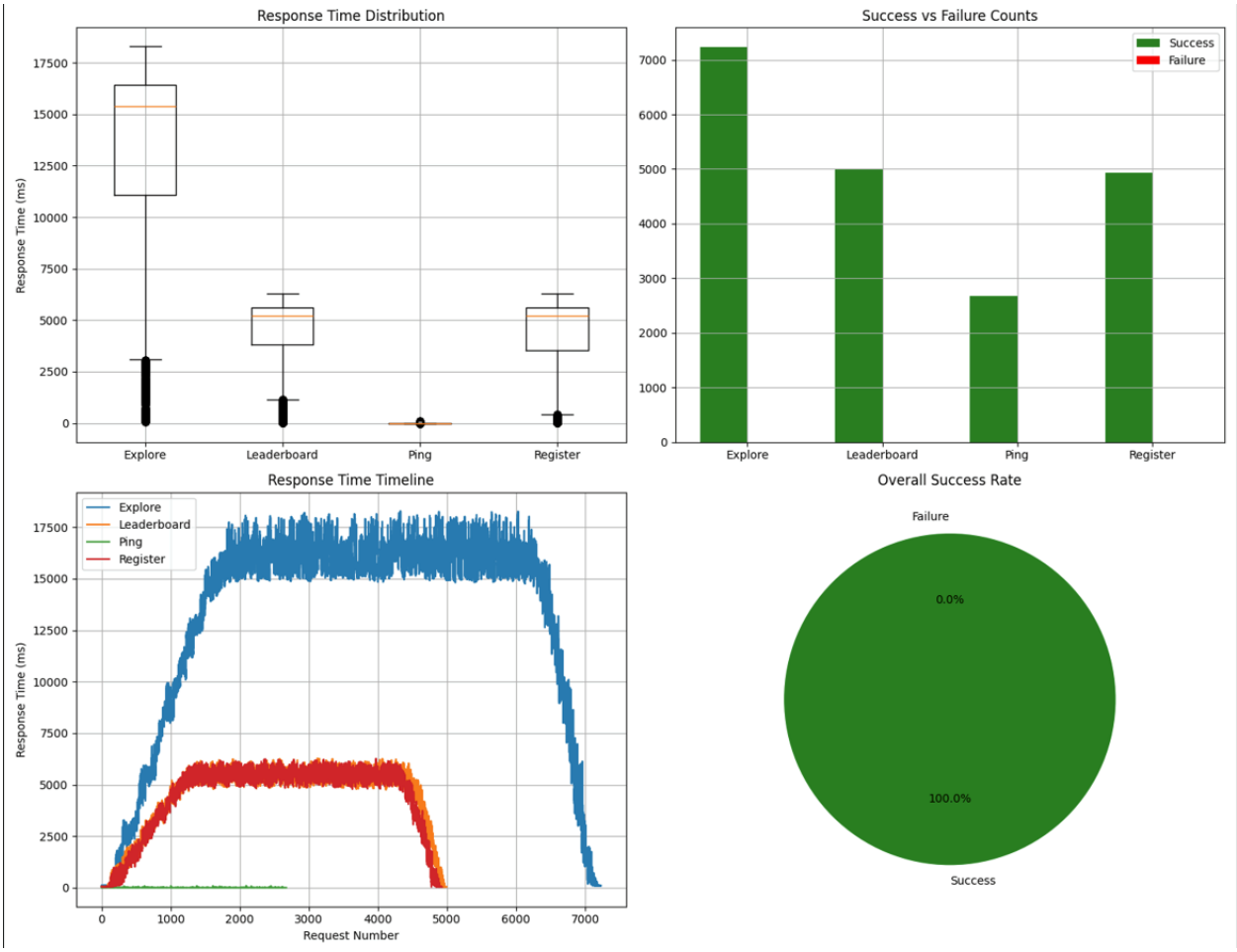


Figure 6.1: EduFlex API Performance Metrics: Response Time, Success Rates, and Load Trends

EduFlex effectively integrates academic, co-curricular, and extracurricular achievement tracking using AI, ML, and automation, offering a balanced and scalable system (Fig. 6.1) for evaluating student progress. The project has made several notable contributions, each of which adds significant value to the educational ecosystem.

