

Student Stress Detection System — Executive Overview

Concise report | December 12, 2025

Abstract (Overview)

A lightweight machine learning system predicts student stress levels (low / medium / high) using an ensemble of Logistic Regression, SVM, Naive Bayes, and Random Forest. The pipeline spans synthetic data generation (10,000 records), cleaning, SMOTE balancing, scaling, model training, evaluation, and a Flask web UI for real-time predictions and motivational suggestions.

Problem Statement

- Stress impacts student well-being and academic outcomes; early detection is essential.
- Manual assessments are slow, subjective, and not scalable.
- Need an automated, data-driven, and actionable prediction tool.

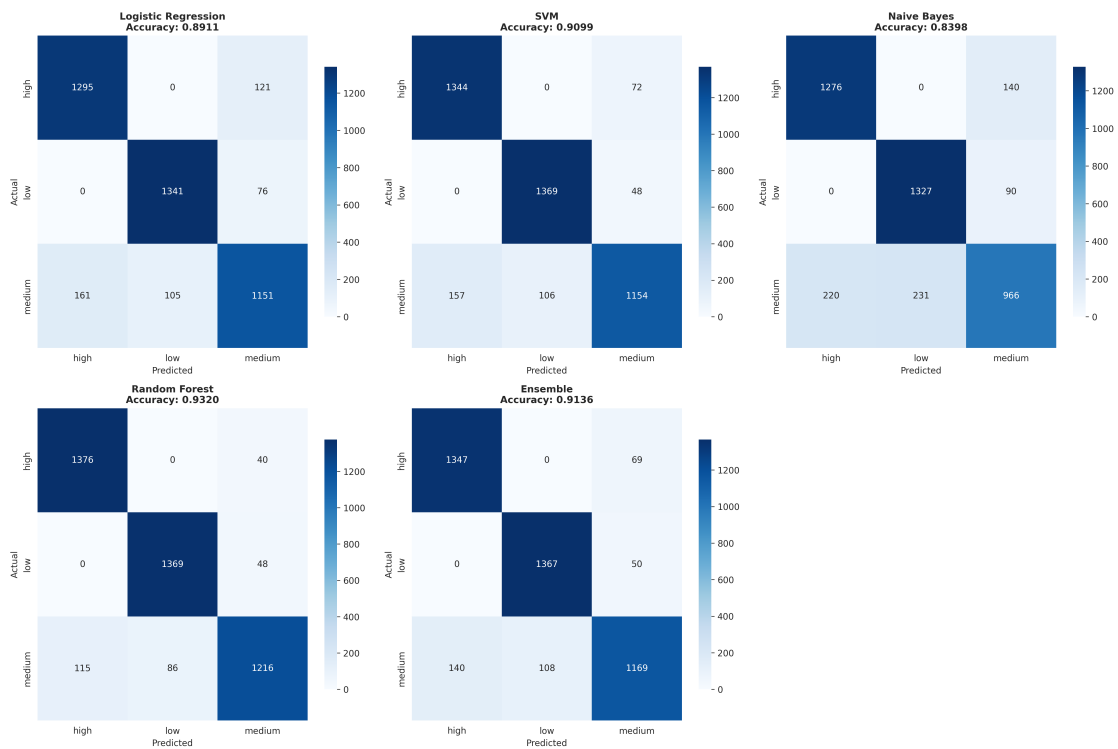
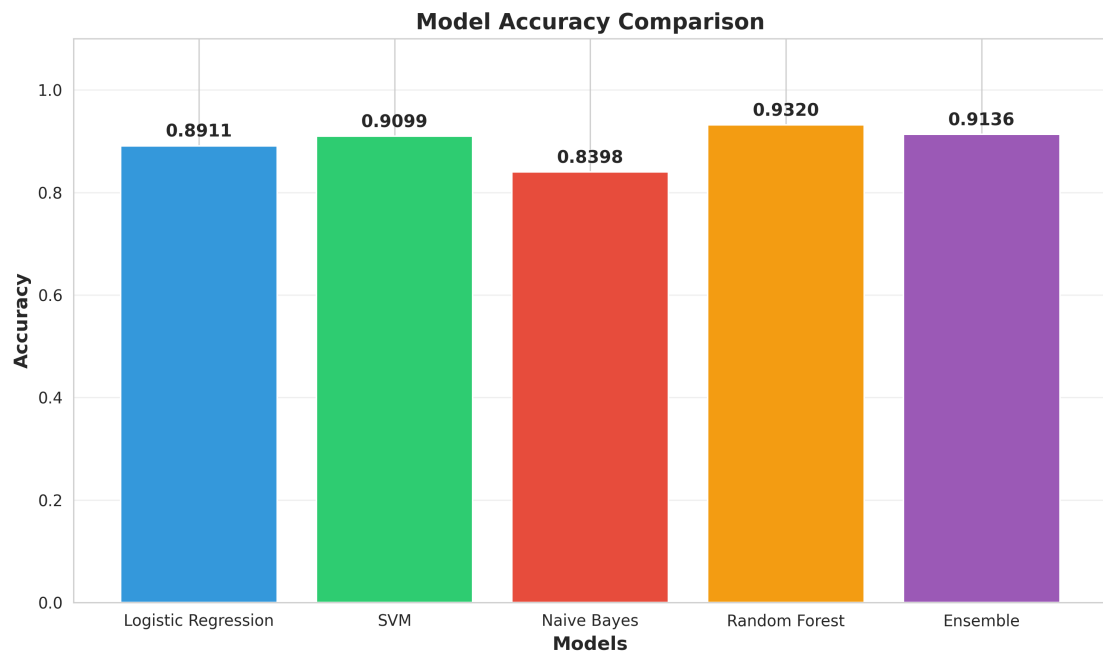
Chosen Model & Rationale

Selected Model: Voting Ensemble (soft voting) combining Logistic Regression, SVM, Naive Bayes, and Random Forest.

- Combines diverse learners to balance bias/variance and improve robustness.
- Soft voting leverages calibrated class probabilities for better uncertainty handling.
- Generalizes better than individual models and reduces overfitting risk.

Results (Overview)

- Accuracy (test): Ensemble 91.36%; strongest individual model (Random Forest) 93.20%.
- Low vs. Medium well separated; some Medium/High confusion (area to tune).
- Outputs class probabilities to gauge prediction confidence.



Conclusion

- End-to-end, reproducible pipeline ready for pilot use.
- Ensemble model offers strong, balanced performance with probability outputs.
- Next focus: threshold tuning for High-stress recall, real-world validation, monitoring.