

Российский государственный педагогический университет им. А. И. Герцена
Институт информационных технологий и технологического образования

Прогнозирование потенциальных показателей загружаемых видеороликов на основе обработки больших данных

Выполнил:

студент

Шумякин Илья Сергеевич

Руководитель:

к. ф.-м. н., доцент кафедры ИТиЭО

Власов Дмитрий Викторович

Санкт-Петербург, 2025

Актуальность

Работа актуальна в связи с увеличением потребности в аналитических подходах к публикации видеороликов среди создателей контента.

Предмет

Прогнозирование потенциальных показателей загружаемых видеороликов на основе обработки больших данных.

Цель работы

Разработка прогнозной методики для расчёта потенциальных показателей загружаемого видеоролика с помощью различных методов обработки больших данных.

Задачи

1. Исследовать особенности работы и структуру платформы YouTube.
2. Проанализировать и синтезировать имеющийся набор данных, импортированный с YouTube Studio.
3. Разработать методику для прогнозирования потенциальных показателей эффективности нового видеоролика, расчёта потенциального количества просмотров, часов просмотра и показов.
4. Провести вычисления на основе имеющихся данных.

Особенности и структура платформы YouTube

Рассмотрены основные технические и юридические стороны платформы, в том числе особенности интерфейса и функциональной алгоритмической базы.

Выявлено наличие автоматизированной системы сбора и анализа информации, основанной на искусственном интеллекте.

Установлено, что выборка видеороликов производится с учётом нюансов персонализации, показателей эффективности и грамотного оформления.

Анализ и синтез имеющегося набора данных

Выявлены критерии эффективности видеороликов, обнаружено наличие временных рядов и проведён анализ имеющихся методов работы с временными рядами.

Сделан вывод о целесообразности проведения обстоятельного математического исследования.

Анализ и синтез имеющегося набора данных

Коэффициент корреляции Пирсона (Видео 1 (X) и 2 (Y) - Возраст - APV)									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) 2	(Y- \bar{Y}) 2	(X- \bar{X})(Y- \bar{Y})			
17.24	15.33	-0.44	-0.81	0.19	0.65	-0.35			
15.33	13.32	-2.35	-2.72	5.51	7.38	6.37			
17.32	15.35	-0.36	-0.69	0.24	0.13	0.47			
16.40	14.90	-1.28	-1.14	1.63	1.29	1.45			
15.92	14.49	-1.76	-1.35	3.09	1.81	2.36			
18.43	17.77	0.75	1.75	1.31	3.01	1.31			
23.10	20.99	5.42	4.95	26.87	29.41	24.54			
Mean:	17.68	16.04	SUM	38.96	40.52	39.15	0.978158		

Коэффициент корреляции Пирсона (Видео 1 (X) и 2 (Y) - Возраст - Views)									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) 2	(Y- \bar{Y}) 2	(X- \bar{X})(Y- \bar{Y})			
12.5	12.1	-1.78	-2.15	3.19	4.61	7.42			
47.2	47.7	32.93	33.37	1084.83	1113.75	1098.63			
20.2	20.2	5.94	5.92	35.16	35.08	35.16			
12.6	11.9	-1.87	-2.43	4.06	5.89	6.79			
6.7	7.29	-7.58	-7.00	53.07	57.52	48.96			
0.35	0.39	-13.93	-13.90	193.65	194.16	193.13			
0.38	0.46	-13.90	-13.03	192.26	193.33	192.19			
Mean:	14.28	14.29	SUM	358.61	3570.27	3592.61	0.999625		

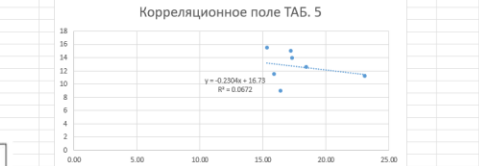
Коэффициент корреляции Пирсона (Видео 1 - Возраст (X) и Страны (Y) - APV)									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) 2	(Y- \bar{Y}) 2	(X- \bar{X})(Y- \bar{Y})			
17.24	15	-0.44	2.30	0.19	5.30	-1.01			
15.33	15.5	-2.35	2.80	5.51	7.84	-6.58			
17.32	14	-0.36	1.29	0.46	0.13	0.47			
16.40	8.95	-1.28	-3.71	1.63	13.74	4.73			
15.92	11.5	-1.76	-1.14	3.09	1.29	2.00			
18.43	12.6	0.75	-0.09	0.57	0.01	-0.07			
23.10	11.2	5.42	-1.47	26.87	2.15	-7.96			
Mean:	17.68	12.66	SUM	-9.34	40.52	32.03	-0.259164		