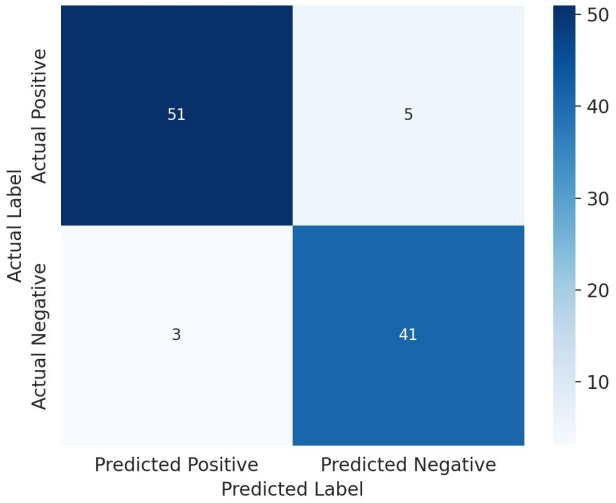


Figure 6.2: Confusion matrix

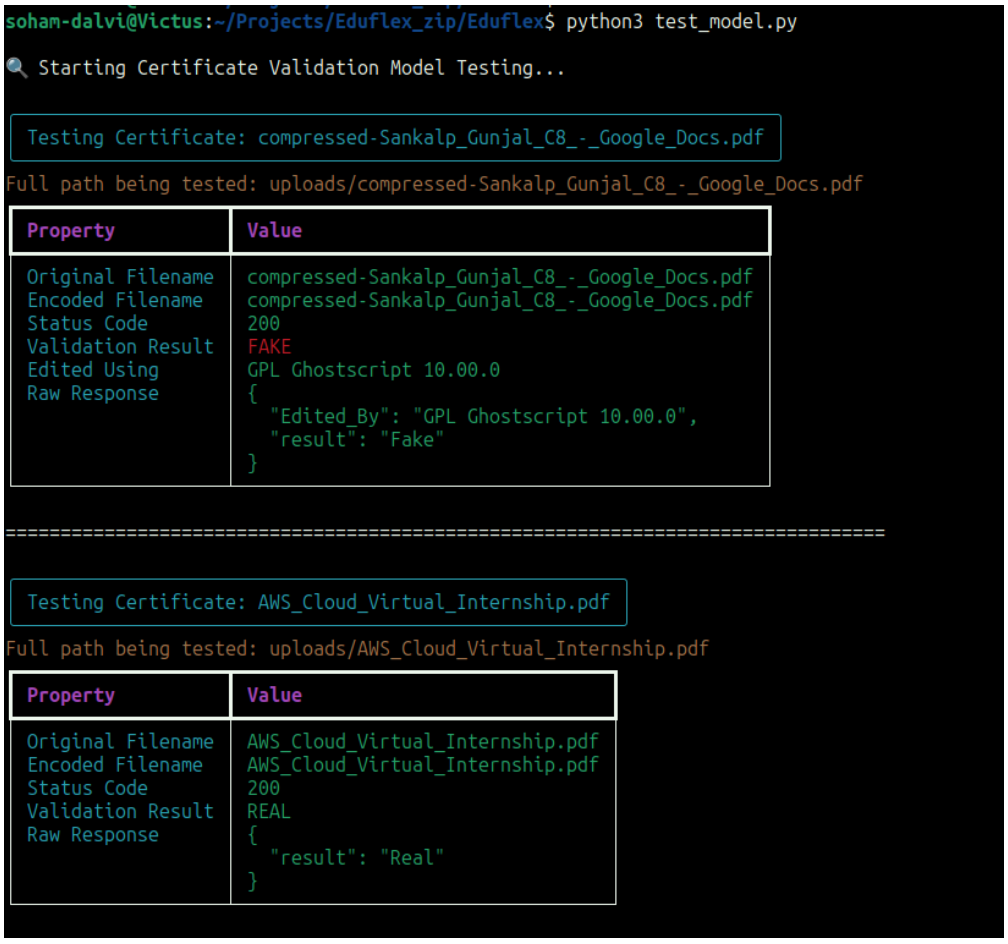


Figure 6.3: Model Prediction

A major contribution of the project is the certificate verification model, which employs Random Forest and Convolutional Neural Networks (CNN) to detect tampered certificates. Achieving an accuracy of 92% (Fig. 6.2), this model surpasses traditional methods of certificate verification. By analyzing features like font size, text positioning, metadata, and image count, EduFlex ensures the authenticity of student credentials, providing a powerful tool for preventing certificate fraud (Fig. 6.3) This model stands out as a significant advancement in applying machine learning to document verification.

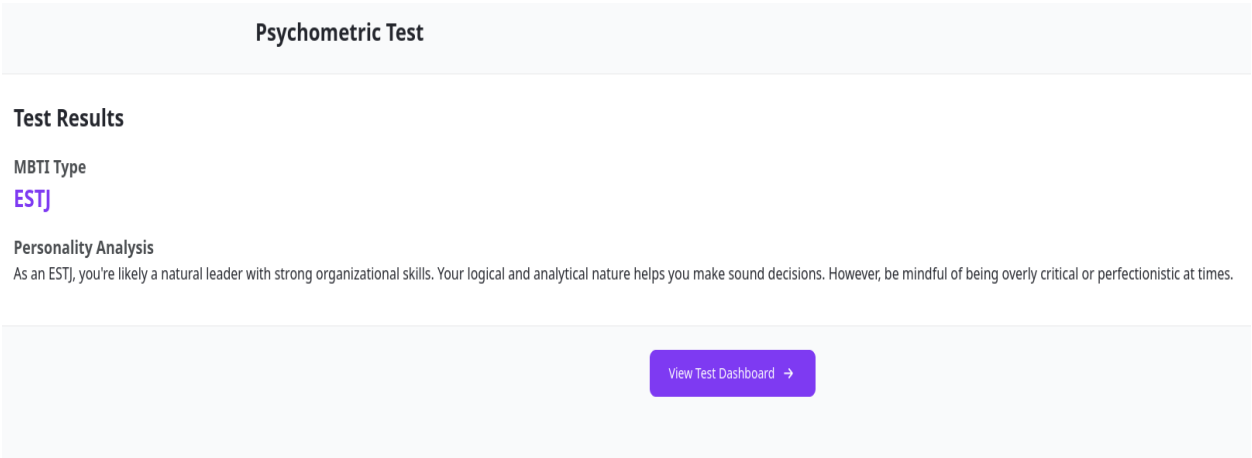


Figure 6.4: Psychometric Test Result

EduFlex offers a comprehensive evaluation framework enabling mentors to validate certificates, assign skill points, and monitor student progress through real-time dashboards. It includes a psychometric testing system powered by AI (Fig. 6.4), where questions dynamically adapt based on previous responses, offering deep insights into student aptitude and behavior. This integrated approach enhances transparency, facilitates informed decision-making, and fosters accurate, data-driven assessments that significantly improve learning outcomes.

To streamline profile management, the system automates Credly badge fetching, ensuring student achievements are always up-to-date with minimal manual effort. This not only enhances engagement by providing a clear view of accomplishments but also maintains a comprehensive academic and extracurricular record. Despite initial challenges like conflicting mentor approvals and high resource usage during large data processing, these were addressed with algorithm optimizations and fallback manual verification strategies.

EduFlex is built on a scalable tech stack comprising Node.js, Flask, and MongoDB, which ensures high performance and adaptability as the platform grows. Looking ahead, the focus will shift to expanding analytics capabilities through advanced visualizations and predictive insights, as well as launching a mobile application to offer flexible access for both students and mentors. With its continuous development, EduFlex stands poised to redefine how educational progress is tracked, assessed, and rewarded.

Chapter 7

Conclusion

EduFlex is developed as an AI-driven platform that redefines student assessment by integrating academic, co-curricular, and extracurricular achievements into a unified system. Traditional assessment methods often overlooked a student's holistic development, focusing primarily on academics. By leveraging machine learning, automation, and AI-driven insights, EduFlex achieved a more balanced and inclusive evaluation process, addressing the limitations of conventional grading systems.

One of the key innovations implemented in EduFlex is its certificate verification system, which utilized ML models to detect tampered certificates, ensuring the authenticity of student achievements. Additionally, real-time feedback, psychometric testing, and AI-powered assessments enabled students to recognize their strengths and areas for improvement. The integration of gamification, digital credentialing, and adaptive learning paths enhanced engagement, making learning a more interactive and rewarding experience.

EduFlex also empowered educators by providing data-driven insights into student performance, allowing them to offer personalized guidance and targeted interventions. The platform's scalable architecture, built on Node.js, Flask, and MongoDB, ensured seamless access across devices, making it a reliable tool for modern educational institutions. The inclusion of secondary academic and skill validation, following the model's initial assessment, further strengthened its credibility through an added layer of manual verification performed by mentors.

By bridging the gap between traditional assessments and modern educational needs, EduFlex set a new benchmark in student evaluation, engagement, and skill recognition, shaping the future of AI-driven education.

Chapter 8

Future Scope

The future development of EduFlex will focus on integrating advanced analytics tools to provide deeper insights into student performance trends. These enhancements will enable educators and institutions to make data-driven decisions, ultimately improving learning outcomes and offering personalized support tailored to individual student needs. Additionally, a mobile application is in development to ensure seamless access for both students and educators, allowing them to interact with the platform anytime, anywhere.

