

Whatever hack

Задача “не смешно”

Команда “Дедушкин ковер” (1-е место)

Постановка задачи

Таги:

face:,eyebrow:,nose:,per
son:,:facial hair:



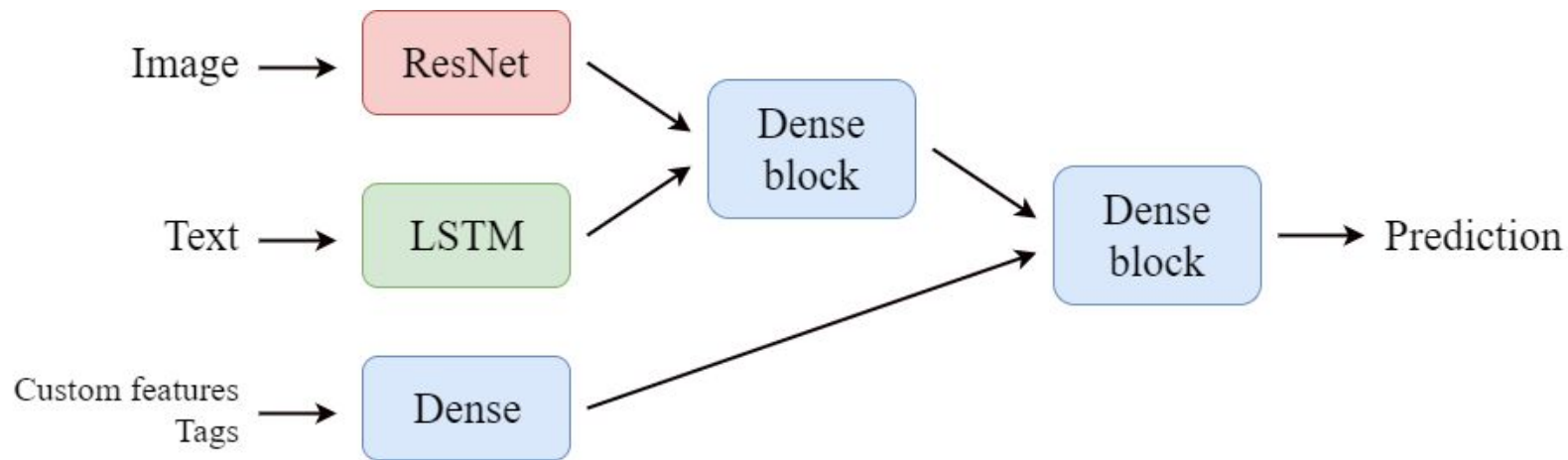
TEXT	LABEL
PULL HIS OLD TIME.	0
LET ME TAKE YOU TO SHORE.	0
PUT SOME HAIR ON IT!	0
LOOKS GREAT IN A UNIFORM.	0
HE LIKES TO SIN.	1

“Честное” решение v.1

LightGBM на признаках:

- TF-IDF - на словах (с биграммami), на тегах, на буквах по словам
- Длина текста, число тегов, количество знаков препинания
- Количество чисел, количество пробельных символов, точка в конце
- Количество подписей к той же картинке

“Честное” решение v.2



Качество “честных” решений

Модель	ROC-AUC Локально	ROC-AUC, Public LB
LGBM	0.74	0.72
NN	0.76	0.66
Logistic Regression (TF-IDF)	0.73	0.64

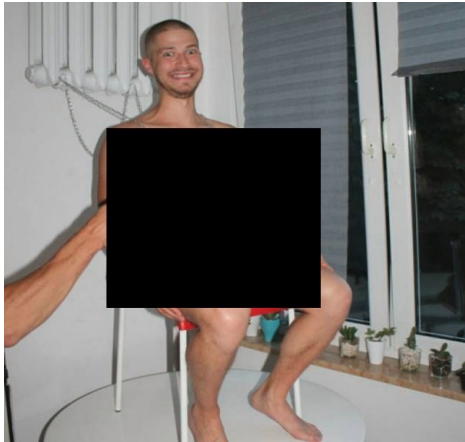
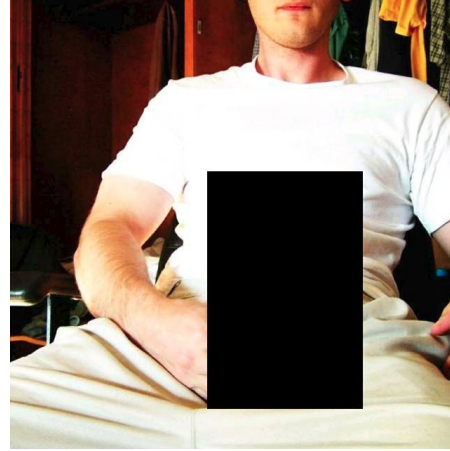
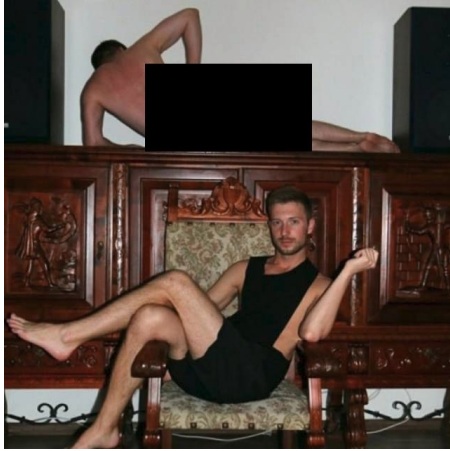
Первый лик

