

# COMP3125 Individual Project

---

Jacob Kabongo  
Department of Computer Science

## Abstract

This project investigates factors influencing procrastination among students. Using a custom dataset collected via surveys, we analyze how variables like workload, stress, time management, and academic major impact procrastination. Statistical tests and machine learning models are applied to uncover patterns, determine relationships, and build predictive tools to help students manage and overcome procrastination.

Keywords—Procrastination, Students, Stress, Workload, Time Management, Machine Learning

## I. Introduction

Procrastination is the intentional delay of tasks despite knowing the negative consequences. Among students, it is often driven by factors like stress, heavy workload, poor time management, and lack of motivation. This project explores these determinants and seeks to provide practical recommendations to reduce procrastination. We aim to answer four central questions using a mix of qualitative and quantitative analysis.

## II. Datasets

### A. Source of dataset

The dataset was created by surveying students. The survey collected responses about workload, stress levels, time management skills, procrastination frequency, academic major, and year level.

### B. Character of the datasets

- Features: Courses taken, hours on assignments, stress levels (1-10), use of planners, missed deadlines, procrastination behaviors, academic major, year level.
- Format: CSV file with structured responses.
- Preprocessing: Cleaned missing data, encoded categorical variables, normalized numerical fields.

## III. Methodology

Key Questions

1. What are the most common reasons for procrastination among students? (Qualitative)
2. What is the relationship between workload and procrastination? (Quantitative)
3. Can we predict procrastination based on stress levels and time management skills? (Machine Learning)
4. Are there differences in procrastination between majors or year levels? (Statistical Testing)

### Model Selection

For prediction, we used a Random Forest Classifier from Scikit-learn.

### Steps Taken

- Preprocessed data: imputed missing values, encoded variables.
- Split data into training/test sets.
- Trained Random Forest model to classify procrastination levels (low, medium, high).
- Evaluated using accuracy, precision, recall, and F1-score.
- Analyzed feature importance.

## IV. Results

The model showed promising accuracy with balanced classification across procrastination levels. Feature importance analysis revealed that lack of motivation, distractions, and poor time management were the strongest predictors.

## V. Discussion

One limitation is the small dataset size which may reduce generalizability. Also, using only binary features may not capture the complexity of human behavior. Future improvements include expanding the dataset with more student responses, integrating more features like GPA or digital distractions, and testing alternative models like logistic regression or neural networks.

## VI. Conclusion

This project demonstrates that machine learning can be effectively used to analyze and predict procrastination behavior among students. By identifying key factors like motivation and time management, we can provide targeted recommendations and interventions for students.

## ACKNOWLEDGMENT

Jacob Kabongo thanks the course instructor and peers for feedback and support during the project.

## REFERENCES

- [1] Scikit-learn Developers, 'Scikit-learn: Machine Learning in Python', <https://scikit-learn.org/>
- [2] Wikipedia, 'IMRAD format', <https://en.wikipedia.org/wiki/IMRAD>
- [3] Google Colab, <https://colab.research.google.com>