Engineering programs are consistently identified as among the most academically demanding fields in higher education, characterized by intensive workloads, complex problem-solving requirements, and sustained exposure to high-pressure evaluation environments. Previous studies have shown that engineering students report higher levels of academic stress and burnout compared to peers in other disciplines, factors that have been linked to reduced academic engagement and performance decline. Despite this, early identification of students at risk of declining performance remains challenging, as traditional indicators such as grade point averages often fail to capture underlying behavioral and mental drivers.

To address this gap, this study adopts a data-driven approach grounded in data science and educational analytics. Data science refers to the systematic extraction of actionable insights from structured and unstructured data through the combined use of statistical analysis, computational techniques, and domain-specific knowledge. Within the educational context, data science has increasingly been applied to model student behavior, identify risk factors for academic failure, and support evidence-based institutional decision-making.

Universities generate large volumes of data related to student attendance, assessment outcomes, learning behaviors, and self-reported well-being. When analyzed appropriately, such data can reveal patterns that are not apparent through descriptive observation alone. Prior research in educational data mining suggests that psychosocial and behavioral variables—such as stress, sleep quality, and motivation—can be strong predictors of academic outcomes, sometimes surpassing purely academic metrics especially for long term predictions.

Building on this body of work, the present study investigates the determinants of declining academic performance among engineering students at l'Université Internationale De Casablanca (UIC). Rather than focusing solely on grades, this paper examines a multidimensional set of variables encompassing psychological well-being, academic behaviors, and perceived learning environment. The primary objective is to identify the factors most strongly associated with performance decline and to evaluate the feasibility of predicting such decline using machine learning techniques.

Research Objectives

The specific objectives of this study are as follows:

To identify key psychosocial, behavioral, and academic factors associated with declining academic performance among engineering students.

To analyze the relative influence of well-being-related variables compared to academic behavior indicators.

To develop and evaluate a machine learning model capable of predicting the likelihood of academic performance decline based on student-reported data.

2. Data Collection and Variable Selection

2.1 Data Collection

Data for this study were collected through a structured survey administered to a sample of 114 engineering students enrolled at Université Internationale De Casablanca (UIC). Participation was voluntary, and responses were collected anonymously to reduce social desirability bias and encourage honest reporting. The survey was designed to capture both objective academic behaviors and subjective perceptions related to well-being and learning experience.

While several variables rely on self-reported measures, such instruments are widely used in educational and psychological research and have been shown to provide meaningful insights when interpreted with appropriate caution.

2.2 Variable Design and Categorization

The dataset consists of 13 variables selected based on prior literature on academic performance, student burnout, and educational psychology. These variables were organized into three main categories: the outcome variable, psychosocial well-being metrics, and academic behaviors and environmental factors.

2.3 Outcome Variable

The primary dependent variable of this study is academic performance trend:

Performance Score: A binary variable indicating whether a student's academic performance is stable or improving (0) versus declining (1). This variable serves as the target outcome for both statistical analysis and machine learning modeling.

2.4 Psychosocial and Well-being Metrics

Given the high-pressure nature of engineering education, several variables were included to capture students' internal psychological state:

Stress Level (1–10): A self-reported measure of perceived daily stress intensity.

Burnout Score (1–5): An indicator of emotional and physical exhaustion related to academic demands.

Motivation Score (1–10): A self-assessed measure of academic drive.

Sleep Habits: Average number of hours slept per night: included due to its established relationship with cognitive performance, memory consolidation, and emotional regulation.

These variables aim to quantify dimensions of student well-being that are often overlooked in traditional academic performance analyses.

2.5 Academic Behaviors and Learning Environment

To capture external behaviors and contextual factors influencing academic outcomes, the following variables were included:

Class Presence Rate (%): A measure of lecture and laboratory attendance consistency.

Study Consistency: A categorical variable distinguishing between regular daily study and last-minute studying prior to examinations.

Schedule Load Perception: A binary variable indicating whether students perceive their academic schedule as excessively heavy and potentially demotivating.

Feeling Lost in Class: A binary indicator capturing whether students frequently feel unable to follow course material, a factor potentially associated with reduced self-efficacy and impostor syndrome.

Troubled Modules Count: The number of modules a student reports struggling with.

Primary Difficulty Source: A categorical variable identifying whether difficulties stem primarily from subject complexity, instructional factors, or a combination of both.

By integrating these academic and environmental variables with psychosocial metrics, the dataset enables a holistic examination of factors contributing to academic performance decline.

This multidimensional variable framework provides the foundation for subsequent exploratory data analysis, hypothesis testing, and predictive modeling aimed at identifying at-risk students and informing potential academic support interventions.

3. Data Preparation, Exploratory Analysis, and Visualization

3.1 Data Preparation Pipeline

Raw survey responses were initially collected and stored in Microsoft Excel due to its suitability for structured data entry, preliminary validation, and error checking. Excel was used to verify data completeness, remove duplicate entries, and ensure consistency in variable formats prior to analysis.

Following initial validation, the dataset was imported into Power BI for data cleaning, transformation, and exploratory analysis. Power BI was selected due to its robust data preparation capabilities, including handling missing values, categorical variable encoding, and feature restructuring. These preprocessing steps were essential to ensure that the dataset was suitable for both exploratory data analysis and subsequent machine learning modeling. The final cleaned dataset was structured to allow seamless export for downstream analytical tasks.

3.2 Exploratory Data Analysis (EDA)

An exploratory data analysis was conducted to examine distributions, relationships, and group-level differences across the selected variables. Visual analytics techniques were employed to identify preliminary patterns between psychosocial metrics, academic behaviors, and academic performance trends. Rather than relying solely on descriptive statistics, visualization enabled the identification of non-linear relationships and interaction effects that may not be immediately apparent through numerical summaries alone.

3.3 Visualization Using Power BI
Power BI was utilized as the primary visual analytics tool to support the exploratory analysis phase. Selected visualizations were generated to illustrate key relationships identified, including comparisons of stress levels, burnout scores, sleep duration, and attendance rates across performance trend categories. These visualizations served to contextualize statistical patterns and guide the formulation of research hypotheses.

To maintain analytical clarity, only visualizations directly supporting the research objectives were included in the main body of this paper. Additional dashboards and interactive visual elements are provided as supplementary material to enhance transparency and reproducibility. Figures referenced in this section represent processed and cleaned data rather than raw survey outputs.

3.4 Data Readiness for Predictive Modeling
Beyond visualization, Power BI was employed to prepare the dataset for predictive modeling. This included transforming categorical variables into machine-readable formats, filtering incomplete records, and ensuring consistent scaling where necessary. The resulting dataset was then exported for use in supervised machine learning models aimed at predicting academic performance decline.

This structured data preparation and exploratory analysis process ensures that subsequent hypothesis testing and modeling efforts are grounded in a well-curated and analytically sound dataset.

