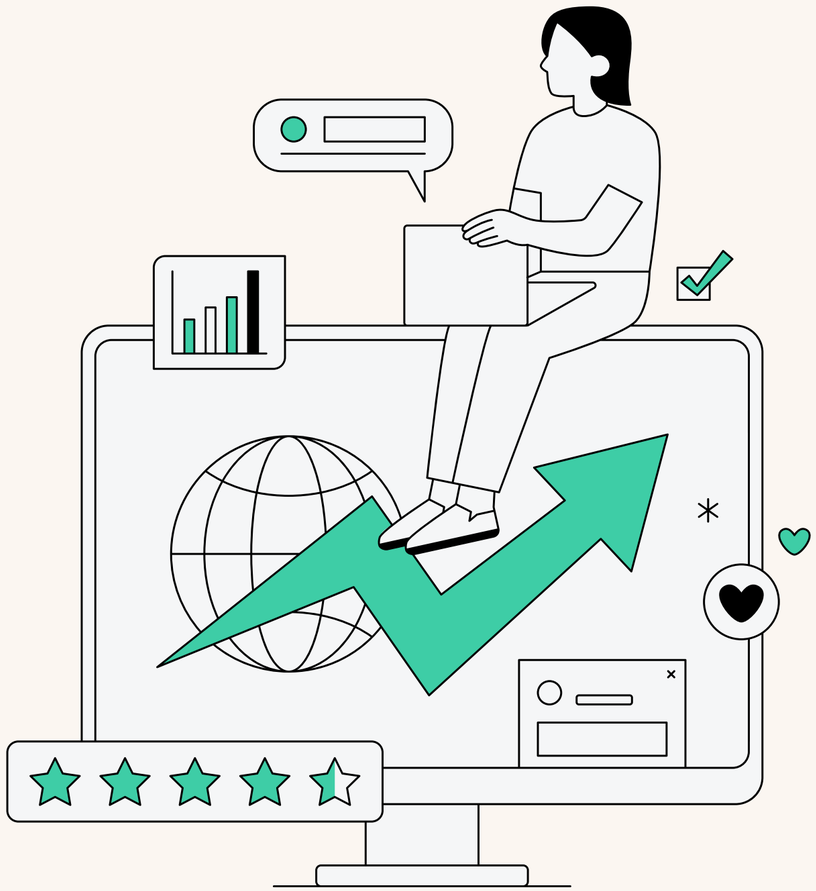
Student Stress Level

Data Mining



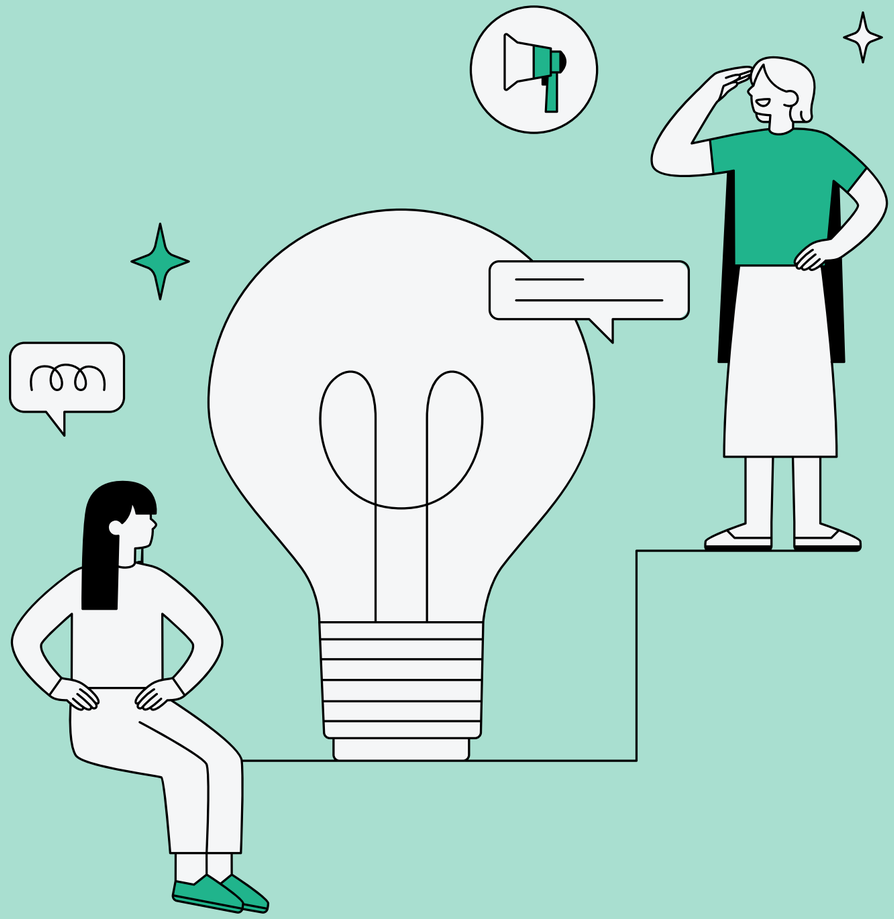
Content

82%

Of Students around the world have a stress

Problem

Student life is fraught with challenges that extend beyond academic responsibilities, encompassing social pressures, time management difficulties, and the need for personal development. The prevalent stress experienced by students can significantly hinder their mental health and academic performance. Despite this, there remains a lack of focused, data-driven analysis on the varied sources of this stress and their direct impacts on students' daily lives and educational outcomes.



Dataset

We applied our data mining tasks on data set consisting of:

- Number of objects: 1100
- Number of attributes: 21

which are :

