HOCAM

Easier monitoring of learning journey for teachers.

Group Name: Ada Lovelace

Melike Enginsoy

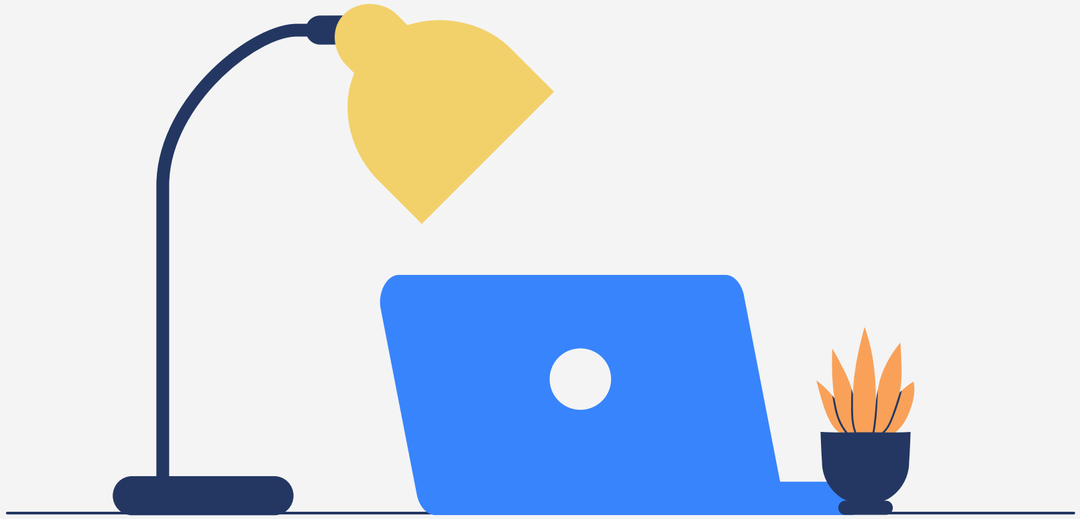
Birgül Ceryan

Neslihan Uzun

Ebra Çepni

Problems

Why should I use artificial intelligence when monitoring students?



Too many students

There are too many student characters and test results teachers need to understand.

Taking too much time

It takes time to make sense of all the results.

Possible prejudices

Rapid developments may be overlooked.

Difficult to understand the source of the failure

Class-wide and student-specific analyzes are difficult.

Solutions

Why should I use artificial intelligence when monitoring students?

Monitor from one tool

Get student insights in one channel.

Saving time

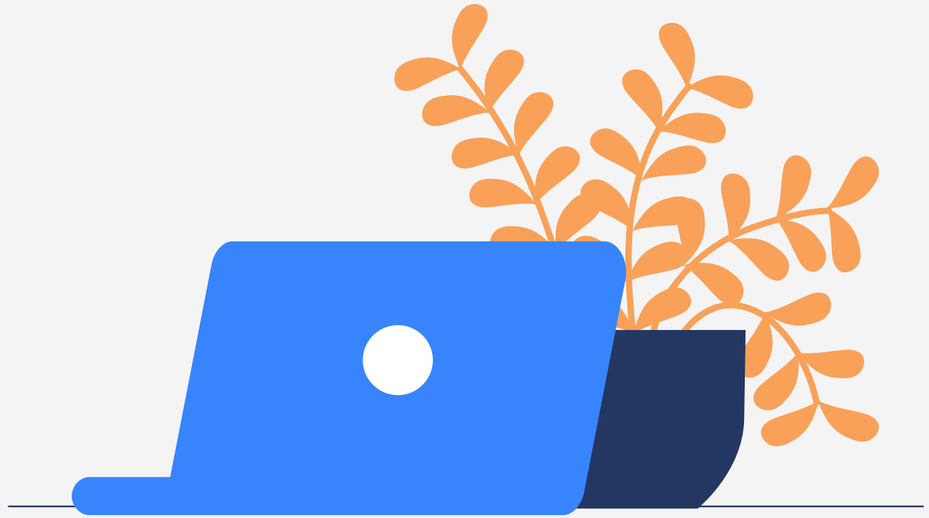
Useful for time management.

Unbiased view

Ability to monitor unexpected student progress.

Quickly understand the source of failure

Makes class-wide and student-specific analyzes more easier.



Competitors

dijitalim Öğretmen

Pakodemy

Poll Everywhere

School CPD TRACKER

PROGRESS | DEVELOP | LEARN

Roadmap

What did we do as part of the project?

- Defining the question
- Collecting the data
- Cleaning the data
- Exploratory data analysis
- Feature engineering
- Apply the ML Model
- Test the outputs

Failures leading to success

We had many unsuccessful attempts before building a successful machine learning model. But now we know better what not to do!

Apriori / ARL / Random Forrest

We couldn't find any meaningful results.

Unsupervised Learning

We didn't have the required metrics.

Go Live

We didn't have enough time.

Achievements!

We designed an algorithm for the teachers that predicts which topic students should work on.

Let's start modeling part!

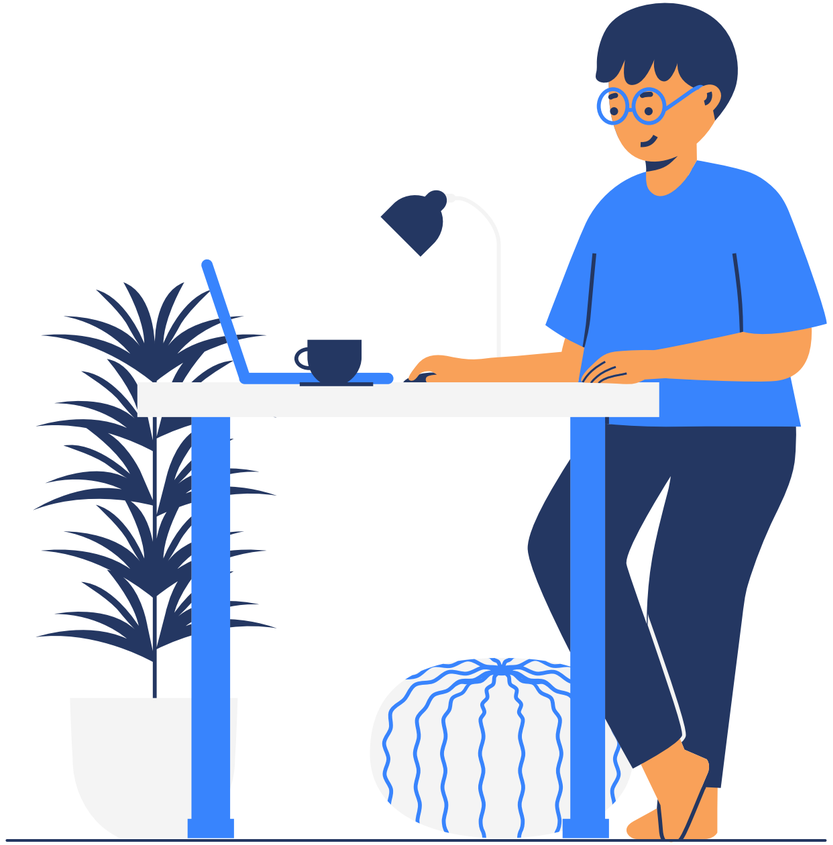
How to make it easier to identify students?

Teachers ;

-what level the students are at,
-which subjects they do not understand,
-linking incomprehensible issues,

takes a lot of time to understand!

We created a data set exceeding 3500 rows, in which we indexed the answers given to each question on a student basis.



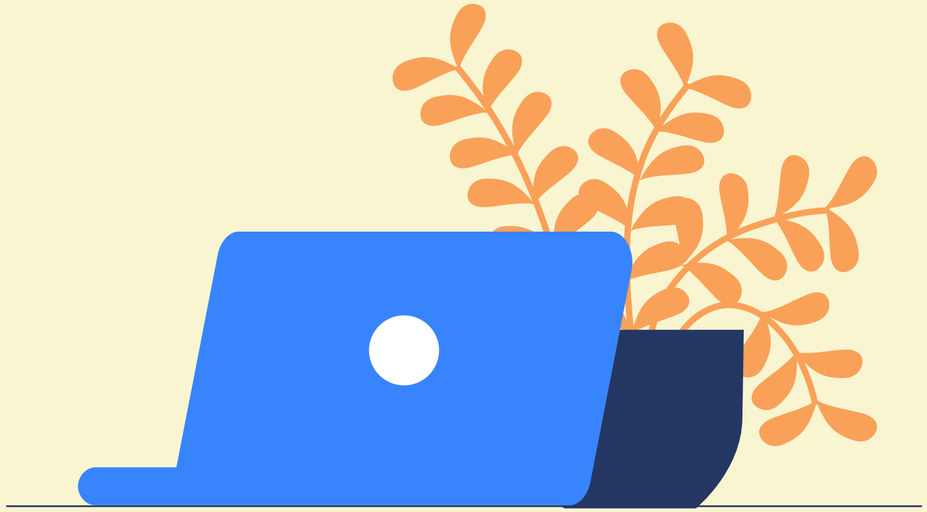
Security first!

Created random and unique ID.

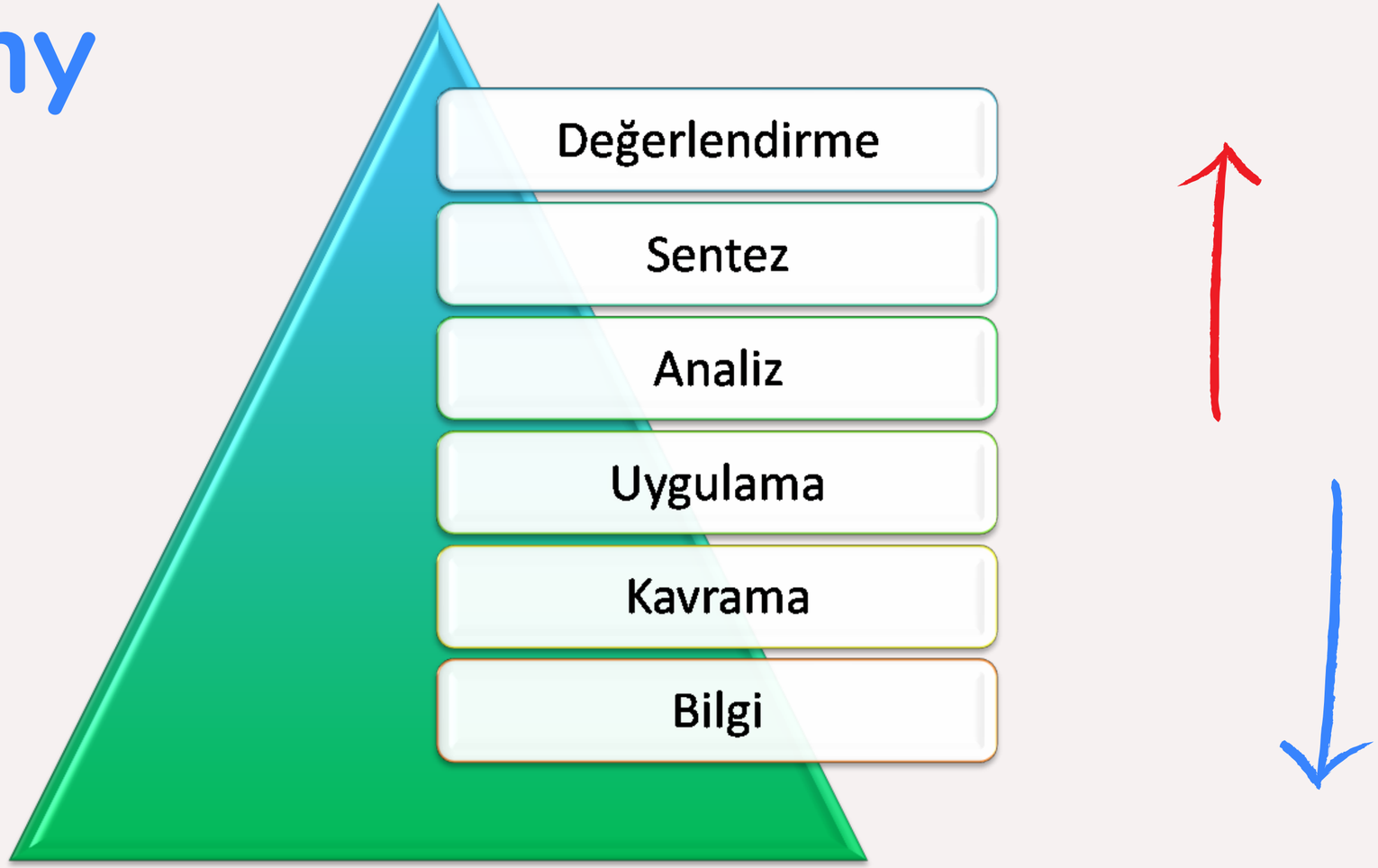
In order to protect personal data, we removed student names and surnames from the data set with a function.

