

# AI-Based System to Predict Student Performance, Risk Level, and Weak Areas

Tech Team: Mayank Rajguru, Faraaz Kurawle & Yash Mhparle

NEXTEDGE: Data Analytics & Cyber Security Showcase



# The Unseen Struggle: Why Traditional Systems Fall Short



## Academic Challenges are Complex

Students often face difficulties silently, leading to delayed support and missed opportunities.



## Missing the Bigger Picture

These rules often fail to detect subtle, underlying patterns in student data that indicate early signs of struggle.



## Reactive, Not Proactive

Current systems rely on rigid, threshold-based alerts (e.g., attendance < 75%, failing an exam).



## The Consequence

Delayed interventions, increased dropout rates, and unfulfilled student potential.

As students, we've witnessed these challenges firsthand among our peers. We believe in leveraging advanced technology to empower both students and educators.

# Our Solution: A Multi-Faceted AI for Student Success

Our "Student Success Predictor" is designed to empower proactive, personalized interventions for every student, directly aligning with the National Education Policy (NEP 2020) emphasis on holistic development and adaptive learning.



## **Predict Student SGPA**

Forecasting their future academic standing with precision.



## **Assess Academic Risk Level**

Categorizing students into High, Medium, or Low risk of academic difficulty.



## **Identify Weak Course Type**

Pinpointing specific areas (e.g., Theory, Practical, Programming) where a student is likely to struggle.





# Beyond Rules: The Power of Data-Driven Insights

## Limitations of Traditional Methods:

- **Static & Inflexible:** "If attendance < 75%, flag" – these rules are rigid and can't adapt to individual student contexts.
- **Generic & Disconnected:** Off-the-shelf systems often use broad datasets, failing to understand our college's unique environment.

## Our AI Advantage:

- **Learns from Real, Local Data:** Builds predictive power from our college's authentic historical student performance.
- **Adaptive & Localized:** Specifically tuned to the nuances of our academic environment, ensuring highly relevant and accurate predictions.
- **Uncovers Hidden Patterns:** AI can identify complex correlations and subtle indicators that human-defined rules simply can't.

# Our Data: The Foundation of Accurate Predictions

## Our Foundation:

**Source:** Authenticated records from our college, spanning two semesters.

**Dataset Size:** Currently 360 student records (robust early-stage dataset, with plans for expansion!).

**Accuracy:** 100% real and accurate, meticulously collected and verified.

## Key Data Points (Features):

- semester (1 or 2)
- job\_hours (hours/week a student works)
- prev\_sgpa (SGPA of the previous semester)
- prev\_course\_score (Avg score in previous practical/theory/programming courses)
- attendance\_percentage (Current semester attendance)
- course\_type\_score (Current semester's average per course type)

## Preprocessing & Feature Engineering Insights:

**Handling Missing Values:** prev\_sgpa and prev\_course\_score for Semester 1 students are set to -1, combined with the semester flag, to robustly inform the model.

**Categorical Encoding:** Transforming labels into numerical formats for optimal model input.

**Feature Creation:** Calculating avg\_theory\_score, avg\_practical\_score, avg\_programming\_score for deeper insights.



# The Brains Behind the Predictions: Our Random Forest Models



## Random Forest

Our algorithm of choice.



## Handles Tabular Data

Excellently suited for structured student records.



## Robust with Missing Values

Tolerant to initial gaps, like prev\_sgpa for new students.



## Strong Performance

Delivers reliable results even with 360 records – crucial for early-stage development.



## Interpretability

Allows understanding of features driving predictions, aiding actionable insights.