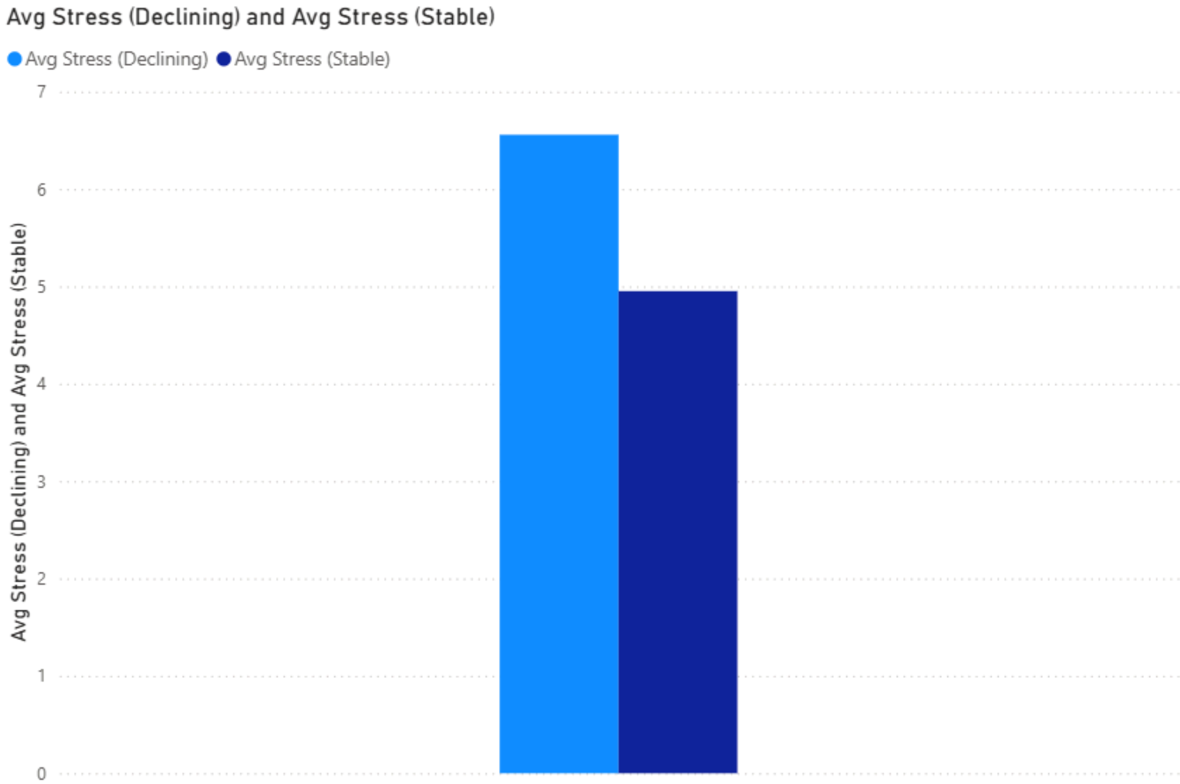
The insights obtained from the exploratory data analysis informed the formulation of research hypotheses, which are presented in the following section.

4.Hypotheses Formulation and Exploratory Validation

4.1 Rationale for Hypothesis Analysis
Following the exploratory data analysis, a hypothesis approach was adopted to systematically examine the relationships between student well-being, academic behaviors, and academic performance scores. Initial hypotheses were formulated based on intuitive assumptions and prior literature, and subsequently evaluated using descriptive statistics and visual analytics. PowerBI visualizations were employed to assess whether observed data patterns supported or contradicted each hypothesis. This iterative process allowed for progressive refinement toward identifying the most influential factors associated with academic performance decline.

4.2 Elevated Stress Levels Are Associated with Academic Decline

First Hypothesis (H1) :Students reporting higher stress levels are more likely to experience a decline in academic performance

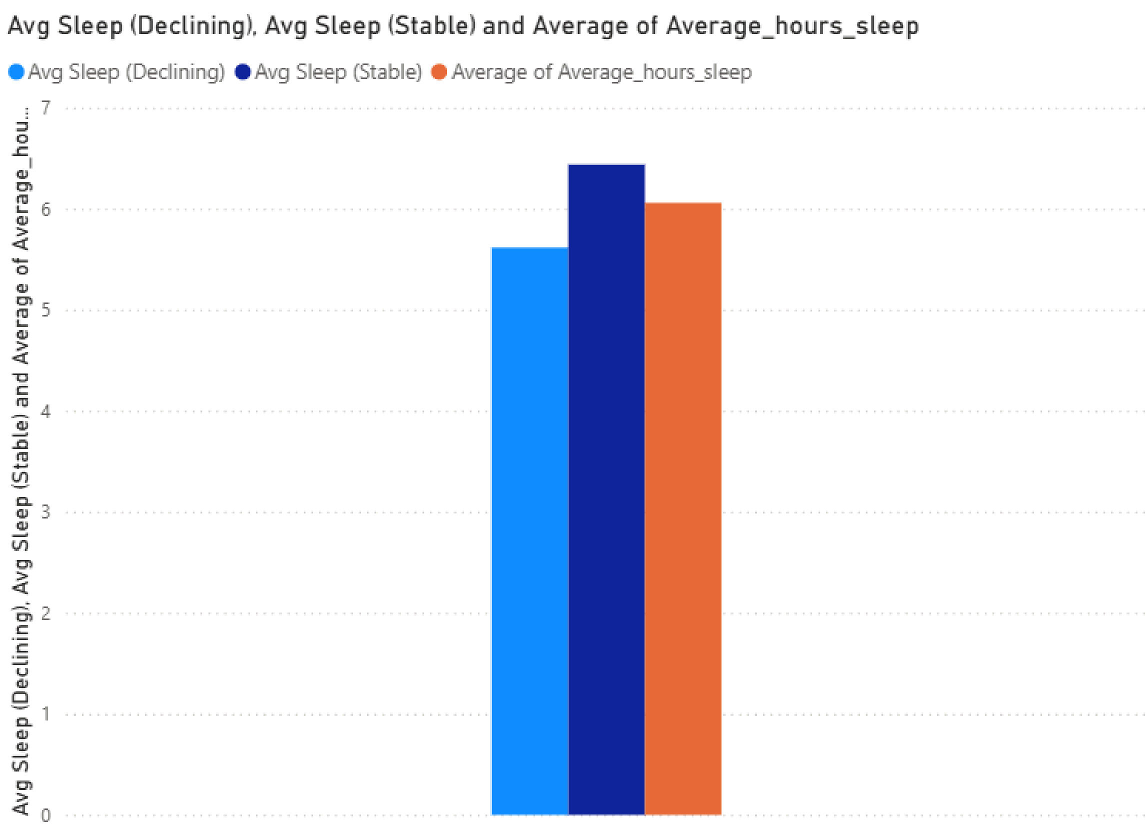


The visualization reveals that students classified as declining report significantly higher average stress levels (6,56) compared to those whose performance is stable or improving (4,95). Although some overlap exists between groups, the overall separation indicates that stress is strongly associated with academic decline. This suggests that heightened stress acts as a risk amplifier rather than a singular determinant. In conclusion H1 is **supported**, indicating a clear association between elevated stress levels and academic performance decline.

4.3 H2 – Reduced Sleep Duration Correlates with Academic Decline

Hypothesis Statement:

H2: Students with lower average sleep duration are more likely to experience academic performance decline.



Explanation:

Declining students report sleeping fewer hours on average (5,62 hours) than students in the stable or improving group (6,44). Adequate sleep appears to function as a protective factor, likely due to its role in cognitive performance, emotional regulation, and stress mitigation.

Conclusion

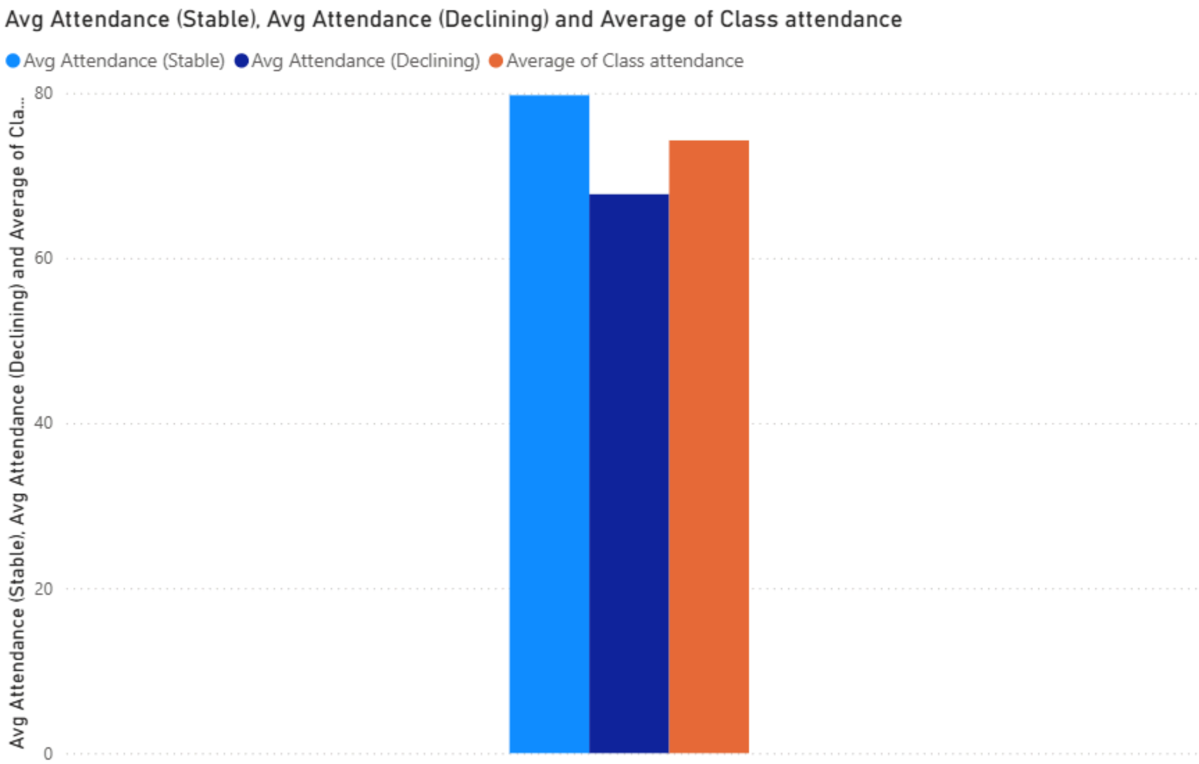
H2 is **supported**, suggesting that insufficient sleep is associated with increased vulnerability to academic decline.

