

Презентация

На тему: “Прогнозирование
и анализ музыкальных
композиций при помощи
машинного обучения”



МОСКОВСКИЙ
ПОЛИТЕХ

Д. А. Волченков
Группа 181-341

Актуальность исследования

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

Актуальность выбранной темы обусловлена ростом рынка и растущим спросом на алгоритмы машинного обучения, которые все теснее внедряются в музыкальную индустрию. При этом анализ и прогнозирование популярности музыкальных композиций крайне востребованная и часто используемая область в современных стриминговых платформах.

Объем рынка музыкальных записей в России, согласно прогнозам, вырастет в два раза к 2023 году и достигнет 160 млн долл. (для сравнения: в 2018 году объем рынка составлял 81 млн долл.), демонстрируя среднегодовые темпы роста в 14,6%. Интенсивное развитие рынка будет происходить преимущественно за счет увеличения выручки музыкальных стриминговых сервисов как от пользовательских платежей, так и от рекламы. Все стриминговые платформы используют алгоритмы, большинство машинное обучение

Цель исследования

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

Целью исследования в рамках выпускной квалификационной работы является понимание эффективности использования машинного обучения и инструментов анализа в музыкальной индустрии.

Применение различных инструментов для понимания алгоритмов формирования популярности определенных треков и непопулярности других.

Объектом исследования являются – музыкальные записи и их влияние на людей и особенности популярной музыки.

В результате исследования будут выделены параметры как наиболее влиятельные факторы на популярность музыкальных композиций

Задачи исследования

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

В рамках выпускной квалификационной работы ожидается выполнение следующих надлежащих ведущих задач:

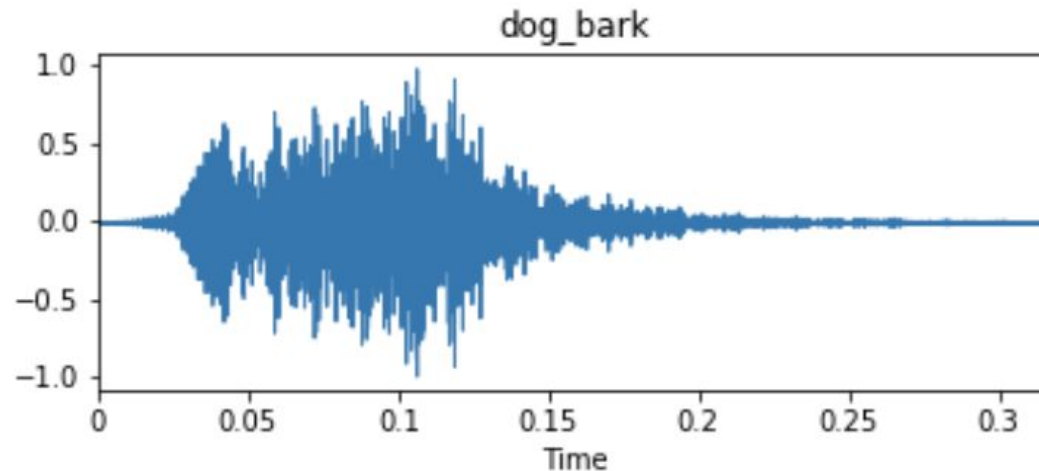
- 1) изучение основ музыкального творчества, способов записи музыки и алгоритмов подготовки цифровых записей;
- 2) исследование алгоритмов музыкальных рекомендательных систем и их математических характеристик песен;
- 3) анализ интеллектуального музыкального производства и влияния искусственного интеллекта на создание музыки;
- 4) изучение методологии музыкального анализа;
- 5) проведение анализа популярности музыки и создание алгоритмов;
- 6) изучение роли алгоритмов машинного обучения в музыке.

Методологическими инструментами научных исследований являются аналитический метод, метод сравнения, классификация, экономико-математический метод, графический метод обработки информации, алгоритмы машинного обучения.