813974	WHERE'S *YOUR* SUPPORT STRESS.	0
813974	MIGHT BE MY LITTLE RIGHT?	0
813974	AND I ARE A STRANGE PRICE TO MAKE AN EMERGENCY.	0
813974	WE HAVE AN UPDATE.	1
813974	BACK YOUR WINTER COAT, HONEY.	0
813974	COACH, THIS IS A WOMAN'S PARADISE.	0
813974	CORRESPONDENT, KNOW BLAH! SHE'S NOT NATURAL.	0
813974	SCHOOL STARTS OVER HERE.	0
813974	ANTI-FOX CAN BE IN THIS HOUSE.	0
813974	WELL, TAKE OFF OUR BEER?	0
813974	PRESIDENT SCAN YESTERDAY.	0
827472	GOOD LORD SAY OR BAD DOGS.	1
827906	ALCOHOL, WHOSE RESORT.	1
840120	ONCE WILDCAT, HAS A GUN.	0
840120	BUT YOU GET YOUR SEAT.	0
840120	SING IT WAS MEAT.	0
840120	LATER WE'LL JUMP OFF THE STREET.	0
840120	BUT WE WERE JUMPING WITH MONEY.	0
840120	WE DRINK ACCORDING TO THE WIND.	0
840120	PLAY LASER TAG ONCE A WEEK.	0
840120	BUT THIS IS BASEBALL.	0
840120	SAGE EXCHANGED THEIR BEDTIME.	0
840120	BACK DEEP INTO A SEA OF PORK.	0
840120	PIECES OF THE PUZZLE.	1
844856	FITALIN IN THE BAG, DINNER IN THE OVEN.	0
844856	GET OFF THE FARM NOW.	0
844856	HARVEST YOUR OWN EXTINCTION.	0

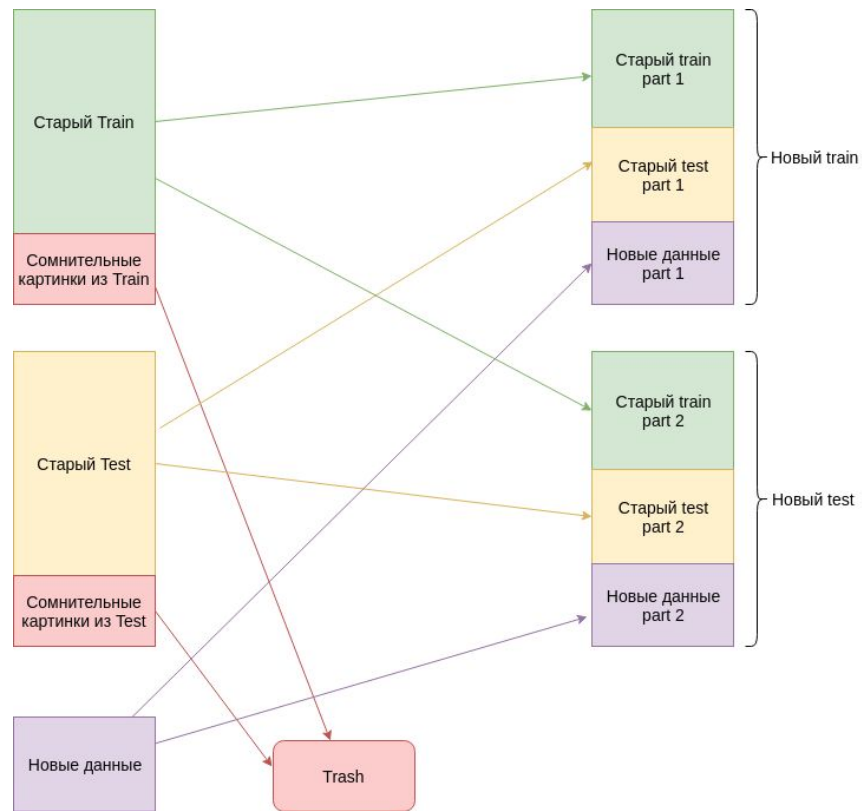
LGBM на:

1. Номер записи с конца, для данной картинки
2. Число подписей для данной картинки

Не все картинки были “удачными”...



... организаторам пришлось принимать меры



Финальный алгоритм

1. Находим все объекты из старого трейна в новом тесте. Размечаем
2. Находим все те, которые были в старом тесте и есть в новом. Для них делаем предсказания бустингом на двух признаках
(Используем первый лик)
3. На оставшихся объектах(~5-10%) делаем предикт первой моделью.

Финальные результаты

Модель	ROC-AUC, Локально	ROC-AUC, Public LB	ROC-AUC, Private LB
LGBM	0.74	0.72	
NN	0.76	0.66	
Logistic Regression (TF-IDF)	0.73	0.64	
Решение с ликами		0.995	0.955
Решение второго места		~0.931	~0.93