📘EduPredict - Developer Documentation

Last Updated: Sept 8, 2025

Developed by: Elisha Noel, Muhammad Shees, Utban Anwar, Vinesh Kumar

📌 Table of Contents

🔍 Project Overview

EduPredict is a predictive analytics platform that utilizes machine learning to forecast student outcomes such as **dropout, enrolled, or graduate** status. It also predicts future grades and detects anomalies in student data.

🏗️ System Architecture

[Frontend (Streamlit UI)]
↓
[App Logic (eduPredict.py)]
↓
[Model Layer (.pkl models)]
↓
[Data Layer (edu_cleaned.csv)]

User Roles:

- Student
- Teacher
- Counselor

Each role has role-based access to features and visualizations.

🧰 Tech Stack

Component	Tools Used
Language	Python 3.10+
ML Libraries	scikit-learn, XGBoost
Web App	Streamlit
Visuals	Plotly, Matplotlib, Seaborn
Auth System	Streamlit session_state
Report Export	Pandas, HTML2PDF (custom)

📁 Folder Structure

```
EduPredict/
│
├── Documentation/            → User & Developer manuals
├── Feedback&Status/          → Google Form feedback
├── GithubRepo/               → Links to GitHub and live app
├── Main_Project/
│   ├── app/
│   │   └── eduPredict.py     → Main Streamlit app
│   ├── data/
│   │   ├── edu_cleaned.csv   → Preprocessed dataset
│   │   └── edu_raw.csv       → Raw UCI dataset
│   ├── model/
│   │   ├── rf_model.pkl
│   │   ├── trend_model.pkl
│   │   ├── tuned_xgboost_model.pkl
│   │   └── anomaly_detector_model.pkl
│   ├── notebooks/
│   │   ├── AcademicED.ipynb  → EDA + insights
│   │   ├── Modeling.ipynb    → Model training + comparison
├── SnapShot/                 → UI screenshots
├── Video/                    → Demo walkthrough
└── README.md
```

🤖 Modeling Details

Model Type	Purpose	File
Random Forest	Outcome classification	`rf_model.pkl`
XGBoost	Tuned final classifier	`tuned_xgboost_model.pkl`
Logistic Regression	Model comparison baseline	`tuned_logistic_regression_model.pkl`
Grade Trend Model	Predict next semester grade	`trend_model.pkl`
Anomaly Detection	Identify unusual student data	`anomaly_detector_model.pkl`

Evaluation Metrics Used:

- Accuracy

- F1 Score
- Confusion Matrix
- ROC-AUC

▶️ How to Run

Step 1: Clone the repo
cd EduPredict/Main_Project

Step 2: Install dependencies

pip install -r requirements.txt

Step 3: Run the Streamlit app

streamlit run app/eduPredict.py

🔐 Authentication System

Streamlit's `session_state` is used for managing login sessions.

Role	Username	Password
Student	student	$7ud3n7
Teacher	teacher	73@ch3r
Counselor	counselor	C0un$3l0r

Passwords are stored in memory (for now). Future versions can integrate hashed storage or Firebase Auth.

🧩 Feature Breakdown

Feature	Description
Role-based Dashboard	Different views for students, teachers, counselors
Predict Student Outcome	Predicts: Dropout / Enrolled / Graduate
Grade Trend Prediction	Predicts next semester grades
Anomaly Detection	Flags outliers in data (grade, attendance, etc.)

Feature	Description
Visualization	Violin, scatter, pie charts, and heatmaps
Downloadable Reports	CSV and PDF export of student analysis
Feedback Integration	Links to Google Forms for user feedback

➕ Adding New Features

To add a new model or chart:

1. Train the model in `notebooks/Modeling.ipynb`
2. Save the model using `joblib.dump()`
3. Add logic in `eduPredict.py` to load and use the model
4. Create a new section in the Streamlit sidebar or main page
5. Use Plotly/Matplotlib for new visualizations

🐞 Common Issues & Debugging

Issue	Fix
Model not found	Ensure `.pkl` files exist in `/model`
Streamlit crashes at start	Check Python version (use Python 3.10+)
Session timeout	Use `st.session_state` correctly
Charts not rendering	Check if data is loaded properly (edu_cleaned.csv)