Were the questions difficult HOCAM?

This is the most common question that teachers hear from parents.



Bloom Taxonomy



examples

Aşağıda elemanları verilen üçgenlerden hangisi çizilemez?

A) m($\hat{A}$)=60° , m($\hat{B}$)=70° , |BC|=10 cm

B) m($\hat{A}$)=45° , m($\hat{B}$)=63° , |BC|=8 cm, |AC|=4 cm

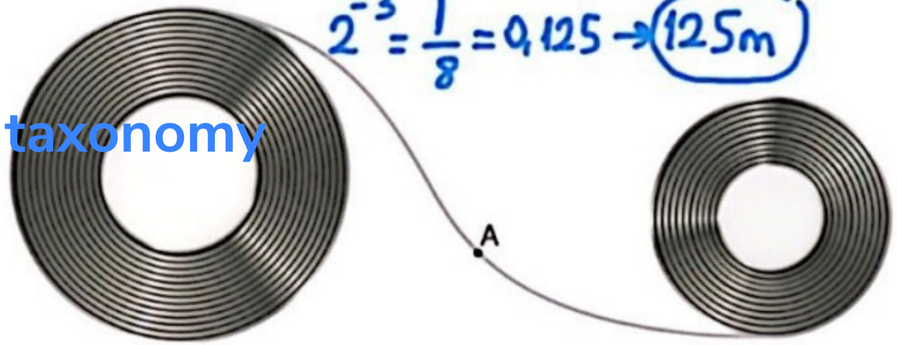
C) m($\hat{C}$)=65° , |AB|=|AC|=7 cm

D) |AB|=7 cm , |BC|=4 cm , |AC|=4 cm

Bloom taxonomy

Bir tel iki ucundan şekildeki gibi sarılmıştır. A ile gösterilen noktanın solundaki kısmın uzunluğu 2^{-3} km, sağındaki kısmın uzunluğu 5^{-2} km'dir.

$2^{-3} = \frac{1}{8} = 0{,}125 \rightarrow 125m$



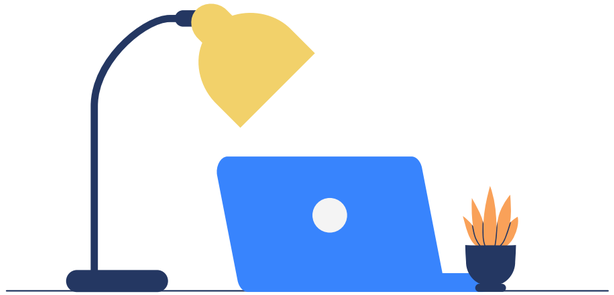
$5^{-2} = \frac{1}{25} = \frac{4}{100} = 0{,}04$

$0{,}04 \rightarrow$ 40metre

Bu telin tamamı metre cinsinden tam sayı olacak şekilde eşit uzunlukta parçalara ayrıldığında elde edilen parça sayısının bir asal sayı olduğu bilinmektedir.

Buna göre elde edilecek parça sayısı en fazla olduğunda bir parçanın uzunluğu kaç metre olur?
(1 km = 1000 m)

A) 33 B) 15 C) 11 D) 3

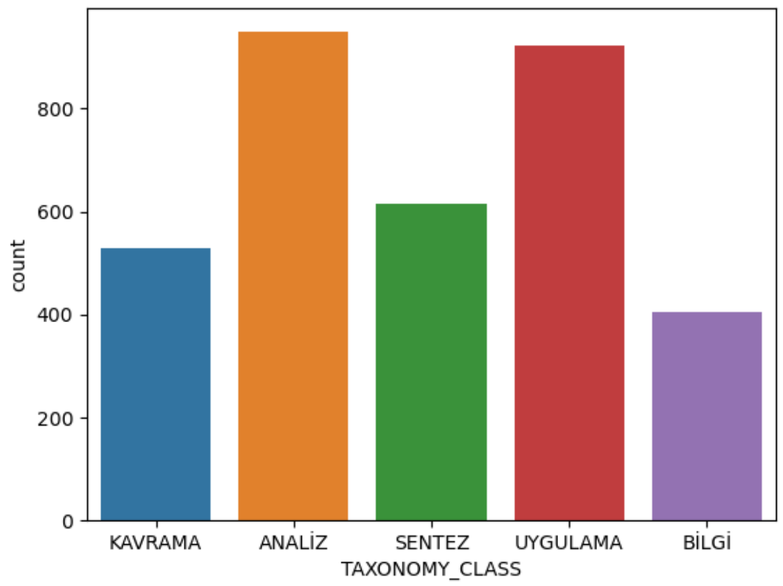


Literature

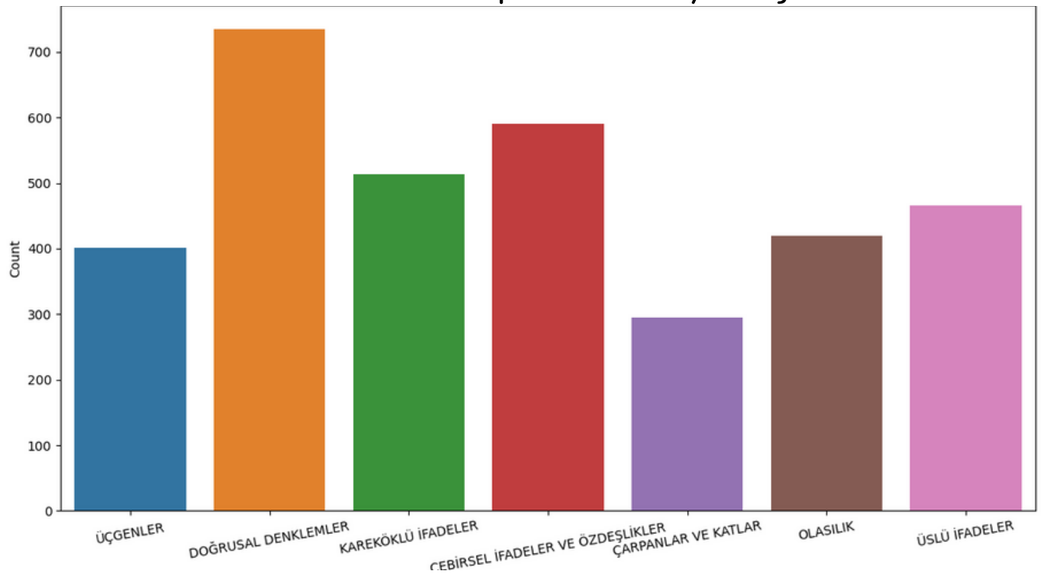
- Bloom Taxonomy was presented in 1956 by Benjamin Bloom, a professor at the University of Chicago.*
- In the literature, knowledge and comprehension questions are seen more frequently in 6th and 7th grades. However, analysis and synthesis questions are increasing in 8th grade books. The distribution of our data set is in accordance with the literature.**
- The way of perception and learning style of each person differs within themselves. Some of us learn by **reading**, while some of us learn by **solving**. At the same time, it is seen that the way of learning of an **intelligent person** and a **normal person** is different.
- The main thing that makes up the Bloom model; It is the ability to **control the quality of teaching** by **keeping the personal perception differences** that affect the student under control.

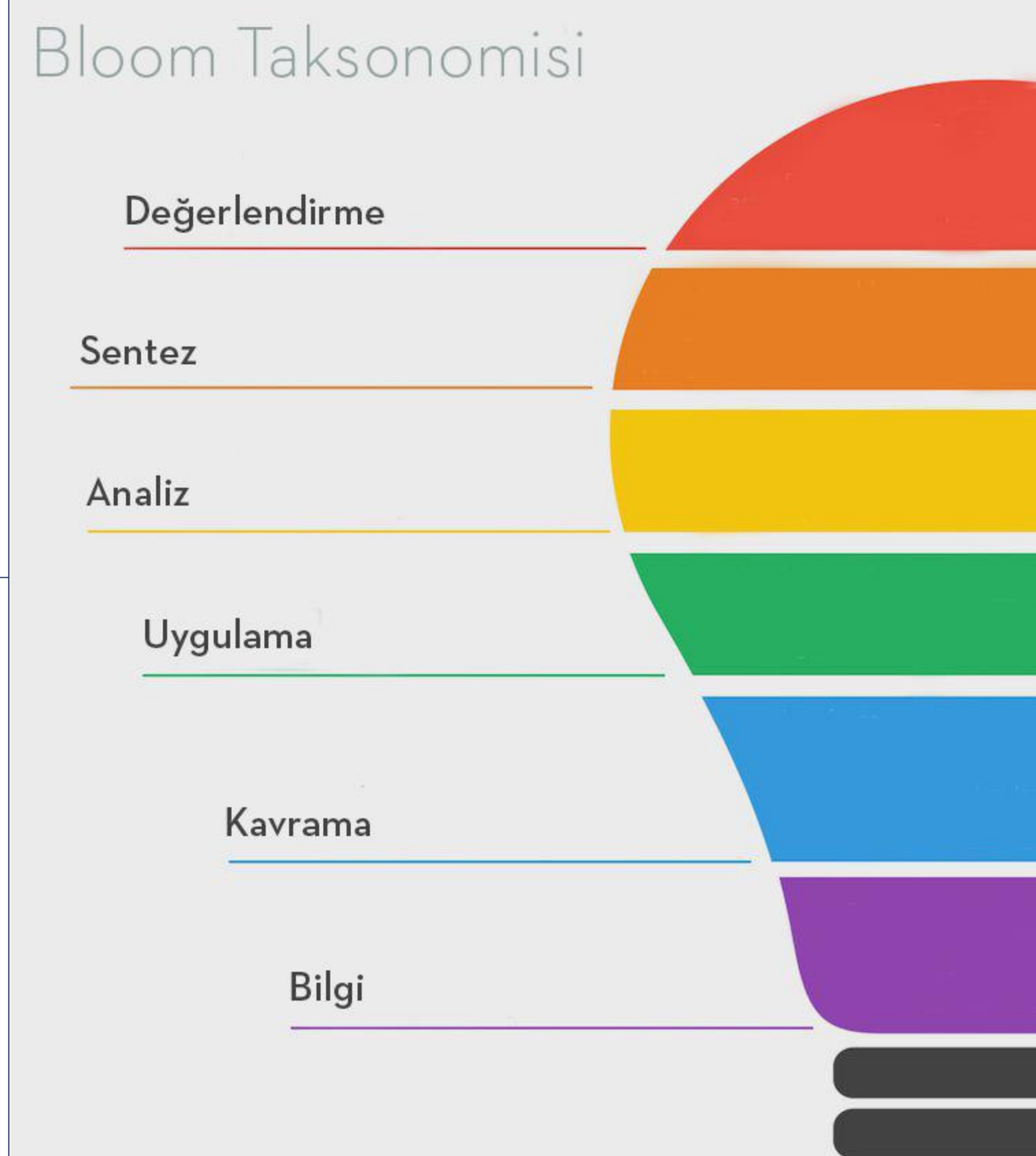
***http://serifeyurtseven.blogspot.com/2016/04/bloom-taksonomisi.html**
****Abdullah Çağrı BİBER, Abdulkadir TUNA, Ortaokul Matematik Kitaplarındaki Öğrenme Alanları ve Bloom Taksonomisine Göre Karşılaştırmalı AnaliziOndokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi 2017, 36(1), 161-174**

