Students food choices

Learn how important is nutrition for student's success

The Why?

Away from mom and dad, 35% of the students would have an unbalanced diet. For lack of time, money or organization... Yet your plate influences day after day your shape, your weight, your morale and even your academic abilities!

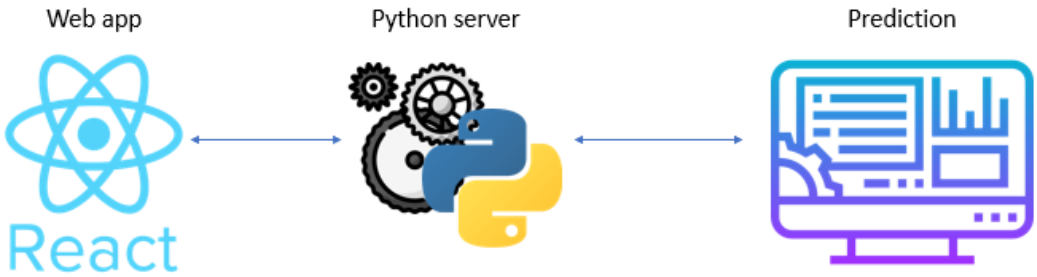
This last subject required our attention. Thus, our first project of machine learning is dedicated to nutrition of students and their impact on their academic abilities.

To do so, we have selected some points that seems to be essential for student academic success. Some features seem basic like take the breakfast or take a coffee. Breakfast is the most skipped meal, yet it is the most important because it breaks the long night time fast during which we deplete all our reserves. In a balanced diet, it should represent 20 to 25% of the calories and nutritional needs of the day.

Others features like eating fries, eating fruit or having veggie days will be taking in consideration in order to evaluate how food impact student's working and their ability to focus, and finally, their success in exams.

Our project

The idea here is to use machine learning in order to predict if a student will succeed at school according to his nutrition. After having defined our global idea, let us present how we have implemented this idea with machine learning. The heart of prediction will be in a program in python and the final deliverable would be a web application coded in Javascript and using React JS. These two parts are linked with a python server.



The dataset

Data comes from kaggle. This dataset includes information on food choices, nutrition, preferences, childhood favorites, and other information from college students.

https://www.kaggle.com/borapajo/food-choices/kernels

After that, we decided to use Collaboratory from Google in order to develop programs with Python.

Now, let's talk about the python program in few points:

1. Data Cleaning
2. Visualization
3. Model Training

Data Cleaning

There are 126 responses from students. Data is raw and uncleaned. Cleaning is in the process and as soon as that is done, additional versions of the data will be posted. We was interested only by the GPA to determine an influence between the success of the students and what they eat.

First, we droped a lot of data we don't mind until have the eleven we care. We have done our study with this criteria:

	GPA	Gender	breakfast	coffee	eating_out	exercise	fries	fruit_day	soup	sports	veggies_day
0	2.400	2	1	1	3	1.0	2	5	1.0	1.0	5
1	3.654	1	1	2	2	1.0	1	4	1.0	1.0	4

Second, We have replace data which was missing by the mean of the column (when its possible) until have all our case completed.

Third, we have drop all missing value.

```
bdd.isnull().sum()
```

```
GPA           0
Gender        0
breakfast     0
coffee        0
eating_out    0
exercise      0
fries         0
fruit_day     0
soup          0
veggies_day   0
dtype: int64
```

We determine the success with a GPA=1 and the failure with a GPA=0.

So we have replaced all the GPA >3 by a 1 and all GPA<3 with a 0. We have 121 rows.

Our date are cleaned, we can now make our Model.

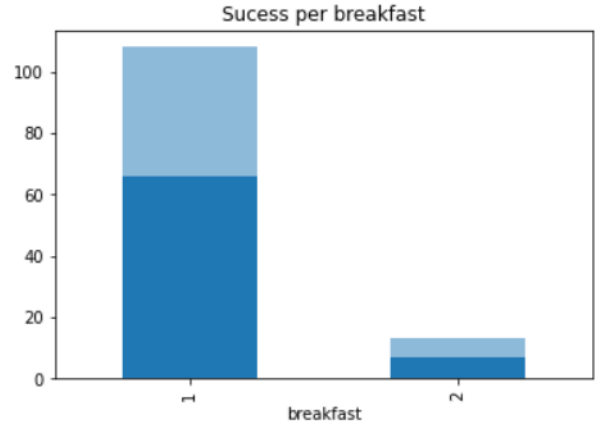
Visualization

Now, we had to visualize our data by representing them on graphs or tables in order to explore data analysis before applying machine learning models.

Here, we compare the percentage of students who take a breakfast (1) and the percentage of student who do not take a breakfast (2), taking account of the student's GPA.

Therefore, we can see that 61,11% of student who take a breakfast succeed at school however 53,85 of students who don't take a breakfast succeed at school.

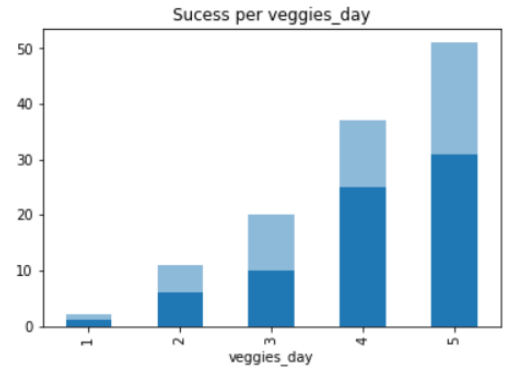
breakfast	total	GPA	percentage
1	108	66.0	61.11
2	13	7.0	53.85



Here, we compare the percentage of students who have one veggie a day (1), two veggie a day (2), three veggie a day (3), four veggies a day (4) and 5 veggies a day (5), taking account of the student's GPA.

Therefore, we can see that 60,78% of student who take 5 veggies a day succeed at school, however 50% of students who take 2 veggies a day succeed at school.

veggies_day	total	GPA	percentage
1	2	1.0	50.00
2	11	6.0	54.55
3	20	10.0	50.00
4	37	25.0	67.57
5	51	31.0	60.78



Model Training

To predict the success of our students, we make a model based on logistic Regression. Logistic regression is a statistical technique used to predict probability of binary response based on one or more independent variables. It means that, given a certain factor, logistic regression is used to predict an outcome which has two values such as 0 or 1, pass or fail, yes or no etc.

We have split the dataset into a training dataset and test dataset. We use 70% of our data to train and the rest 20% to test.

We have instantiate the logistic regression using '*LogisticRegression*' function and fit the model on the training dataset using 'fit' function.

We calculate the probabilities of the success for the test dataset.

We calculate the model accuracy on the test dataset using '*score*' function. We can see the accuracy of 64%.

To finish, we predict the success with all of our criteria.

Conclusion

This article tried to introduce with some details our project, especially how we implemented it.

The analysis of dataset represents a good insight of food habit of successful students. However, the dataset is not trustworthy because of its size. Also, it has a lot of missing data. Therefore, the chance to get almost optimal solution is not very good.

However, an analysis was possible and thanks to machine learning we are convinced that nutrition has an impact on student academic abilities.

You can find more information in our github repository and our youtube video to have a realistic view.