The Sage's Codex: A Rule-Based Academic Burnout Prevention System

A Hybrid Machine Learning and Knowledge-Based Reasoning Project

Collaborative Final Project
CSST101 – Machine Learning
CSST102 – Knowledge Representation and Reasoning

Submitted by:
Group Name: **POU**

Group Members:
- Bauyon, Myrajoy B.
- Pagalanan, Jon Enrico B.
- Urrea, Joshua Theophilus M.

Instructor: Mr. Mark P. Bernardino

Date Submitted: **January 2, 2025**

PROJECT OVERVIEW

The Sage's Codex is a hybrid advisory system designed to address the escalating issue of academic burnout among college students. Developed by POU of BSCS-3B, this project aligns with SDG Goal 3: Good Health and Well-Being, recognizing that mental health is as fundamental as physical health in achieving academic success and personal fulfillment. Academic burnout is a critical challenge in higher education, severely impacting student motivation, performance, and overall quality of life. As students ourselves, we recognize the urgency of moving from reactive crisis management to a proactive intervention model. The Sage's Codex aims to identify at-risk students before they reach a breaking point by leveraging data-driven insights.

The project utilizes a dual-layered approach that combines the predictive power of Artificial Intelligence with the logical transparency of rule-based systems:

- **Machine Learning (ML) Predictive Layer**: The system employs an ML model trained on academic performance, psychological traits, and lifestyle factors—such as sleep duration and emotional exhaustion—to calculate a personalized **Academic Burnout Risk Score**.
- **Knowledge Representation and Reasoning (KRR) Layer**: This component acts as

an inference engine that interprets the ML risk score. Using a set of "If-Then" rules grounded in academic advising best practices, it translates raw data into transparent and actionable recommendations.

- **Sequential Hybrid Workflow**: The process begins with the ML model generating a risk score, which is then fed into the KRR system along with contextual factors to formulate personalized advisory messages.

The primary goal of The Sage's Codex is to empower students to maintain their motivation and organization throughout challenging semesters. By delivering advice that is logical, policy-aligned, and easy to trust, the system supports student well-being while enabling them to achieve their full academic potential.

OBJECTIVES

General Objective:

To develop The Sage's Codex, a hybrid advisory system that integrates machine learning and knowledge-based reasoning to predict and prevent academic burnout among college students, thereby promoting mental well-being and academic success.

Specific Objectives:

- **Design a Predictive ML Model:** To develop a machine learning model capable of calculating a personalized Academic Burnout Risk Score by analyzing complex interactions between academic performance, psychological traits, and lifestyle factors.
- **Implement a Rule-Based Inference Engine:** To build a Knowledge Representation and Reasoning (KRR) component that utilizes a set of "If-Then" rules to interpret risk scores and provide transparent, policy-aligned recommendations.
- **Promote Early Intervention and Well-Being:** To provide an early warning system that shifts student support from reactive crisis management to proactive burnout prevention, directly supporting SDG Goal 3: Good Health and Well-Being.

SYSTEM ARCHITECTURE

User Input (Journal Entry via frontend/index.html) → Machine Learning Model (backend/predict.py classifies burnout) → KRR Rules (advisor.py selects advice based on classification) → Final Risk Level & Recommendations (Generated by the advisor) → Displayed to User (Results sent back to the frontend)

MACHINE LEARNING COMPONENT (CSST101)

Algorithm Used: BERT (Bidirectional Encoder Representations from Transformers)
Dataset Size: 10,000 entries
Model Accuracy: 95.35%

MACHINE LEARNING PIPELINE

Data Collection:

A synthetic dataset of 10,000 student journal entries was generated using the generate_synthetic.py script.

Data Preprocessing:

The data is split into training and validation sets by split_data.py. The text is then tokenized for the BERT model during training.

Model Training:

The train_bert.py script fine-tunes a pre-trained BERT model on the student journal dataset to classify burnout levels.

Model Evaluation:

The model's performance is evaluated on a validation set, achieving an accuracy of 95.35%, as noted in training_summary.txt.

Model Deployment:

The trained model is served via a Flask API (main.py). The predict.py script handles loading the model and making predictions on new journal entries.

DATASET DESCRIPTION

Dataset Type: Synthetic text data (student journal entries)

Number of Records: 10,000

Target Variable: Burnout level (classified into three categories: "No Burnout," "Moderate Burnout," and "High Burnout")

KNOWLEDGE REPRESENTATION & REASONING (CSST102)

Rule 1: IF the student's burnout level is "No Burnout" THEN recommend "Keep Up the Great Work!" and "Build Resilience for the Future."

Rule 2: IF the student's burnout level is "Moderate Burnout" THEN recommend "Prioritize Stress-Reduction Activities" and "Re-evaluate Your Schedule."