To further enhance engagement and motivation, a redeemable points system will be introduced, enabling students to exchange earned points for rewards. This gamification approach will foster a greater sense of achievement and active participation. Moreover, an automated resume-building service will be integrated, allowing EduFlex to extract academic and extracurricular records, helping students generate professional resumes effortlessly, streamlining career preparation.

Future enhancements will also include an AI-driven system for personalized test preparation and mental well-being assessment. Through psychometric testing, the system will evaluate a student's mental stability, providing insights that help identify areas of cognitive and emotional development. Additionally, AI-powered assessments will assist students in recognizing and improving their academic weak areas. Furthermore, AI-based personalized course generation will be introduced, tailoring learning paths based on each student's needs to enhance their learning experience. These innovations aim to create a dynamic, engaging, and personalized learning ecosystem, leveraging technology to support both academic growth and mental well-being.

Bibliography

[1] Ivica Pesovski, Ricardo Santos, Roberto Henriques, and Vladimir Trajkovik. Generative AI for customizable learning experiences. *Sustainability*, 16(7):3034, 2024. doi:10.3390/su16073034.

[2] Dunusinghe A. V., Ranasinghe T. K. S. A., Gamage J. G. A. C. H., Perera K. G. D. T., Samantha Thelijjagoda, and Poojani Gunatilake. AI-powered smart and personalized education platform. In *Proceedings of the 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, pages 1–7, 2023. doi:10.1109/ICSES60034.2023.10465439.

[3] Elena Wolz, Matthias Gottlieb, and Hans Pongratz. Digital credentials in higher education institutions: A literature review. In *Innovation Through Information Systems: Volume III: A Collection of Latest Research on Management Issues*, pages 125–140. Springer, 2021. doi:10.1007/978-3-030-86800-0_9.

[4] Adebowale Owoseni, Oluwaseun Kolade, and Abiodun Egbetokun. Enhancing personalised learning and student engagement using generative AI. In *Generative AI in Higher Education: Innovation Strategies for Teaching and Learning*, pages 123–150. Springer Nature Switzerland, Cham, 2024. doi:10.1007/978-3-031-60179-8_5.

[5] Rui Huang, Albert D. Ritzhaupt, Max Sommer, Jiawen Zhu, Anita Stephen, Natercia Valle, John Hampton, and Jingwei Li. The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 68:1875–1901, 2020. doi:10.1007/s11423-020-09807-z.

[6] Shamsudeen Mohammed, Lois Nwobodo, and Njoku Ekene. Certificate fraud verification model using clustered-based classification approach. 2024.

[7] Alejandra Castelblanco, Jesus Solano, Christian Lopez, Esteban Rivera, Lizzy Tengana, and Martín Ochoa. Machine learning techniques for identity document verification in uncontrolled environments: A case study. In Karina Mariela Figueroa Mora, Juan Anzurez Marín, Jaime Cerda, Jesús Ariel Carrasco-Ochoa, José Francisco Martínez-Trinidad, and José Arturo Olvera-López, editors, *Pattern Recognition*, pages 271–281. Springer International Publishing, 2020. doi:10.1007/978-3-030-49076-$8_2$6.

[8] V. R. Sadasivam, V. Sanjaykumar, A. Sheik Imam, and R. Vasanthan. Guided student profile system: Mentor feedback integration. In *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, pages 1–5, 2024. doi:10.1109/ISCS61804.2024.10581291.

[9] Moritz Breit, Martin Brunner, Antoine Fischbach, Rachel Wollschläger, Ulrich Keller, Sonja Ugen, and Franzis Preckel. Academic profile development: An investigation of differentiation processes based on students' level of achievement and grade level. *Journal of Educational Psychology*, 2024. doi:10.1037/edu0000856.

[10] Manyata Yadav. A study of the effects of an individual's personality and characteristics on job behavior using the Myers-Briggs type indicator. *International Journal of Innovative Research in Engineering and Management*, 10:49–52, June 2023. doi:10.55524/ijirem.2023.10.3.9.

Appendices

Appendix-A: Eduflex Download and Installation

1. Download Eduflex from git :

   ```
   git clone https://github.com/Soham01011/Eduflex.git
   ```

2. Make sure that NodeJS is downloaded, is above version 16.

3. Let's install the packages and modules which are required. Thus run the following commands for it

   ```
   cd backend
   npm install
   pip install -r requirements.txt
   ```

4. Run Llama3.1:8b model locally :

   ```
   ollama pull llama3.1:8b
   ```

5. Set up .env file in backend folder:

   ```
   USE_PYTHON_SERVER=""
   USE_CREDLY_BADGES=""
   USE_VALIDATE_CERT=""
   USE_ADD_MENTES=""
   USE_LLAMA_MODEL=""
   SERVER_SEC_KEY=""
   MONGODB_CONN_URI=""
   ```

6. Run the project by running the following commands in the backend folder :

   ```
   node server.js
   python3 automation_server.py
   ```

Publication

Paper entitled **"EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements"** is presented at **"ICMRACC-2025"** by **"Sankalp Gunjal, Soham Dalvi, Siddharth Devare, Sumit Mesta, Sonal jain, Kiran Deshpande".**

Here is the attached research paper:

EduFlex: ML-Driven Cross-Platform Application for Displaying and Acknowledging Student's Achievements

Sankalp Gunjal
Department of Information Technology
A. P. Shah Institute of Technology
Thane, India
sankalpgunjal286@apsit.edu.in

Soham Dalvi
Department of Information Technology
A. P. Shah Institute of Technology
Thane, India
sohamdalvi219@apsit.edu.in

Siddharth Devare
Department of Information Technology
A. P. Shah Institute of Technology
Thane, India
siddharthdevare324@apsit.edu.in

Sumit Mesta
Department of Information Technology
A. P. Shah Institute of Technology
Thane, India
sumitmesta273@apsit.edu.in

Sonal Jain
Department of Information Technology
A. P. Shah Institute of Technology
Thane, India
sajain@apsit.edu.in

Kiran Deshpande
Department of Information Technology
A. P. Shah Institute of Technology
Thane, India
kbdeshpande@apsit.edu.in