Stress_Level

Depression

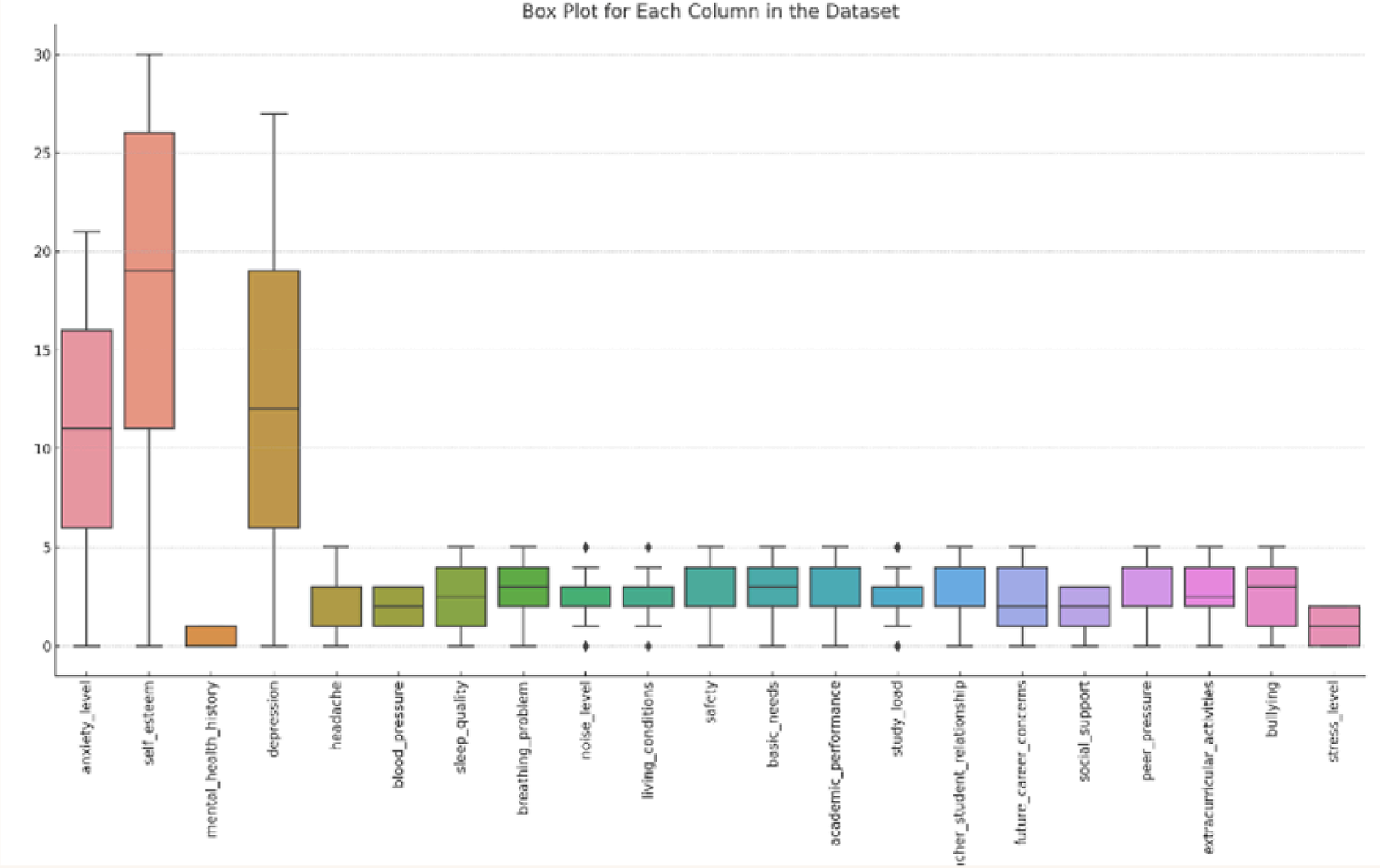
Sleep Quality

Anxiety Level

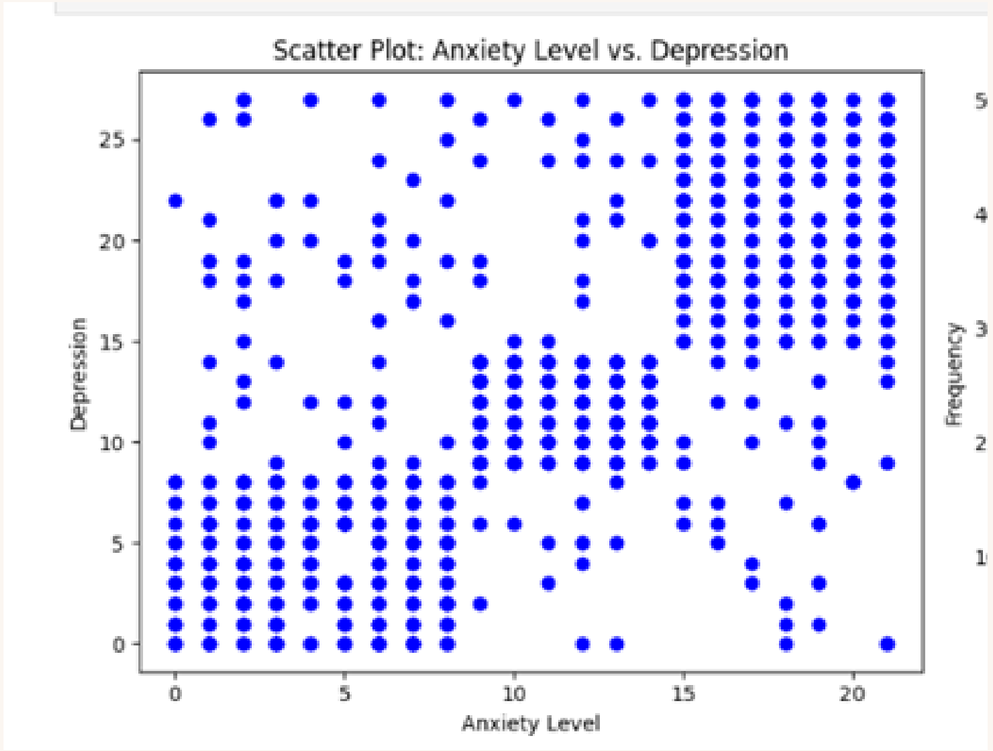
Academic Performance

Future Career Concerns

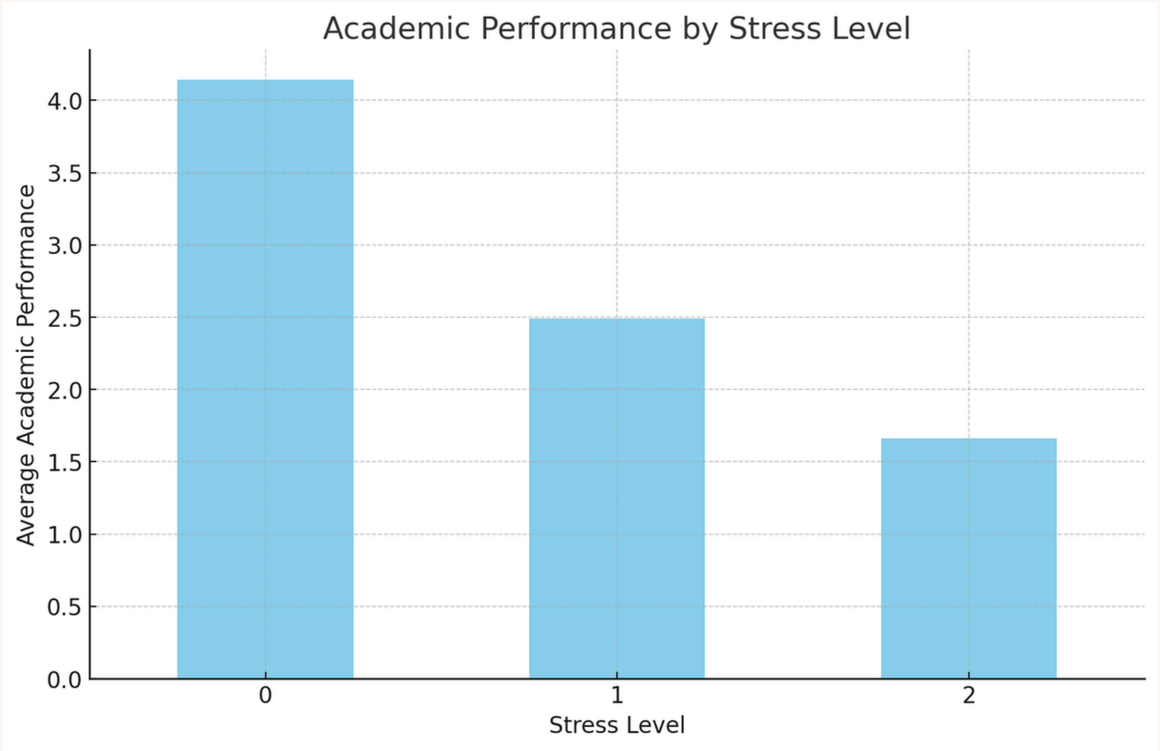
Dataset Graph



Dataset Graph



Scatter Plot (Anxiety Level vs. Depression)



Stress Level vs. Academic Performance

Main challenges identified

01.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

02.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

03.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Data preprocessing

In the data preprocessing, we applied various techniques to enhance the dataset's suitability and improve it

Preprocessing techniques used:

Data cleaning

Data transformation

Data cleaning

checking if there is any missing value and outlier

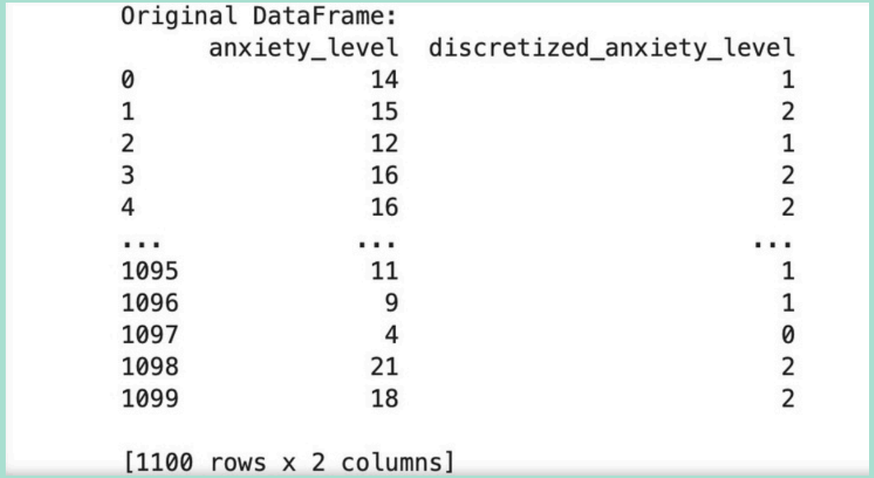
we didn’t find any missing value and outlier in our dataset

```
Missing values in each column:
 anxiety_level                   0
self_esteem                      0
mental_health_history            0
