

# SMART INDIA HACKATHON 2024

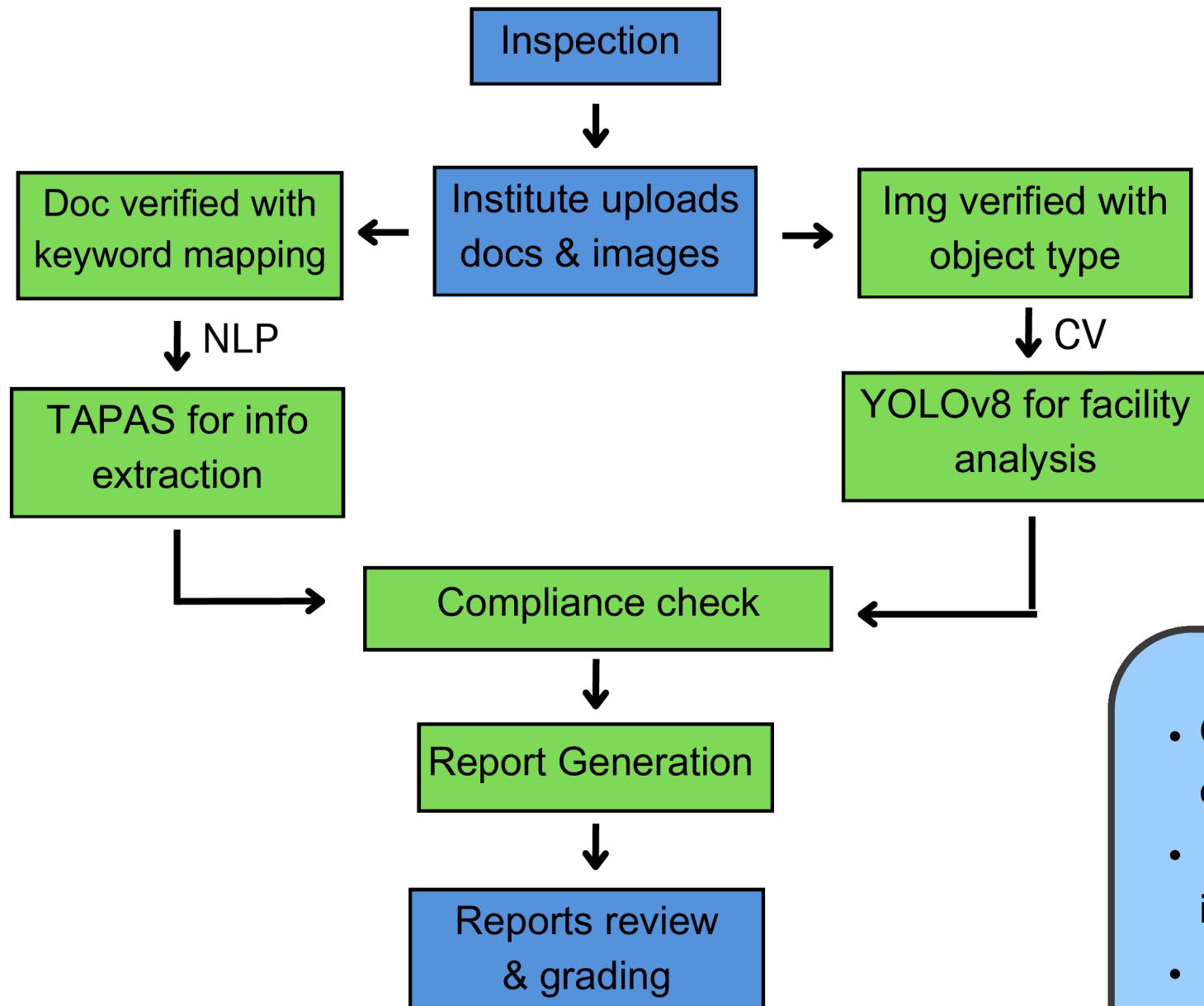


- **Problem Statement ID** - 1730
- **Problem Statement Title** - AI Driven Inspection of Institutes
- **Theme** - Smart Automation
- **PS Category** - Software
- **Team ID** - 2718
- **Team Name** - Neural\_Nexus

| Team Members               |                  |
|----------------------------|------------------|
| Sara Nambiar (Team Leader) | Isha Shah        |
| Nikita Bhedasgaonkar       | Param Joshi      |
| Atharva Kulkarni           | Shree Mengshetti |

