



**Институт
интеллектуальных кибернетических систем
Кафедра №22 «Кибернетика»**

Отчет

по дисциплине «Анализ данных и машинное обучение»

**На тему: «Анализ результатов ЕГЭ школ Москвы и выявление динамики
степени обучения за последние 5 лет»**

Выполнил	Старков С.В.
Группа	М21-524
Преподаватель	Киреев В.С.
Оценка	

Москва 2022

Содержание

1.Цель проекта	2
2.Источники данных	2
3.Используемые методы	3
4.Анализ и результаты	3
Приложение 1 (ссылка на google.colab).....	6
Приложение 2а (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для первого кластера).....	6
Приложение 2б (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для второго кластера).....	6
Приложение 2в (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для третьего кластера).....	7
Приложение 2г (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для четвертого кластера)	7

1 Цель проекта

Образование в школе постоянно модернизируется. Единый государственный экзамен претерпевает изменение, как и подготовка к нему и уровень знаний подготавливаемых учеников. В данной работе рассматривается динамика изменения процентного количества учеников, сдающих ЕГЭ хорошо и плохо за разные учебные года школ Москвы.

2 Источники данных

Для анализа были использованы данные, взятые из открытого источник <https://data.mos.ru/opendata/7719028495-rezultaty-ege-dogm>

Предоставленные источником данные имели следующие данные

- Название государственного учреждения
- Учебный год
- Административный округ Москвы
- Район

- Количество выпускников, сдавших ЕГЭ по сумме 3 экзаменов больше чем на 220 баллов
- Количество выпускников, сдавших ЕГЭ по сумме 3 экзаменов меньше чем на 160 баллов

Пример данных предоставлен ниже.

Рис.1. Пример необработанных данных. Статистика сдачи ЕГЭ школами Москвы

3 Используемые методы

Для первоначальной обработки данных использовались:

- pandas
- numpy

В рамках кластерного анализа использовались метод Kmeans Для построения графиков использовался matplotlib.pyplot

4 Анализ и результаты

Алгоритм обработки данных представлял собой следующее:

- найти школы, которые предоставляли данные о результатах ЕГЭ весь временной отрезок, представленный в файле

	A	B	C	D	E	F	G	H
1	ID	EDU_NAME	global_id	YEAR	AdmArea	District	PASSES_OVER_220	PASSER_UNDER_160
2	8236	Государственное бюджетное общеобразовательное учреждение города Москвы «Школа № 390 имени ген	797946035	2016-2017	Восточны	район Бог	5	19
3	8237	Государственное казенное общеобразовательное учреждение города Москвы кадетская школа-интернат	797949150	2016-2017	Восточны	район Бог	14	47
4	8238	Государственное казенное общеобразовательное учреждение города Москвы «Специальная (коррекцион	797949275	2016-2017	Восточны	район Вел	0	1
5	8239	Частное учреждение общеобразовательная организация «Московская интернациональная школа «Плане	797949425	2016-2017	Восточны	район Вел	0	6
6	8240	Автономная некоммерческая организация средняя Общеобразовательная школа «Колибри»	797949570	2016-2017	Восточны	район Вос	1	5
7	8241	Государственное бюджетное общеобразовательное учреждение города Москвы «школа № 664»	797949715	2016-2017	Восточны	район Вос	8	27
8	8242	Государственное бюджетное образовательное учреждение города Москвы школа № 922	797949790	2016-2017	Восточны	район Ива	16	41
9	8243	Государственное бюджетное общеобразовательное учреждение города Москвы центр образования № 14	797949791	2016-2017	Восточны	район Изг	2	7
10	8244	Общеобразовательная автономная некоммерческая организация «Лицей духовной культуры во имя преп	797949792	2016-2017	Восточны	район Изг	0	0
11	8245	Государственное бюджетное общеобразовательное учреждение города Москвы «Школа № 1602»	797949793	2016-2017	Восточны	район Кос	7	30

- перевести численные показатели выпускников, сдавших ЕГЭ больше чем на 220 баллов, и выпускников, сдавших ЕГЭ меньше чем на 160 баллов, в процентное отношение
- кластеризовать данные за первый предоставленный учебный год
- проследить динамику изменения центров кластеров за предоставленный временной период