***Abstract*— Students often struggle to reflect their overall growth, as conventional evaluation methods emphasize academic scores while neglecting co-curricular achievements and soft skills. To address this limitation, we propose a comprehensive platform that dynamically generates student profiles by aggregating diverse accomplishments. The platform supports both manual certificate uploads and automated badge fetching via Credly. To ensure authenticity, certificates are validated using Convolutional Neural Networks (CNNs) and Random Forest models, with hyperparameter tuning employed to verify their legitimacy. For personality analysis, the system incorporates the LLaMA model, utilizing an MBTI (Myers-Briggs Type Indicator)-based psychometric assessment, which provides deeper insight into student personas. A trending dashboard, through graphical representation, visualizes weighted achievements, identifies areas of academic concern, and tracks individual progress over time. The certificate validation module achieved an accuracy of 92 percent, effectively reducing the overall manual effort by 60 percent. This platform promotes a holistic, verifiable, and inclusive approach to student evaluation.**

Keywords*— *Machine Learning and Artificial Intelligence, Rectified Linear Unit (ReLu), Convolutional Neural Network (CNN), LLaMA Model, Random Forest Algorithm

I. INTRODUCTION

Conventional methods of assessment in educational institutions focus mainly on examination marks and strict grading systems, ignoring the value of the attainment of soft skills, co-curricular and extracurricular activities. Such a narrow strategy does not reflect a student's overall growth and hence, there is a lack of balance in measuring their actual capacity and accomplishments. In today's fast-changing world, students are getting involved in all kinds of learning experiences. They're not just hitting the books; they're taking internships, joining online courses, and competing in various contests. Plus, they're also part of extracurricular activities like sports and cultural events, which finds no mention in traditional academic records. We urgently need a new system based on smart analytics and digital portfolios that fills this awareness gap.

In spite of all the advancements and innovations in educational technologies, the majority of the evaluation systems in institutions remain fragmented and linear in their assessment models. **This study proposes to fill the critical void by investigating if a single, smart platform can comprehensively assess the performance of the students by bringing together various academic and non-academic accomplishments while also providing secure verification and interpretable outcomes of learning. EduFlex is proposed as an Artificial Intelligence (AI) and Machine Learning (ML) driven system that brings together academic marks, behavioural characteristics, and extracurricular accomplishments to aggregate them in the form of a verifiable student profile in pursuit of providing a comprehensive and future-oriented alternative to the conventional evaluation systems**.

Psychometric assessment is carried out by using the Myers-Briggs Type Indicator (MBTI) through the Ollama Large Language Model Meta AI (LLaMA) 3.1:8b model in analysing personality dimensions, which adds to the in-depth analysis of individual students' personality profiles.

After collecting data on the students through assisted uploads and automated badge retrieval, the system makes use of machine learning models to validate certificates based on visual and metadata analysis through the application of hyperparameter tuning. These data points are integrated into individualized analysis and visualized through graphical dashboards that enable both the students' and their educators' identification of learning trends, strengths, and weaknesses.

The platform employs models like Convolutional Neural Networks (CNNs) and Random Forest classifiers to verify certificates efficiently and accurately. Powered by a strong stack of Node.js, Flask, and MongoDB, EduFlex is engineered to be responsive, scalable, and user-friendly.

Using features like gamification for tracking achievements and live trend visuals on a leaderboard makes it easier for everyone to get involved and stay engaged. EduFlex makes it really simple to assess students. It uses smart analytics to provide deeper insights while also keeping the user experience in mind. It looks at student performance in a clear, complete, and flexible way.

II. Literature Review

The application of blockchain in academic credentialing is picking up as it promises security, decentralization, and verifiability. Ambast and Sumesh [1] constructed a system based on blockchain that allows decentralized verification of academic credentials in PDF, DOCX, JPG, or PNG format. They ensure tamper-proof storage by anchoring credential hashes on-chain while storing document content off-chain, striking a balance between confidentiality and transparency. Supporting this, Ziyi Li et al. [5] examined how educational applications of blockchain experience slow global uptake. Through a layered analysis involving Australian tertiary systems, they screened for technical, policy, and adoption hurdles, and proposed five key qualities for scalable credentialing infrastructures. Taking this domain further, Rashmi et al. [9] proposed an NLP-driven deep-learning approach to validation of certificates. Their IDLSP pipeline, which combined feature extraction and Multi-Layer Perceptron models, recorded 92% accuracy, evidencing the viability of automating academic certification trust. Concurrently, researchers also turned their attention to improving support of, and for, students using AI and NLP-based systems**. Assayed et al. [2] used a domain-tuned BERT model to build a question-answering system for high school-level academic advising. Fine-tuned on a curated educational corpus, their model reported strong ROUGE scores and alleviated computational load using PyTorch-based optimization.** Correspondingly, Sadasivam et al. [7] suggested the Guided Student Profile System (GSPS) that utilizes machine learning in providing real-time academic counseling. GSPS provides role-based access to institutional stakeholders in tracing performance, allocating resources, and mentoring of students. These studies reflect AI's increasing presence in adaptive educational support although differing in scale of deployment and scheme of end-user interaction. Student personality and psychological profiling were also examined using AI**. Liu et al. [3] previously used pre-trained LLMs to substitute psychological assessment questionnaires using the 16PF approach. Through their model, there was enhanced dynamic and contextualized interaction over traditional methods and showed the capability of LLMs in subtle assessments.** Yadav [10] also centered on personality assessment using the Myers-Briggs Type Indicator (MBTI), examining workplace behavior influenced by it. Although qualitative in nature, the study affirms the applicability of psychometric profile assessment to educational and career planning, albeit emphasizing empirical validation. Gamification was also found to be beneficial in enhancing engagement and motivation in students. **Jun and Lucas [4] conducted a meta-analysis of the components of gamification—leaderboards, badges, and point systems—finding their strong effects on self-regulation and academic performance in various educational settings. They were reinforced by the study of Gregg et al. [8], which examined perceptions of digital badges among students in micro-credential schemes. Employing the Situational Motivation Scale (SIMS), it was learnt that intrinsic and extrinsic motivators both drive engagement, especially when badges are unified in platforms such as LinkedIn, for greater professional exposure. For better predictive modeling of outcomes and protection of anonymity, Farooq et al. [6] proposed a federated learning (FL) approach. Through grading of pupils' grades and optimization across schools, FL—in tandem with SVM—was found to yield the best performance in terms of F1-score and accuracy. Balancing data utility and anonymity, this approach promises a direction for multi-institutional analysis in provision of educational outcomes.**

III. Methodology

The platform has two primary user roles: Student and Mentor. Mentors can create groups by adding student's and actively monitor their progress.

Student Side: The platform creates a profile based on academic, co-curricular, and extracurricular accomplishments. Points are given for accomplishments, leading to a ranking leaderboard of student performance. **Academic certificates are categorized as Entry/Foundation (5 points), Intermediate (7 points), and Global/Industry-Specific (10 points). Experience certificates such as Internships, Jobs, and Entrepreneurship are given 6, 8, and 10 points, and extracurricular activities such as Sports, Hackathons, and Seminars are given 5 to 10 points. This fixed scoring system leads to fair assessment and can be edited based on future needs of institutions.**

Student's register by providing their basic data and filling in their profile. The dashboard shows them the distribution of points. The Explore page lists the latest posts by the student's, enabling them to post their achievement in PDF format. Students can also add skills, experience, educational attainment, and projects. An ML model keeps checking the post of the student's for tampered or edited certificates. Also, a mentor is allocated a batch of students to verify the achievement of the student's and give permission to upload the post.