# Nirikshan SahAq

## WORKFLOW:



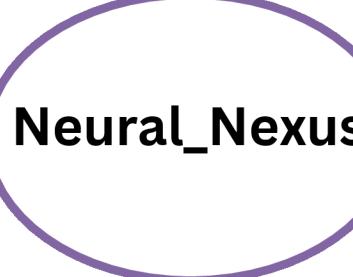
  : automated simplification

## PROPOSED SOLUTION:

1. **Infrastructure analyzed** with object detection models and trained CNNs in **real - time** to provide information on upkeep of institute's infrastructure quality
2. Query based tabular data extraction models + transformers for **analyzing authenticated documents** to verify compliance with standard requirements/norms
3. **Integration** with institutional database to cross-check data with existing records
4. ML algorithms implemented to **analyze trends** in current + historical data, to address potential flags
5. **User-friendly dashboard** with **customizable inspection criteria** to provide **flexibility**
6. Automatically **generated comprehensive reports** include compliance status, faculty analysis, improvement suggestions for any deficiencies found and more

### Innovativeness

- **Counterfeit Prevention with Transparency:** Automated checks on logos, formats, and data ensure authenticity; Uploaded image + CCTV feeds verify **consistency** with uploaded data
- **Inspector Input and Customization:** Inspectors can add **comments**, **view** documents and images, **upload** reports and provided with **discussion forum** amongst inspectors
- **Rubrics for Unbiased Ratings:** A rubric system ensures fair, **consistent** ratings based on **input guidelines by AICTE officials**
- **Reduced Document Workload:** Facility images **replace** certificates, **directly verifying** compliance and reducing manual effort



# Technical Approach



## Technology Used

### Image Recognition for Facility Analysis:

YOLOv8 used to assess infrastructure by object detection

### NLP for Document Analysis:

TAPAS for query based tabular extraction to verify compliance with AICTE norms, and pdfplumber for document verification

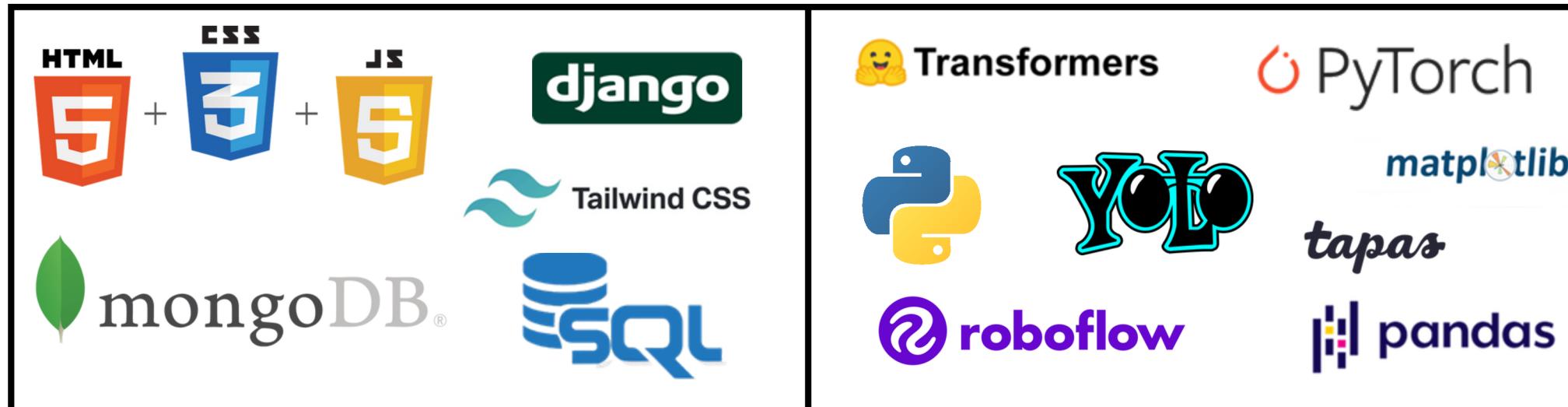
### Pattern Recognition:

Analyzed placement trends using **Random Forest models** and predict institutional compliance using **XGBoost** based on inspection scores