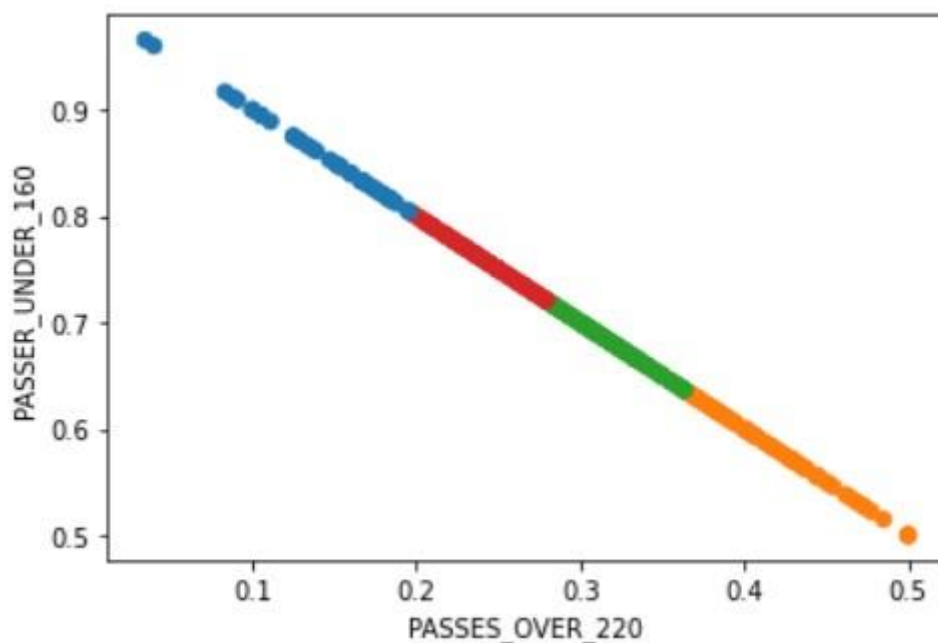


Рис.2. График распределения результатов ЕГЭ школ по кластерам

На графике выше изображены кластеры, получены в ходе использования алгоритма Ксредних (KMeans) с четырьмя выделенными кластерами.

Полученные кластеры можно характеризовать следующим образом:

- Синий кластер представляет обычные среднеобразовательные школы, не ориентирующиеся на показатели ЕГЭ выпускников
- Оранжевый кластер представляет школы с уклоном, лицеи, в чьих интересах получить большее количество выпускников, сдавших ЕГЭ лучше, чем остальные
- Красный и зеленый кластеры представляют собой среднее между синим и оранжевым кластерами.

Далее для полученных кластеров были найдены их новые центры за последующие периоды времени, предоставленные в исходных данных. Общий график изменения положения центров кластеров изображен на рисунке 3. Графики динамики кластеров по годам отдельно друг от друга представлены в приложениях 2а, 2б, 2в и 2г. Изображенные цифры на графике обозначают изменение годов, таким образом 0 обозначает данные за 2016-2017 учебный год, 1 – 2017-2018 учебный год, ..., 4 обозначает данные за 2020-2021 учебный год.

