# A+ in Math

We can say that knowing mathematics has become indispensable if we want our society to progress.

# "Mathematics is the key and

It is the most important key that allows us to understand the functioning of the world.

# Achtung!

But beware! We have a problem.

# PISA 2018

We are falling behind! PISA shows thing are not going right.

PISA is an international assessment that measures the academic performance of 15-year-old students in reading, mathematics, and science.

# PISA 2018

We are in 16th place! Behind the Netherlands, Sweden, Belgium

We need to do something, This cannot remain as it is.

# What is the key to

We need to figure out the key to improve our results.

# Most predictive feature of low

If we know the characteristics of students who have low scores in mathematics.

# Focus especially on guiding

The education System can provide them with greater support in earlier educational stages.

# I know

I know which are the features...

# ...do you want to know?

...do you want to know how did I get them?

# Datase/E.D.A./Regression

# Dataset Kaggle

I took the dataset from Kaggle, unfortunately it is a fictional dataset

# Rows and Columns

It has nineteen thousand rows (after cleaning) and

Fifteen columns

# E.D.A.

The data set contains three different scores of students between 0 and a 100 in Reading Writing and Math.

We can see that in the average Writing and Reading score of Females are higher than male but in math is the opposite.

Avg Practice

Beside the scores I also have information on the Average Sport Practice in three categories: never, sometimes and regulary. [Señalar mientras lo digo] Most of students practice sport sometimes.

Lunch type

There is also the information of lunch type. There are two types: free/reduce and Standard. Here talks a bit about the resourses of students family.

Ethnic group

In the dataset ethnic group are not specified.

Test prep

We also know…

One third has completed the test preparation the other 2/3 haven’t

Rest of the…

We also have other categories that we are not going look at in now in detail.

Regressions

Base in this data I created three models one linear regression, one decision tree and one knn model.

Linear Regression

Results

The result of this regression is quiet good, we can see that the R-squared is 0.842.



*Adj. R-Squared penaliza si pones demasiadas variables. En este caso está bien.*

*F-statistic*

Results

Here the features that have more impact are Gender Female, with eleven poins of diference in comparation with males.

All the EthnicGroup has more or less the same results exept Ethnic Group E that seems to have an advantage with 5.6 points in comparison to group A

We also see that children with free lunch have a disadvantage of minus almost five points in comparison to children with parents that paid for their lunch.

And finally It seems that sport has a great impact on math score. But you shouldn’t over do it.

Decision tree

The Error metric is 0.83-0.82 and that is good. But almost as good as linear regression

Feature Importance

Here we see that by far the most important feature is the reading score, sorpresibly . This is follow by gender and free lunch.

KNN

To complement my analysis I tried a KNN model.

But I saw that the performance was lower than the previous two models.

Overview

Here are all three regressions and we can see that linear regression is the most accurate but there is little difference with decision tree.

Conclusion

What I could see in my analysis is that the two main predictors that negative influence math scores are gender female and free lunch type and

The two variables with positive influence are the ethnic group E and the reading score.

Thank you

Please someone call the minister of education.