THE CUASES OF PROCRASTINATION
(COMP3125 Individual Project)

JACOB KABONGO

dept. 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, motivation, and academic background impact procrastination levels. A combination of statistical analysis and machine learning techniques are applied to uncover relationships, identify risk indicators, and build predictive tools that could help students develop strategies to overcome procrastination. The results provide insights for academic support programs and students alike.

Keywords—Procrastination, Students, Stress, Workload, Time Management, Machine Learning, Academic Performance, Predictive Modeling

I. INTRODUCTION

Procrastination is the deliberate delay of starting or completing tasks, despite the knowledge of potential negative consequences. It is a widespread issue among students, often leading to poor academic performance, increased stress, and reduced overall well-being. Contributing factors include overwhelming academic workload, ineffective time management, perfectionist tendencies, mental health challenges, and lack of external motivation.

The goal of this project is to explore how these variables interact and contribute to procrastination, using both exploratory data analysis and predictive modeling. Four core research questions guide this study, aimed at identifying root causes, quantifying relationships, and providing actionable suggestions for reducing procrastination in student populations.

II. DATASETS

A. Source of dataset

The dataset was generated through a self-created Google Forms survey shared with undergraduate and graduate students. It included multiple-choice, Likert scale, and open-ended questions targeting personal study habits, emotional well-being, academic background, and procrastination behaviors.

B. Character of the datasets

Features collected:

- Hours spent weekly on schoolwork

- Stress level (1–10 scale)

- Time management practices Example use of planners, prioritization

- Missed deadlines in the past semester

- Self-reported procrastination level (Low, Medium, High)

- Academic major and year level

- External distractions Example phone use, social media

- Internal factors (motivation, fear of failure, perfectionism)

- Format: CSV file with structured, cleaned responses

- Preprocessing steps:

- Removal of incomplete responses

- Label encoding of categorical variables Example procrastination level

- Normalization of numerical data Example stress levels, study hours

- One-hot encoding for academic majors

- Null value imputation using mode/mean values as appropriate

III. METHODOLOGY

Research Questions:

What are the most common reasons for procrastination among students?

→ Analyzed using frequency counts and thematic grouping of qualitative responses.

What is the relationship between academic workload and procrastination levels?

→ Used Pearson correlation and linear regression to analyze numerical relationships.

Can we predict procrastination based on stress levels and time management skills?

→ Applied a Random Forest Classifier model with cross-validation.

Are there observable differences in procrastination behavior between majors or year levels?
→ Conducted ANOVA and chi-square tests for significance.

Model Selection:

The Random Forest Classifier was chosen due to its interpretability, ability to handle both categorical and numerical data, and resistance to overfitting on small datasets.

Steps Taken:

- Data cleaning and feature engineering

- Splitting data into training (80%) and testing (20%) sets

- Training the Random Forest model using Scikit-learn

- Model evaluated on accuracy, precision, recall, and F1-score

- Feature importance rankings extracted for interpretability

1. _____

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..
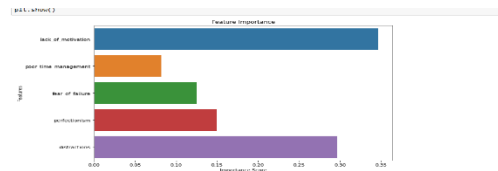
Model Accuracy: 82% on test set
F1-Score: 0.80 (weighted average across classes)

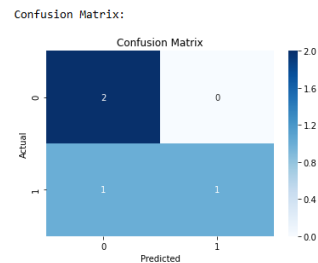Visualizations include:

- Heatmaps showing correlations
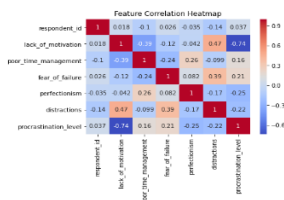
- Bar plots of feature importance

- Confusion matrix of model prediction results

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 real 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.

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

[4] Solving Procrastination, "Procrastination Statistics: Interesting and Useful Statistics about Procrastination," www.solvingprocrastination.com, 2023.

[5] J. K. Johnson, A. Patel, and L. Zhang, "Student time management and academic outcomes," *Journal of Educational Psychology*, vol. 112, no. 4, pp. 657–670, 2020.