### Random Forest Regressor:

**Predicts SGPA:** A precise numerical forecast of the student's academic standing.

### Random Forest Classifier (Risk):

**Predicts Academic Risk Level:**

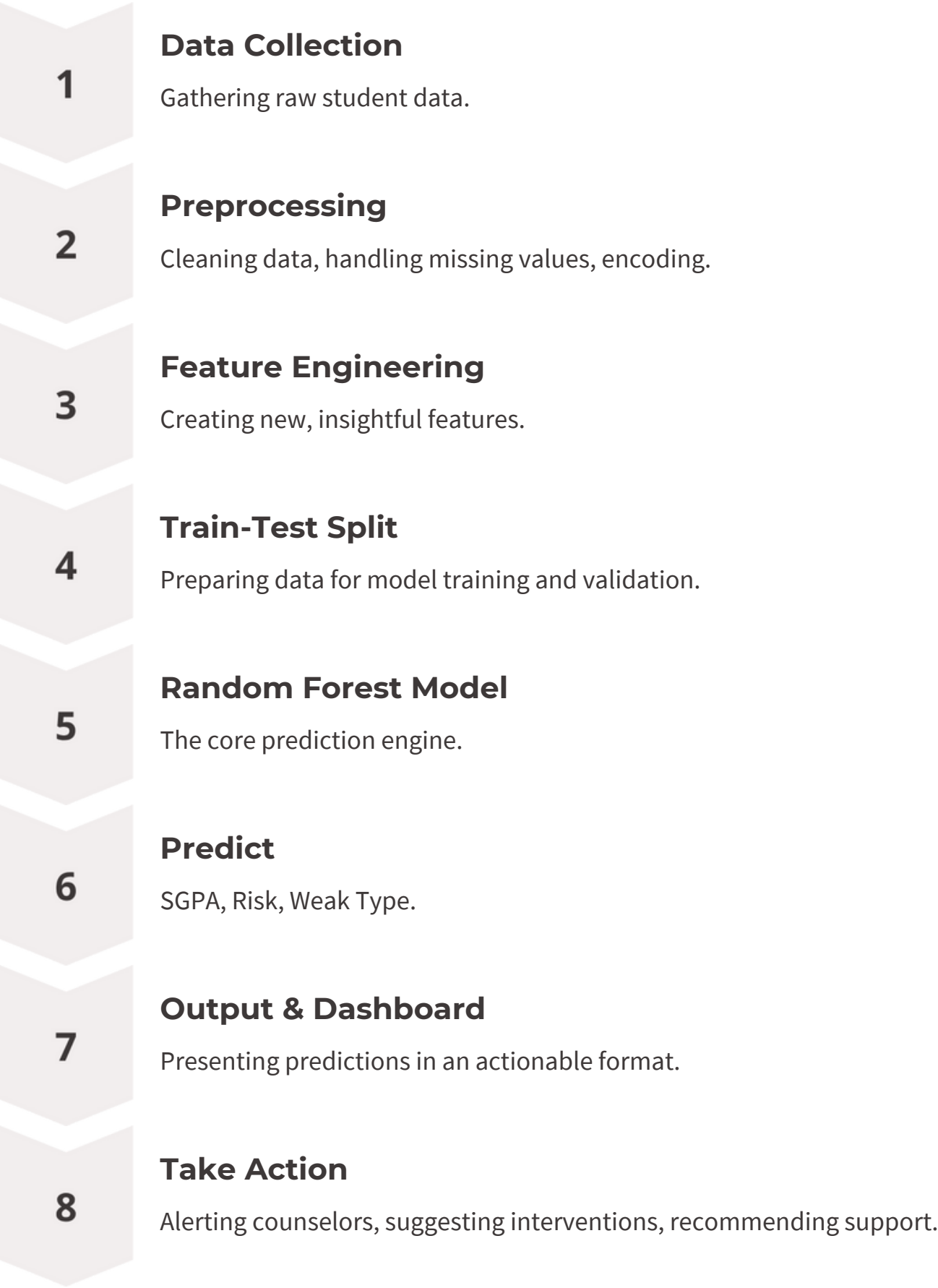
- **High Risk** (Predicted SGPA < 6)
- **Medium Risk** ( $6 \leq$  Predicted SGPA < 7.5)
- **Low Risk** (Predicted SGPA  $\geq$  7.5)