Distribution of questions by taxonomy class

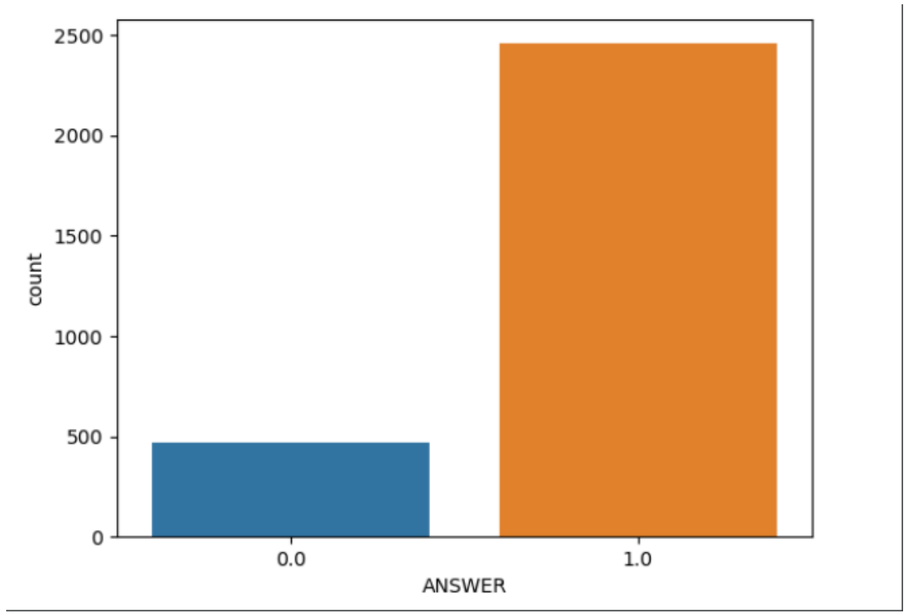


Distribution of questions by subject





True & False Rate



True & False Rate

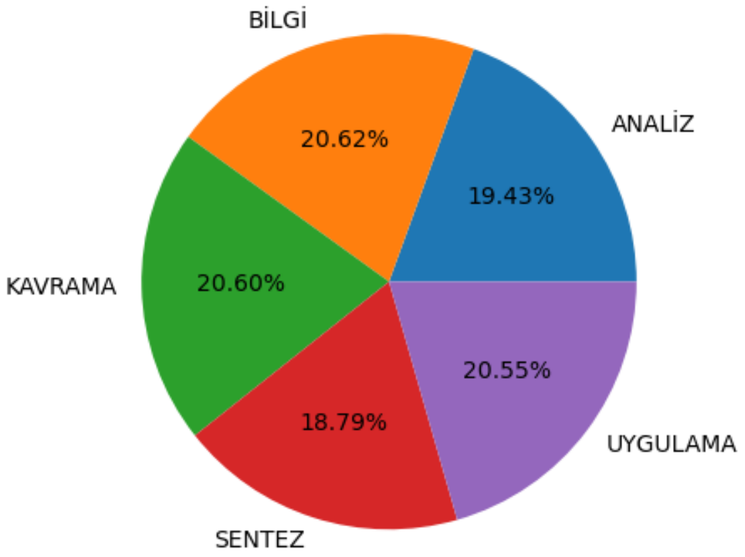
1.000 %83.9 ---->2459

0.000 %16.1 ---->472

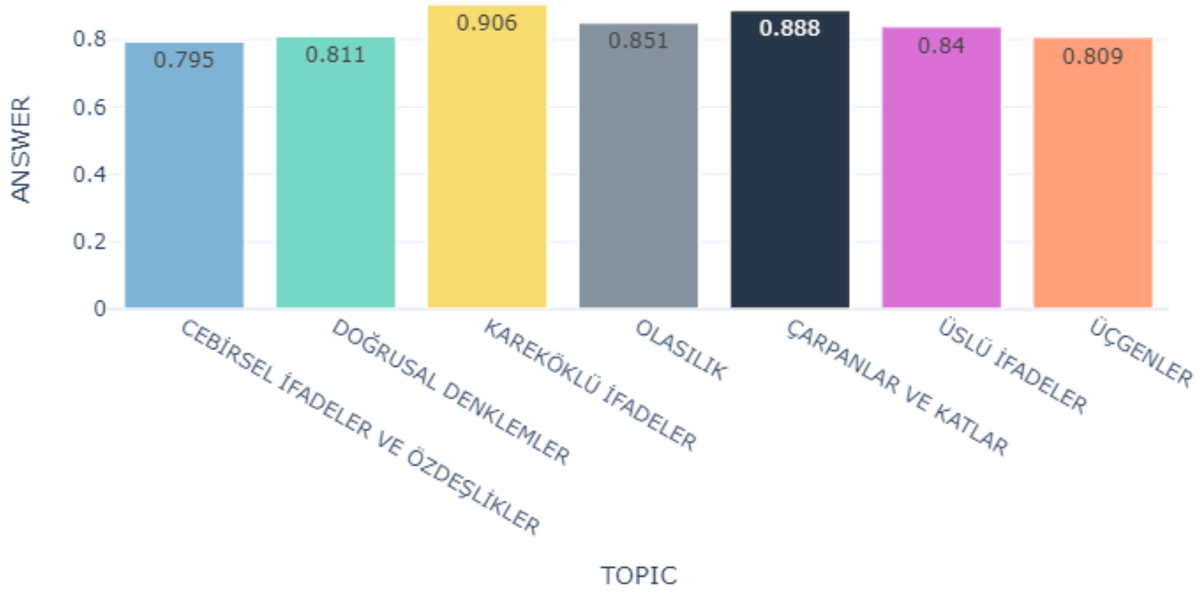
Unique Info

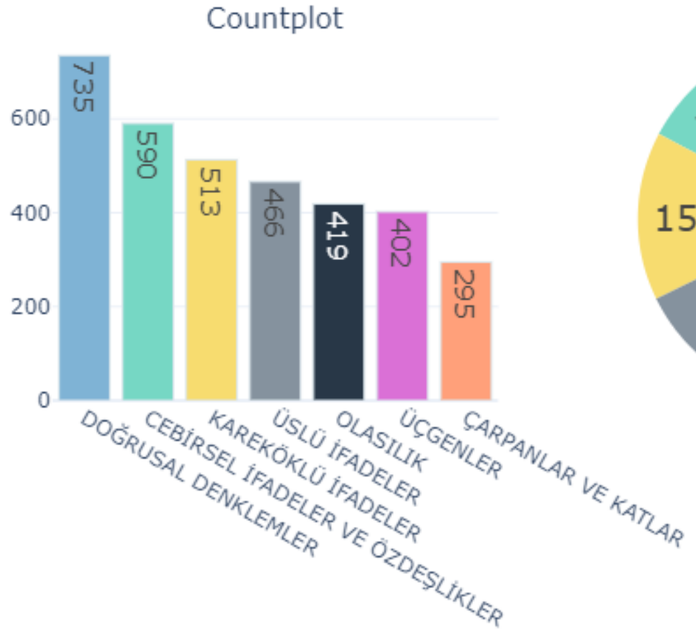
STUDENT_ID	49
CLASS	3
TEST_NUMBER	4
QUESTION_NUMBER	20
TOPIC	7
SUB_TOPIC	28
TAXONOMY_CLASS	5
ANSWER	2
TEST_QUESTION	80

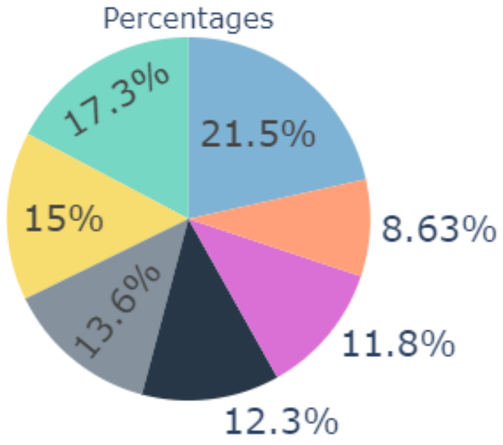
The average of correct answers to the question by taxonomy class



