BRSM PROJECT PROPOSAL

Team: NSTW

N V Raghavendra (2022702010)

Sreejan Patel (2020101084)

Haasa G (2020101135)

Project Title: Leveraging Smartphone Data to understand Student Behavior and Academic Performance.

Dataset description:

Student Life Study is a study performed at Dartmouth university to assess different characteristics of a student's life using smartphone data. The following is a summarized description of the data.

- Number of students: 48
- Time duration over which data was recorded: 10 weeks
- Factors assessed: Mental Health (Depression, Stress levels, etc.), Academic performance (GPA), behavioral trends (sleep, exercise, etc.).

The aim of the study was to understand more about what is going on in students' lives, which usually stays hidden. It was done using a smartphone app which collected data from sensors on the phones along with well known surveys. Computational methods, paired with machine learning algorithms were used to make higher level inferences about sleep, sociability, activity, etc. The following links pertain to the dataset and its detailed description by the study makers.

- Link to Dataset
- Link to Detailed Description

Reason for choosing this dataset:

As students, we understand the importance of knowing how fellow students are coping with the stress and strain arising from academic life. Many students fail to understand this importance which leads them to not making the best out of their college lives. The dataset provides many different perspectives of the events in a student's life over a single term and also provides insights about their reactions to certain events. It gives a detailed deep dive into each student's behavioral trends, activities, academic performance, etc. all of which are interlinked to each other.

Methodology adopted:

Our methodology is split into different activities as described below. Each activity has its own set of sub-activities to be performed.

Data preprocessing:

1. Identifying necessary data to be studied: The first step of preprocessing would be to go through the data and discard unnecessary data.
2. Verify types of data: Understanding the modality of data (eg: text, time series, etc.) is important to make correct analysis and avoid unnecessary errors at a later stage.
3. Consolidate subset of data extracted: After performing the above steps, a subset of the data will be available for the aforementioned analysis. This data will be compiled into a single data which would be a subset of the original dataset.
4. Process the final data: Processing would involve steps such as standardization or normalization depending on requirements.

Exploratory Data Analysis:

Once the subset of the dataset is consolidated, exploratory data analysis will be performed in the following ways.

- Visualization: Plots such as histograms, spider-charts, heatmaps, line-charts, etc. will be used to plot the data. This will help to obtain a preliminary understanding of the dataset at hand and provide us with a direction for analysis. This would also help in understanding the distribution of the data.
- Tests for normality: Normality tests such as the Shapiro-Wilk test will be checked to confirm whether the features in the dataset follow a normal distribution.
- Descriptive Statistics: Calculate descriptive statistics (mean, median, standard deviation, etc.) to summarize key characteristics of the dataset.
- Transformations: Depending on requirements, transformations such as the box-cox transformation will be performed to convert features into a normal distribution and perform statistical tests.

In-depth Data Analysis:

Depending on the type of question, further in-depth analyses will be performed. Few analyses are mentioned below:

- Cluster groups identification - Students can be grouped according to common behavioral characteristics in order to distinguish different clusters and investigate variations in mental health and academic performance among these clusters.
- T-test and ANOVA- Find any meaningful changes in the means of continuous variables (like stress levels and sleep duration) between groups (like high vs. low GPA).
- Correlation test - Examine the connections between various factors (such as stress levels, social contacts, and sleep patterns) and measures of academic success (such as grades and GPA).
- Chi-Squared Test - Examine the degree of association between academic achievement and categorical factors (such as eating habits and app usage).
- Hypothesis Testing - We can gain a deeper knowledge of student behavior and success by developing precise hypotheses and using the right statistical tests to derive meaningful conclusions about the links between various factors and academic performance.

Expected outcomes of the study:

Through various EDA and In-Depth data anaylsis, we expect to gain insights into the following

- What factors affect stress levels of individuals? Are there any particular durations or episodes of increased stress (for e.g.: exams)?
- Does stress variability correlate with sleep variability? Are they dependent on each other?
- How long do students take to recover their regular activities (such as gym) after dropping them for some reason?
- What behavioral factors and mental health indicators show an impact on students' academic performance.
- Identifying patterns and trends in how students behave in reaction to their coursework and outside events.

Conclusion:

Using the proposed methodology, we aim to perform an in-depth analysis to understand student related behavior from the smartphone data captured by Dartmouth university. The outcomes of the study will help us understand the ways in which different factors influence students' lifestyle and mental health.

- [Link to Project Video](#)