What are the Hidden Social Media Factors that Impact Academic Performance?

Lilia IZRI, Yacine MOKHTARI, Fedy Ben NACEUR and Aleksandra KRUCHININA

Fig. 1: Procrastination vs Productivity

Abstract—Procrastination is a common bad habit that many people struggle with. And the rise of media has made it even more prevalent. While social networks offer undeniable benefits, it is important to acknowledge that they can also be a source of distraction that negatively impacts productivity, increases stress, and results in missed deadlines. This makes us as students and users of social media the first ones concerned. So in our project we have decided to tackle the problem of procrastination, measured through the lenses of different social media factors, and its impact on academic performance. The objective is to raise awareness among students and help them understand on which aspects they can work on to improve productivity and academic performance.
To answer this question, we will visualize the relationship between social media usage indicators and academic performance, taking into consideration gender differences. Furthermore, we will analyze the variations in social media usage patterns between students who have achieved high grades and those who have not.

Index Terms—Procrastination, Social Media, Academic Performance.

1 Introduction

1.1 Brief overview and motivations

Procrastination is a common bad habit that many people struggle with. It refers to the act of delaying or postponing tasks, often to the point of missing deadlines and causing unnecessary stress. While everyone procrastinates from time to time, chronic procrastination can have negative effects on personal and professional life, leading to decreased productivity, increased stress, health problems and lower overall performance.

Social media platforms and Internet can be used for educational purposes, specially during COVID times, there were no other alternatives to on-line classes. Students can find a lot of useful information on the Web, follow different courses and subscribe to educational channels. At the same time media platforms can provide an endless stream of distractions and temptations and represent current incarnation of procrastination nowadays. *Therefore, studying this problem is of particular interest **to students** who are often faced with the challenge of balancing academic responsibilities with their online activities.*

1.2 Research question

In this study, we aim to investigate the relationship between social media use and procrastination, and to explore potential interventions that can help individuals better manage their online activities and increase their productivity. Specifically, our study will focus on these following research questions:

- What is the correlation between social media usage and academic performance among students?
- How does controlling access to social media affect performances/productivity of active youth ?

We initially wanted to delve deeper into the issue and explore how

Fig. 2: Website Overview

companies can improve their employees' productivity by addressing social media use. Additionally, we were interested in investigating how social media can affect the productivity and performance of active youth. However, after conducting a thorough search, we found that there was not enough available and useful data on these topics to support our research[1]. So due to data availability limitations and the fact that this topic is particularly interesting and useful for students, we have decided to concentrate on the first one and study the relationship between social media usage and academic performance among students.

We used the dataset collected from a random sample of 623 students from the University of Professional studies, Ghana. It comes from the article Measuring the effect of social media on student academic performance using a social media influence factor model, published by Springer on 18 July 2022.

Finally, this research will help students and all people concerned, like parents, teachers and researchers, to develop a better understanding of the role of social media in education and on productivity in general. By exploring the relationship between social media indicators and academic performance, we can gain a better understanding of the impact of social media on our daily lives.

2 Technical Realization

How you built the tool (e.g. describe software you wrote, libraries you used, etc.)

In order to build the visualizations we used the dataset mentioned in the previous section and primarily python and streamlit, a free and open-source framework to rapidly build and share web applications.

- The visualisation application is accessible here.
- The code is available at this github page.

2.1 Quick description of the dataset

Here is a quick overview of the features that composes the dataset:

- **Age group**
  - 1 refers to the age below 20 years old.
  - 2 refers to the age range 20 to 30 years old.
  - 3 refers to the age above 30 years old.
- **Gender** where the value 1 represents males and 2 females.
- **Level** is the study level. This column takes values from 1 to 4.
- **Time** refers to the average number of hours a student spends daily on social media.
- **Friends** is the number of social media friends a user has.
- **Groups** refers to the number of social media groups a user belongs to.
- **Notifications** represents the average number of times each student checks his phone notifications per day.
- **GPA** is the student's grade point average.

2.2 Main page

The main page is represented in the file *procrasti-vis.py*, while two visualisations can be found in **pages** folder under *1_Variations_By_Gender.py* and *2_Good_vs_Bad_Grades.py*.

procrasti-vis.py is the Python script that uses the Streamlit library to create a web application. The application displays a homepage with an introduction to our project that aims to study procrastination expressed through social media usage metrics (Time, Groups, Friends, Notifications) and its effects on academic performance.