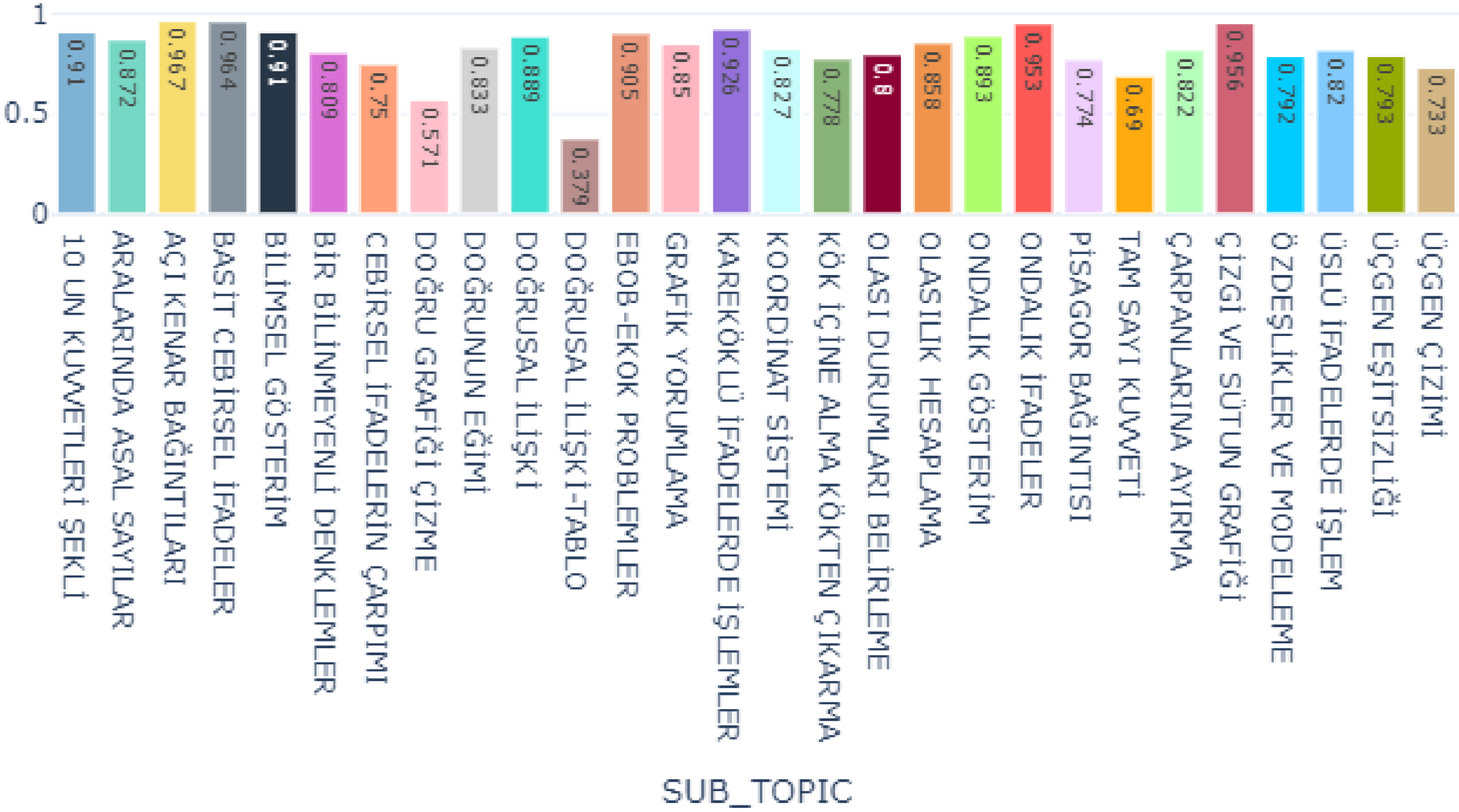
Correct answer rates by topic



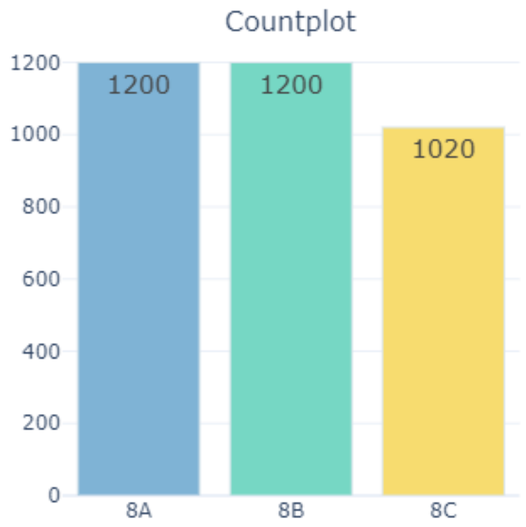


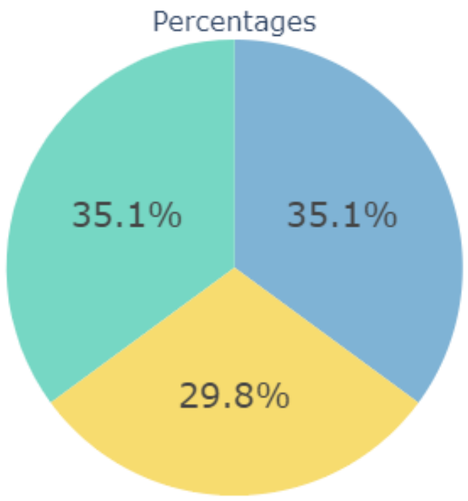


Correct answer rates by sub-topic

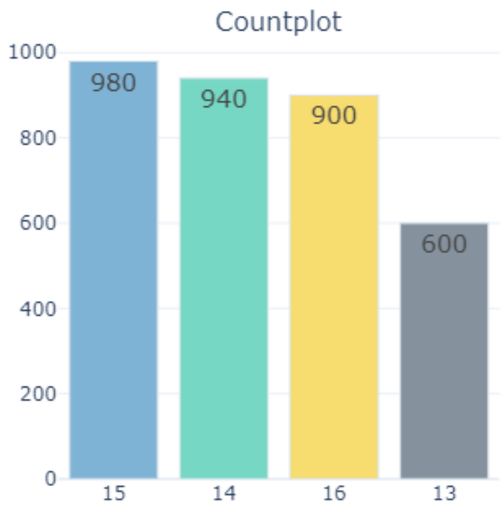


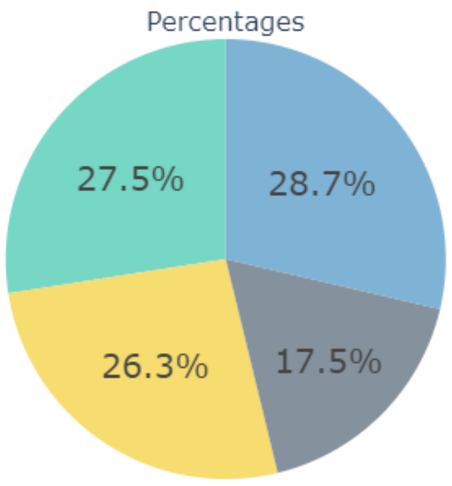
Number of questions answered for each class



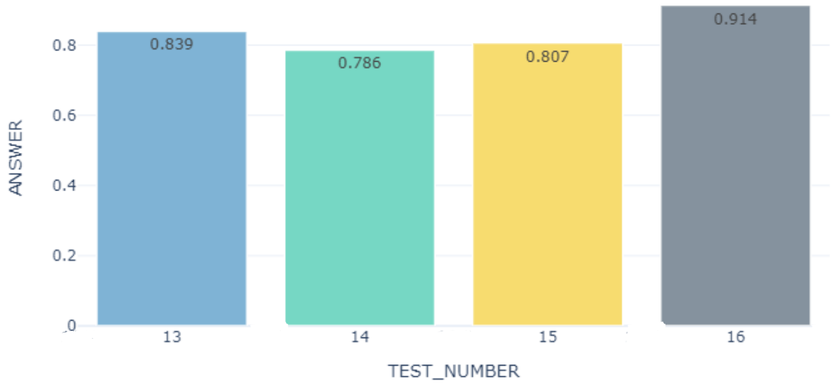


Number of questions answered for each test

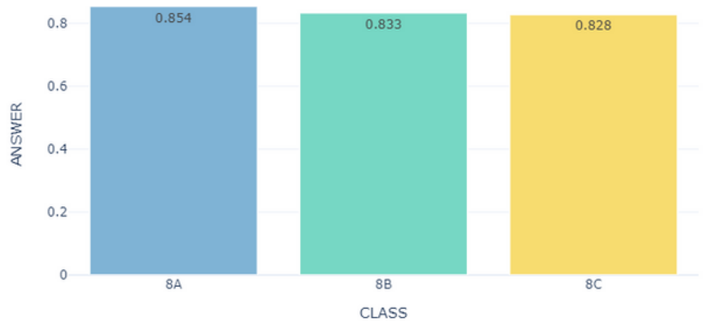




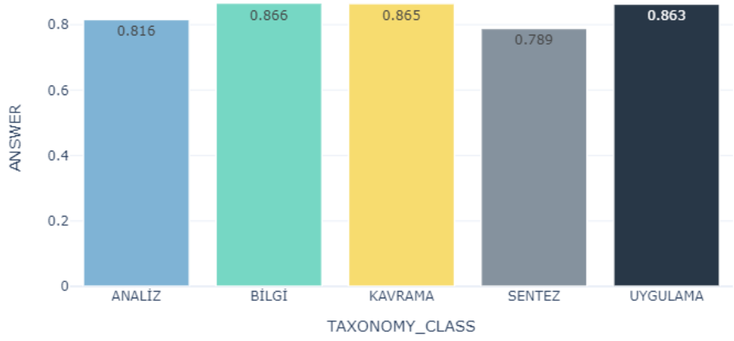
Correct answer rate on test basis



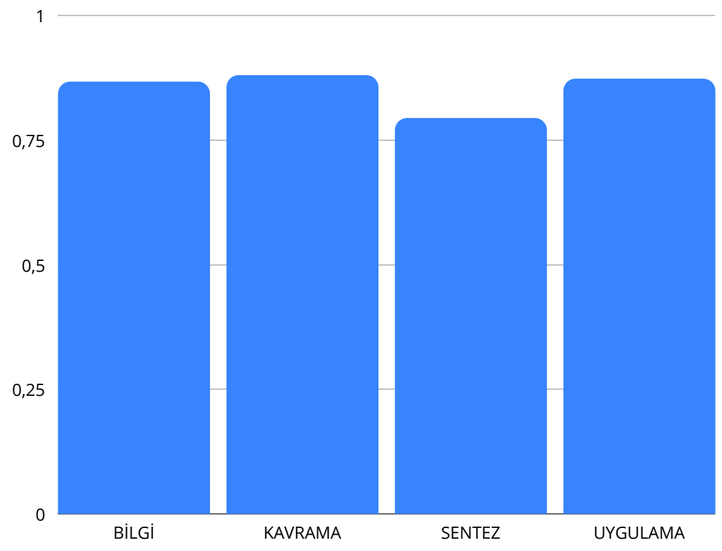
Correct answer rate on class basis



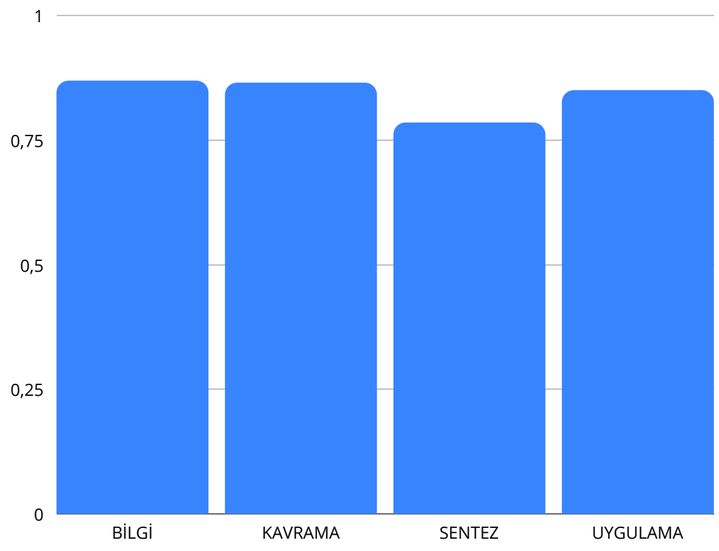
Correct answer rate on taxonomy basis



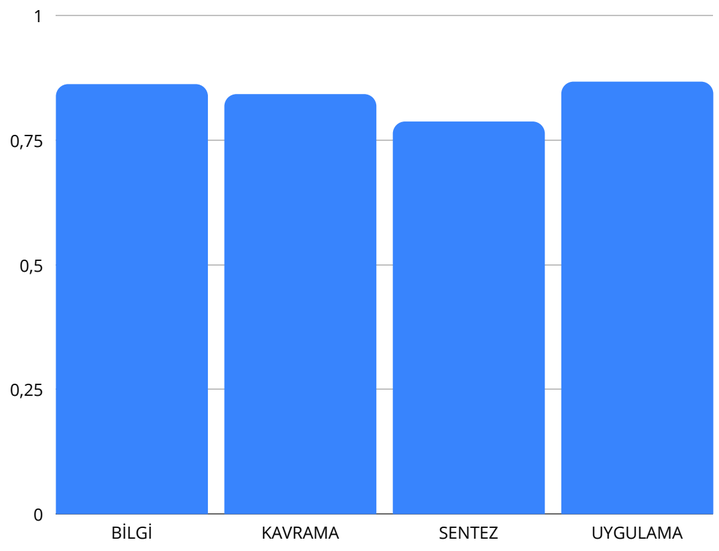
Correct answer rates on the basis of taxonomy