Работа с первичными данными

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

Данные семпла без метаданных



Пример чистых данных для анализа

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

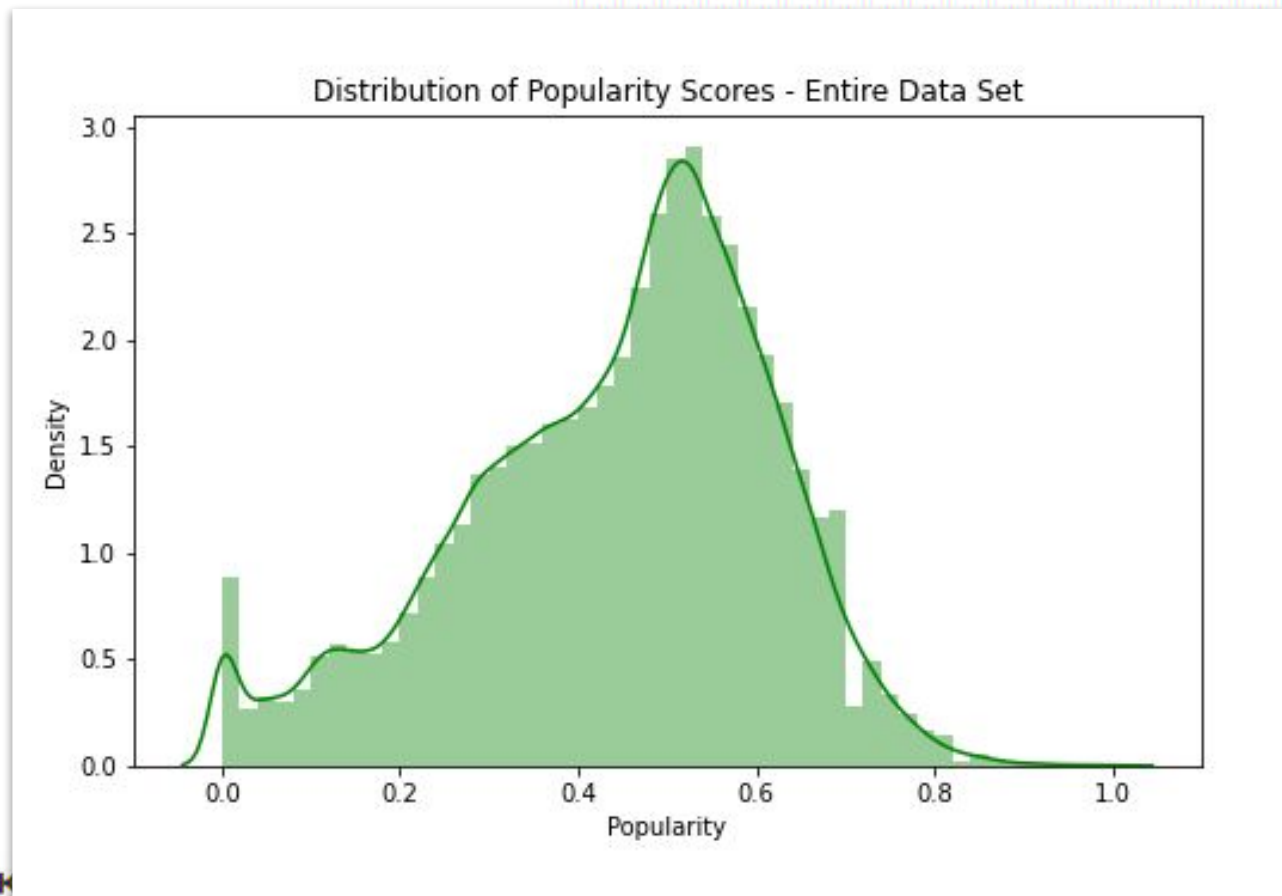
Готовый набор данных SpotifyFeatures в формате csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	genre	artist_name	track_name	trap	popul	acoust	dancea	duration	energy	instrumen	key	liveness	loudness	mode	speechi	tempo	time_si	valence
2	Opera	Giuseppe Verdi	Stiffelio, Act III: Ei fugge	7E	21	0.986	0.313	490867	0.231	0.000431	C#	0.0964	-14.287	Major	0.0547	86.001	04.anp	0.0886
3	Opera	Giacomo Puccini	Madama Butterfly / Act 1	7M	18	0.972	0.36	176797	0.201	0.028	D#	0.133	-19.794	Major	0.0581	131.798	04.anp	0.369
4	Opera	Giacomo Puccini	Turandot / Act 2: Gloria, 7p	7p	10	0.935	0.168	266184	0.47	0.0204	C	0.363	-8.415	Major	0.0383	75.126	03.anp	0.0696
5	Opera	Giuseppe Verdi	Rigoletto, Act IV: Venti s	02i	17	0.961	0.25	288573	0.00605	0	D	0.12	-33.44	Major	0.048	76.493	04.anp	0.038
6	Opera	Giuseppe Verdi	Don Carlo / Act 4: "Ella gi	03	19	0.985	0.142	629760	0.058	0.146	D	0.0969	-23.625	Major	0.0493	172.935	04.anp	0.0382
7	Opera	Giuseppe Verdi	D'amor sull'ali rosee	0G	20	0.99	0.211	334720	0.064	0.0196	G#	0.073	-20.666	Major	0.0534	81.403	03.anp	0.04
8	Opera	Georges Bizet	Waxman : Carmen Fanta	10i	13	0.98	0.341	646813	0.16	0.0146	D	0.113	-19.753	Major	0.0575	171.444	01.anp	0.3
9	Opera	Giuseppe Verdi	4 Pezzi sacri: No. 4. Te De	1ia	19	0.975	0.109	963440	0.018	0.791	F#	0.0712	-26.778	Major	0.0399	82.178	04.anp	0.0359
10	Opera	Vincenzo Bellini	I puritani: Ah! per semp	29	20	0.951	0.169	417987	0.178	0.185	C	0.665	-21.589	Minor	0.054	65.995	04.anp	0.112