Mentors can also monitor and check the performance of their student's. Students can link their Credly profile, and the system automatically fetches and reflects their Credly badges on their profile. Students can search for other student's profile either by search or by looking at the leaderboard. Students can also view trending profile by going through the uploaded achievement. Students can appear for a personality test to assess their soft skills and interview readiness. The AI will generate 20 questions pursuant to previous responses, providing individualized feedback upon completion. Tests can be repeated, adjusting to prior performance to keep improving.

Mentor Side: Upon logging in, mentors fill out their profile to form and organize groups of mentees. They upload lists of students in PDF or Excel format, enabling the system to automatically generate groups. They are also able to view profiles of the students and access the leaderboard to monitor progress.

The platform includes a Mentor-driven Academic along with Skills Validation which is a management interface where mentors validate student credentials and check their skills. It offers real-time performance view of the students through interactive graphs and charts that reflect test performance, skill builds, and activity timelines. Each pupil is assigned to a particular mentor in order to provide a defined structure. Although full-scale

pilot testing is in the offing, the system was tested in-house by using simulated scenarios for functionality and usability.

A. Certificate Verification Model

Verifying digital certificates is difficult in the presence of minor tampering. It adopts a Machine Learning (ML) model, which was trained on their own curated dataset, to verify authenticity. Students upload certificates that are processed by the model to extract important features such as font size, font color, font style, text coordinates, image count, and PDF producer metadata. If the model predicts a low probability score, the certificate is tampered and a mentor verification is requested for final validation.

The dataset includes 64,450 records of imitation records and 59,541 records of original records, each consisting of 7 parameters. These certificates are of various academic and certification organizations. Actual certificates were ethically obtained after obtaining permission from the students, while fake certificates were manually fabricated to mimic tampering using Adobe Acrobat and Microsoft Word. Certificate tampering manipulated fields like names, grades, and institute details without touching any of the actual fields allowing similar structure resembling actual certificates.

It compares the uploaded certificate using the model's feature combination approach to patterns learned during the training stage. It examines each parameter - such as structure, contents, and metadata - for any tampering sign using its learned dataset. It then calculates a prediction percentage based on the analysis, determining certificates less than 50% as fake. The model's feature extraction mechanism is text-independent in nature and rather emphasizes structure and image metadata features, hence it can be made flexible for non-English certificates and different educational systems. It can be fine-tuned using new sets of data for better effectiveness on new types of certificates.

For better detection of forged certificates, a hybridized model consisting of Convolutional Neural Network (CNN) and Random Forest (RF) was used. The main incentive for this design was to make use of the non-linear feature expression capability of CNN together with the decision stability and interpretability of Random Forest. Generally, the overall model structure is designed to deal with the structured metadata that is derived from PDF certificates and also to learn complex patterns that separate authentic certificates from fabricated ones efficiently.

1) Role of CNN as a Non-Linear Feature Transformer

In this devised system, CNN is applied not in the traditional capacity of raw image processing but as a non-linear transformation of features that is learned over structured metadata features derived from PDF certificates. Some of these features are font style, font size, font color, the coordinate positions of text components, the number of embedded images, and PDF producer data. These features, though structured, can bear complex dependencies and fine-grained differences between true and forged certificates that are usually hard to model using standard machine learning models.

The CNN architecture in this approach is developed to support multiple convolution layers that can learn hierarchical representations and interactions between these input features. These transformed features resulting from the convolution layers are fed to fully connected dense layers, which yield a refined and improved feature space. Such output, normally derived by flattening the feature maps of the last dense layer, captures subtle patterns and non-linear associations, providing a greater discriminative representation of the certificates.

We implement a Convolutional Neural Network (CNN), in which the input layer is composed of seven nodes that extract features of PDFs. Data is processed by several layers, followed by 64 nodes in the next, then a dropout to 32 nodes to avoid overfitting. It then ends in an output layer of one node classifying the certificate by providing a prediction, as depicted in "Fig. 1".

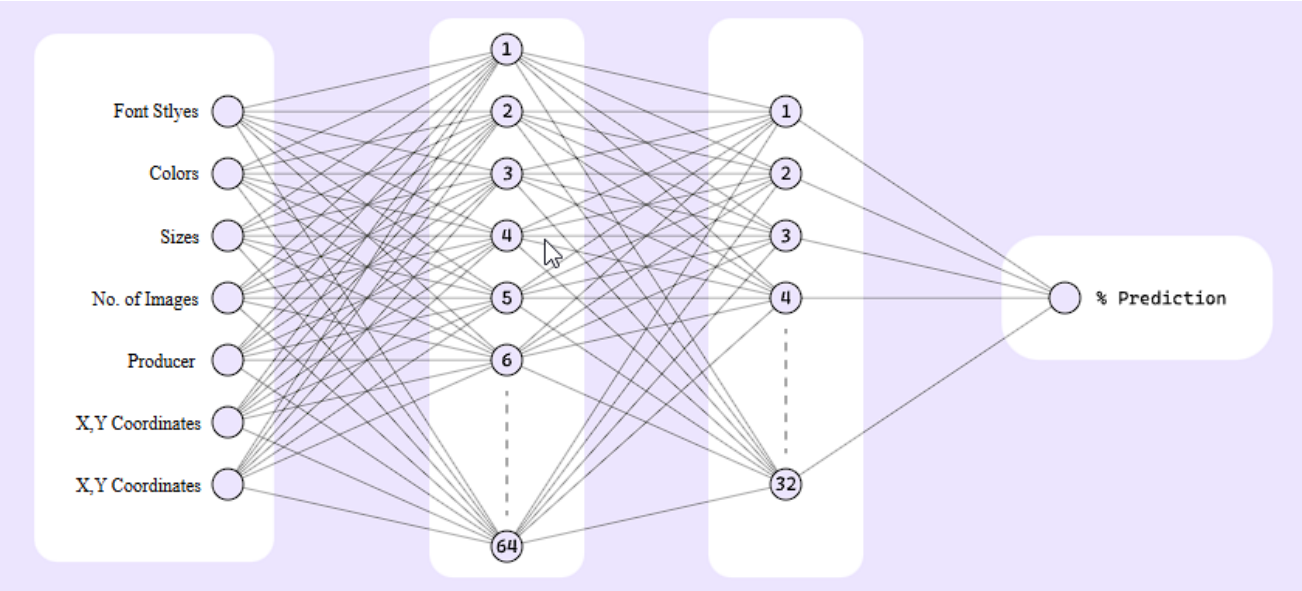


Fig. 1. Convolutional Neural Network (CNN) Architecture

A dropout layer was added in the suggested CNN structure to avoid overfitting and increase model generalization. Training accuracy was 95% and validation accuracy was 92% when using a 0.3 dropout rate. It was also a very high value in comparison to the previous situation without dropout, where the training accuracy was well over 98% and the validation accuracy was almost 85% due to overfitting. Prediction time was increased without any alteration in performance by the dropout layer.

