

# CVlytics – AI-Based Resume Analyzer and ATS Scoring System

**\*\*Project Report\*\***

**\*\*Group Members:\*\*** Muhammad Khuzaima Siddiqui (29), Muhammad Rafay Anwar (17), Saeed (11), Muneeb (19)

**\*\*Program:\*\*** Bachelor of Science in Artificial Intelligence

**\*\*Tools & Technologies:\*\*** Python, Django, Scikit-learn, XGBoost, Pandas, NumPy, NLP, Joblib

[Learn More](#)

[View Demo](#)



Made with **GAMMA**

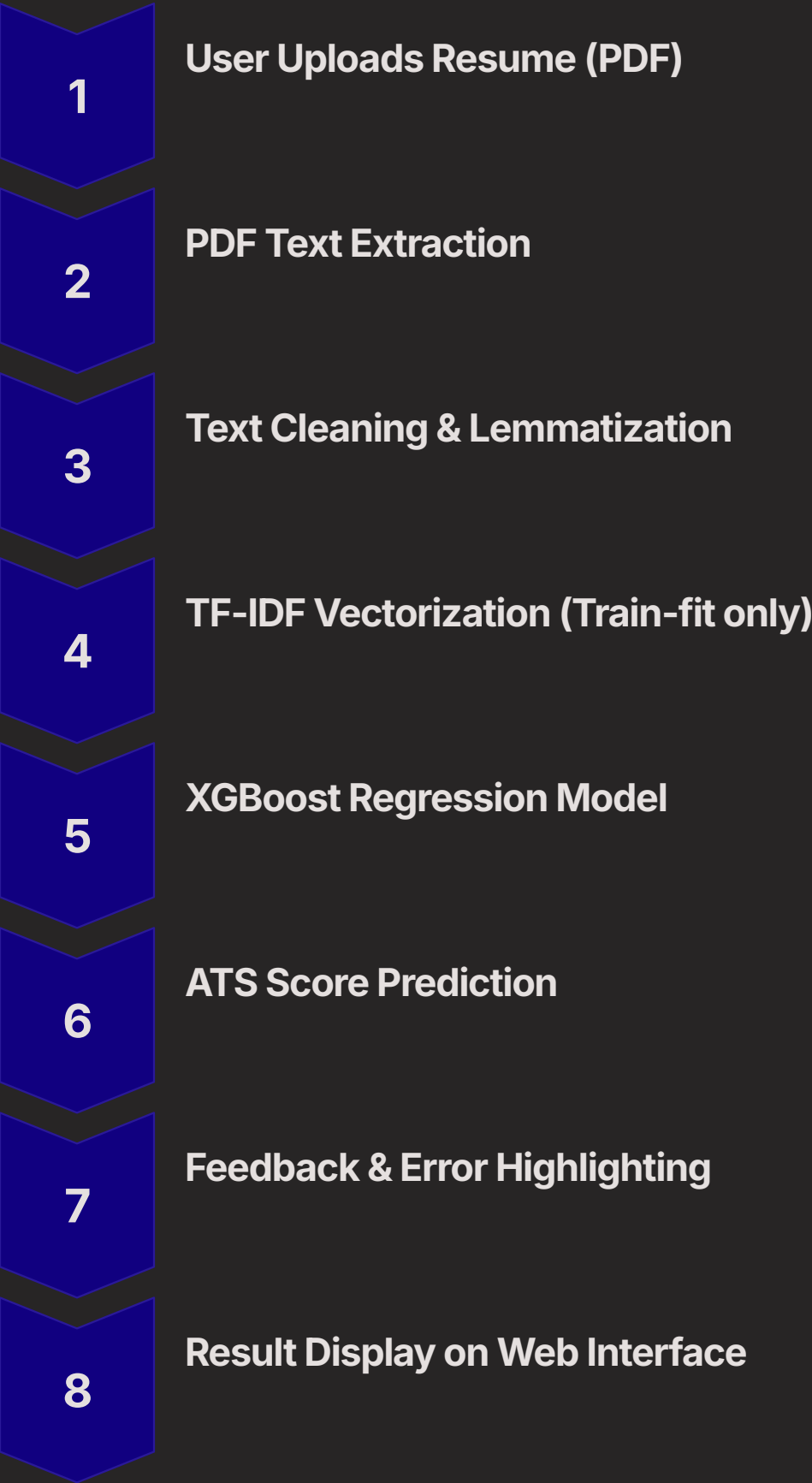
# 1. Project Overview

CVlytics is an AI-powered resume analysis system that predicts the ATS score of a resume using Natural Language Processing and Machine Learning. Users upload a CV in PDF format, which is automatically processed and evaluated to generate a numerical ATS score along with improvement suggestions.

The objective of this system is to provide an automated, fair, and scalable resume evaluation solution for both job seekers and recruiters.

## 2. System Architecture and Pipeline

### Architecture Flow



### System Components

Layer	Technology
Frontend	HTML, CSS, JavaScript
Backend	Django
ML Model	XGBoost Regressor
Vectorization	TF-IDF
Database	Supabase / SQLite
File Handling	PyMuPDF / pdfminer

# 3. Integrated Lab Concepts

## Text Preprocessing

Lowercasing, regex cleaning, lemmatization

## Feature Engineering

TF-IDF with bigrams

## Regression Modeling

ATS score prediction using XGBoost

## Hyperparameter Tuning

GridSearchCV

## Cross Validation

3-Fold CV

## Model Evaluation

MAE,  $R^2$  Score

## Model Serialization

joblib

## Deployment

Django model serving



## 4. Dataset Description and Preprocessing

### Dataset

- CSV File containing resume text and ATS scores
- Columns: text, ats\_score
- Size: 500+ resumes

### Cleaning Process

- Removed rows containing job descriptions instead of resumes
- Converted all text to lowercase
