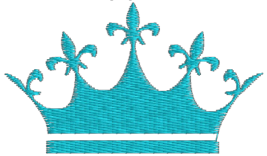
Как стать

принцессой

(*основываясь на данных фильмов про принцесс от диснея и пиксара)

Мы решили посмотреть, есть ли динамика у *принцев* и *принцесс* по стереотипным описаниям.

Статья England D. E., Descartes L., Collier-Meek M. A 2011 Gender Role Portrayal and the Disney Princesses дает следующие признаки:



- готов к исследованиям
- физически сильный
- handsome
- неэмоциональный
- вовлечен в интеллектуальную активность
- лидер

- физически слабая
- beautiful
- заботящаяся
- эмоциональная
- жертва
- подчиняющаяся

По статье Gaucher D., Friesen J., Kay A. C. Evidence that gendered wording in job advertisements exists and sustains gender inequality собрали списки типично “феминных” и “маскулинных” слов

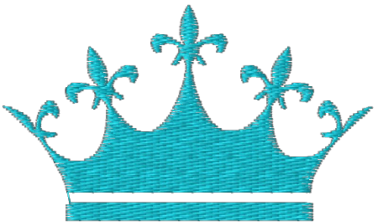
honor
fearless
brave
pride
initiative
reasonable
objective
certain
champion
strong

beautiful
sweety
warm
fragile
weak
caring
amazing
emotional
soft
baby

Предполагается, что:

- в целом у принцесс будет больше феминных характеристик, чем маскулинных, с принцами наоборот

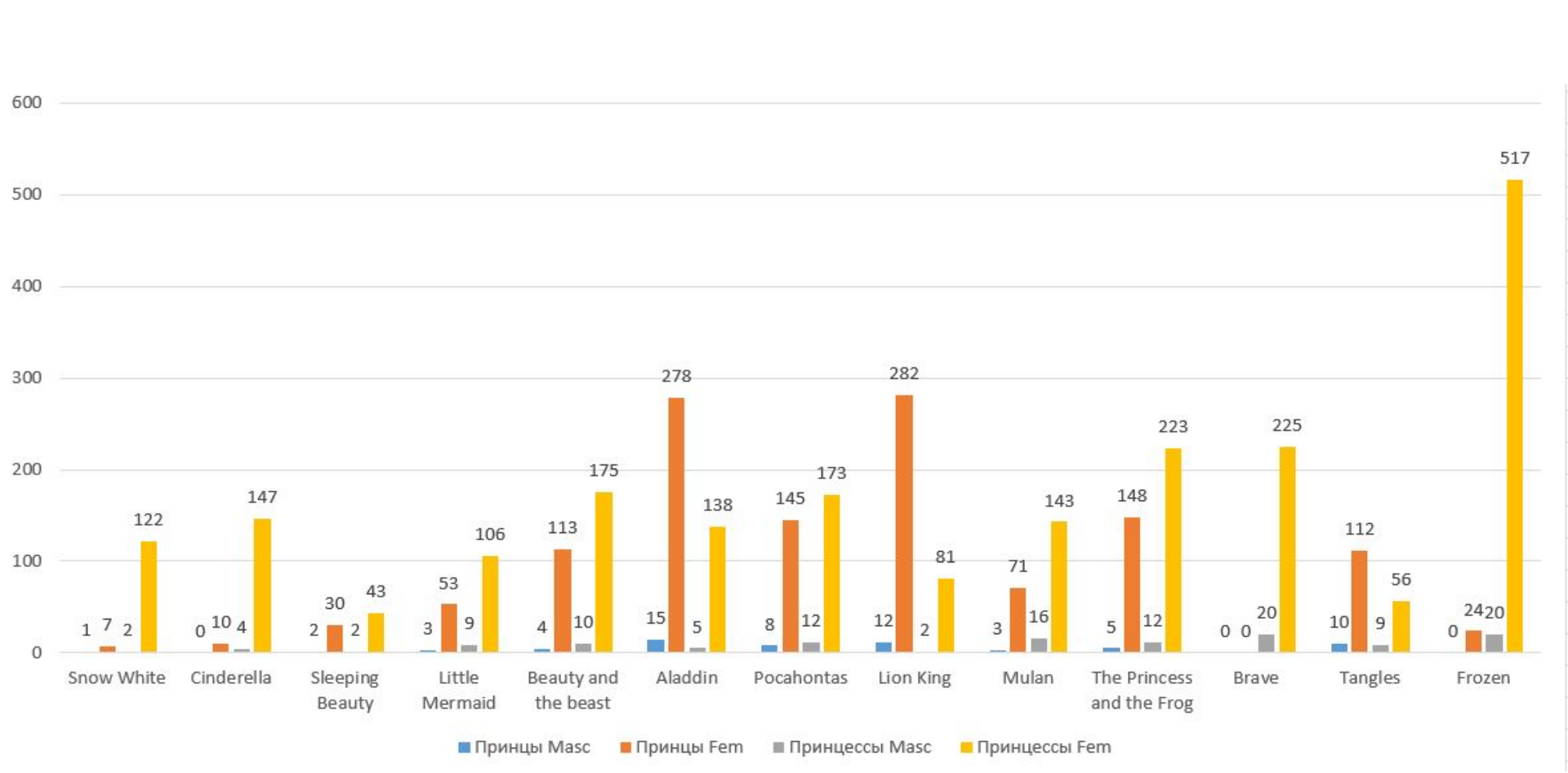
- чем позже выпущен фильм, тем меньше феминных характеристик будет у принцессы