4.4 H3 – Academic Behaviors Influence Performance Outcomes

4.4.1 Attendance Rate

Hypothesis Statement

H3a: Lower class attendance rates are associated with academic performance decline.



Explanation

The results indicate that declining students exhibit lower attendance rates (67,73%) compared to stable or improving students (79,73%) . This finding reinforces the role of consistent class participation as a foundational academic behavior linked to performance stability.

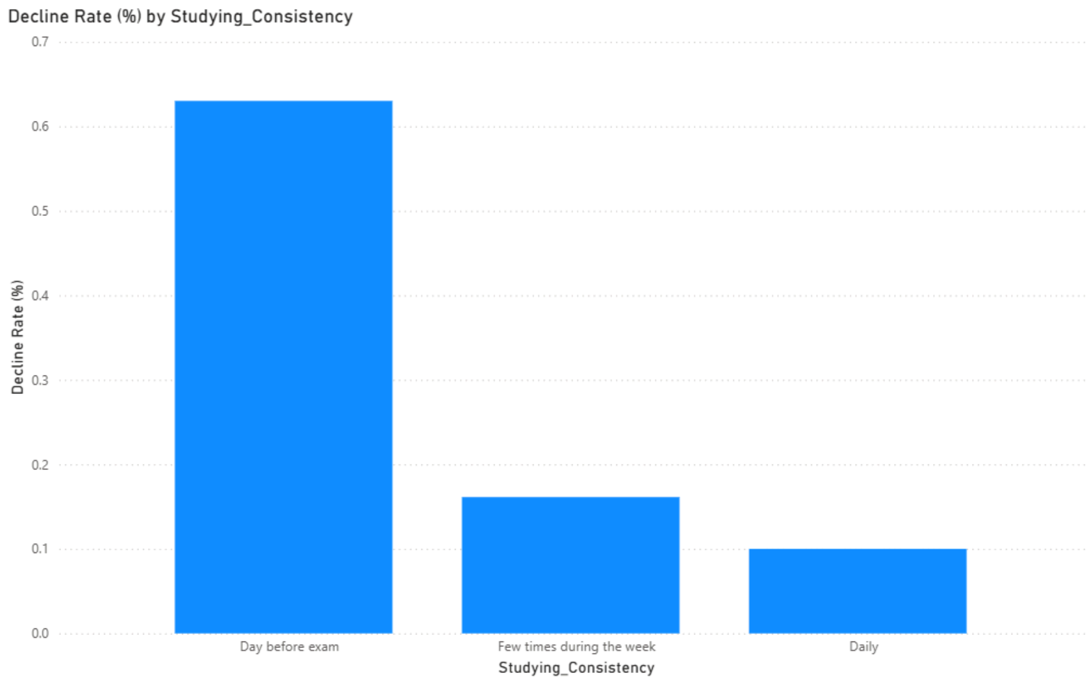
Conclusion

H3a is **supported**, demonstrating that reduced attendance is associated with academic decline.

4.4.2 Study Consistency

Hypothesis Statement

H3b: Students who study primarily shortly before examinations are more likely to experience academic decline.



Explanation

More than 60% of students experiencing academic decline report studying mainly the day before exams, whereas stable or improving students show more consistent study habits. This pattern suggests that cramming-based strategies may contribute to shallow learning and increased performance risk.

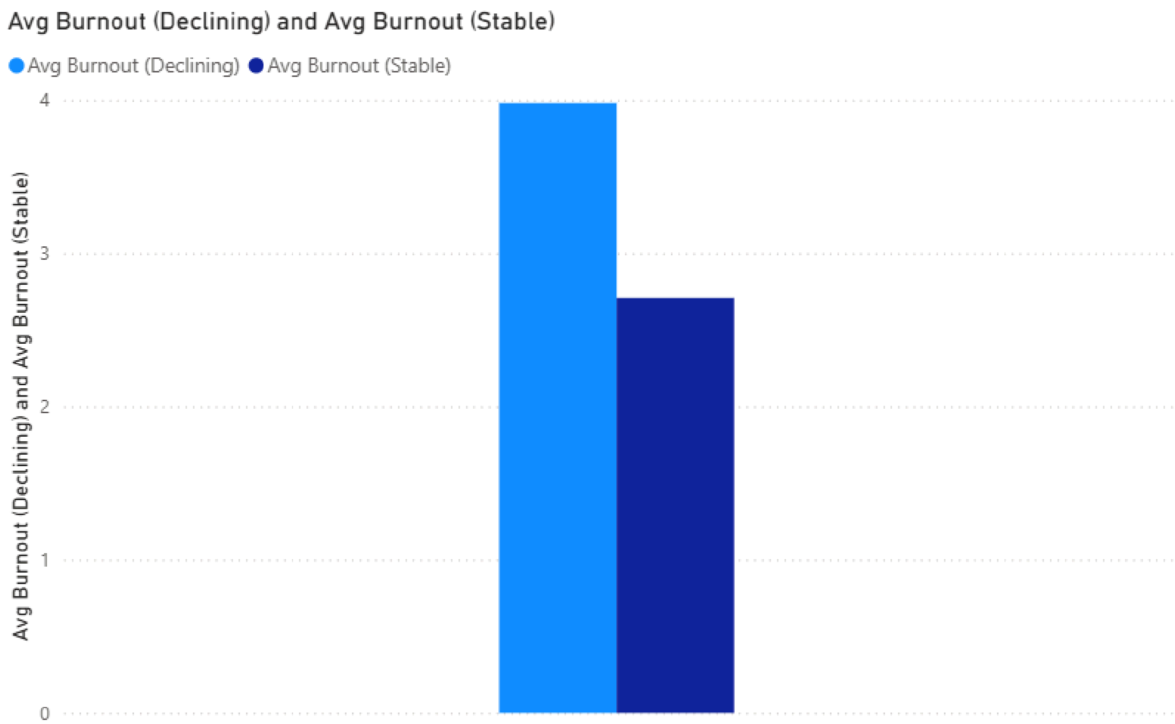
Conclusion

H3b is **strongly supported**, identifying study consistency as a critical behavioral indicator.

4.5 H4 – Burnout Is a Stronger Indicator of Decline Than Stress Alone

Hypothesis Statement

H4: Burnout exhibits a stronger association with academic performance decline than stress alone.



Explanation
Burnout demonstrates a clearer separation between declining (3,94/5) and stable students (2,71/5) than stress levels alone. This suggests that sustained emotional and physical exhaustion may play a more decisive role in long-term academic disengagement and decline.