2) Random Forest for Final Classification

The feature vector that is transformed by the CNN is also fed to a Random Forest classifier, the last decision-making module of the hybrid model. Random Forest, consisting of many decision trees, is well-established in processing structured tabular data. In this paper, the Random Forest was set using 100 estimators, Gini impurity as the splitting value, and free choice of tree depth in order to identify any fine distinctions between the transformed features.

The use of Random Forest in this hybrid architecture is especially important in addressing noisy or outlier data, typical in real-world certificates databases. Additionally, Random Forest's ensemble characteristics also serve to discourage overfitting, which means that the model will generalize well beyond what it was exposed to. It also makes the model predictions easier to interpret, a critical aspect of working with sensitive data like that of certificates' authenticity.

Example: IF Font Size > 12 AND Font Color = 0.02 THEN Possibly Tampered ELSE IF PDF Producer = "Adobe" AND No. of Photos = 0 THEN Not Tampered.

$$P(\text{Tampered}) = \frac{\text{No. of Trees that Predict "Tampered"}}{\text{Total No. of Trees}} \quad (1)$$

$$P(\text{Tampered}) = \frac{6}{7} \approx 0.857\ (85.7\ \%)$$

3) ReLU Activation Function in Feature Extraction

The ReLU activation function is utilized in the CNN layers to induce non-linear interaction among the model's features, which can be learned by the model more efficiently. Through the utilization of ReLU, only the portions of activation that are deemed significant or relevant are passed on to the network layers. Along with enhancing computational efficiency, it also inhibits the vanishing gradients issue, enabling faster and stable training of the CNN module. ReLU is mathematically given as:

$$f(x) = \max(0, x) \quad (2)$$

It outputs the input if positive and zero otherwise, enhancing computational efficiency and learning speed. Its key roles in our model include:

- **Introduces Non-Linearity:** Enables CNN to model complicated document feature relationships.
- **Avoiding Vanishing Gradients:** Keeps gradients strong to ensure stable and faster training.
- **Efficient Feature Extraction:** Removes unwanted activations, supporting correct classification.
- **Sparse Activation & Efficiency:** Outputs zero for negative inputs, avoiding superfluous calculations.

4) CNN-Random Forest Pipeline

The first step in the proposed pipeline is the extraction of pertinent metadata features from the given PDF certificates by the pdfminer module. The pdfminer module efficiently examines through the PDFs as it picks out important structural features like font style, font size, font color codes, text location coordinates (x0, y0, x1, y1), number of images included in the document, and PDF producer details. These extracted features constitute the structured tabular data that is used as the input for the machine learning pipeline that follows. In contrast to regular image-based CNN applications, in this scenario, the CNN is used to deal with these structurally tabular features, and the model can learn complex dependencies and relationships that are perhaps beyond observation using regular methods.

Once the features are extracted via pdfminer, they are fed into the Convolutional Neural Network (CNN), which is a non-linear feature transformation block. The CNN layers extract subtle interactions among various features and map the input to a higher-dimensional feature space. This high-feature vector, after flattening the output of the CNN, is fed as input to the Random Forest (RF) classifier. Random Forest model, made up of ensemble decision trees, processes this transformed data further to conduct the end-classification of certificates as Real or Fake. CNN allows the model to learn non-linear patterns in structured data, while Random Forest provides robustness, improved decision boundary handling, and resistance to noisy or outlier data. The hybrid model efficiently marries the deep-learning strengths of CNN for feature transformation and ensemble learning strengths of Random Forest for stable and successful classification in the certificate verification problem.

5) Credly badges fetching

This module provides a built-in feature in the platform that automatically fetches badges from a user's Credly profile. It does away with the effort that the student's need to enter badge credentials manually. Through application of web scraping automation, the system gets badge information directly from the profile URL that the user inputs during profile setup. After initializing, the platform regularly fetches and syncs the badge information, which keeps the badge records of the users up to date in their profile without them needing to make any manual effort.

6) Fetching Process

In support of this automation, a separate Flask server runs independently of the application server, hosting a standalone /fetchbadges endpoint. It accepts the user's Credly profile link as argument. When the request is received, the backend web scraper starts the extraction, which focuses on publicly available badge data like badge title, issuing organization, and issue date. It then maps the pulled data directly to the relevant user entry in the database, supporting real-time update.

Implementation of fetching badge data from Credly was done using web scraping methods as there was no public API to fetch user-specific badge data. Extraction of badge details through scraping is done in a strict mode of extracting only public badge details from the Credly profile of a student if the student has made badges public. As the fetched data is public, there are no issues of rate limiting or breach of privacy. Proper error-handling mechanisms in the scraping module ensure that any failure, such as changes in the page structure or network errors, does not disrupt the overall verification flow. If it fails, the badge field is set as null without disturbing the core flow of the verification of certificates. Further, the functionality is made optional and configurable, enabling system administrators to switch it on or off as per organizational requirements without disturbing the core activities of the platform.

Additionally, ethical concerns over web scraping methods were thoroughly considered. The scraper functions strictly in the domain of publically published information volunteered by users on their Credly pages. No sensitive or private data beyond this domain is scraped or stored. We've set up our request rates to help stop any bad scraping activities, which keeps the Credly system running smoothly and well-balanced. Still, in the interest of long-term compliance and sustainability, the platform was designed in a modular fashion — making it easy to switch to a properly licensed API-based approach, if in the future Credly were to supply authorized API access.

B. Psychometric Test

The psychometric testing capability built into the platform utilizes the LLaMA (Large Language Model Meta AI) that performs dynamic MBTI (Myers-Briggs Type Indicator) personality testing. The test is designed to cater to the academic and social life of engineering students, analysing parameters such as learning style, stress management, type of decision-making, and style of interaction. The model takes the form of adaptive testing where each subsequent question depends upon the response of the student in the previous one, enabling a more precise and individualized personality assessment. It allows mentors to better understand the students and give them direction based upon their strengths and weaknesses.