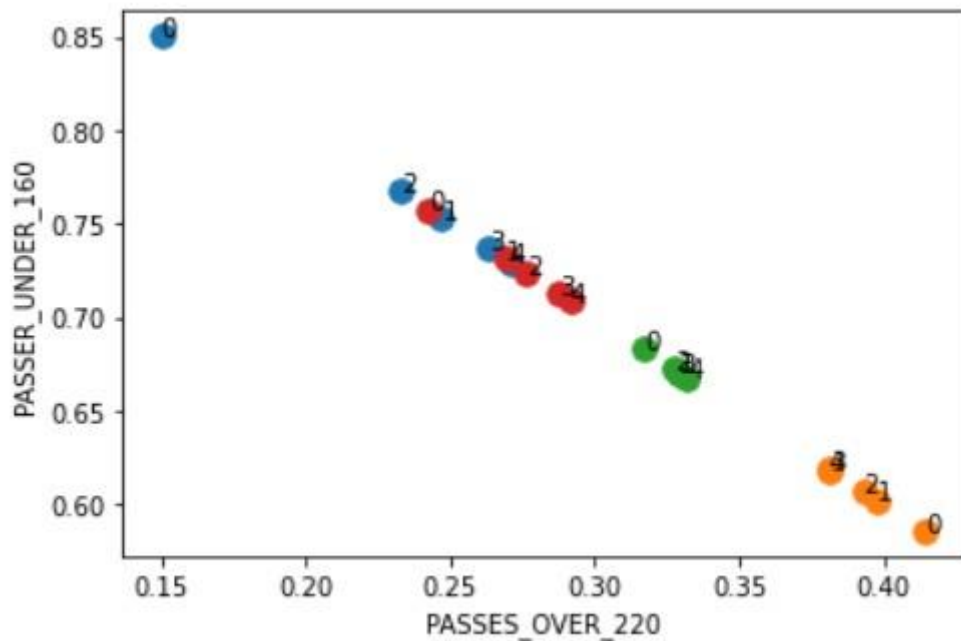


Рис.3. Динамика изменения показателей школ по сдаче ЕГЭ

Интерпретировать полученные результаты можно следующим образом:

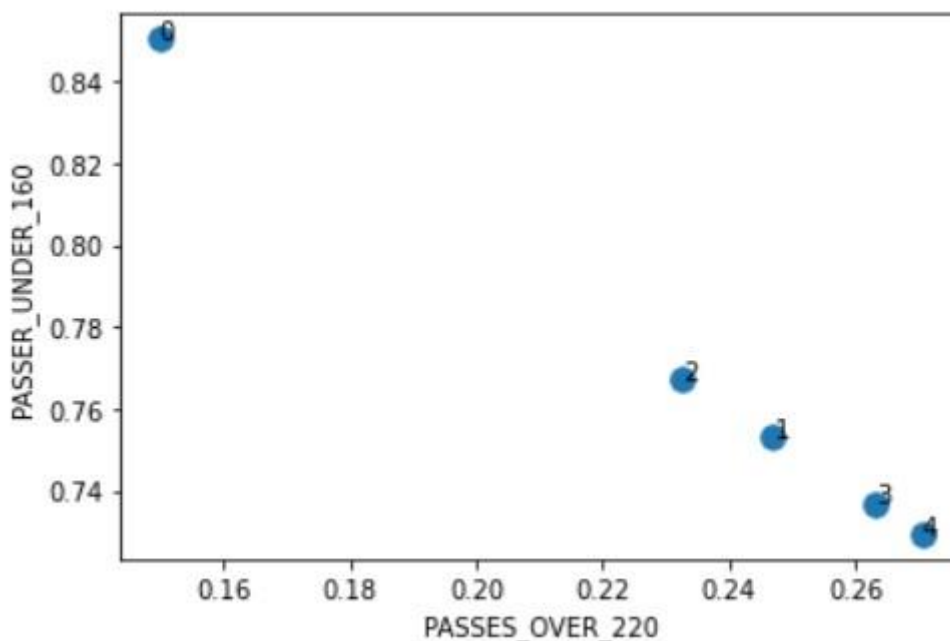
- Школы, попавшие в синий кластер, заметно улучшили свои показатели по ЕГЭ за последние пять лет и добрались до ближайшего по уровню оценки ЕГЭ красному кластеру.
- Школы с углубленным изучением и лицеи, представленные в оранжевом кластере наоборот же, немного ухудшили результаты выпускников ЕГЭ.
- Школы, попавшие в красный кластер, улучшили результаты ЕГЭ примерно так же, как и школы из синего кластера.
- Школы из зеленого кластера оставили уровень подготовки ЕГЭ на прежнем уровне.

Основной вывод по проведенному исследованию можно сделать следующий: школы в Москве в среднем усредняют знания школьников по разным школам.