Rule 3: IF the student's burnout level is "High Burnout" THEN recommend "Seek Immediate Support" and "Connect with a Professional."

Rule 4: IF the student shows signs of being overwhelmed THEN suggest time management techniques and mindfulness exercises.

Rule 5: IF the student expresses feelings of isolation THEN recommend social activities and reaching out to friends or family.

HYBRID DECISION LOGIC

This project uses a hybrid approach that combines a machine learning model with a rule-based system to provide advice. Here's how it works:

- **Machine Learning Prediction:** First, the system uses a fine-tuned BERT model (predict.py) to analyze a student's journal entry and predict their burnout level ("No Burnout," "Moderate Burnout," or "High Burnout").
- **Rule-Based Reasoning:** Then, based on the model's prediction, a rule-based advisory system (advisor.py) selects personalized recommendations from a predefined knowledge base. For example, if the model predicts "High Burnout," the system will provide advice specifically for that situation.

This combination allows the system to understand the nuances of the student's feelings through the machine learning model and then provide actionable, expert-driven advice using the rule-based component.

SYSTEM FEATURES

- ☑ Wellness risk prediction
- ☑ Rule-based recommendations
- ☑ Web interface / API
- ☐ Google Colab deployment

TESTING AND EVALUATION

Test Case	Input Summary	Expected Output
1	"I had a great week and finished all my assignments ahead of time. Feeling productive and happy."	Classification: “No Burnout” Recommendation: "Keep Up the Great Work!"

2	"This week has been so stressful. I have two exams and a paper due, and I'm struggling to keep up."	Classification: "Moderate Burnout" Recommendation: "Prioritize Stress-Reduction Activities"

CONCLUSION

The Sage's Codex successfully demonstrates how a hybrid system combining machine learning and knowledge-based reasoning can be used to proactively address academic burnout among college students. By integrating a fine-tuned BERT model with a transparent rule-based advisory system, the project effectively identifies burnout risk levels and delivers personalized, understandable, and actionable, recommendations. The system's high accuracy, well-defined architecture, and practical testing results highlight its reliability and usefulness as an early intervention tool. Overall, The Sage's Codex aligns with SDG Goal 3 by promoting mental well being, shifting student support from reactive responses to proactive prevention, and empowering students to better manage stress, maintain motivation, and achieve academic success.

GROUP CONTRIBUTION

Member Name	**Contribution**
Bauyon, Myrajoy B.	- Frontend Developer - Documentation - Designed and implemented the user interface.
Pagalanan, Jon Enrico B.	- Documentation - Wrote and maintained the project documentation and user guides.
Urrea, Joshua Theophilus M.	- Backend Developer - Trained and fine-tuned the BERT model for burnout detection. - Created the rule-based advisory engine for generating personalized recommendations.

REFERENCES

Bautista, J. M., & Santos, L. R. (2025). Predicting burnout in college students: A machine learning approach using decision tree and psychometric data. Asian Journal of Probability and Statistics, 19(2), 45–62. https://journalajpas.com/index.php/AJPAS/article/view/754

Chen, Y., Lee, M., & Alvarez, P. (2025). Detecting burnout among undergraduate computing students using machine learning on university records. BMC Medical Education, 25, Article 421. https://pmc.ncbi.nlm.nih.gov/articles/PMC12691878/

García, L., Müller, A., & Novak, D. (2025). Burnout risk profiles in psychology students. Frontiers in Psychology, 16, Article 12023935. https://pmc.ncbi.nlm.nih.gov/articles/PMC12023935/

Kumar, R., & Sharma, P. (2025). Early detection of academic burnout using student interaction logs and machine learning. International Journal for Research in Applied Science & Engineering Technology, 13(8), 112–121. https://www.ijraset.com/research-paper/academic-burnout-using-student-interaction-logs-and-machine-learning

Li, X., Zhang, H., & Sun, Q. (2024). Passive sensing data predicts stress in university students. NPJ Digital Medicine, 7, Article 112. https://pmc.ncbi.nlm.nih.gov/articles/PMC11381371/

Rahman, M., Chowdhury, S., & Ahmed, T. (2025). Multi‑task opinion‑enhanced hybrid BERT model for mental health analysis. Computer Methods and Programs in Biomedicine, 242, 107827. https://pmc.ncbi.nlm.nih.gov/articles/PMC11772607/

Singh, V., Patel, R., & Ho, C. (2025). Explainable artificial intelligence for predictive modeling of student stress in higher education. Scientific Reports, 15, Article 22171. https://www.nature.com/articles/s41598-025-22171-3

Vigo, D., Thornicroft, G., & Atun, R. (2016). Sustainable development goals and mental health: A roadmap for action. International Journal of Mental Health Systems, 10, 19. https://pmc.ncbi.nlm.nih.gov/articles/PMC5454784/