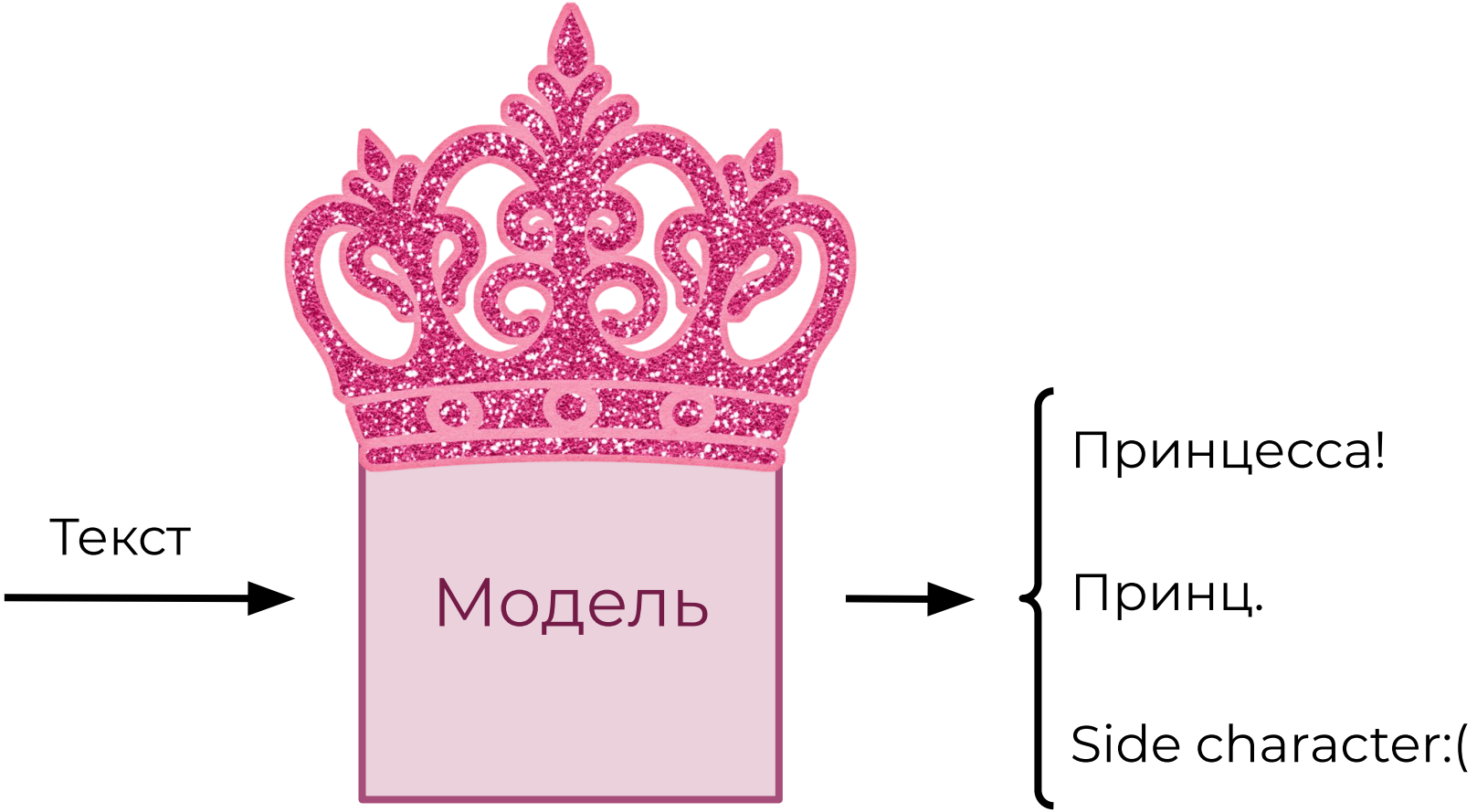
		Принцы		Принцессы	
Фильм	Год	Masc	Fem	Masc	Fem
EARLY					
Snow White	1937	1	7	2	122
Cinderella	1950	0	10	4	147
Sleeping Beauty	1959	2	30	2	43