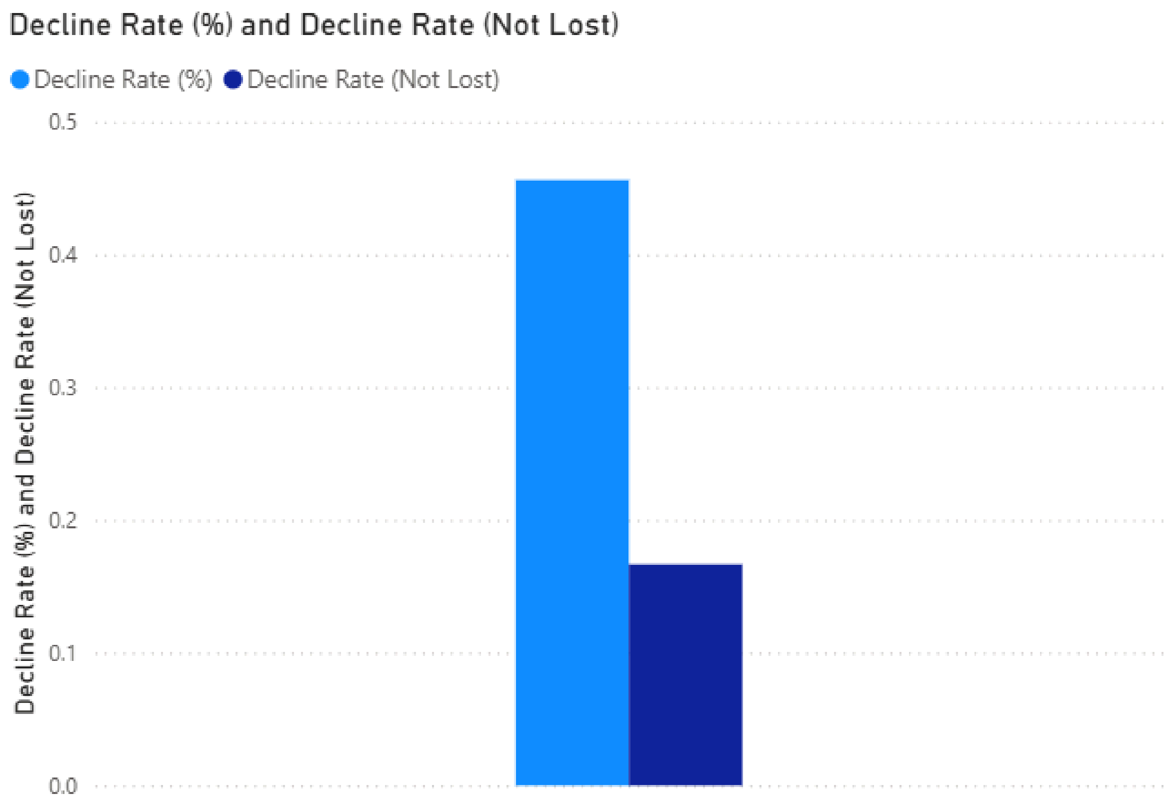
Conclusion

H4 is **strongly supported**, positioning burnout as a central contributor to academic performance decline.

4.6 H5 – Feeling Lost in Class Increases the Likelihood of Decline

Hypothesis Statement

H5: Students who frequently feel lost during lectures are more likely to experience academic decline.



Explanation

Students who report frequently feeling lost exhibit substantially higher decline rates (46% vs 17%) . This factor likely interacts with stress and burnout, compounding academic difficulty and increasing disengagement.

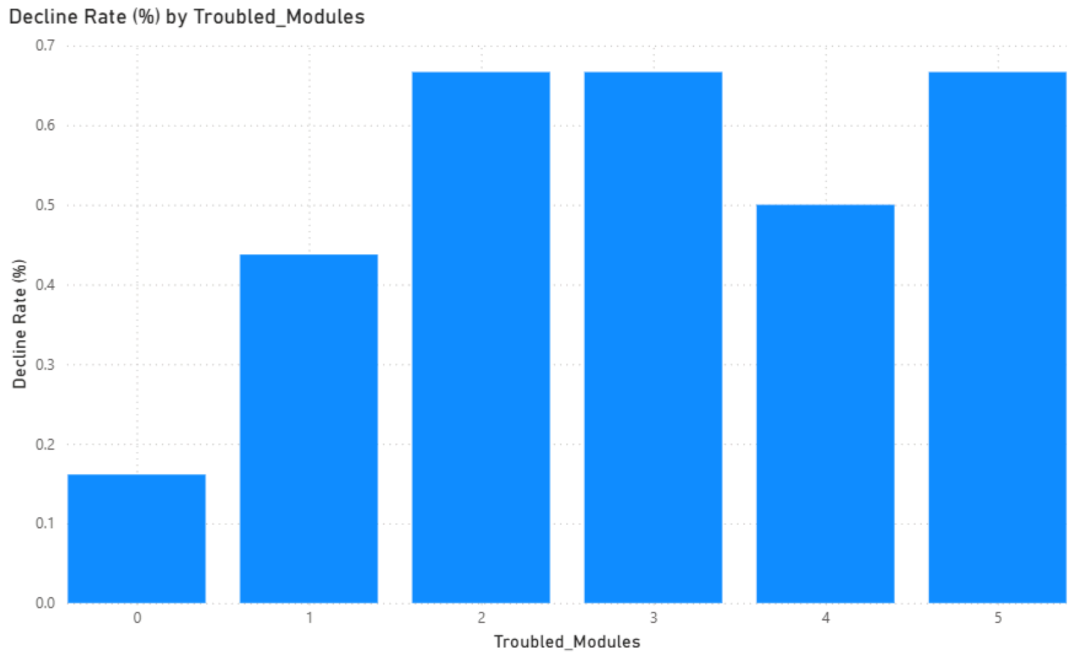
Conclusion

H5 is **supported**, indicating that perceived academic disorientation is a meaningful risk factor.

4.7 H6 – Multiple Troubled Modules Create a Performance Tipping Point

Hypothesis Statement

H6: The probability of academic decline increases non-linearly with the number of troubled modules.



Explanation

The visualization reveals a threshold effect: decline rates remain relatively low for students struggling with zero (16%) or one module (44%) , but increase sharply once students report difficulties in two or more modules(≥50%) . This suggests cumulative academic overload that accelerates performance decline.