Распределение данных

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

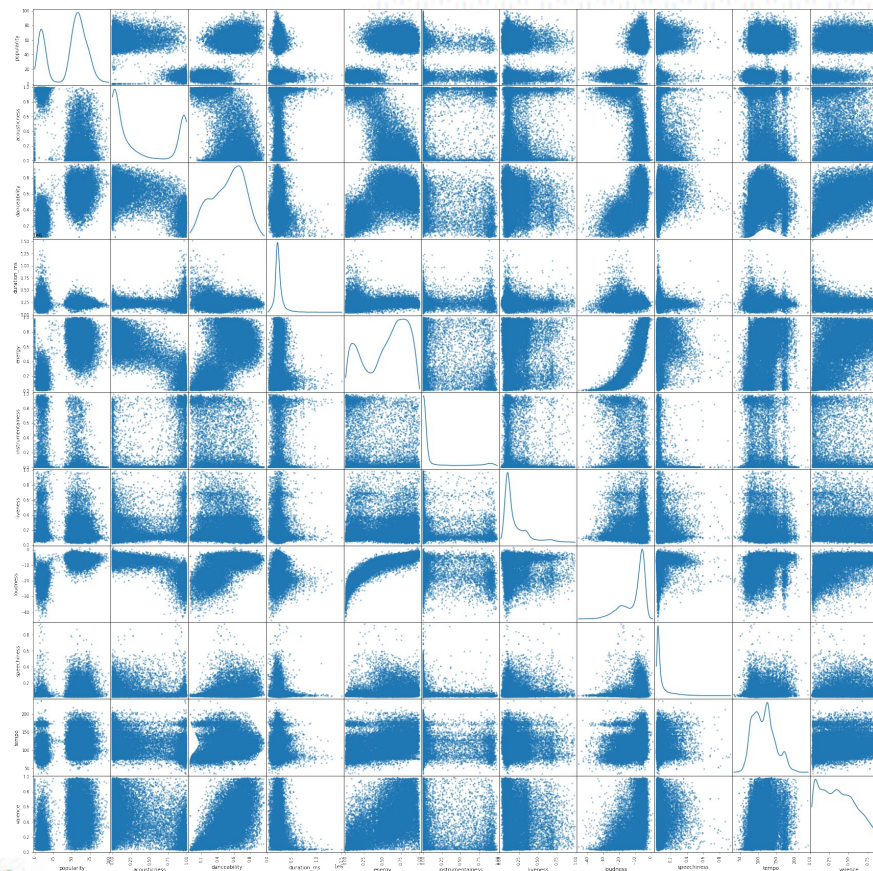
При анализе данных датасета оказалось, что данные распределены по популярности с очень большим перекосом в сторону песен со значением популярности 0.4-0.6. Это можно увидеть на графике представленном ниже



