Predict Student Performance from Game Play

MRCC

Table of contents

01

Project Goal & Data Overview

The goal of this competition is to predict student performance during game-based learning in real-time.

Data Source

Kaggle Dataset

Data Type (22 features)

Numerical: 7 (e.g. elapsed_time)

Categorical: 9 (e.g. level, event name)

Binary: 3 (e.g. hq, music)

Textual: 1

Explanation

The original data includes multiple rows for a session_id which stand for different events

Level group is relevant to questions: correct ordering in time

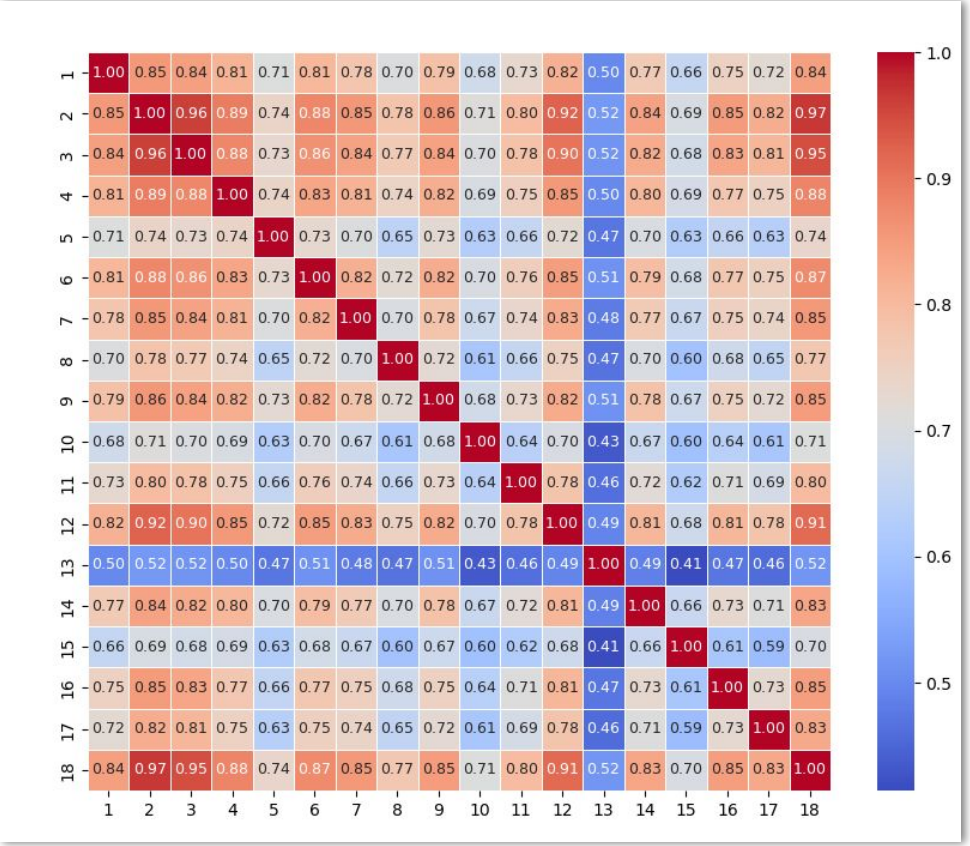
Data Manipulation

Elapsed_time isn't always monotonically increasing

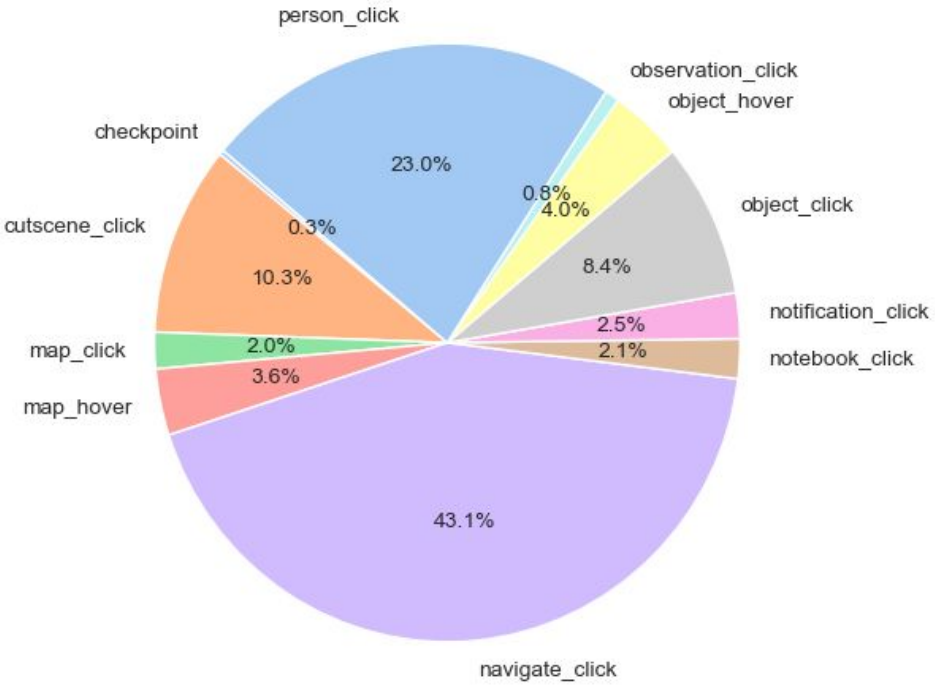