depression                       0
headache                         0
blood_pressure                   0
sleep_quality                    0
breathing_problem                0
noise_level                      0
living_conditions                0
safety                           0
basic_needs                      0
academic_performance             0
study_load                       0
teacher_student_relationship     0
future_career_concerns           0
social_support                   0
peer_pressure                    0
extracurricular_activities       0
bullying                         0
stress_level                     0
dtype: int64

Rows with missing values:
 Empty DataFrame
Columns: [anxiety_level, self_esteem, mental_health_history, depression, headache, blood_pressure, sleep_quality, breathing_pro
blem, noise_level, living_conditions, safety, basic_needs, academic_performance, study_load, teacher_student_relationship, futu
re_career_concerns, social_support, peer_pressure, extracurricular_activities, bullying, stress_level]
Index: []

[0 rows x 21 columns]
```

```
Column: anxiety_level, Number of outliers: 0
Column: self_esteem, Number of outliers: 0
Column: mental_health_history, Number of outliers: 0
Column: depression, Number of outliers: 0
Column: headache, Number of outliers: 0
Column: blood_pressure, Number of outliers: 0
Column: sleep_quality, Number of outliers: 0
Column: breathing_problem, Number of outliers: 0
Column: noise_level, Number of outliers: 0
Column: living_conditions, Number of outliers: 0
Column: safety, Number of outliers: 0
Column: basic_needs, Number of outliers: 0
Column: academic_performance, Number of outliers: 0
Column: study_load, Number of outliers: 0
Column: teacher_student_relationship, Number of outliers: 0
Column: future_career_concerns, Number of outliers: 0
Column: social_support, Number of outliers: 0
Column: peer_pressure, Number of outliers: 0
Column: extracurricular_activities, Number of outliers: 0
Column: bullying, Number of outliers: 0
Column: stress_level, Number of outliers: 0
Total rows with outliers: 0
```