8A

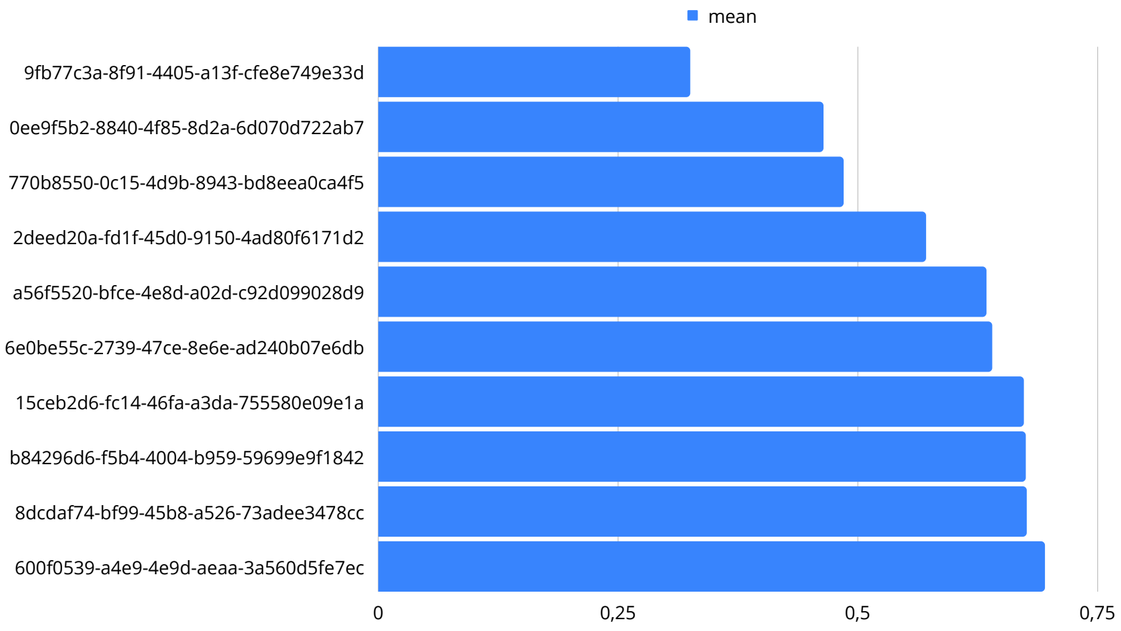


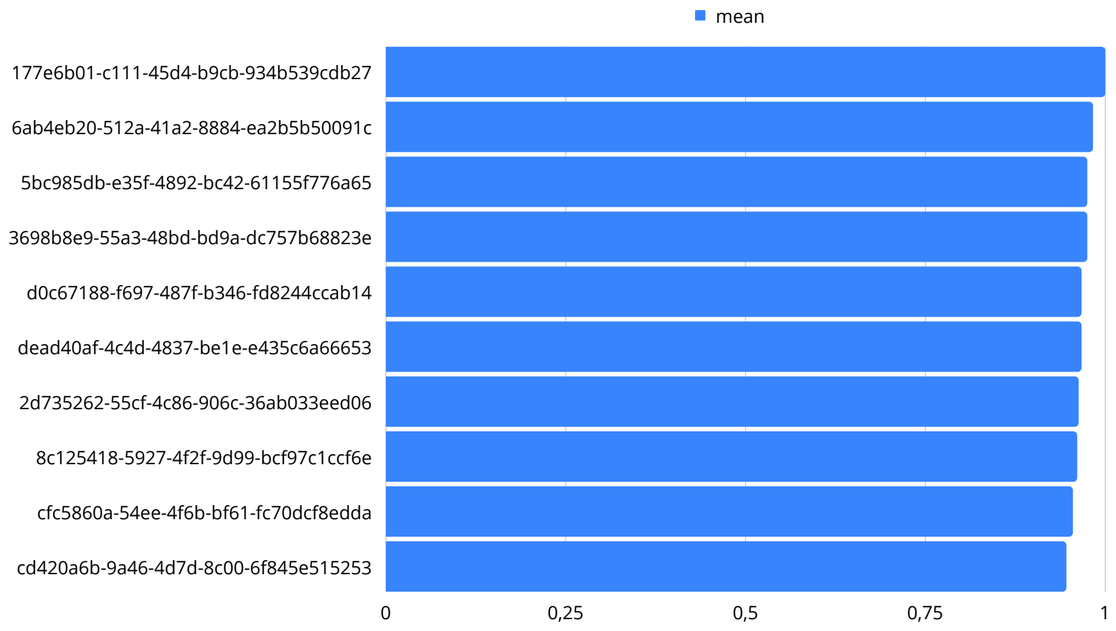
8B



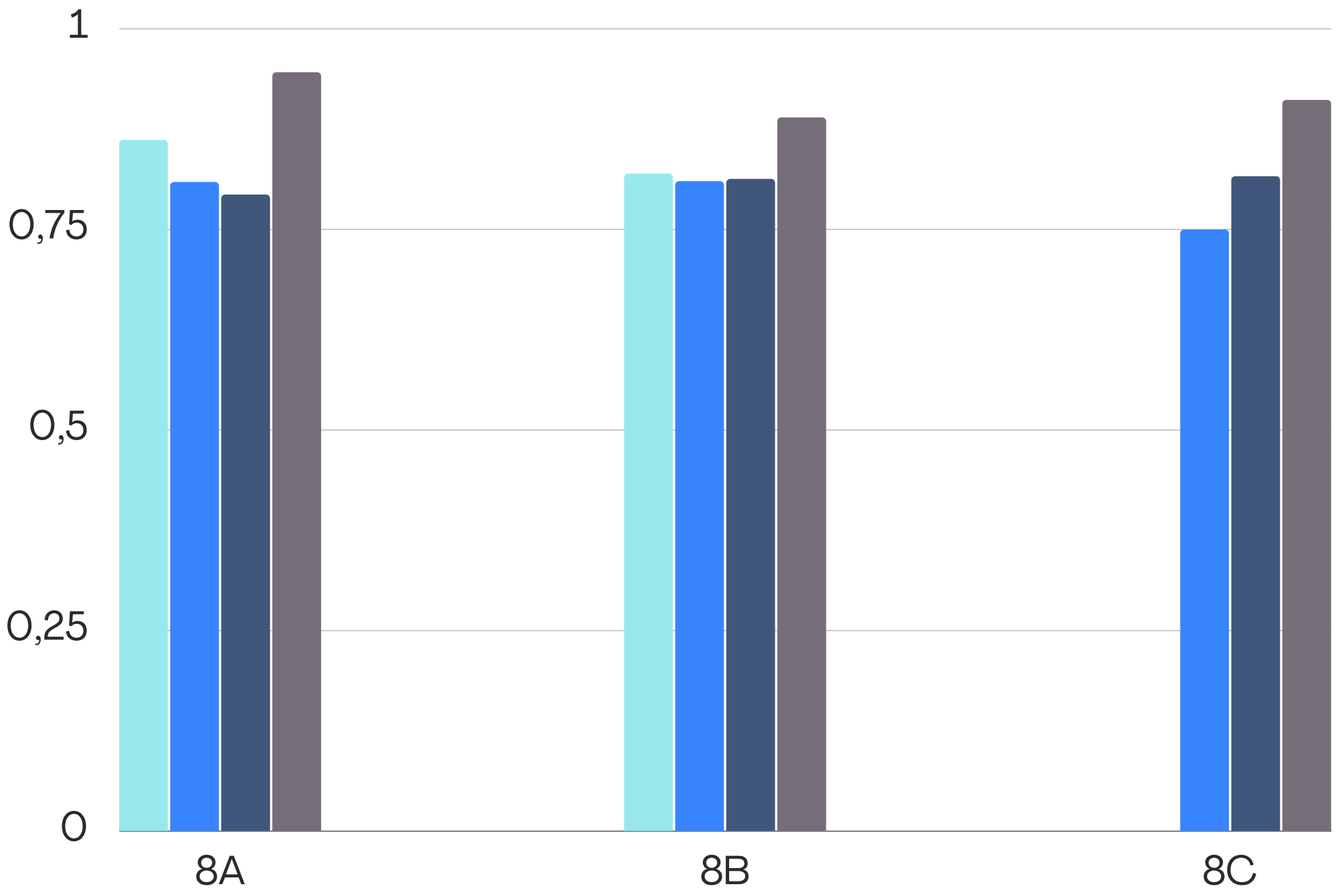
8C

Top 10 successful and unsuccessful students



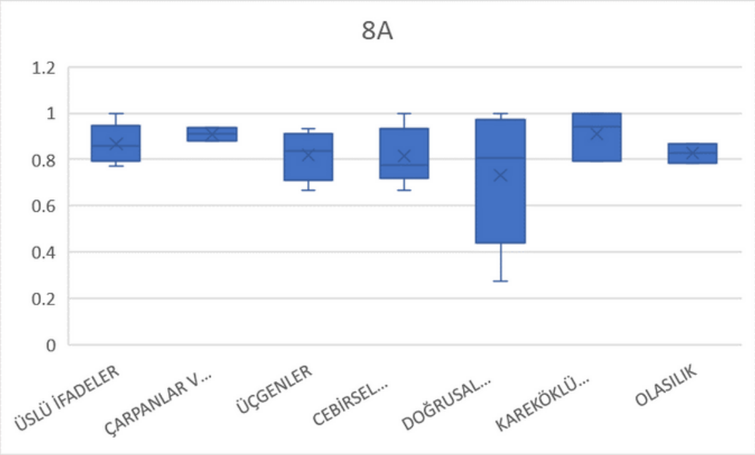


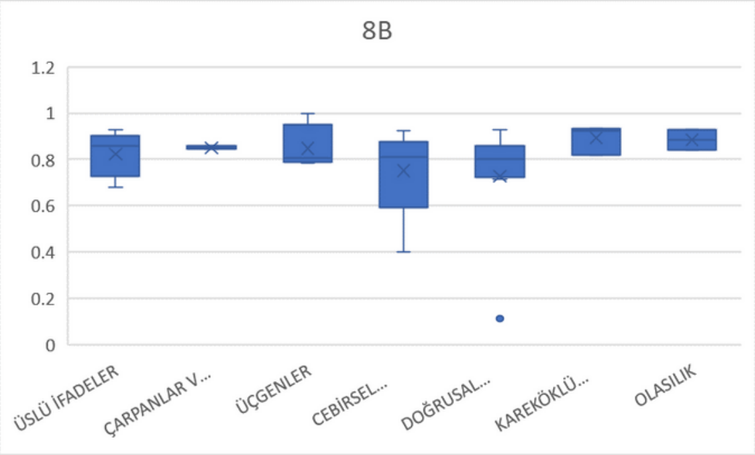
Class Average On Test Basis

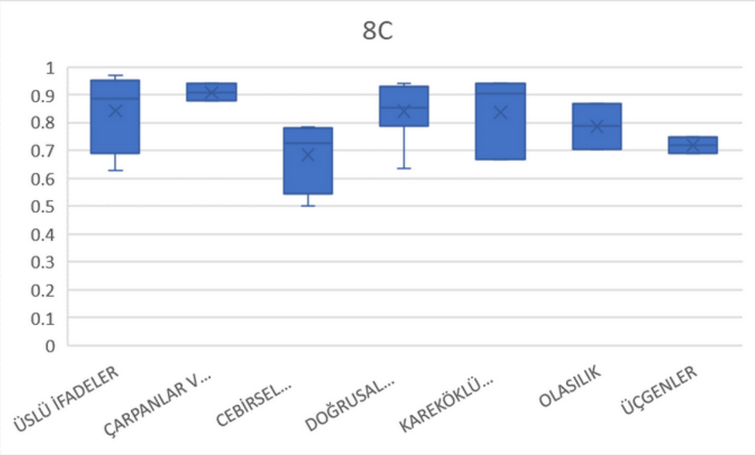


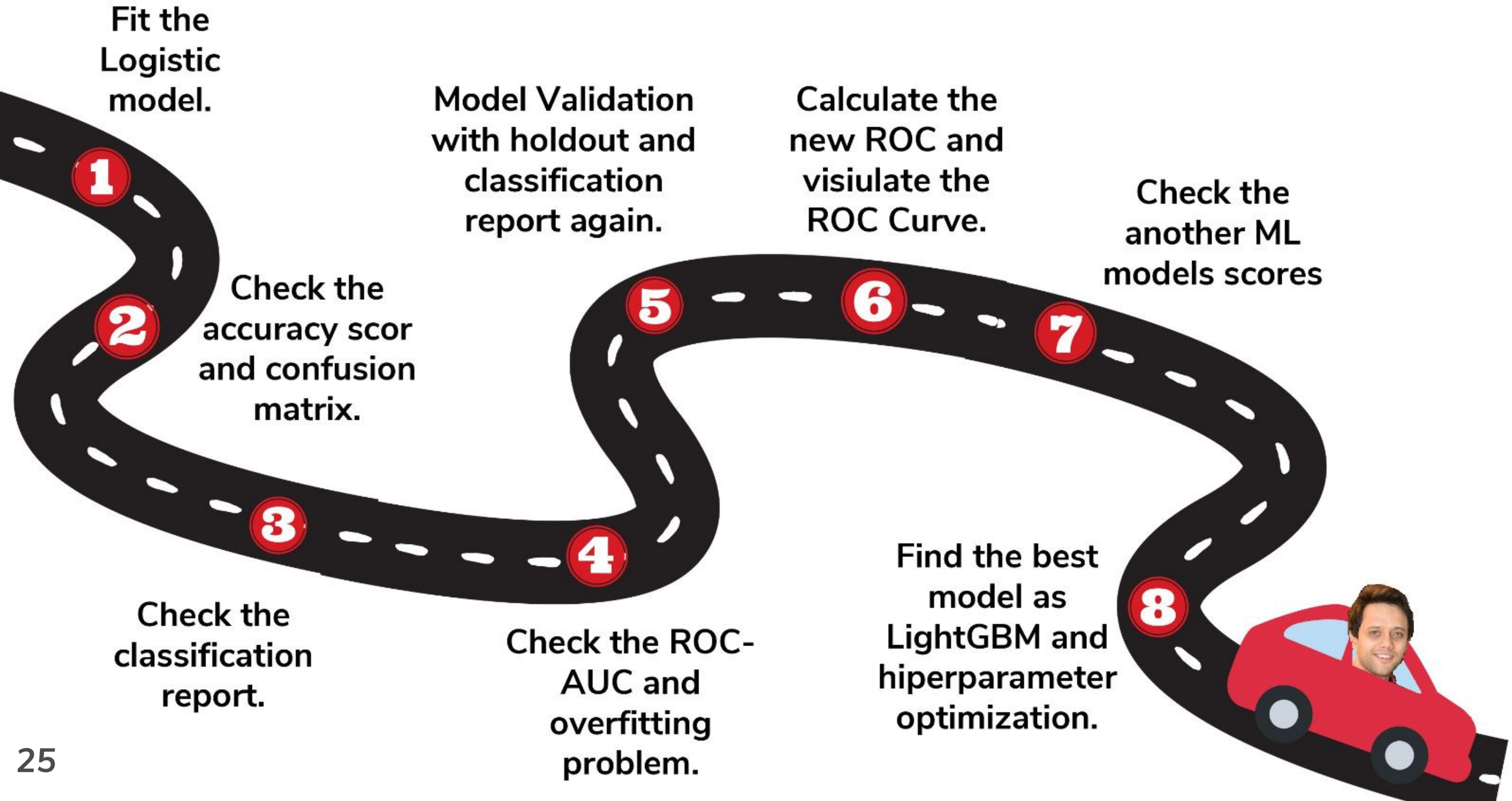
Categorize Function

In order for the teachers to be easily evaluated, we took a printout of the taxonomy class of the students with an average of less than 0.50.








Fit the Logistic model.