Диапазон значений r	Линейная корреляционная зависимость Y от X
0-0.1	практически отсутствует
0.1-0.3	слабая
0.3-0.5	умеренная
0.5-0.7	заметная
0.7-0.9	сильная
0.9-0.99	очень сильная
0.99-1	практически функциональная

Типы сравнимых шкал		Коэффициент корреляции	
1	2		
Дихотомическая	Дихотомическая	Дихотомический (ϕ)	
Дихотомическая	Порядковая (ранговая)	Рангово-бисериальный (r_{pb})	
Дихотомическая	Интервальная	Точечно-бисериальный (r_{tpb})	
Ранговая	Ранговая	Коэффициент Пирсона (r_p)	
		Коэффициент Спирмена (r_s)	
		Коэффициент Кендалла (τ)	
Ранговая	Интервальная	Коэффициент Пирсона (r_p)	
Интервальная	Интервальная	Коэффициент Пирсона (r_p)	

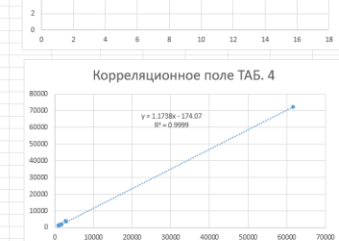
Коэффициент корреляции Пирсона (Видео 1 (X) и 2 (Y) - Страны - APV)									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) 2	(Y- \bar{Y}) 2	(X- \bar{X})(Y- \bar{Y})			
14.96	12.97	2.48	1.70	4.20	6.13	2.88			
13.95	11.47	0.97	0.20	0.19	0.93	0.04			
8.95	7.9	-4.03	-3.37	13.61	16.27	11.38			
11.52	10.67	-1.46	-0.80	0.88	2.14	0.36			
12.57	9.67	-0.41	-1.60	0.66	0.17	0.27			
11.19	10.51	-1.79	-0.76	1.37	3.22	0.58			
11.25	10.55	-1.73	-0.72	1.25	3.00	0.52			
17	14.86	4.02	1.39	14.41	16.19	12.88			
Mean:	12.98	11.27	SUM	39.71	51.93	33.72	0.949135		

Коэффициент корреляции Пирсона (Видео 1 (X) и 2 (Y) - Страны - Views)									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) 2	(Y- \bar{Y}) 2	(X- \bar{X})(Y- \bar{Y})			
61664	72197	52288.67	62551.89	323922275.36	24074788.44	2912738803.57			
2947	3307	-5418.33	-6338.11	3441998.70	2938336.11	4017652.46			
2836	3592	-5529.33	-6053.11	3346969.04	30573527.11	36640154.12			
1801	1832	-6564.33	-7813.11	51287865.70	43090472.11	61044705.23			
1608	1464	-6757.33	-8181.11	55282494.81	45661533.78	66930579.01			
1373	1423	-6992.33	-8222.11	57491741.59	48892725.44	67603111.12			
1136	1108	-7229.33	-8537.11	61717621.59	52263280.44	72882266.12			
988	939	-7377.33	-8706.11	64277883.70	54425047.11	75766370.68			
935	944	-7430.33	-8701.11	64652155.89	55209633.44	75709334.57			
Mean:	8365.33	9645.11	SUM	3756403706.67	3202222644.00	4809516976.89	0.999969		