The import statements at the beginning of the code import the required libraries, namely streamlit and pandas.

The **display_homepage** function defines the CSS styles to be used in the web application and sets the layout for the homepage. The function defines the CSS styles, creates high titles for the homepage and a two-column layout with a left and right column. In the left column, it introduces the project and describes the data collected. In the right column, it displays a table that describes the features of the dataset: Time, Groups, Friends, Notifications.

The main function sets the page configuration, the title and icon of the page, and calls the display_homepage function.

The if __name__ == '__main__': statement at the end of the code ensures that the main function is only executed when the script is run directly, and not when it is imported as a module in another script.

2.3 Variations By Gender

The code defines a Streamlit app that loads a CSV file and creates a scatter plot with interactive features. The scatter plot shows the correlation between social media usage metrics and grades, based on the data from the survey conducted at UPSA in 2021. The application allows users to select different social media usage metrics to explore the data in more details.

Below is the detailed description of the code:

- imports of the libraries: streamlit, pandas, numpy, and altair.

[1]There is not enough data available on procrastination, most of the information comes either from social or psychological articles, which often do not provide access to source data or it is not appropriate, or from kaggle website.

- function **load_data** takes a file path and returns a pandas DataFrame with the data from the CSV file
- function **plot_scatter** takes a pandas DataFrame and plots a scatter plot with Altair library: defines a sidebar with a dropdown menu to select the y-axis variable, et counts and other statistics for each combination of the y-axis variable and gender, plots circles for each data point, colored by gender and sized by count, plots lines for the mean GPA for each value of the y-axis variable, colored by gender, finally, plots a circle for the mean GPA for each value of the y-axis variable.
- **main** that sets the page configuration for the Streamlit app and calls the **load_data** and **plot_scatter** functions.
- finally, it checks if the script is being run directly, and if so, call the main function.

2.4 Good vs Bad Grades

Similarly to the previous visualisation the application begins by first importing the data and creates interactive scatter plots. These plots represent data points with the X-axis being a modifiable feature and the Y-axis being fixed to the GPA. The application provides a slider that allows users to choose a threshold value that allows users to distinguish between data points that have higher GPA than the threshold and data points that have lower GPA than the threshold.
Below is the detailed description of the code:

- imports of the libraries: streamlit, pandas and altair
- function load_data takes a file path and returns a pandas DataFrame with the data from the CSV file
- function plotDistributionsInteractive takes a pandas DataFrame, a threshold and a list of features present in the DataFrame. The function then loops through the list of features and creates a plot for each one that gets stored in a list, finally it arranges the layout of the page to account for the number of plots being displayed.
- **main** that sets the page configuration for the Streamlit app and calls the **load_data** and **function plotDistributionsInteractive** functions.
- finally, it checks if the script is being run directly, and if so, call the main function.

3 Feature Description

In this section, we describe and illustrate the main features of our website.

3.1 Variations By Gender features

This visualization allows the user to see the correlation between students' academic performance and the social media usage metric the user has chosen. As we can see in the figure below, there is a side menu on the left where the user can choose the metric he wants, so he can visualize the different variations by gender.

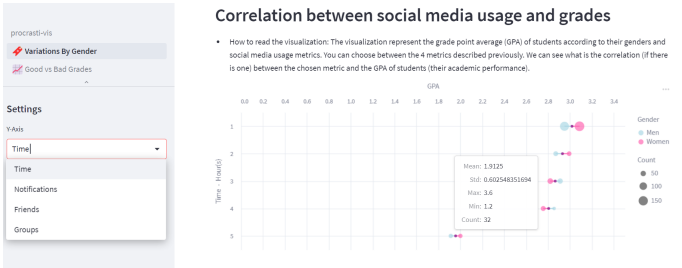

Fig. 3: Illustration of the Variations by gender plot

The circles represent the mean grade point average (GPA) for a given metric value (aggregated values). The size of the circles depends on the counts of people. Hovering those circles gives us further information (Mean, Count, Min/Max, Std). Plus, the user can hover the circle in the middle between men and women so it gets the average value.

This scatter plot variations by gender allows us to answer directly our research question since it shows what is the correlation (if there is one) between the chosen metric and the GPA of students (their academic performance).