02

EDA

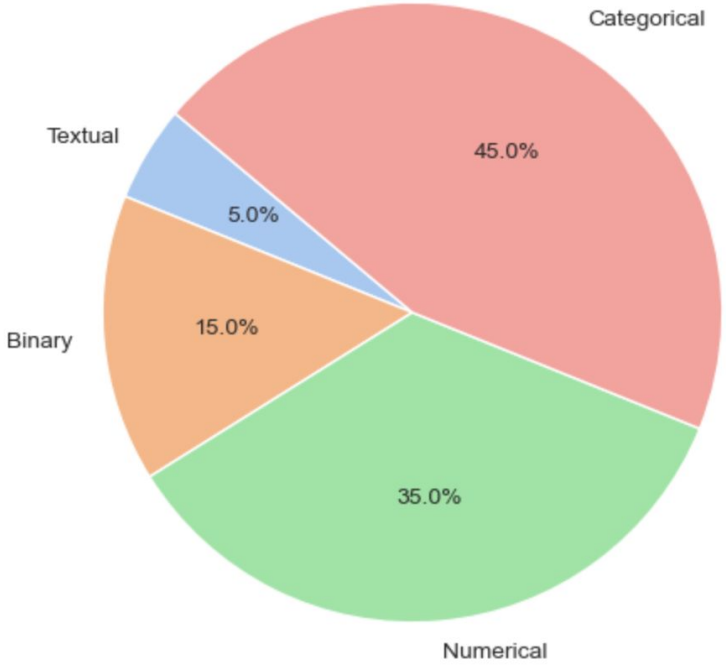
Correlation Between Each Question



Event Distribution



Distribution of Data Types



Hypothesis Testing

H0: questions correctness and fullscreen/hq/music are independent.

Chi-squared test p-value: 0.07411
Fail to reject the null hypothesis - there is no evidence of an association between listening to music and correctness rate.

Chi-squared test p-value: 0.00021
Reject the null hypothesis - there is an association between high quality mode of game and correctness rate.

Chi-squared test p-value: 0.51023
Fail to reject the null hypothesis - there is no evidence of an association between fullscreen and correctness rate.