Discretization

we performed a discretized operation on the 'anxiety_level' column



Original DataFrame:

	anxiety_level	discretized_anxiety_level
0	14	1
1	15	2
2	12	1
3	16	2
4	16	2
...	...	...
1095	11	1
1096	9	1
1097	4	0
1098	21	2
1099	18	2

[1100 rows x 2 columns]

each 'anxiety_level' value is assigned to one of three bins (0, 1, or 2), based on its magnitude relative to the distribution of the data in 'anxiety_level'.

Encoding

The goal of encoding is to convert categorical variables in the dataset into a numeric format.

	anxiety_level	self_esteem	mental_health_history	depression	headache
0	14	20	0	11	2
1	15	8	1	15	5
2	12	18	1	14	2
3	16	12	1	15	4
4	16	28	0	7	2
...	...	...	...	...	...
1095	11	17	0	14	3
1096	9	12	0	8	0
1097	4	26	0	3	1
1098	21	0	1	19	5
1099	18	6	1	15	3

Because there is no categorical columns in the dataset, we assuming 'mental_health_history' is a categorical column with textual data like 'yes' and 'no'.

Data mining Technique:

Classification

Technique: Decision Trees

Why: Decision trees are intuitive and can handle both numerical and categorical data, making them suitable for many classification tasks.

How: Utilize the DecisionTreeClassifier from the scikit-learn package in Python. This package offers efficient implementation and various parameters for tuning the model.

Clustering

Technique: K-Means Clustering

Why: K-Means clustering is widely used for exploratory data analysis and identifying patterns in unlabeled datasets.

How: Implement K-Means clustering using the KMeans module from scikit-learn. This module allows specifying the number of clusters and provides methods for evaluating clustering performance.

External and internal challenges

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

Findings

- We applied several plotting methods sto illustrate the data so it can help us understand our data

- We applied several preprocessing techniques that improve the efficiency of the data

- We applied data transformation, so we normalized and discretized some attributes to give them equal weight and to facilitate handling the data during data mining tasks.

- We applied the data mining tasks such :

Findings

For Classification :

- We evaluate our model by measuring the accuracy of different testing size.

Accuracy	90 % training set 10% testing set:	80 % training set 20% testing set	70 % training set 30% testing set
Information Gain (IG)	84%	85%	86%
Gini Index	89%	86.3%	86.6%

- The model that has the best accuracy was 90% training data and 10% test data.

Findings

For Clustering :

In our clustering analysis, we employed the K-means algorithm with varying numbers of clusters (K) to identify the optimal configuration. The evaluation revealed the following results:

- K=2, the average silhouette width was 0.4885.
- K=3, it decreased to 0.4499.
- K=4, it slightly increased to 0.4640.

After examining these findings, it's clear that the model with K=2 clusters had the highest average silhouette width , we can conclude that the 2-Mean clustering model is the most suitable for this dataset.

Thank you !

Any Question?

Presented by Layan Alsaykhan , Ghaida Altamimi ,Renad Alnassar