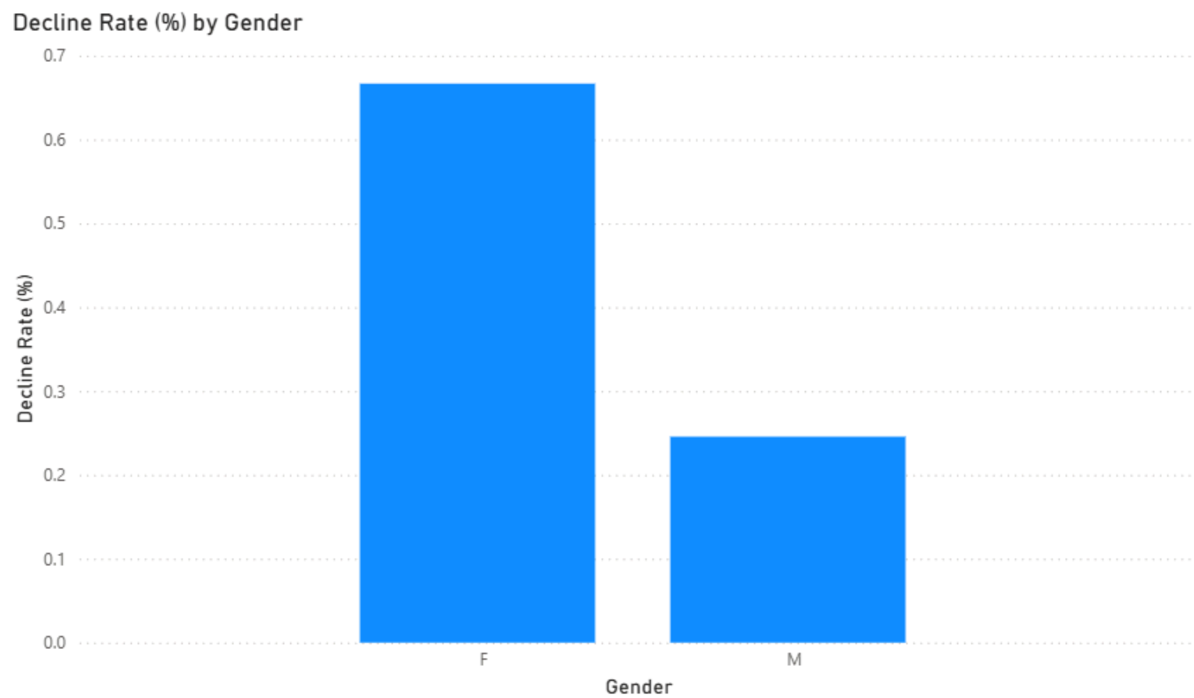
Conclusion

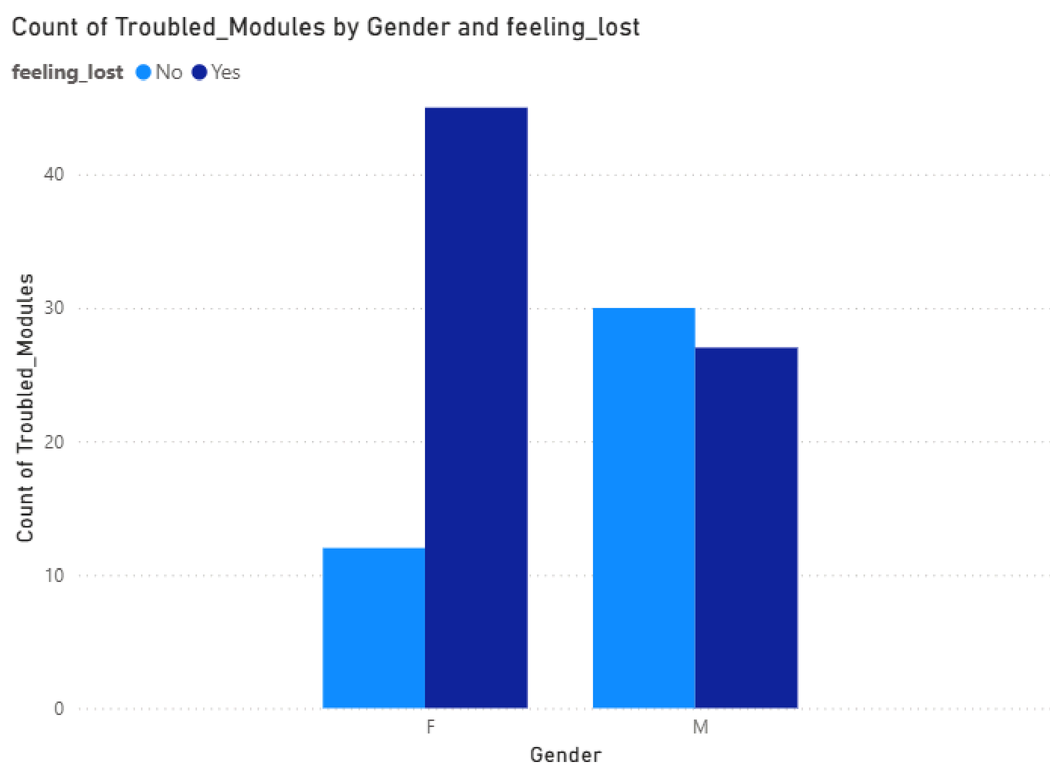
H6 is **strongly supported**, identifying multiple troubled modules as a critical structural risk factor.

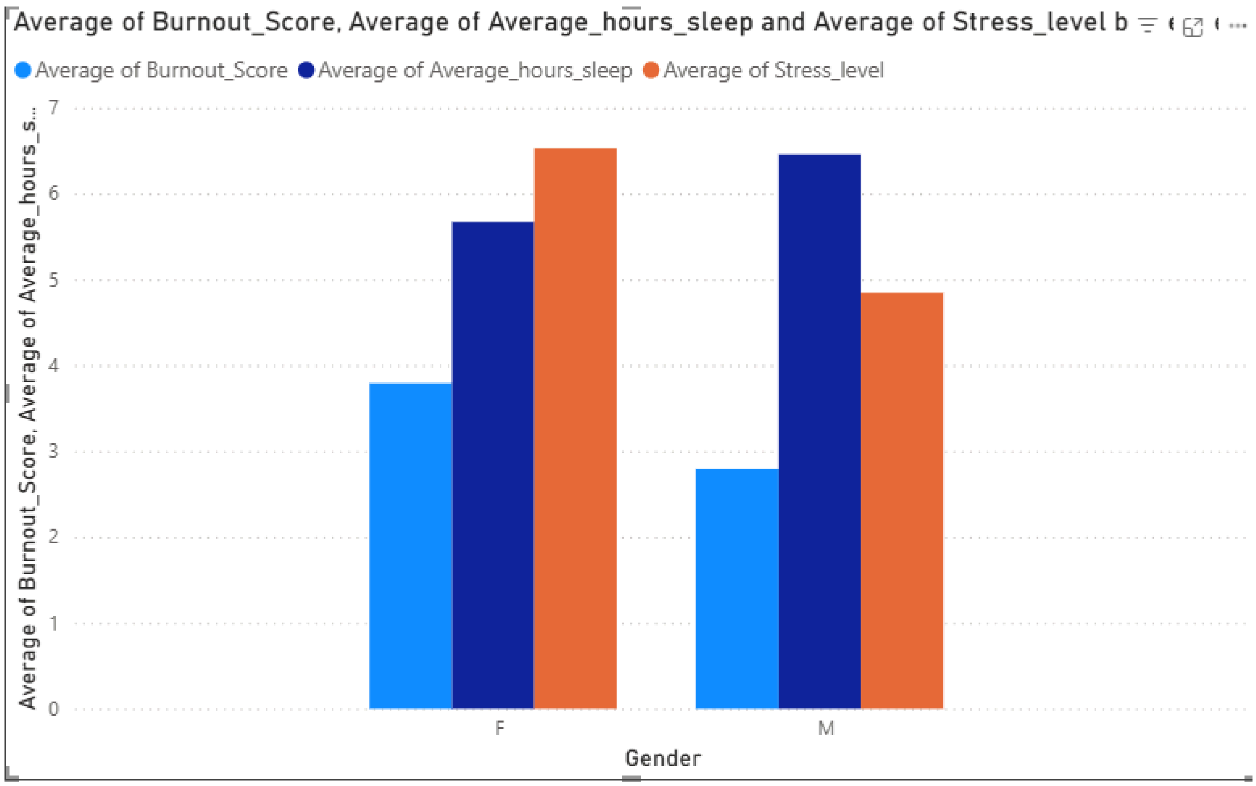
4.8 Gender-Based Exploratory Analysis

Hypothesis Statement

H7: Gender-based differences exist in academic decline rates and associated psychosocial indicators.


Decline Rate (%) by Gender
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Decline Rate (%)
F
M
Gender


Count of Troubled_Modules by Gender and feeling_lost
feeling_lost
No
Yes
40
30
20
10
0
Count of Troubled_Modules
F
M
Gender



Explanation

In this scenario female students display a higher proportion of academic decline (67%) and consistently higher average stress (65%) and burnout levels (3,79/5) across performance categories. Female students report feeling lost more frequently than male students. This suggests that decline patterns may differ by gender, with emotional exhaustion playing a more prominent role than perceived cognitive difficulty.

Conclusion

H7 is exploratorily supported. These findings are descriptive in nature and do not imply causality.

4.9 Section Summary

The hypotheses validated in this section indicate that academic performance decline among engineering students in UIC is not driven by isolated factors, but rather by a network of interrelated psychosocial, behavioral, and structural influences. Elevated stress, reduced sleep duration, increased burnout, inconsistent study habits, and academic overload were each found to be associated with higher decline rates.

Importantly, the exploratory analysis suggests that these factors may reinforce one another. For instance, reduced sleep duration is associated with higher stress levels, which in turn may contribute to emotional exhaustion and burnout. As burnout intensifies, students may disengage from academic behaviors such as regular attendance and consistent studying, thereby increasing the likelihood of struggling across multiple modules. This pattern reflects a cumulative or cascading effect, wherein initial stressors may propagate through interconnected domains of student well-being and academic behavior.

While the present analysis does not establish causal pathways, the observed associations are consistent with a domino-like progression of risk factors that collectively increase vulnerability to academic performance decline. Recognizing this interconnected structure underscores the importance of adopting multivariate and predictive modeling approaches rather than examining individual variables in isolation.

Consequently, these findings motivate the use of supervised machine learning techniques in the following section to quantify individual risk, model interactions between variables, and identify the most influential contributors to academic decline.

5. Predictive Modeling Using Machine Learning

5.1 Introduction to Machine Learning

Machine Learning (ML) is a subfield of artificial intelligence that focuses on developing algorithms capable of learning patterns from data and making predictions or decisions without being explicitly programmed for each specific case. Unlike traditional rule-based systems, machine learning models infer relationships directly from historical data and generalize these patterns to unseen observations.
In the context of educational analytics, machine learning is particularly valuable for identifying students at risk of academic decline by leveraging multiple interdependent variables simultaneously.

5.2 Labelled vs Unlabelled Data

Machine learning tasks are commonly categorized based on the availability of labeled outcomes:

- Labelled data contains both input features and a known target outcome.
- Unlabelled data consists only of input features, with no predefined output variable.