03
Feature Engineering

Basic Aggregation Features (29 features)

ALL Group By [‘session_id’, ‘level_group’]

Categorical Columns: nunique(), count()
Numerical Columns: mean(), std(), sum()
Binary Columns: same value in each group

session_id	level_group	fqid_nunique	room_coor_x_mean	hover_duration_mean	hover_duration_std	person_click_sum
20090312431273200	0-4	30	7.701275	2389.500000	3227.370757	22.0
20090312431273200	13-22	49	-130.347170	899.925926	1305.088265	123.0
20090312431273200	5-12	39	14.306062	969.333333	1316.408315	104.0
20090312433251036	0-4	22	-84.045960	1378.750000	2114.876406	18.0
20090312433251036	13-22	73	-30.762282	720.384921	1990.705518	145.0
20090312433251036	5-12	45	50.284171	824.096774	1836.236232	97.0

Other Features (46 features)

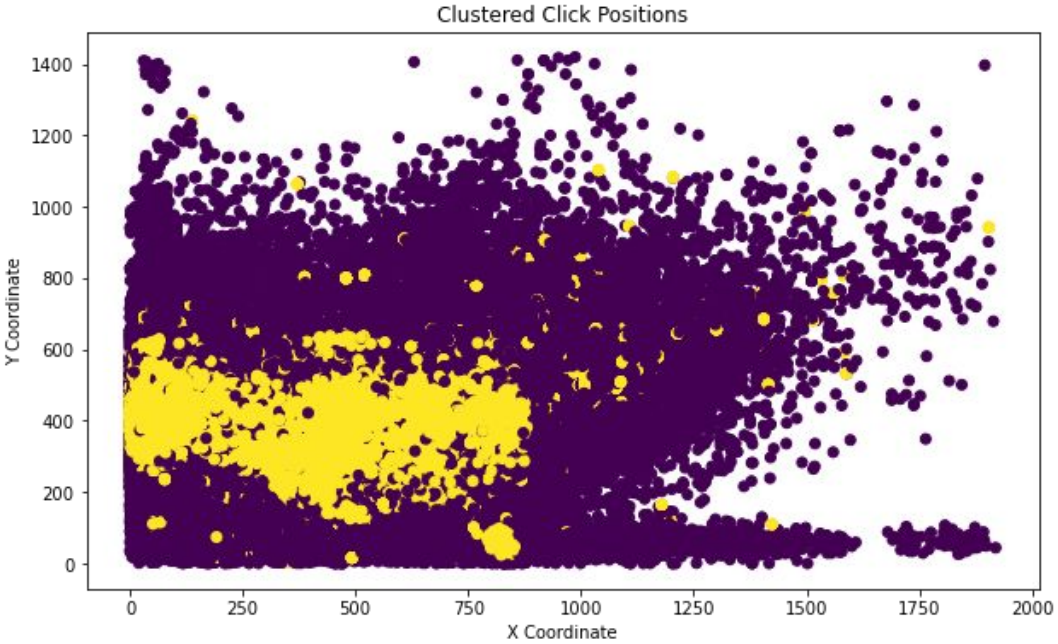
ALL Group By [‘session_id’, ‘level_group’]

- Elapsed Time (e.g. aggregation of time diff(), text reading speed)
- Ratio of Change (e.g. swapping rate, room change rate)
- Bingo

session_id	level_group	elapsed_time_over10000_ratio	level_swapping_ratio	room_change_ratio	bingo_time_mean	first_bingo_elapsed_time
22090614264755148	13-22	0.001637	0.0	0.056511	1.709010e+05	1813262.0
22090618154452916	13-22	0.022573	0.0	0.067873	3.304335e+05	2045962.0
22090618545074750	13-22	0.021164	0.0	0.068966	3.110298e+05	2537157.0
22090619362224080	13-22	0.019678	0.0	0.098566	1.191636e+06	9665985.0
22100212552203824	13-22	0.009560	0.0	0.048804	1.580739e+07	159336240.0
22100213081672770	13-22	0.001742	0.0	0.048866	1.976499e+05	1247161.0
22100215460321130	13-22	0.002203	0.0	0.086093	2.605925e+05	2003283.0
22100221145014656	13-22	0.005263	0.0	0.080369	4.473640e+05	5205501.0

Special Feature - Click Pattern

DBScan



04

Modeling

Step 1: Model selection - XGBoost

- Our dataset is not that big, which makes it easy to be influenced by the noise. Our main goal is to construct a robust model.
- The output of tree-based models can be easily explained and understood, enhancing the interpretation of our model.
- Furthermore, our job is to build a binary classification model. Tree-based models perform well on such realms.