Проверка данных, определение корреляции

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

Матрица линейных корреляций



Результат обучения модели линейной регрессии

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

OLS Regression Results

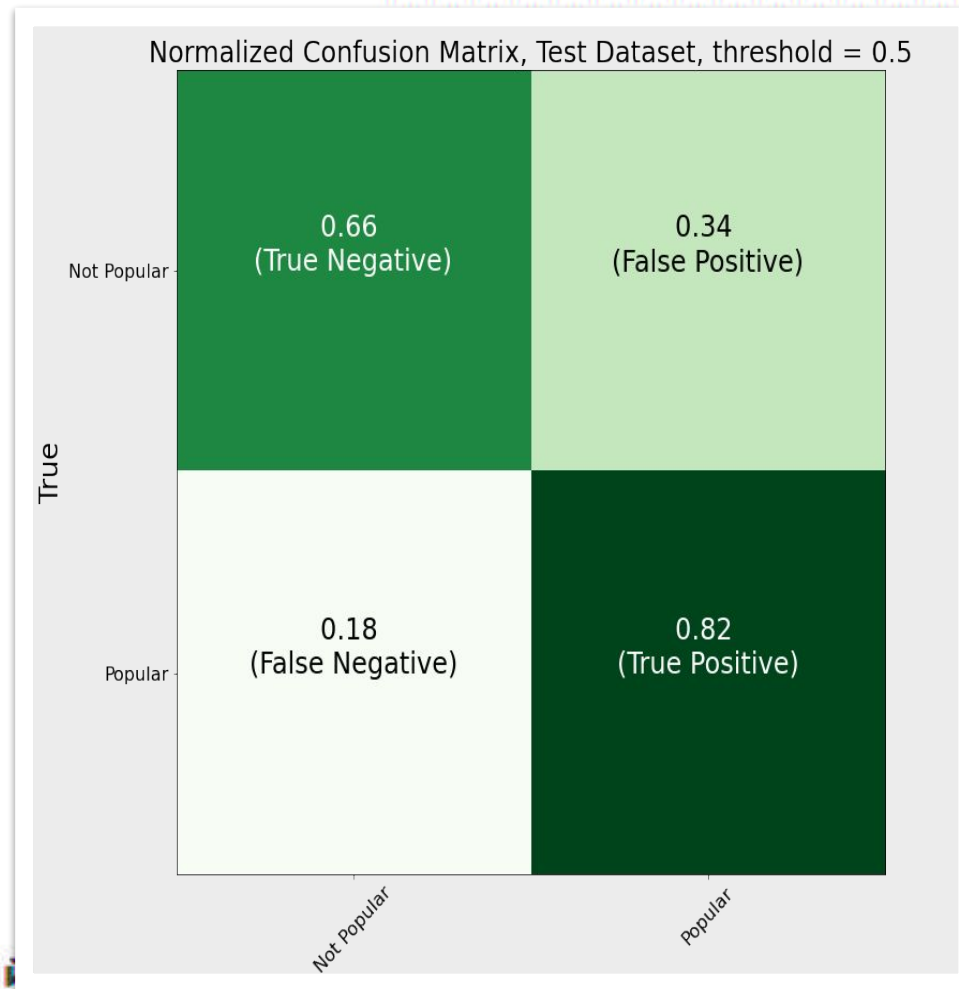
```
=====
Dep. Variable:          popularity    R-squared:                0.283
Model:                  OLS          Adj. R-squared:           0.283
Method:                 Least Squares  F-statistic:              6915.
Date:                   Fri, 17 Jun 2022  Prob (F-statistic):       0.00
Time:                   06:51:41      Log-Likelihood:           -9.3595e+05
No. Observations:      228159        AIC:                      1.872e+06
Df Residuals:          228145        BIC:                      1.872e+06
Df Model:               13
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	53.3941	0.233	229.507	0.000	52.938	53.850
acousticness	-11.2482	0.146	-76.786	0.000	-11.535	-10.961
danceability	15.4351	0.230	67.021	0.000	14.984	15.886
duration_ms_std	-0.2489	0.031	-8.007	0.000	-0.310	-0.188
energy	-8.1027	0.261	-30.989	0.000	-8.615	-7.590
instrumentalness	-3.5746	0.129	-27.747	0.000	-3.827	-3.322
key	-0.0709	0.009	-8.015	0.000	-0.088	-0.054