1
Check the accuracy scor and confusion matrix.
2
Check the classification report.
3
Check the ROC-AUC and overfitting problem.
4
Model Validation with holdout and classification report again.
5
Calculate the new ROC and visiulate the ROC Curve.
6
Check the another ML models scores
7
Find the best model as LightGBM and hiperparameter optimization.
8

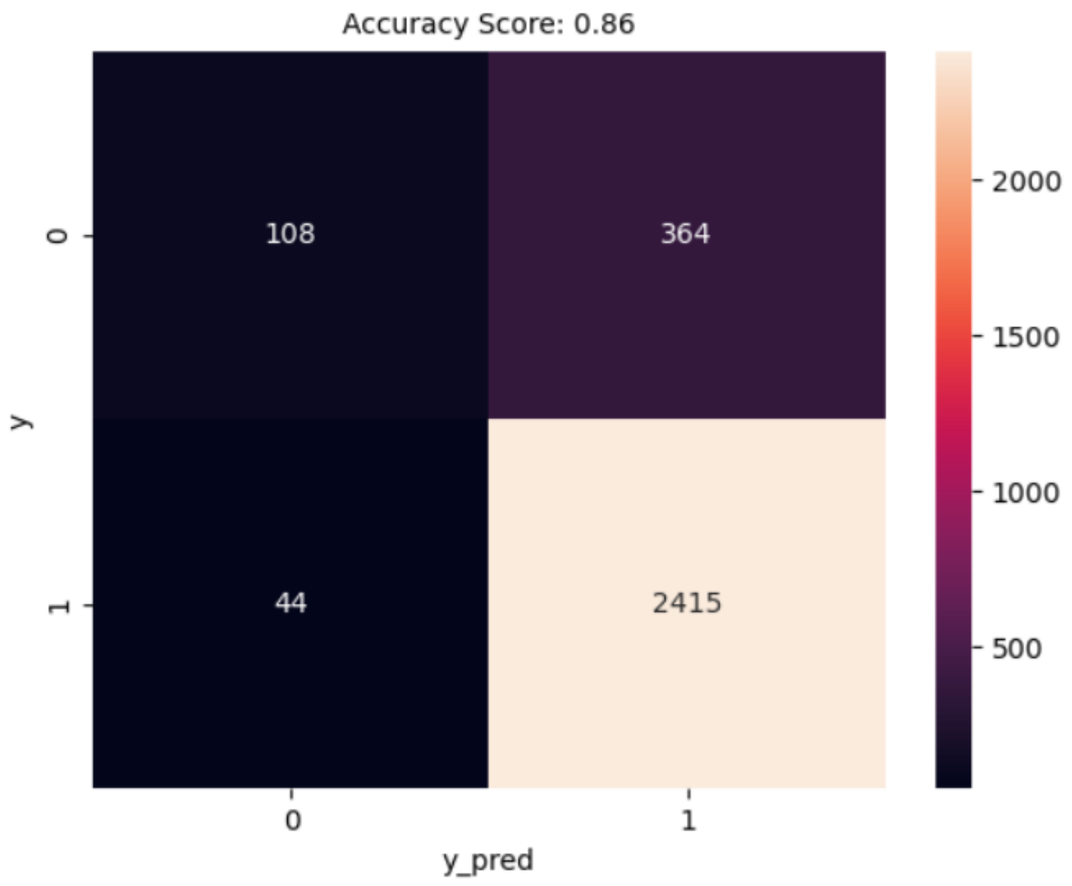
Confusion Matrix

Logistic Regression Model Evalution Step

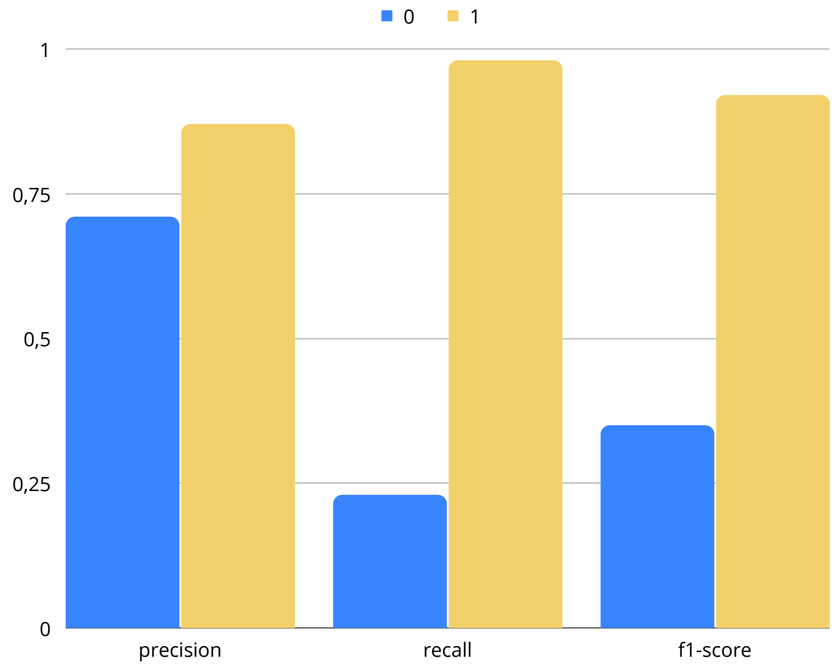
```
In [15]: roc_auc_score(y, y_prob)
Out[15]: 0.8591769425355491
```