For data protection and confidentiality, all psychometric assessments are taken in full within the local infrastructure of the institution. The LLaMA model is solely generating the test questions and not accessing or storing any sensitive or identifiable data of the students. All the data of the students, including their answers and results of the test, is stored in the database of the institution in a secure manner. No data is transmitted to any external server, ensuring confidentiality in full. The result of the test is made available to the individual student and their mentor-only, ensuring ethical use of AI and keeping the end assessment and feedback under the purview of humans.

IV. SYSTEM ARCHITECTURE

"Fig. 2", the system architecture diagram, gives a comprehensive overview of the system components, the interactions between them, and the flow of data, and how each of the various components interacts together to enrich the experience of the user.

EduFlex platform architecture involves various components providing a secure and flawless experience for users. User Access & Authentication allows students and educators to log in together using web browsers, and users gain access to their learning dashboard and analytics. Frontend involves a profile management dashboard that is personalized, enabling student's to monitor achievements, certificates, and learning. Leaderboards encourage engagement by ranking student's on the basis of participation, test scores, and accomplishments. An exclusive test page provides AI-driven assessment that conducts psychometric tests, and a feature of certification upload enables student's to upload certificates, which get verified using machine learning.

The Credly is also integrated on the platform, fetching badges in real-time to ensure legitimacy. The database keeps track of student information, certificate details, and test scores in an structured way. It features a Student Database for storing profile data, a Test Database for keeping a tab on AI-generated test results, and a Certificate Database storing authenticated

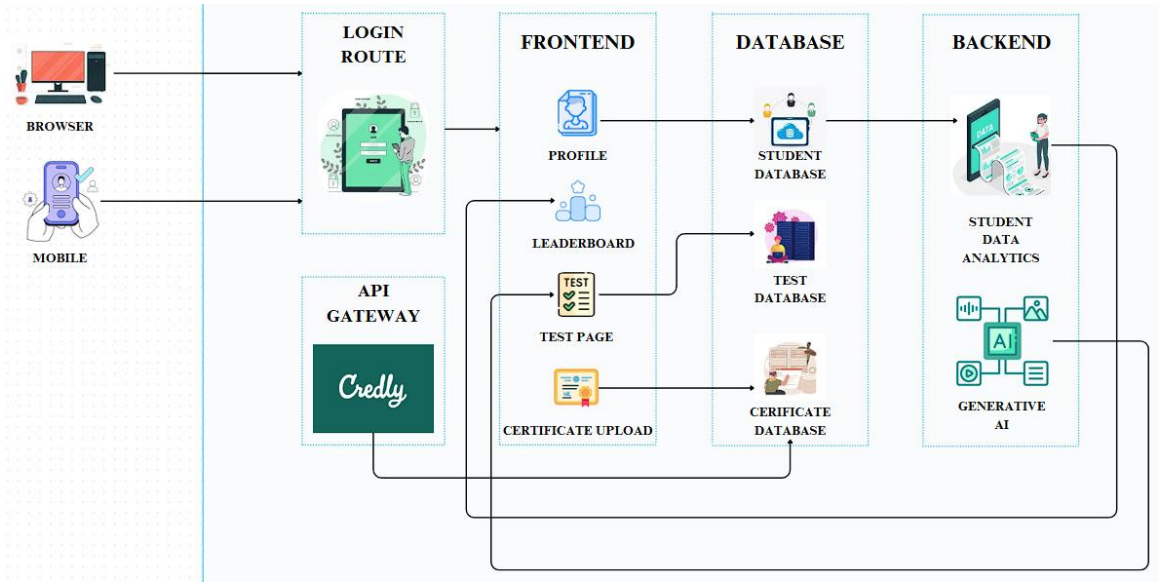


Fig. 2. System Architecture Diagram

certificates & badges. The Backend includes ML-driven verification. In addition, ML-based verification of certificates ensures that certificates uploaded are authentic, checking for fraudulent submissions by using machine learning models to cross-authenticate certificates prior to inclusion in student profiles. Complete architecture modernizes the assessment of students by using AI, ML-based automation, and secure validation, providing a comprehensive and well-rounded educational experience.

V. RESULTS AND FINDINGS

EduFlex is meant to give a smart and scalable infrastructure for validating the accomplishments of students in academics, co-curricular, and extracurricular activities.

Massive load testing thus asserted the scalability of EduFlex. It processed a maximum of 6000 requests "Fig. 3" in stress testing, which had response times of 17 seconds in the worst-case scenario. But considering the application is aimed in college environments - where there is a capped load of approximately 1000 students - the system remains well under performance thresholds. Further, normal deployment methods such as load balancing and horizontal scaling can disperse traffic efficiently in the event of anticipated greater usage too.

For further improving model performance, various architectural experiments were carried out, mainly to handle overfitting. One of the notable enhancements was the inclusion of a dropout in the CNN model. Various dropout levels between 0.2 and 0.5 were experimentally varied, of which 0.3 provided the optimum balance. Without dropout, overfitting occurred - the model achieved over 98% on the train set but ~85% on unseen samples. After using dropout, the learning of the model stabilized to 95% in the training set, whereas the validation improved to 92% "Table I", which demonstrates the model's capacity to learn better without affecting computational efficiency.

Table I. Confusion Matrix

	Predicted Genuine	**Predicted Fake**
Actual Genuine	51(True Negative)	5(False Positive)
Actual Fake	3(False Negative)	41(True Positive)

Regarding real-time performance, the model's prediction time averages between 800–900 milliseconds per certificate. This speed is reasonable given the extensive feature checks performed during classification.

By using AI, ML, and automation, the system enhances verification efficiency and accuracy in certificates. Its core detection model, which makes use of a CNN and a Random Forest classifier, was able to achieve a level of 92% accuracy on the in-house training data in identifying tampered certificates, yet still had an average of 85% accuracy when tested on unseen, real-world samples of certificates. The provided dataset was ethically acquired, consisting of both authenticated and fake student certificates that were made using different tools that replicated real-world environments of font modifications, text movement, metadata manipulation, and object insertions.