- Removed special characters using regex
- Lemmatized each token using WordNet
- No stopword removal to preserve ATS-relevant keywords

# 5. Model Design and Training

## Feature Extraction

Parameter	Value
Max Features	3000
N-Gram Range	(1,2)
Min DF	3
Max DF	0.9

## Model: XGBoost Regressor

Hyperparameter	Value
n_estimators	200, 300
max_depth	4, 6
learning_rate	0.05, 0.1
subsample	0.8
colsample_bytree	0.8

## Training Setup

- Train-Test Split: 80% / 20%
- GridSearchCV with 3-fold cross-validation
- Objective: reg:squarederror

## Evaluation Metrics

Metric	Result
Mean Absolute Error (MAE)	~5.2
R <sup>2</sup> Score	~0.87



## 6. Challenges, Limitations and Ethics

### Challenges

- ATS scores were manually estimated
- Resume formatting noise
- PDF parsing inconsistencies

### Limitations

- Model only predicts generic ATS score
- No job-specific resume optimization

### Ethics & Fairness

- No demographic features used
- Scoring purely based on resume content
- No personal data stored after processing

# 7. Setup and Execution Instructions

## Environment Setup

```
conda create -n cvlytics python=3.10
conda activate cvlytics
pip install -r requirements.txt
```

## Run Django Server

```
python manage.py makemigrations
python manage.py migrate
python manage.py runserver
```

## Hardware Requirements

Component	Specification
RAM	8 GB
CPU	Intel i5 or equivalent
OS	Windows / Linux



## 8. Sample API Command



**POST /analyze\_cv/**



**Input: resume.pdf**



**Output: Predicted ATS Score  
+ Resume Feedback**

## 9. Conclusion

CVlytics is a complete AI-driven ATS scoring platform that uses TF-IDF and XGBoost regression to predict resume quality. The system demonstrates how real-world machine learning models can be trained, evaluated, deployed, and served within a full-stack Django application.