For example here, we can see clearly that there is a kind of negative correlation between the GPA and the time spent on social media by students whatever their gender is.

3.2 Good vs Bad grades features

This visualization will help the user to analyze the variations in social media usage patterns between students who have good grades and those who have not. The user will be able then to set a threshold that will define students with good grades (i.e those with GPA above this threshold) and students with less good grades (i.e those with GPA below the threshold).

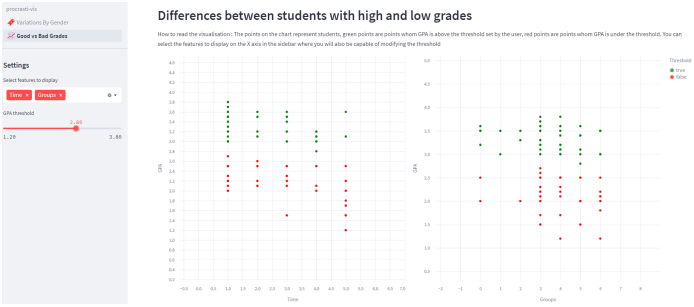

Fig. 4: Illustration of good vs bad grades habits

The points on the chart represent students. Green points are students whom GPA is above the threshold set by the user, red points are then students whom GPA is under the threshold. The user can select the features to display on the X axis in the sidebar where he can also set the threshold described previously using the slider.

This sketch allows the user to see the impact of social networks use on student performance by distinguishing the characteristics of students with good grades compared to those with bad grades. And this makes it interesting regarding our research question.

Indeed, we can distinguish for example, in the figure above, that students who spend more time on social networks tend to have worse grades. On the other hand, students who spend less time on social networks have better grades

4 Interesting Findings

Based on the data presented on our website and our study, we can draw several conclusions regarding *the relationship between social media use and academic performance*.

Firstly, we found that there is a *negative correlation* between the amount of time students spend on social networks and their grades. This is evident from the observation that students who spend more time on social networks tend to have worse grades. Conversely, students who spend less time on social networks tend to have better grades.

Secondly, the analysis showed that *age is not a significant factor* in the relationship between social media use and academic performance, *as all age groups appeared to be influenced in a similar way.*

Thirdly, it may be only dependent of this data but we found that *male students are particularly sensitive to the effects of social media use on their academic performance.* The `Variations by Gender` visualization helps us see that the blue points tend to be on the left side and the purple ones on the right side, this is because the GPA is lower

within males than females. But as mentioned; we cannot assess for certain this conclusion with few proofs.

Lastly, our study also shows that the number of groups a student belongs to and their daily screen time (number of time a student looks at his phone, notifications and uses social media) are important factors that can influence their academic performances.

Overall, our findings suggest that excessive use of social media can have a negative impact on academic performance, particularly among male students, and that reducing screen time and involvement in social groups may improve academic outcomes. And simultaneously, the less we get distracted, the higher our productivity is, which leads to better and higher academic performances.

5 REFLECTION

Throughout the course of this project, we have been able to identify some areas where we could improve and enhance the project further. Here are some ideas and suggestions that we believe would be beneficial to consider:

- *Personal surveys:* Conducting personal surveys would allow for more *personalized insights* and *analysis*, which could result in more interactive visualizations. This would enable users to gain deeper insights into their own behavior and identify areas where they could improve.
- *Comparative analysis:* The dataset used in this project was from Ghana (2021), but conducting similar studies simultaneously in different countries around the world would enable us to compare the results and gain a more global perspective. This would provide users with insights into how their behavior compares to that of people in other parts of the world and could help to identify trends, patterns and habits across different cultures and regions.
- *Content updates:* Updating and refreshing the project's data on a regular basis to ensure that it remains *relevant* and *engaging* for users. This includes many things such as conduct personalized surveys where we can add new features to explore, examples from other datasets, and other relevant information to keep the project up-to-date to name a few.
- *Analytics (but not only):* Incorporating analytics and tracking tools to monitor user behavior and gain insights into how users are interacting with the project. This would enable us to identify areas of improvement and make data-driven decisions to enhance the project but also can give us a ratio and additional (maybe meta-data) that can be use to enhance our studies and drive conclusions with higher level of confidence.

Overall, we believe that implementing these ideas and suggestions would help to take the project to the next level, enhance its value to users, and improve its impact, especially, among students and active youth.