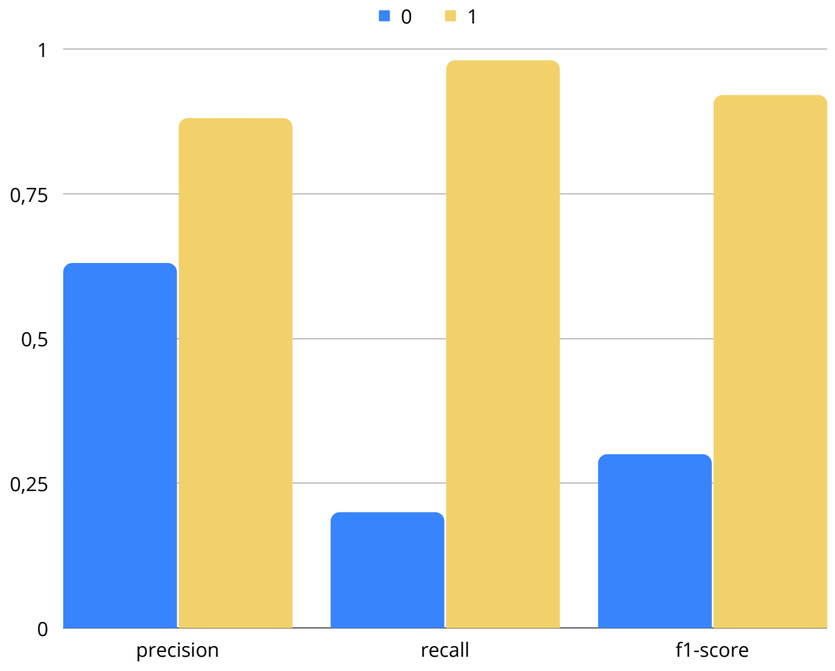
	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

- True Positives (TP): Positive predicted and it's true.
- True Negative (TN): Negative predicted and it is correct.
- False Positive (FP): Positive predicted and this is incorrect.
- False Negative (FN): Negative predicted and this is incorrect.



Logistic Regression Classification Report Holdout



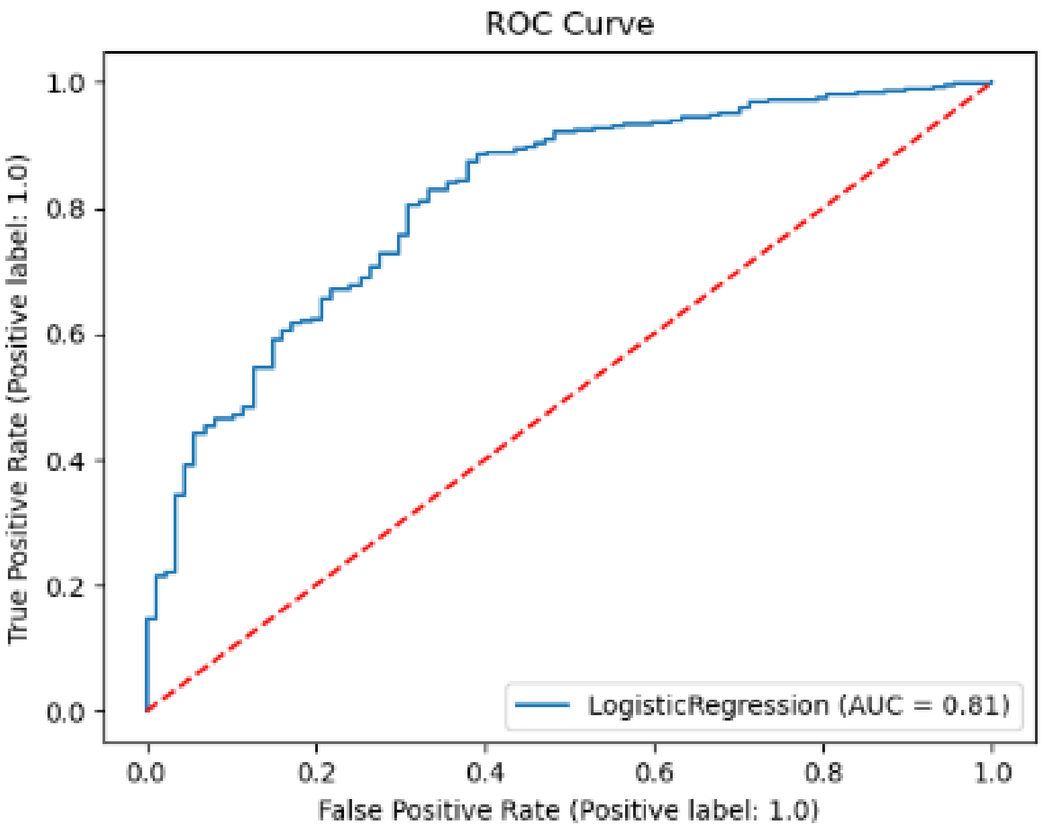


Logistic Regression Classification Report Before Holdout

	precision	recall	f1	support
0	0.71	0.23	0.35	472
1	0.87	0.98	0.92	2459
accuracy			0.86	2931
macro avg	0.79	0.61	0.63	2931
weighted avg	0.84	0.86	0.83	2931

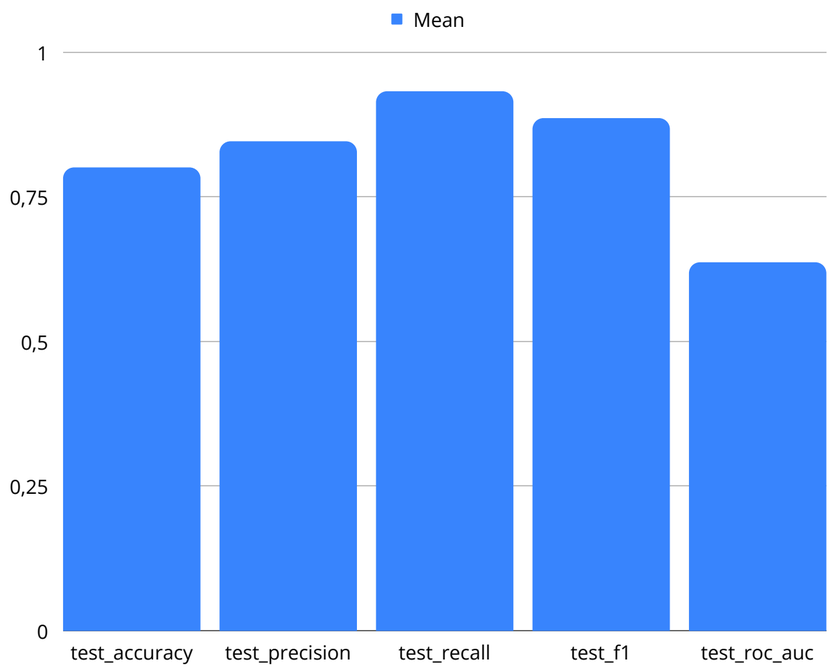
Logistic Regression Classification Report After Holdout

	precision	recall	f1-score	support
0	0.63	0.2	0.3	87
1	0.88	0.98	0.92	500
accuracy			0.86	587
macro avg	0.75	0.59	0.61	587
weighted avg	0.84	0.86	0.83	587



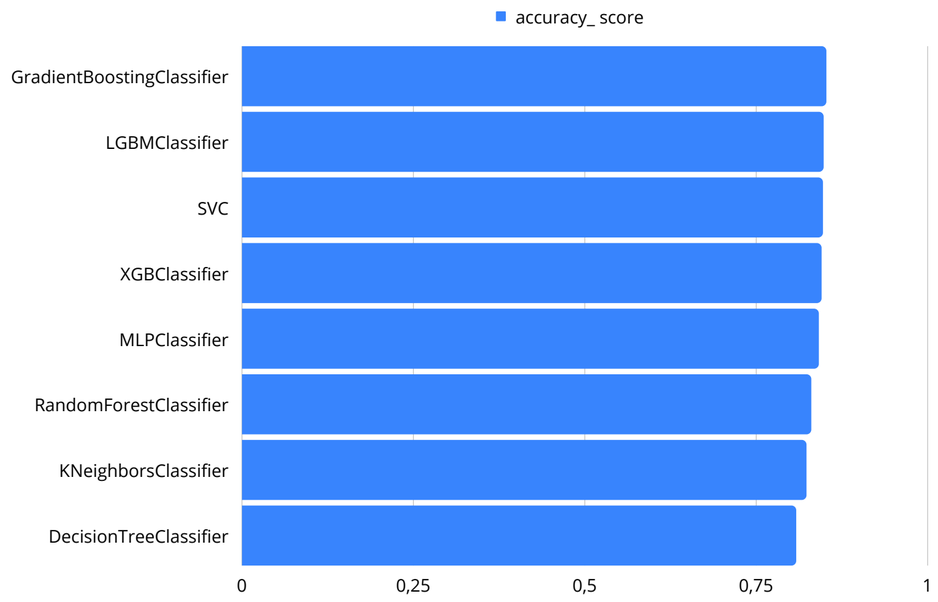
After Logistic Regression model validation, the score decreased to 0.814, but we prevented overfitting.

Logistic Regression Classification Report After 5-Fold Cross Validation



Logistic Regression Classification Report After 5-Fold Cross Validation	
Score	Mean
test_accuracy	0.800395951
test_precision	0.845584792
test_recall	0.932035998
test_f1	0.885621353
test_roc_auc	0.636448024