In this study, the dataset is **labelled**, as each student record includes an explicit academic performance outcome classified as Declining or Stable/Improving. This outcome variable serves as the ground truth for model training and evaluation.

5.3 Supervised vs Unsupervised Learning

Supervised Learning
Supervised learning algorithms learn a mapping between input features and a known target variable. These models are commonly used for:

- Classification (e.g., decline vs no decline)
- Regression (e.g., predicting continuous outcomes)

Unsupervised Learning
Unsupervised learning algorithms operate without labeled outcomes and aim to uncover hidden structures within the data, such as clusters or latent patterns.

Learning Paradigm Used in This Study
Given the presence of a labeled outcome variable and the objective of predicting academic performance decline, this study adopts a supervised learning classification approach.

5.4 Problem Definition and Modeling Objective

The predictive task is formulated as a binary classification problem, where the goal is to estimate the probability that a student will experience academic performance decline based on psychosocial, behavioral, and academic features.

Formally:

- **Input (X):** Stress level, burnout score, sleep duration, attendance rate, study consistency, troubled modules, feeling lost, gender, and related variables.
- **Output (Y):** Academic performance trend (Declining = 1, Stable/Improving = 0)

The model aims to learn a decision boundary that best separates declining students from non-declining students.

5.4 Feature Selection Based on Exploratory Analysis

Features included in the model were selected based on insights from Section 4. Variables demonstrating strong associations with academic decline—such as burnout, stress, sleep duration, study consistency, and number of troubled modules—were prioritized to improve predictive performance and interpretability.

This approach ensures alignment between exploratory findings and predictive modeling.

5.5.1 Logistic Regression

Logistic regression is a linear classification algorithm that estimates the probability of a binary outcome using a logistic (sigmoid) function. It is highly interpretable and allows direct examination of feature coefficients, making it well-suited for understanding which factors most strongly influence academic decline.

5.5.2 Decision Tree Classifier

Decision trees model decision rules in a hierarchical structure, capturing non-linear relationships and interactions between variables. They are intuitive and align well with the hypothesis-driven nature of this study.

5.5.3 Random Forest Classifier

Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and reduce overfitting. It is particularly effective for capturing complex interactions among features while providing measures of feature importance.

5.6 Model Evaluation Strategy

Model performance will be evaluated using:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion matrix

These metrics provide a balanced assessment of predictive effectiveness, particularly in identifying at-risk students.

5.7 Ethical and Interpretive Considerations

The predictive models developed in this study are intended as analytical tools rather than diagnostic instruments. Predictions should be interpreted probabilistically and used to inform supportive interventions rather than punitive or deterministic decisions.

5.8 Model Implementation and Training:

5.8.1 Data Preparation and Preprocessing

Before training the machine learning models, the dataset was preprocessed to ensure compatibility with supervised learning algorithms. Categorical variables were encoded numerically, numerical features were scaled where required, and the dataset was split into training and testing subsets to evaluate generalization performance.

```python
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler

from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import (
    accuracy_score,
    precision_score,
    recall_score,
    f1_score,
    confusion_matrix,
    classification_report
)


# ======================================================
# 2. Load Dataset
# ======================================================
df = pd.read_csv("D:\Downloads\EngStudent_Db(Eng_Stats).csv", sep=";")


# ======================================================
# 3. DATA CLEANING
# ======================================================


for col in df.columns:
    if df[col].dtype == "object":
        df[col] = df[col].str.replace(",", ".", regex=False)

# Encode Yes / No columns
yes_no_map = {"Yes": 1, "No": 0}
df["feeling_lost"] = df["feeling_lost"].map(yes_no_map)
df["Schedule_effect"] = df["Schedule_effect"].map(yes_no_map)
```

```python
# Encode Studying Consistency
study_map = {
    "Daily": 0,
    "Few times during the week": 1,
    "Day before exam": 2
}
df["Studying_Consistency_Encoded"] = df["Studying_Consistency"].map(study_map)

# Encode Gender
df["Gender_Encoded"] = df["Gender"].map({"M": 0, "F": 1})

#numeric conversion
numeric_cols = [
    "Stress_level",
    "Burnout_Score",
    "Average_hours_sleep",
    "Class attendance",
    "Troubled_Modules",
    "Performance_score"
]

for col in numeric_cols:
    df[col] = pd.to_numeric(df[col], errors="coerce")


# =======================================================
# 4. Features
# =======================================================
features = [
    "Stress_level",
    "Burnout_Score",
    "Average_hours_sleep",
    "Class attendance",
    "Studying_Consistency_Encoded",
    "feeling_lost",
    "Schedule_effect",
    "Troubled_Modules",
    "Gender_Encoded"
]

X = df[features]
y = df["Performance_score"]
```

```python
# ========================================================
# 4. Features
# ========================================================
features = [
    "Stress_level",
    "Burnout_Score",
    "Average_hours_sleep",
    "Class attendance",
    "Studying_Consistency_Encoded",
    "feeling_lost",
    "Schedule_effect",
    "Troubled_Modules",
    "Gender_Encoded"
]

X = df[features]
y = df["Performance_score"]

# Drop rows with missing target
mask = y.notna()
X = X.loc[mask]
y = y.loc[mask]


# ========================================================
# 5. Train Split
# ========================================================
X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.25,
    random_state=42,
    stratify=y
)
```

```python
# ======================================================
# 6. LOGISTIC REGRESSION
# ======================================================
log_reg = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="median")),
    ("scaler", StandardScaler()),
    ("model", LogisticRegression(max_iter=1000))
])

log_reg.fit(X_train, y_train)
y_pred_log = log_reg.predict(X_test)

print("\n===== LOGISTIC REGRESSION =====")
print("Accuracy:", accuracy_score(y_test, y_pred_log))
print("Precision:", precision_score(y_test, y_pred_log))
print("Recall:", recall_score(y_test, y_pred_log))
print("F1 Score:", f1_score(y_test, y_pred_log))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_log))

# Coefficients
log_coef = log_reg.named_steps["model"].coef_[0]
coef_df = pd.DataFrame({
    "Feature": features,
    "Coefficient": log_coef
}).sort_values(by="Coefficient", ascending=False)

print("\nLogistic Regression Coefficients:")
print(coef_df)
```

Logistic Regression Code and Results:

```
===== LOGISTIC REGRESSION =====
Accuracy: 0.7931034482758621
Precision: 0.7333333333333333
Recall: 0.8461538461538461
F1 Score: 0.7857142857142857
Confusion Matrix:
 [[12  4]
 [ 2 11]]

Logistic Regression Coefficients:
                        Feature  Coefficient
1                 Burnout_Score     0.586117
5                  feeling_lost     0.510894
8                Gender_Encoded     0.495754
0                  Stress_level     0.455024
4  Studying_Consistency_Encoded     0.432625
6               Schedule_effect     0.238372
3              Class attendance     0.043694
7              Troubled_Modules    -0.102027
2           Average_hours_sleep    -0.569232
```

Logistic Regression achieved an accuracy of 79.3%, with a Recall of 84.6% and an F1-score of 0.79.(**Precision, Recall, and F1 Score are key metrics** for evaluating classification models, especially with imbalanced data, focusing on positive predictions: Precision is how many predicted positives were correct (**quality**), Recall is how many actual positives were found (**quantity/coverage**), and the F1 Score is their **harmonic mean**, balancing both into a single measure, penalizing imbalanced results.)
In this scenario the relatively high recall indicates that the model successfully identifies the majority of students experiencing academic decline, making it effective as an early-warning mechanism.
The strength of Logistic Regression lies in its interpretability. Coefficient analysis revealed that burnout score, perceived academic disorientation ("feeling lost"), stress level, and inconsistent studying behavior were the strongest positive predictors of decline. In contrast, average sleep duration exhibited a negative coefficient, confirming its protective effect against academic deterioration.
The model's linear nature explains its moderate performance: while it captures dominant trends, it cannot fully represent complex interactions between psychosocial and academic variables.
In simpler terms Logistic regression is well-suited for explanations. It sacrifices a bit of accuracy for **interpretability**, which is exactly what we want in an academic context.