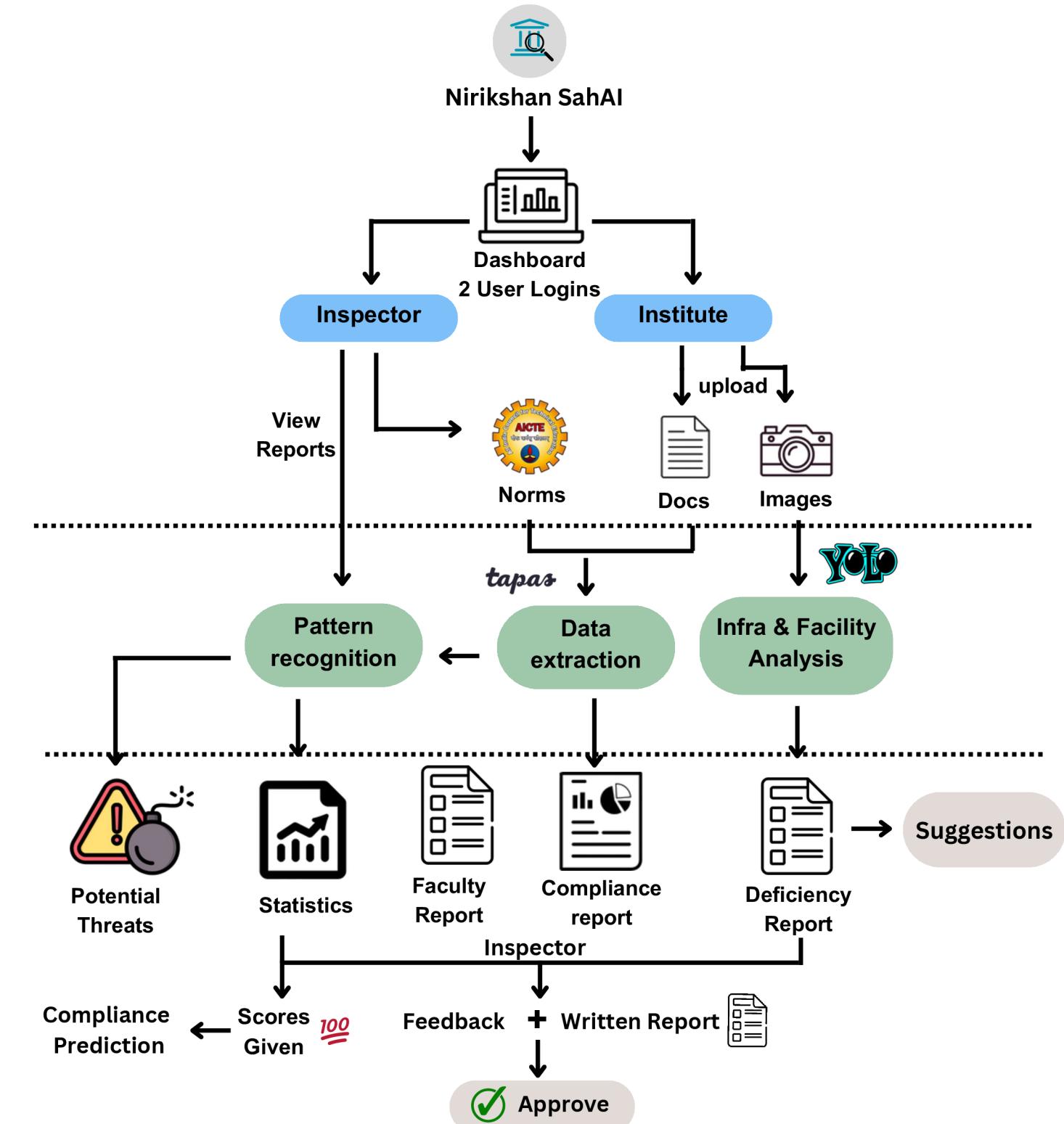
### Frontend & Backend Development:

Built an interface using **HTML, CSS, JS** and backend powered by **Django, PostgreSQL** for structured data, and **MongoDB** for unstructured data storage

## Technology Stack



## Program Architecture



# Feasibility and Viability

| Potential Risks                     | Strategies  |
|-------------------------------------|---|
| Input Data Manipulation by Colleges | Geotagged images, data verification using <b>keyword mapping &amp; anomaly detection</b>                              |
| Resistance to rely on AI            | Controlled <b>access</b> to docs & images with added options for <b>feedback + comments &amp; report modification</b> |
| Data Privacy and Security Concerns  | Incorporation of <b>AES + Cloud encryption</b> solutions & <b>Role Based Access Control</b>                           |
| Fear of biased ratings              | Inclusion of <b>master table</b> for admin; access to institute for <b>verification</b> of grading criteria           |