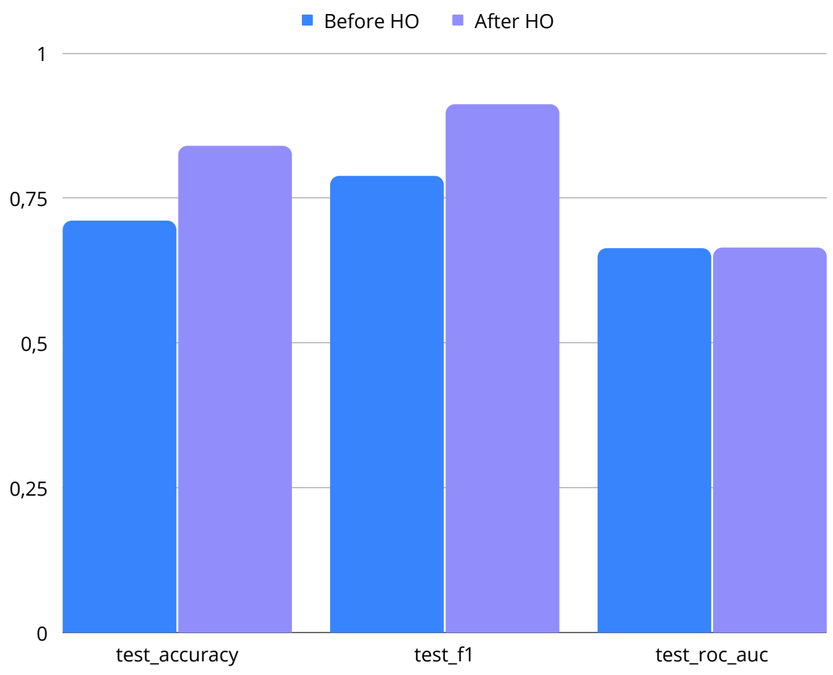
ML Models Accuracy Scores



ML Models Accuracy Scores

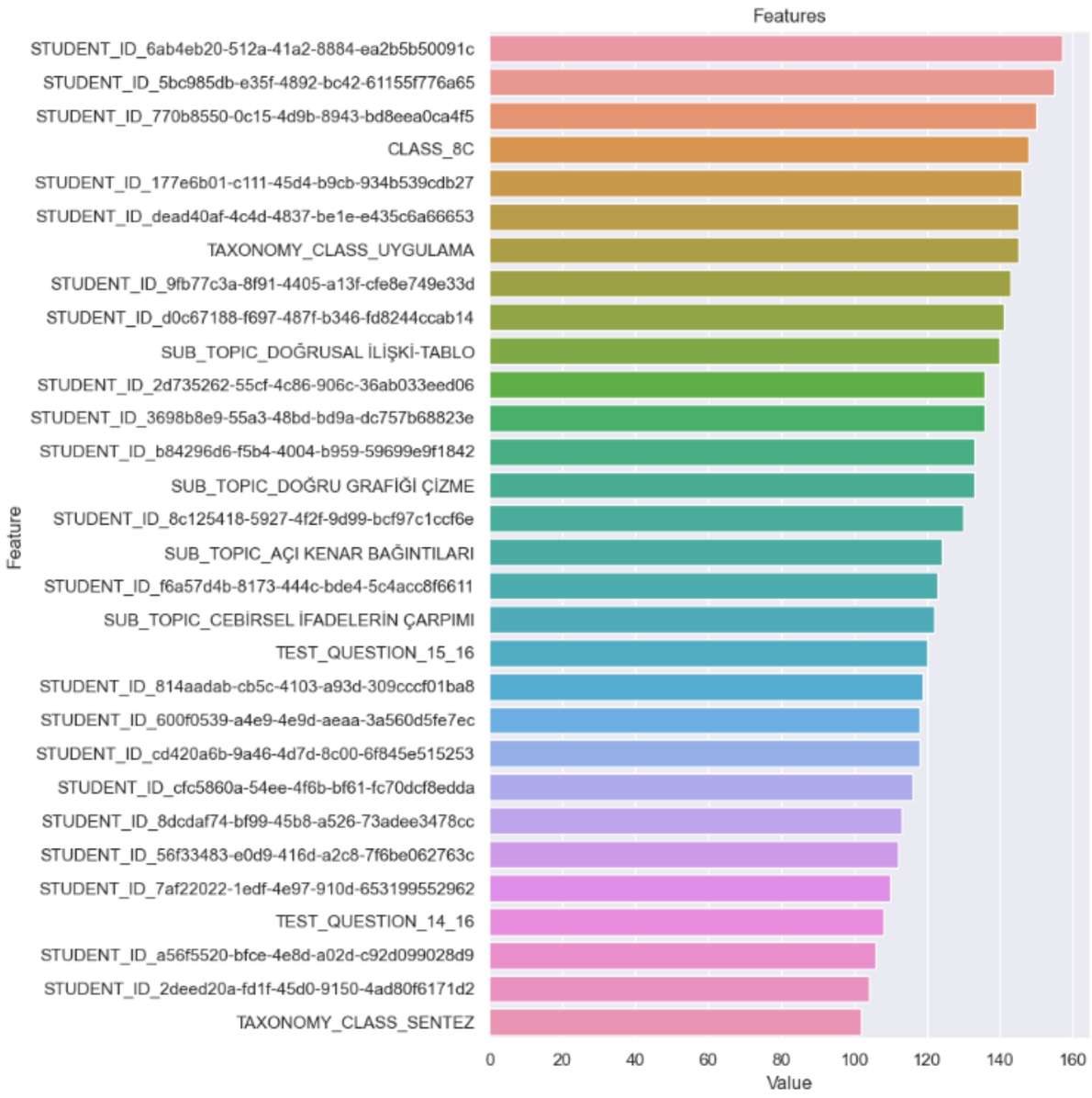
Model	accuracy_ score
GradientBoostingClassifier	0.852
LGBMClassifier	0.848
SVC	0.847
XGBClassifier	0.845
MLPClassifier	0.841
RandomForestClassifier	0.83
KNeighborsClassifier	0.823
DecisionTreeClassifier	0.808

Light GBM Classifier Scores Before After Hyperparameter Optimization

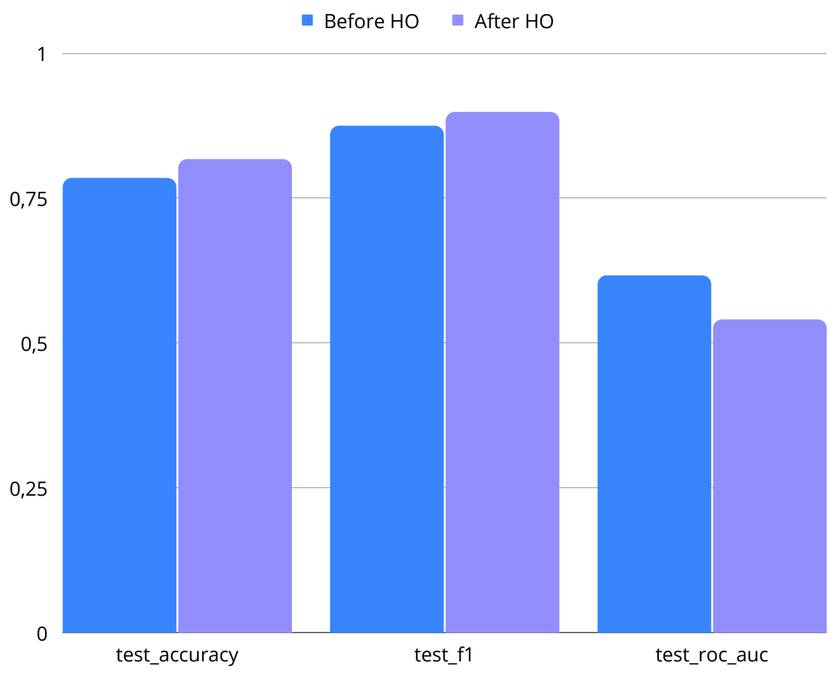


Light GBM Classifier Scores		
Score	Before Hyperparameter Optimization	After Hyperparameter Optimization
test_accuracy	0.71063776	0.83964393
test_f1	0.78783625	0.91145538
test_roc_auc	0.66291382	0.66398666

Feature Importance - LightGBM

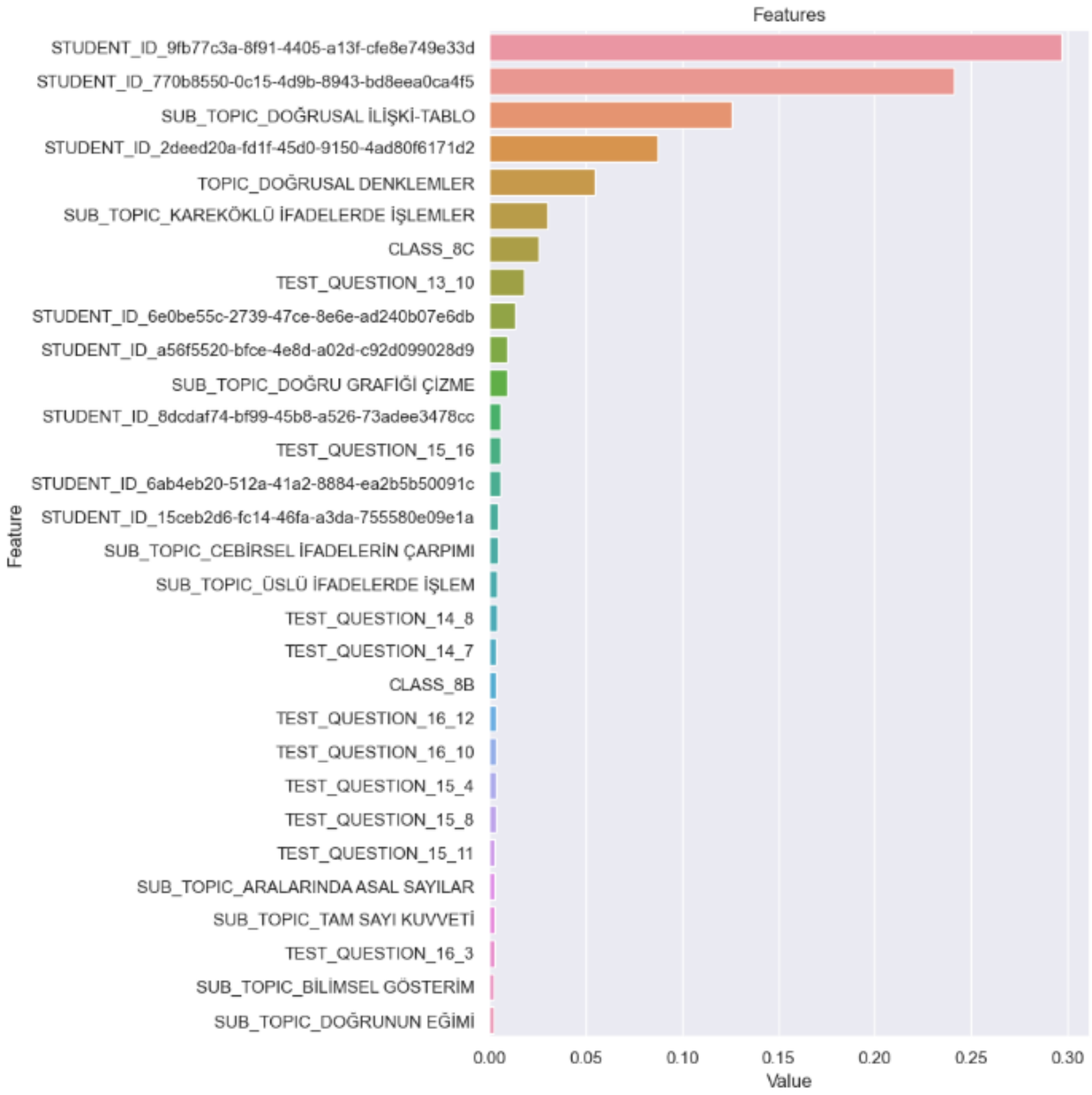


GBM Classifier Scores Before After Hyperparameter Optimization



GBM Classifier Scores		
Score	Before Hyperparameter Optimization	After Hyperparameter Optimization
test_accuracy	0.78435383	0.81677878
test_f1	0.87440491	0.89827907
test_roc_auc	0.61612311	0.54027143

Feature Importance - GBM



Quality Assurance Test

```
random_student = X.sample(5, random_state=4)
log_model.predict(random_student)
#770b8550-0c15-4d9b-8943-bd8eea0ca4f5
```

MEET THE TEAM!

Thank
you