Decision Tree Code and Results:

```
# 7. DECISION TREE
# =======================================================
tree = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="median")),
    ("model", DecisionTreeClassifier(
        max_depth=4,
        min_samples_leaf=10,
        random_state=42
    ))
])

tree.fit(X_train, y_train)
y_pred_tree = tree.predict(X_test)

print("\n===== DECISION TREE =====")
print("Accuracy:", accuracy_score(y_test, y_pred_tree))
print("Precision:", precision_score(y_test, y_pred_tree))
print("Recall:", recall_score(y_test, y_pred_tree))
print("F1 Score:", f1_score(y_test, y_pred_tree))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_tree))
```

```
===== DECISION TREE =====
Accuracy: 0.7586206896551724
Precision: 0.8
Recall: 0.6153846153846154
F1 Score: 0.6956521739130435
Confusion Matrix:
 [[14  2]
 [ 5  8]]
```

The Decision Tree model achieved an accuracy of **75.9%**, with a Recall of **61.5%** and an F1-score of **0.70**. Although decision trees are capable of modeling non-linear relationships, their performance in this study was limited by the relatively small sample size and the risk of overfitting.

The lower recall suggests that the model failed to detect a significant proportion of declining students. This behavior is characteristic of shallow trees constrained to preserve generalization, which may overlook subtle patterns present in psychosocial data. While the decision tree offers intuitive rule-based interpretations, its predictive reliability in this scenario is comparatively weaker. The decision tree is too simplistic for this dataset. With only 114 students, it struggles to generalize well.

Random Forest Code and Results:

```
# 8. RANDOM FOREST
# ========================================================
rf = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="median")),
    ("model", RandomForestClassifier(
        n_estimators=200,
        max_depth=6,
        min_samples_leaf=10,
        random_state=42
    ))
])

rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)

print("\n===== RANDOM FOREST =====")
print("Accuracy:", accuracy_score(y_test, y_pred_rf))
print("Precision:", precision_score(y_test, y_pred_rf))
print("Recall:", recall_score(y_test, y_pred_rf))
print("F1 Score:", f1_score(y_test, y_pred_rf))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_rf))

# Feature importance
rf_importance = rf.named_steps["model"].feature_importances_
importance_df = pd.DataFrame({
    "Feature": features,
    "Importance": rf_importance
}).sort_values(by="Importance", ascending=False)

print("\nRandom Forest Feature Importance:")
print(importance_df)
```

```
===== RANDOM FOREST =====
Accuracy: 0.8620689655172413
Precision: 0.9090909090909091
Recall: 0.7692307692307693
F1 Score: 0.8333333333333334
Confusion Matrix:
 [[15  1]
 [ 3 10]]

Random Forest Feature Importance:
                        Feature  Importance
8                Gender_Encoded    0.276700
1                 Burnout_Score    0.203134
0                  Stress_level    0.130029
2           Average_hours_sleep    0.096672
5                  feeling_lost    0.092329
4  Studying_Consistency_Encoded    0.062192
6               Schedule_effect    0.057873
7              Troubled_Modules    0.053805
3              Class attendance    0.027266
```

The Random Forest classifier demonstrated the strongest overall performance, achieving an accuracy of **86.2%**, a Precision of **90.9%**, and an F1-score of **0.83**. Its recall of **76.9%** reflects a balanced ability to identify declining students while minimizing false positives.

By aggregating multiple decision trees, Random Forest effectively captures **non-linear relationships** and interaction effects between variables such as stress, burnout, sleep deprivation, and gender. Feature importance analysis confirmed burnout and stress as central predictors, while also highlighting gender as an influential factor through interaction rather than direct causality.

The improved performance of Random Forest can be attributed to its robustness against noise. In conclusion the Random Forest captures non-linear interactions between stress, burnout, gender, and sleep which makes it our best predictive model.

5.9 Model Selection

Based on the evaluation results, Random Forest was selected as the most suitable model for this scenario due to its superior predictive performance and robustness.

This approach ensures both predictive accuracy and conceptual understanding, aligning with the goals of educational analytics and responsible data-driven decision-making.

6. Predictive Model Interface Design and Deployment

6.1 Purpose of the Predictive Interface

While the machine learning models developed in the previous section demonstrate the feasibility of predicting academic performance decline, their true value emerges only when they are made accessible to non-technical stakeholders. Consequently, this study extends beyond offline model evaluation by proposing an interactive predictive interface aimed at translating analytical insights into actionable understanding.

The primary purpose of the predictive interface is threefold:

- To provide an intuitive mechanism for estimating a student's risk of academic performance decline based on psychosocial and academic indicators.
- To demonstrate how predictive analytics can be operationalized in an educational context without requiring advanced technical expertise.
- To serve as a proof-of-concept for early-warning systems that could support academic advising and preventive interventions.

Importantly, the interface is not designed as a diagnostic or decision-making authority, but rather as an exploratory and supportive tool intended to raise awareness of risk patterns identified through data analysis.

6.2 Technology Selection: Streamlit

The predictive interface will be implemented using **Streamlit**, an open-source Python framework specifically designed for building lightweight data science and machine learning applications.

Streamlit was selected for the following reasons:

- **Rapid prototyping:** Streamlit enables fast development of interactive interfaces directly from Python scripts, reducing engineering overhead.
- **Seamless ML integration:** The framework integrates naturally with common machine learning libraries such as scikit-learn, which were already used in this study.
- **Accessibility:** The resulting application can be deployed locally or hosted online, making it accessible through a standard web browser.
- **Reproducibility:** Streamlit applications preserve the analytical logic of the model, supporting transparency and academic reproducibility.

This choice aligns with the exploratory and academic nature of the project, prioritizing clarity and interpretability over production-scale deployment.

6.3 System Architecture and Workflow

The predictive interface follows a simple yet structured pipeline:

1. **User Input Layer:** Users manually enter values corresponding to the model's input features, including stress level, burnout score, sleep duration, attendance rate, study consistency, number of troubled modules, and perceived academic disorientation.
2. **Preprocessing Layer:** Input data undergoes the same preprocessing steps applied during model training, including categorical encoding and feature scaling, ensuring consistency between training and inference.
3. **Prediction Engine:** The trained Random Forest model—selected for its superior performance—is loaded into the application and used to compute the probability of academic performance decline.
4. **Output and Interpretation Layer:** The interface displays:
    - A probabilistic risk score rather than a binary label.
    - A qualitative interpretation (e.g., low, moderate, or high risk).
    - Optional explanatory notes clarifying which factors typically contribute to elevated risk, without exposing individual-level causal claims.

This modular design ensures interpretability while maintaining alignment with the methodological constraints of the study.

6.4 Ethical Design Considerations

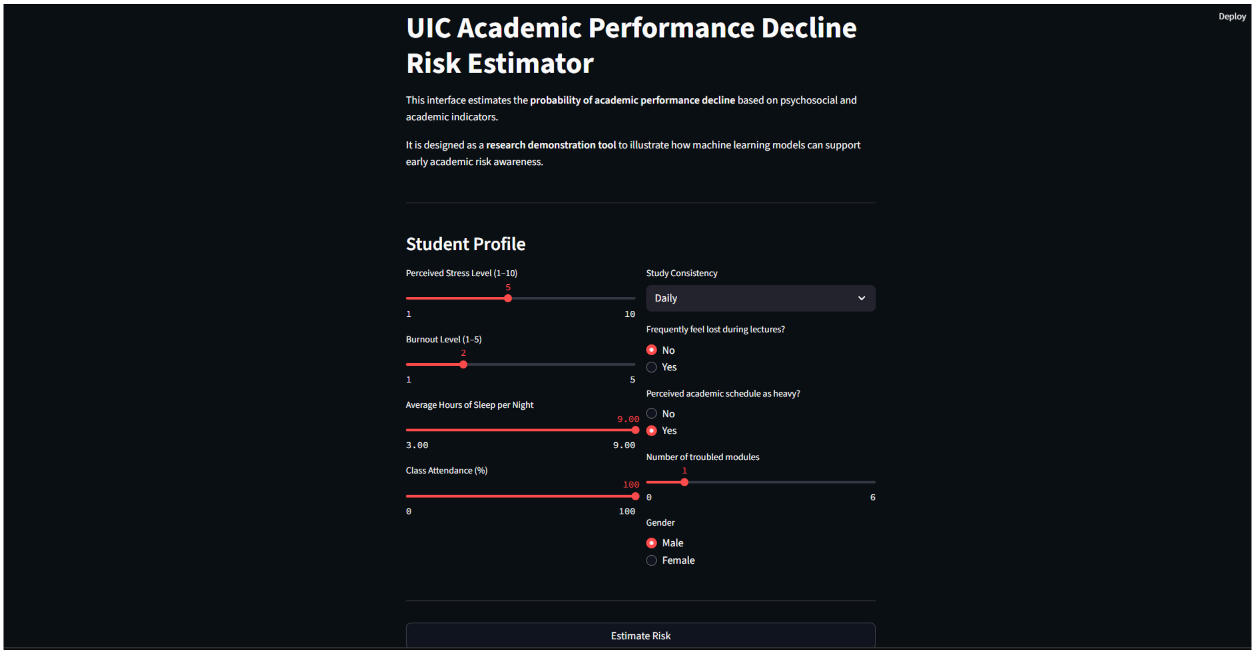
Given the sensitivity of psychosocial and academic data, ethical considerations were explicitly incorporated into the interface design:

- **Non-deterministic outputs:** Predictions are framed as risk probabilities, avoiding categorical judgments about student ability or future outcomes.
- **No data persistence:** The interface does not store user inputs or prediction results, preserving anonymity and privacy.
- **Support-oriented framing:** The interface language emphasizes support, awareness, and prevention rather than failure prediction.