Критические значения коэффициентов корреляции Пирсона и Спирмена

Критические значения коэффициентов корреляции Пирсона и Спирмена									
Степени свободы n-2	Уровень значимости	Степени свободы n-2		Степени свободы n-2		Степени свободы n-2		Уровень значимости	
		0.95	0.99	0.95	0.99	0.95	0.99	0.95	0.99
5	0.75	0.87	20	0.42	0.54	60	0.25	0.33	
6	0.71	0.83	21	0.41	0.53	70	0.23	0.30	
7	0.67	0.80	22	0.40	0.52	80	0.22	0.28	
8	0.63	0.77	23	0.40	0.51	90	0.21	0.27	
9	0.60	0.74	24	0.39	0.50	100	0.20	0.25	
10	0.58	0.71	25	0.38	0.49	125	0.17	0.23	
11	0.55	0.68	26	0.37	0.48	150	0.16	0.21	
12	0.53	0.66	27	0.37	0.47	200	0.14	0.18	
13	0.51	0.64	28	0.36	0.46	300	0.11	0.15	
14	0.50	0.62	29	0.36	0.46	400	0.10	0.13	
15	0.48	0.61	30	0.35	0.45	500	0.09	0.12	
16	0.47	0.59	35	0.33	0.42	700	0.07	0.10	
17	0.46	0.58	40	0.30	0.39	900	0.06	0.09	
18	0.44	0.56	45	0.29	0.37	1000	0.06	0.09	
19	0.43	0.55	50	0.27	0.35				



Geography	Average percentage viewed (%)					Views					Watch time (hours)				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
US	14.96	12.86	12.63	13.55	16.19	61664	72197	67618	66435	61742	13471	13325	13119	13714	12120
CA	15.46	12.97	12.94	13.3	16.92	2947	3307	3381	3096	2819	487	416	612	627	587
GB	13.95	11.47	11.64	12.31	15.11	2836	3592	3216	3018	2871	492	591	575	569	526
IN	8.95	7.9	7.84	8.57	10.26	1801	1832	1992	1813	1757	201	208	240	237	219
MX	11.52	10.67	10.36	11.71	14.1	1608	1464	1444	1234	1232	230	224	230	220	211
PH	12.57	9.67	9.92	10.75	13.78	1373	1423	1294	1064	1282	215	197	197	174	214
ID	11.19	10.51	12.08	12.53	13.45	1136	1108	894	948	923	158	167	166	181	151
PL	11.25	10.55	10.46	10.12	13.31	988	939	822	963	770	138	142	132	148	124
DE	17	14.86	15.28	16.28	18.65	935	944	810	866	778	198	201	190	215	171

Race	Age/Gender	Average percentage viewed (%)					Views (%)					Watch time (hours) (%)				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	13-17 years	17.24	15.23	15.39	16.75	20.38	12.52	12.14	11.45	12.35	12.56	13.33	12.93	12.39	13.2	13.55
2	18-24 years	15.33	13.32	13.52	14.79	17.5	47.21	47.66	49.2	46.54	47.91	44.71	44.39	46.18	43.91	44.37
3	25-34 years	17.32	15.35	16.04	16.92	20.51	20.22	20.21	20.63	21.26	20.6	21.64	21.69	22.3	22.93	22.36
4	35-44 years	16.4	14.9	14.92	16.42	19.81	12.61	11.86	11.02	11.57	10.89	12.78	12.35	11.41	12.12	11.42
5	45-54 years	15.92	14.69	13.39	14.35	19.32	6.7	7.29	7.48	7.56	6.94	6.59	7.49	6.95	6.92	7.1
6	55-64 years	18.43	17.77	9.97	19.63	21.23	0.35	0.39	0.49	0.32	0.46	0.4	0.48	0.34	0.4	0.51
7	65+ years	23.1	20.99	19.07	20.26	20.18	0.38	0.46	0.33	0.39	0.64	0.54	0.67	0.44	0.51	0.68
1	Male	16.24	14.29	14.35	15.68	18.72	97.04	97.39	97.49	97.11	97.33	97.24	97.54	97.76	97.64	97.2
0	Female	15.27	13.41	12.76	12.74	19.67	2.96	2.61	2.51	2.89	2.67	2.78	2.46	2.24	2.36	2.8

Коэффициент корреляции Пирсона (Видео 1 - Возраст (X) и Views (Y))									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) 2	(Y- \bar{Y}) 2	(X- \bar{X})(Y- \bar{Y})			
17.24	12.1	-0.44	-2.15	0.19	4.61	-0.95			
15.33	47.21	-2.35	32.93	5.51	1084.83	-77.28			
17.32	20.22	-0.36	5.94	0.24	0.13	0.47			
16.40	12.61	-1.28	-1.67	1.63	1.29	2.14			
15.92	6.7	-1.76	-7.58	3.09	57.52	13.33			
18.43	0.35	0.75	-13.93	0.57	193.65	-10.49			
23.10	0.38	5.42	-13.90	26.87	194.16	-75.40			
Mean:	17.68	14.28	SUM	-149.06	40.52	1570.27	-0.590917		