## Feasibility

- Scalability :** Cloud services, containerization, and microservices enable seamless scaling
- Accuracy:** ML-Ops ensures models are regularly updated for high precision
- Proven AI:** Utilizes YOLOv8 and TAPAS for reliable object detection and document analysis
- User Adoption:** AI-driven workflow reduces manual effort and speeds up inspections
- Data Security:** Implementing robust security measures ensures sensitive institutional data is protected, complying with data protection laws
- Operational Scalability:** Institutions are already familiar with digital submissions, and integrating AI for inspection processes fits into their existing workflows

## Viability

- Scalable cloud infrastructure:** Optimized one time setup cost for infrastructure and models
- Lower Variable Costs:** Automation reduces manual work, decreasing long-term inspection costs
- Data Driven Insights:** Enhanced operational efficiency with quality improvement efforts
- Scalability for National Standards:** Inclusion of standardized, nation-wide inspection framework

# Benefits and Impact



vs



**Assessment is thorough & consistent, reducing human error**



Infrastructural analysis using CV **reduces the counterfeit actions**



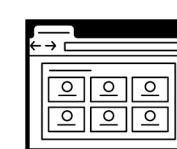
**Reduces** the time required for inspection drastically



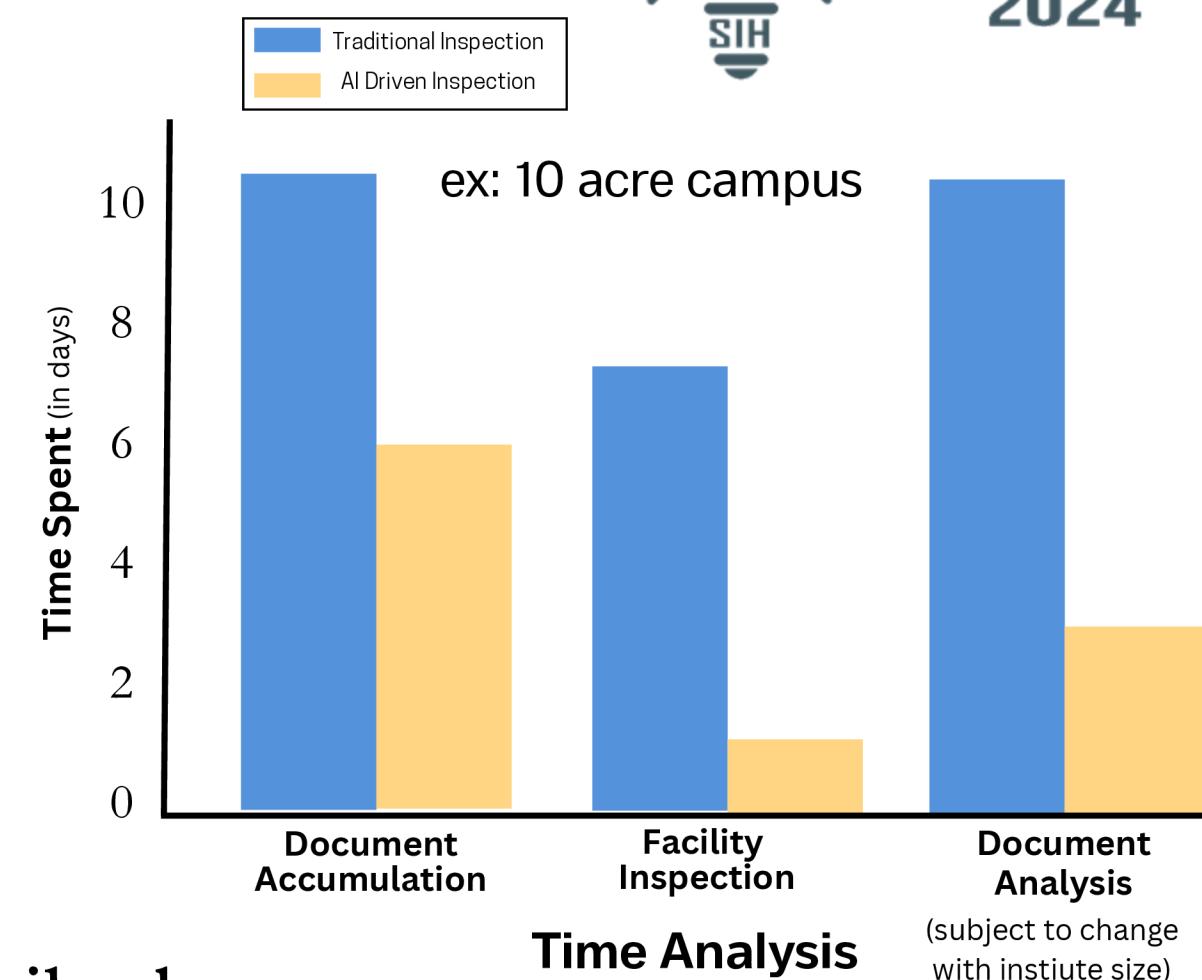
Analyzing **placement data** helps assess whether the **curriculum is up-to-date** and relevant