The platform of EduFlex makes use of aspects of gamification, such as leaderboards, that have proven to increase student engagement, self-improvement, and healthy competition. An internet connection for the entire duration of the process is needed to use the platform,

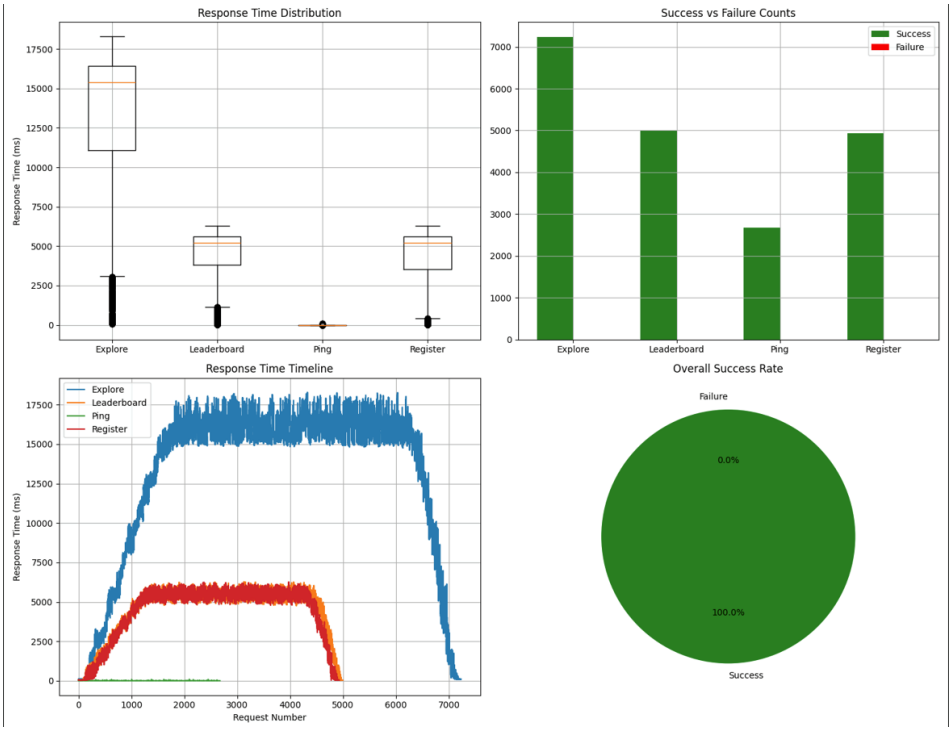


Fig. 3. Stress Testing Results

which is a strong limitation. However, the user interface provides for individuals without much technical knowledge. The technical stack for the platform makes the system scalable and efficient. Feedback by the users suggests satisfaction with the functionality and experience of the platform.

VI. FUTURE SCOPE

EduFlex is looking ahead to adding some cool new analytics features. These will help us really dive into how students are doing and support personalized learning for everyone. **Also, there is planned a mobile application using Flutter for cross-platform support, allowing the features of upload of certificates, verification status, and performance tracking to be accessed by students and instructors on-the-go. Its current backend architecture, developed using the modular REST APIs, makes the incorporation of the mobile aspect technically viable by making fewer changes, allowing it to scale and provide a flawless experience to the end user.**

In addition to that, a redeemable points mechanism will be introduced where the points earned by student's can be redeemed for rewards to ensure active participation. **An automated resume builder will also be included in EduFlex. Student's will be able to select one of the pre-designed resume template options, and according to the choice made by them, the system will automatically retrieve their academic records, certificates, and extracurricular activities from their profile and build a professional resume in downloadable formats such as PDF or DOCX. It'll make it quicker and easier to create resumes that fit your needs, helping you get ready for your career.**

Future developments will also encompass an Artificial Intelligence-driven system of customized test preparation and mental well-being assessment. Based on psychometric assessment, the system will assess a student's mental stability in terms of their cognitive and emotional development, providing them with insights that enable identification of where there is room for improvement.

Further, assessments using Artificial Intelligence will help the student identify and develop their academic weaknesses. We're working to create a lively and personalized learning space that uses technology to support both academic success and mental well-being.

VII. CONCLUSION

EduFlex offers an ML-based platform to reimagine student evaluation by merging academic, co-curricular, and extracurricular accomplishments in a comprehensive and smart assessment system. The application employs machine learning for verification of certificates using both actual and synthetically created data in tandem to identify tampering and verify the authenticity of student records. Focused on deployment and flexibility, EduFlex meets the needs of institutions and engages users in a game-based leaderboard and reward system while providing automated resume-building capabilities that compile professional resumes based on verified student records. **Through the application of AI models, automation, and real-time analysis, EduFlex adds a new dimension to valid, complete, and data-driven learning assessment - establishing foundations for future advancements in AI-based learning environments.**

REFERENCES

[1] Ambast, S.K., Sumesh, T.A.: A blockchain based credential verification system using ipfs. In: 2022 IEEE 19th India Council International Conference (INDI-CON), pp. 1–5 (2022). https://doi.org/10.1109/INDICON56171.2022.10039743

[2] Assayed, S.K., Alkhatib, M., Shaalan, K.: A transformer-based generative ai model in education: Fine-tuning bert for domain-specific in student advising. In: Basiouni, A., Frasson, C. (eds.) Breaking Barriers with Generative Intelligence. Using GI to Improve Human Education and Well-Being, pp. 165–174. Springer, Cham (2024)

[3] Liu, Z., Kang, Y., Li, X.: Research on psychological test based on large language model. In: 2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), pp. 503–510 (2024). https://doi.org/10.1109/RAIIC61787.2024.10670817

[4] Jun, M., & Lucas, T. (2024). Gamification elements and their impacts on education: A review. Multidisciplinary Reviews, 8(5), 2025155. https://doi.org/10.31893/multirev.2025155

[5] Ziyi Li, Z., Joseph, K.L., Yu, J., Gasevic, D.: Blockchain-based solutions for education credentialing system: Comparison and implications for future development. In: 2022 IEEE International Conference on Blockchain (Blockchain), pp. 79–86 (2022). https://doi.org/10.1109/Blockchain55522.2022.00021

[6] Farooq, U., Naseem, S., Mahmood, T. et al. Transforming educational insights: strategic integration of federated learning for enhanced prediction of student learning outcomes. J Supercomput 80, 16334–16367 (2024). https://doi.org/10.1007/s11227-024-06087-9

[7] Sadasivam, V.R., Sanjaykumar, V., Sheik Imam, A., Vasanthan, R.: Guided student profile system: Mentor feedback integration. In: 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS), pp. 1–5 (2024). https://doi.org/10.1109/ISCS61804.2024.10581291

[8] Gregg, A., Park, J., Fenton, C., Lang, D., Handley, M.: Exploring the "why" of micro-credentials and digital badges: Engineering students' motivations for and perceived utility of learning outside of class. In: 2022 IEEE Frontiers in Education Conference (FIE), pp. 1–7 (2022). https://doi.org/10.1109/FIE56618.2022.9962376

[9] Rashmi, M.J., Manoj Kumar, N.V., Suguna, K.: Experimental evaluation of student certificate validation system using improved deep learning strategy with prediction principles. In: 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), pp. 470–476 (2024). https://doi.org/10.1109/ICIPCN63822.2024.00083

[10] Yadav, M.: A study of the effects of an individual's personality and characteristics on job behavior using the myers-briggs type indicator. International Journal of Innovative Research in Engineering and Management 10, 49–52 (2023) https://doi.org/10.55524/ijirem.2023.10.3.9