### Random Forest Classifier (Weak Areas):

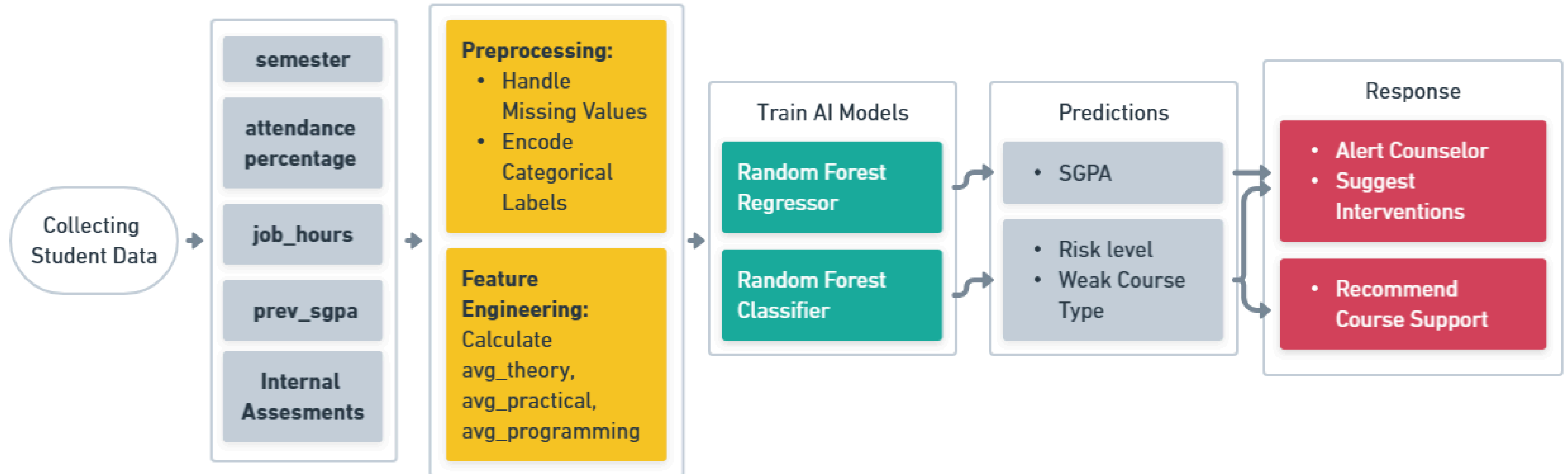
**Predicts Weak Course Type:** Identifies specific areas for intervention:

- Theory-focused courses
- Programming-focused courses
- Practical-focused courses

# Our Seamless Pipeline: Data to Intervention



# Model Workflow:





# Ensuring Reliability & Driving Real Change

## Rigorous Evaluation:

- **SGPA Prediction:** Evaluated using RMSE (Root Mean Squared Error) to quantify prediction accuracy.
- **Risk Level & Weak Course Type:** Evaluated using Accuracy (with future plans for Precision, Recall, F1-score to handle class imbalances more rigorously).

## Robust Validation (Cross-Validation):

Helps us:

- Assess how well models generalize to unseen data.
- Reduce risk of overfitting, especially with early-stage dataset.
- Provide reliable estimate of model performance across data subsets.

## Why Our System is Unique & Impactful:

### Localized & Relevant

Trained on our college's data, ensuring immediate applicability and trust.  
Clear vision for broader adoption.

### Proactive & Actionable

Moves beyond reactive flags to predictive insights, enabling timely interventions.

### Customized for India & NEP 2020

Directly supports personalized learning and outcome-based education.

### More Adaptive

Outperforms rigid, threshold-based rules by learning from complex data patterns.

# Overcoming Hurdles & Charting the Future

Challenge	Mitigation Strategy (Current & Future)
Data Availability & Scale	Expanding data collection to include more semesters and student cohorts. Exploring partnerships with other institutions for broader datasets.
Data Privacy & Security	Implementing robust anonymization techniques. Adhering to strict data protection regulations (e.g., GDPR, local equivalents). Encrypted storage and access controls.
Model Explainability	Using SHAP/LIME for clearer insights into model decisions. Developing user-friendly dashboards for educators to understand predictions.
Intervention Effectiveness	Longitudinal studies to track student outcomes based on interventions. A/B testing different intervention strategies. Feedback loops with counselors and students.
Integration with Existing Systems	Developing APIs for seamless integration with university management systems (e.g., SIS, LMS). User-friendly interface design.
Bias in Data & Model	Regular audits of data for bias. Implementing fairness metrics in model evaluation. Exploring debiasing techniques in training data and algorithms.

# Future Enhancements

We envision improving the capabilities to provide even more comprehensive and integrated solutions.