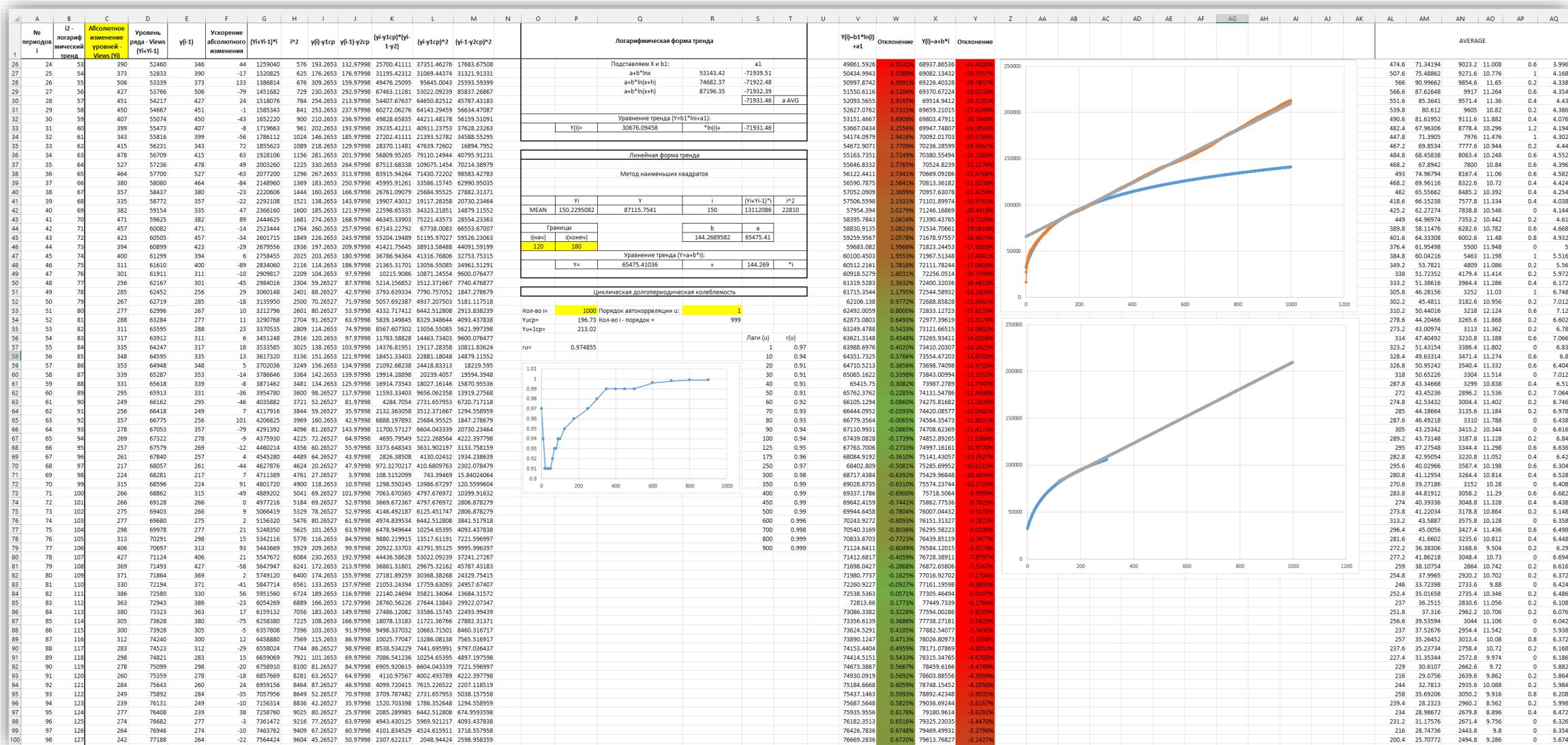
Коэффициент корреляции Пирсона (Видео 1 - Возраст - Views (X) и Watch Time (Y))									
X	Y	X- \bar{X}	Y- \bar{Y}	(X- \bar{X}) ²	(Y- \bar{Y}) ²	(X- \bar{X})(Y- \bar{Y})			
12.52	13.33	-1.76	-0.95	1.68	3.11	0.91			
47.21	44.71	32.93	30.43	1081.79	1044.10	925.72			
20.22	21.64	5.94	7.36	35.28	35.28	54.11			
12.61	12.78	-1.67	-1.50	2.52	2.80	2.26			
6.7	6.59	-7.58	-6.69	58.36	57.52	59.20			
0.35	0.4	-13.93	-13.88	193.47	194.16	192.77			
0.38	0.54	-13.90	-13.73	191.10	190.33	188.91			
Mean:	14.28	14.28	SUM	1492.58	1570.27	1423.88			0.99819