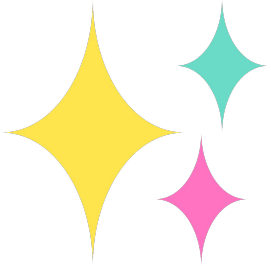
		Принцы		Принцессы	
Фильм	Год	Masc	Fem	Masc	Fem
MIDDLE					
Little Mermaid	1989	3	53	9	106
Beauty and the Beast	1991	4	113	10	175
Aladdin	1992	15	278	5	138
Pocahontas	1994	8	145	12	173
Lion King	1994	12	282	2	81
Mulan	1998	3	71	16	143

		Принцы		Принцессы	
Фильм	Год	Masc	Fem	Masc	Fem
LATE					
The Princess and the Frog	2009	5	148	12	233
Brave	2012	0	0	20	225
Tangled	2012	10	112	9	56
Frozen	2013	0	24	20	517


600
500
400
300
200
100
0
1 7 2 122
0 10 4 147
2 30 2 43
3 53 9 106
4 113 10 175
15 278 5 138
8 145 12 173
12 282 2 81
3 71 16 143
5 148 12 223
0 0 20 225
10 112 9 56
0 24 20 517
Snow White
Cinderella
Sleeping Beauty
Little Mermaid
Beauty and the beast
Aladdin
Pocahontas
Lion King
Mulan
The Princess and the Frog
Brave
Tangles
Frozen
Принцы Masc
Принцы Fem
Принцессы Masc
Принцессы Fem

Можем ли мы тогда научить машину различать принцев и принцесс?


Текст
Модель
Принцесса!
Принц.
Side character:(



Модели

1. Разбиение:

A. по фразам
B. по предложениям

2. Токенизатор:

A. по леммам
B. по стемам

3. Векторайзер:

A. count vectorizer
B. tf-idf
C. hashing

Обучающая выборка

		countvec	tf-idf	hashing
phrase	lem	0.848	0.804	0.751
	stem	0.846	0.806	0.746
sentence	lem	0.771	0.724	0.688
	stem	0.771	0.726	0.693

Тестовая выборка

		countvec	tf-idf	hashing
phrase	lem	0.622	0.651	0.657
	stem	0.62	0.652	0.656
sentence	lem	0.651	0.665	0.655
	stem	0.651	0.661	0.663

@DisneyGuesserBot