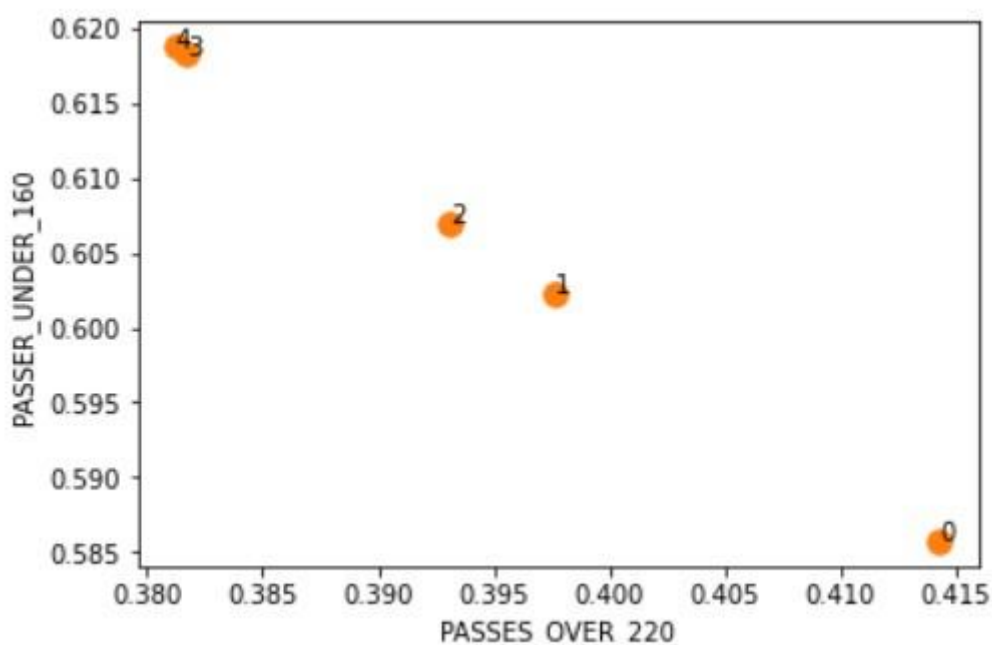
Приложение 1 (ссылка на google.colab)

https://colab.research.google.com/drive/1xrPwmvqAZT2xFmM-k539C2zIVrsO51_g?usp=sharing

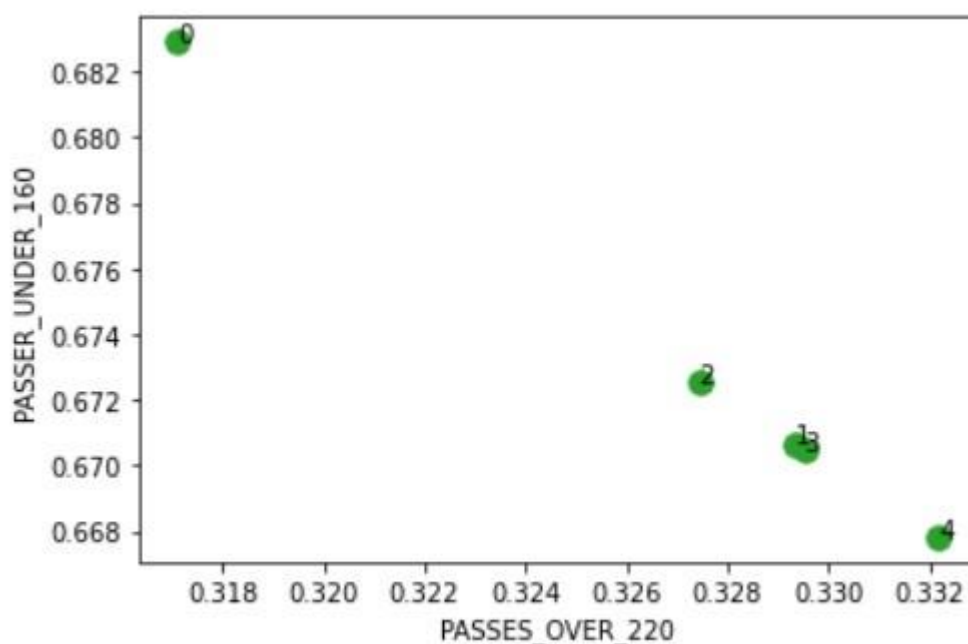
Приложение 2а (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для первого кластера)



Приложение 2б (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для второго кластера)



Приложение 2в (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для третьего кластера)



Приложение 2г (динамика изменения процентного изменения выпускников, сдавших ЕГЭ больше 220, и выпускников, сдавших ЕГЭ меньше 160, за предоставленные 5 лет для четвертого кластера)