Прогнозирование показателей эффективности

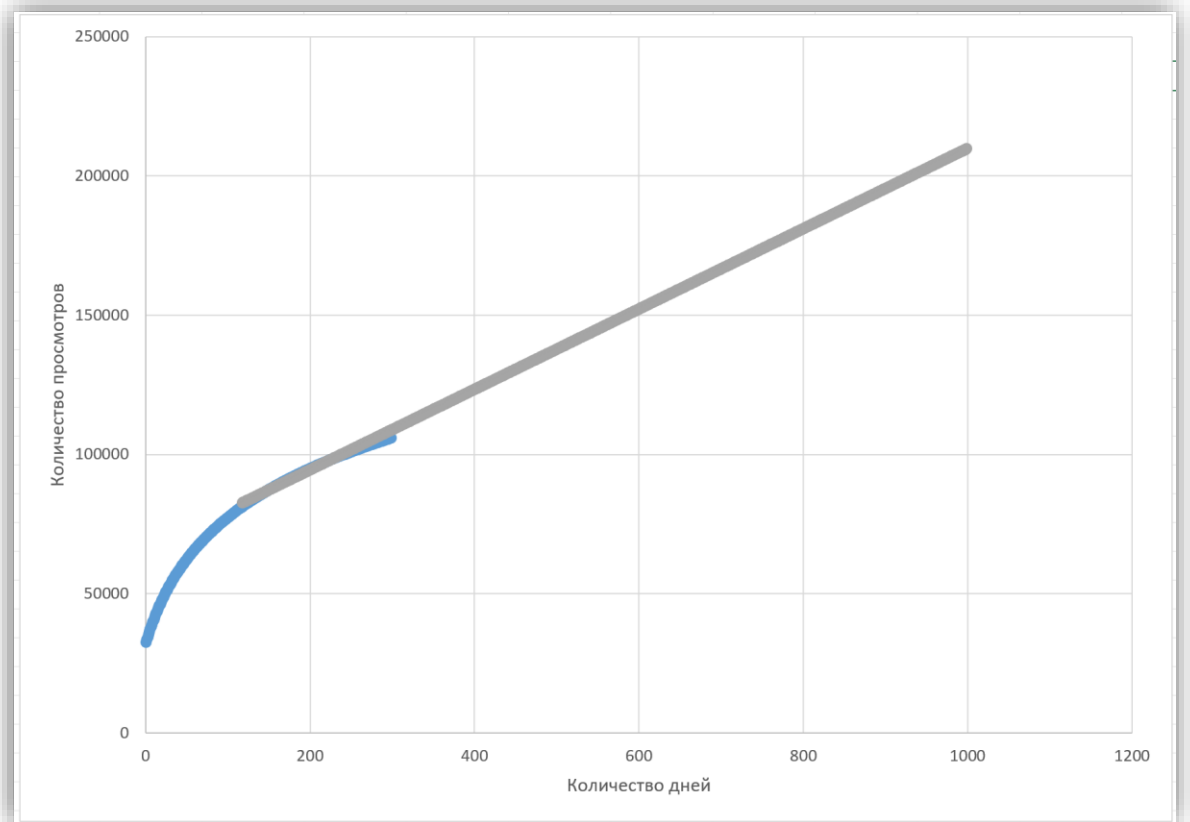
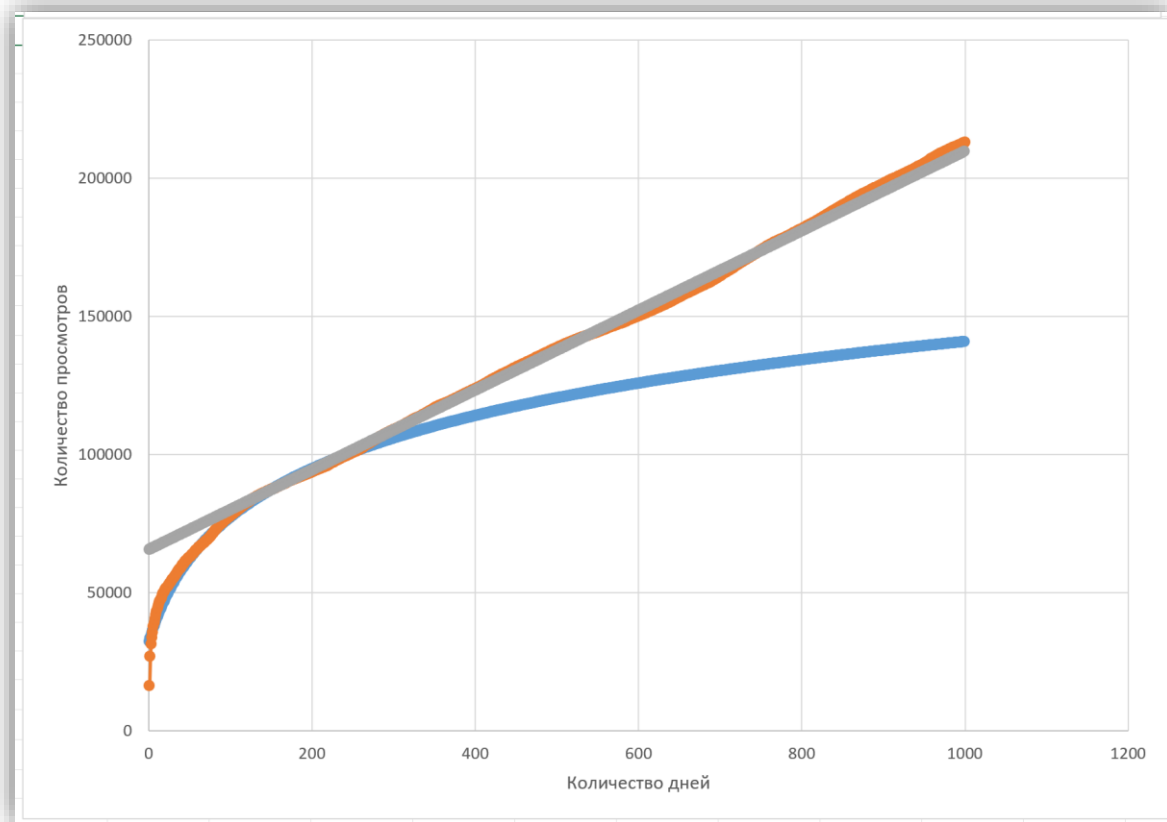
Разработана методика для прогнозирования потенциальных показателей эффективности нового видеоролика, расчёта потенциального количества просмотров, часов просмотра и показов.

Спроектированы и разработаны универсальные таблицы для идентификации предполагаемых уравнений тренда и построения кривых или прямых тренда.

Проведение вычислений



Результат



Публикации автора по теме ВКР

1. Шумякин И.С. Фундаментальные особенности работы системы алгоритмов рекомендаций видеоконтента на платформе YouTube // «Цифровая экосистема педагогического образования. Актуальные вопросы. Достижения. Инновации» / А. М. Атаян, Е. А. Барахсанова, К. О. Вехова и др.; Под науч. ред. Е. З. Власовой, Е. А. Барахсановой. – СПб.: Изд-во НИЦ АРТ, 2022. – 148 с.
2. Шумякин И.С. Извлечение и анализ показателей эффективности видеороликов с помощью сервиса YouTube Studio. Методы прогнозирования временных рядов // Современное образование: традиции и инновации. 2023. № 2. С. 230–234.