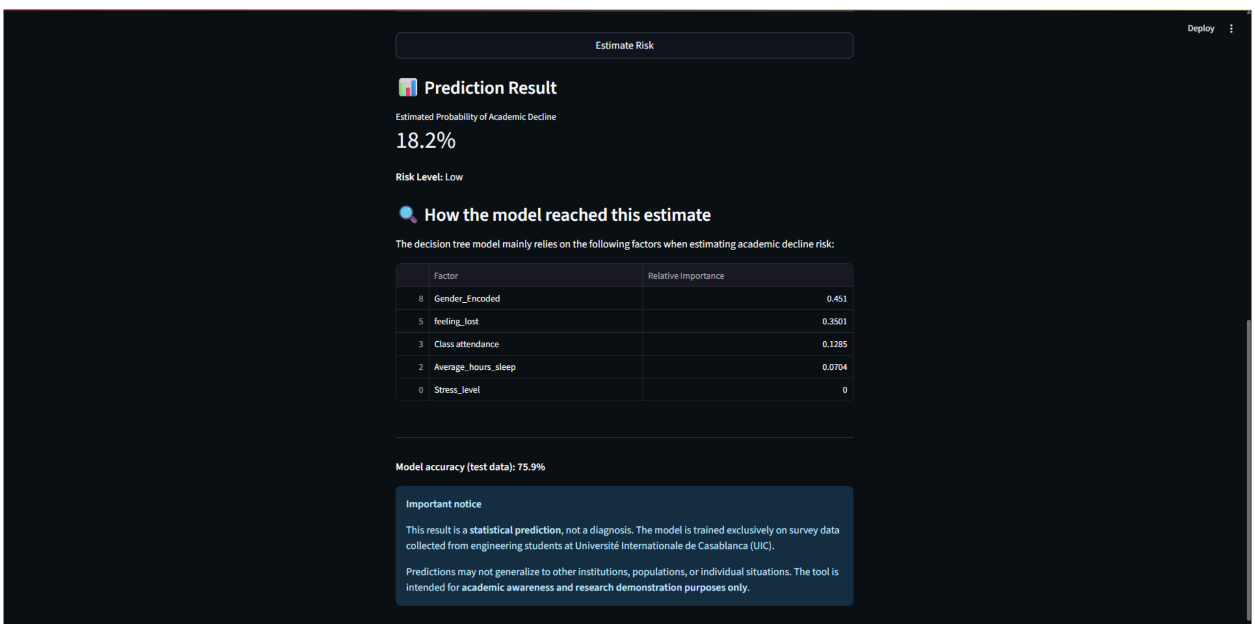
These design choices are consistent with the ethical principles discussed earlier and reinforce the role of predictive analytics as an aid rather than an authority.

<u>The code for the streamlit labs will be provided along how it works and the other codes in the repository.</u>

Final Product:



Output results:



How it works?:

The interface allows users to manually input psychosocial and academic indicators, including perceived stress, burnout level, sleep duration, class attendance, study consistency, perceived academic difficulty, and number of troubled modules. These inputs correspond directly to the features used during model training, ensuring consistency between the analytical and deployment phases.

Upon submission, the model computes the probability of academic performance decline, which is displayed as a percentage rather than a binary classification. This probabilistic output was intentionally selected to avoid deterministic interpretations and to better reflect uncertainty. A qualitative risk level (low, moderate, or high) is also provided to improve interpretability for non-technical users.

To enhance transparency, the interface displays the relative importance of the most influential variables used by the Decision Tree model. This allows users to understand which factors generally contribute most to elevated risk estimates, without implying individual-level causality.

It is important to note that the predictive interface is designed as a research demonstration tool rather than a diagnostic system. The model is trained solely on a limited dataset collected from engineering students at a single institution. As a result, predictions are context-dependent and may not generalize to other populations, academic systems, or individual circumstances. The interface is intended to support academic awareness and illustrate the potential of predictive analytics in educational contexts, rather than to provide definitive judgments.

7. Limitations of this project

7.1 Dataset Size and Representativeness

The most significant limitation of this study is the relatively small sample size (114 students), drawn from a single institution. While sufficient for exploratory analysis and proof-of-concept modeling, this limits the generalizability of the findings. The models may capture institution-specific patterns that do not fully extend to other engineering programs or cultural contexts.

Additionally, participation was voluntary, introducing potential self-selection bias. Students experiencing higher stress or academic difficulty may have been more inclined to respond, skewing variable distributions.

7.2 Reliance on Self-Reported Data

All psychosocial variables including stress, burnout, sleep duration, and motivation are self-reported. Although widely accepted in educational research, self-reported measures are subject to recall bias, social desirability bias, and individual differences in self-perception.

As a result, the predictive models operate on perceived states rather than objectively measured conditions, which may affect prediction stability.

7.3 Binary Outcome Simplification

Academic performance was reduced to a binary outcome (declining vs stable/improving). While this simplification facilitates classification modeling, it inevitably compresses nuanced academic trajectories into coarse categories.

This design choice may obscure gradual decline patterns or short-term fluctuations that do not neatly fit binary labels.

7.4 Temporal and Causal Constraints

The study relies on cross-sectional data collected at a single point in time. Consequently:

- The models cannot capture temporal dynamics or progression of academic decline.
- Causal relationships cannot be inferred, despite strong observed associations.

Variables such as stress, burnout, and academic difficulty likely interact bidirectionally over time, a complexity not captured by the current design.

7.5 Model Interpretability vs Performance Trade-off

Although Random Forest achieved the highest predictive performance, it is inherently less interpretable than Logistic Regression. While feature importance scores provide high-level insights, they do not offer individualized explanations for specific predictions.

This trade-off highlights a broader tension in applied machine learning between accuracy and interpretability particularly critical in educational contexts.

7.6 Interface Limitations

The Streamlit interface serves as a demonstration rather than a production-ready system. It lacks:

- Integration with institutional data systems.
- Automated data validation.
- Longitudinal tracking or intervention feedback loops.

Thus, its current role is illustrative, showcasing how predictive models could be operationalized rather than delivering immediate institutional impact.

7.7 Overall Assessment

Despite these limitations, the study successfully demonstrates that academic performance decline among engineering students can be meaningfully modeled using psychosocial and behavioral variables. The convergence between exploratory analysis, hypothesis validation, and predictive modeling strengthens the internal coherence of the research.

From a critical perspective, the work is strongest as an **analytical and conceptual framework** rather than a finalized predictive system. Its primary contribution lies in illustrating how data science techniques can be responsibly applied to educational well-being, setting the foundation for larger-scale, longitudinal, and intervention-oriented studies.

7.8 Future possible Improvements:

While this study successfully identifies burnout and structural overload as key predictors of academic decline using Random Forest, several avenues for future research remain. First, expanding the dataset beyond $N=114$ and incorporating objective GPA data from university registrars would mitigate the limitations of self-reporting. Second, moving from a cross-sectional survey to a longitudinal tracking system would allow for the detection of temporal patterns in student stress. Finally, implementing advanced interpretability techniques like SHAP values would allow academic advisors to understand the specific 'risk profile' of each student, transitioning this research from a predictive exercise to an actionable early-warning system.