liveness	-10.0417	0.189	-53.221	0.000	-10.411	-9.672
loudness_std	5.0597	0.063	79.726	0.000	4.935	5.184
mode	-1.1412	0.065	-17.641	0.000	-1.268	-1.014
speechiness	-9.3737	0.218	-42.928	0.000	-9.802	-8.946



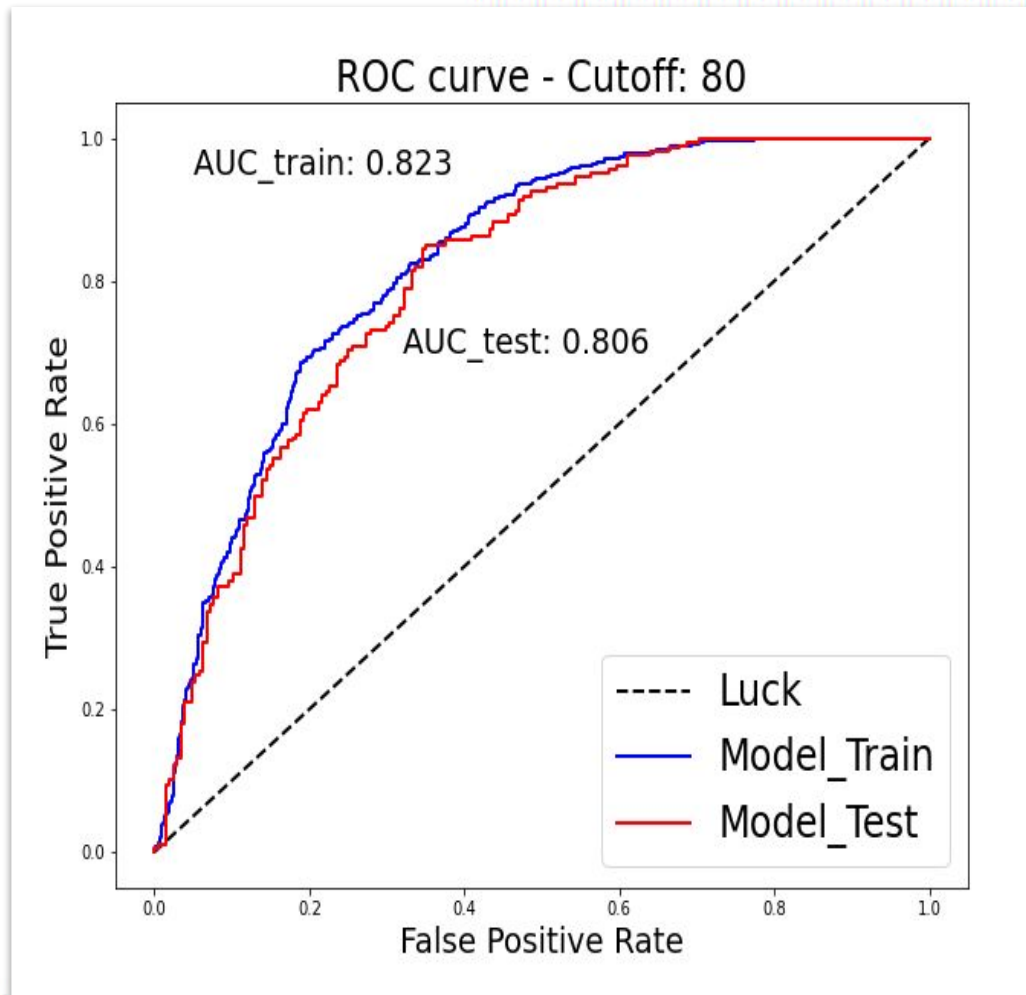
Построение матрицы несоответвий

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения



Построение ROC – кривой

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения



Заключение

Прогнозирование и анализ популярности музыкальных композиций про помощи машинного обучения

Таким образом подводя итоги исследования, можно сделать вывод , что удалось спрогнозировать и выявить определенную закономерность в становлении музыкальных композиций популярными. Так же в ходе исследования были рассмотрены и использованы технологии и математические методы современного анализа больших данных. В данном исследовании приведено сравнение двух алгоритмов машинного обучения. Так же были затронуты темы современных алгоритмов, которые используются в крупных айти компаниях. Результаты исследования можно считать начальной точкой и использовать для дальнейшего анализа, так же и само исследование, как доказательство целесообразности математического анализа музыкальных композиций.

Спасибо
за внимание!

Д.А. Волченков
Группа 181-341



МОСКОВСКИЙ
ПОЛИТЕХ