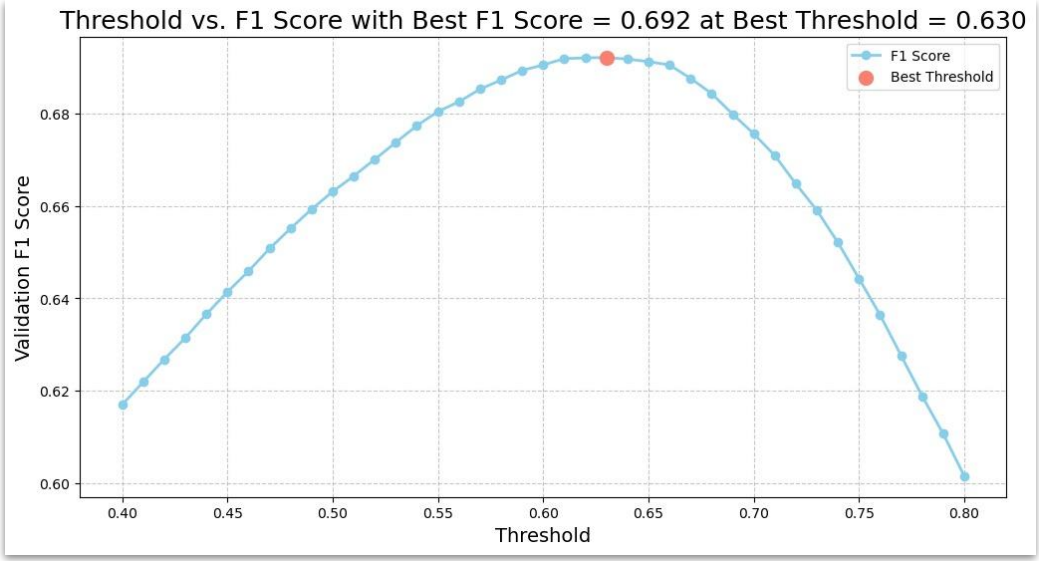
Step 2: Model Training Code Structure Design

- Take the features we construct as the input, the binary correctness of the question as the response
- Train and evaluate XGBoost models in one run for each question(total: 18) through for loop

Step 3: Hyperparameter Tuning (Grid Search)

- learning rate
- max_depth
- n_estimators

Step 4: Best Threshold Finding



05

Evaluation

Feature Importance

FEATURE	Scores (1-the most important)
elapsed_time_mean	1.0
tunic.wildlife.center_room_elapsed_time	2.0
elapsed_time_over200000_ratio	2.4
first_bingo_elapsed_time	3.0
elapsed_time_over100000_ratio	3.4
navigate_click_sum	3.5
map_click_sum	4.0
object_click_sum	4.0
tunic.capitol_0.hall_room_elapsed_time	4.5
tunic.flaghouse.entry_room_elapsed_time	5.0

FEATURE	Scores (1-the least important)
hq	1.0
level_swapping_ratio	1.4
bingo_time_mean	1.7
elapsed_time_over500000_ratio	2.0
tunic.capitol_0.hall_room_elapsed_time	2.6
first_bingo_elapsed_time	2.7
tunic.capitol_2.hall_room_elapsed_time	2.7
tunic.capitol_1.hall_room_elapsed_time	3.7
tunic.flaghouse.entry_room_elapsed_time	3.9
tunic.historicalsociety.closet_room_elapsed_time	4.6

Model Performance

F1 score for each model and each question:

Question1	0.662827
Question2	0.504746
Question3	0.499783
Question4	0.669210
Question5	0.619797
Question6	0.642213
Question7	0.629237
Question8	0.548865
Question9	0.628417

Question10	0.558173
Question11	0.614161
Question12	0.496265
Question13	0.446155
Question14	0.605364
Question15	0.528049
Question16	0.446719
Question17	0.543879
Question18	0.494333

Overall F1 score:

0.693