Detailed reports and feedback, helps institutes **address deficiencies**



Intuitive interface makes model **easy to use**



## Scalability [more details - here](#)

- **Data security** through encryption, data masking, role-based access control (RBAC) and regular audits
- **Scale PostgreSQL** for **structured** data using sharding, read replicas, and partitioning to improve query performance, and use **MongoDB** for **unstructured** data for horizontal scaling
- **Comprehensive Master Table of rubrics** for Admin; Access to college authorities as well to get verification of grading criteria

- Deployment on **cloud platforms** like AWS, Azure, would leverage managed services for databases, storage, AI/ML models, and computing resources
- Leveraging a **SaaS model** to offer flexible, cost-effective, and scalable solutions to institutions and regulatory bodies
- **API Support** : Can integrate with other educational management systems, enhancing functionality and enabling broader adoption

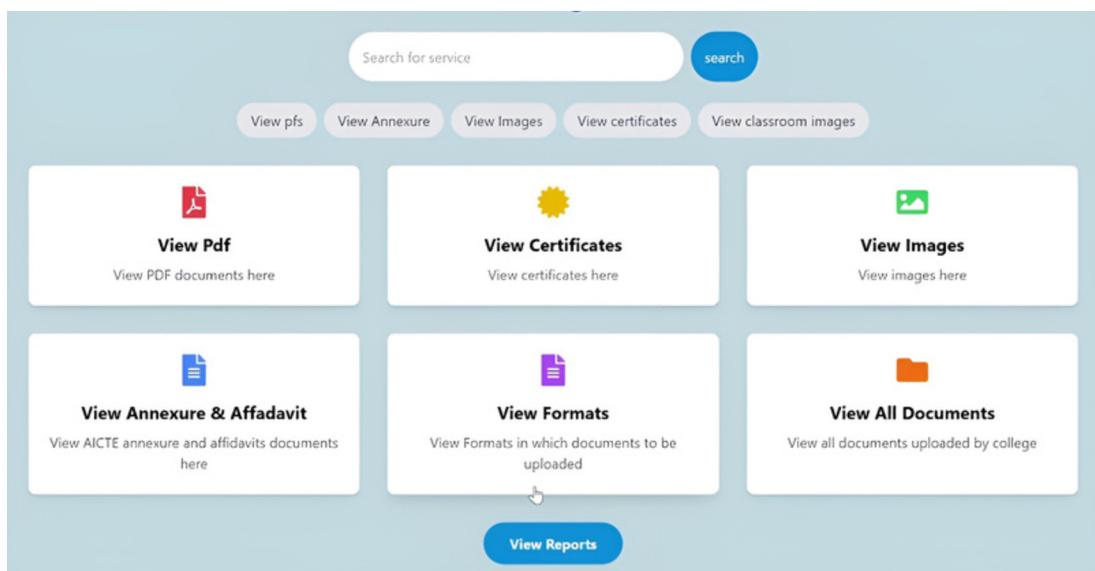


## Documentations:

- [PyMuPDF \(Fitz\) for PDF Text Extraction](#)
- [Natural Language Processing \(NLP\) - spaCy](#)
- [Roboflow for dataset management](#)
- [Data Extraction - Camelot](#)
- <https://www.nltk.org/>

## Research Papers:

- [https://arxiv.org/pdf/cs/0205028](https://arxiv.org/pdf/cs/0205028.pdf)
- [https://arxiv.org/pdf/1506.02640](https://arxiv.org/pdf/1506.02640.pdf)
- <https://aclanthology.org/2020.acl-main.398.pdf>
- [https://www.researchgate.net/publication/377135877 XGBoost and Random Forest Algorithms Analysis in Depth Analysis](https://www.researchgate.net/publication/377135877)



## Prototype Video



## Github



## Generated Reports

