



FAKE NEWS PREDICTION





Валишева Аделя 11-001, Мартазова Ольга 11-002



. .



набор данных из 72 134 новостных статей, из которых 35 028 — настоящие, а 37 106 — поддельные



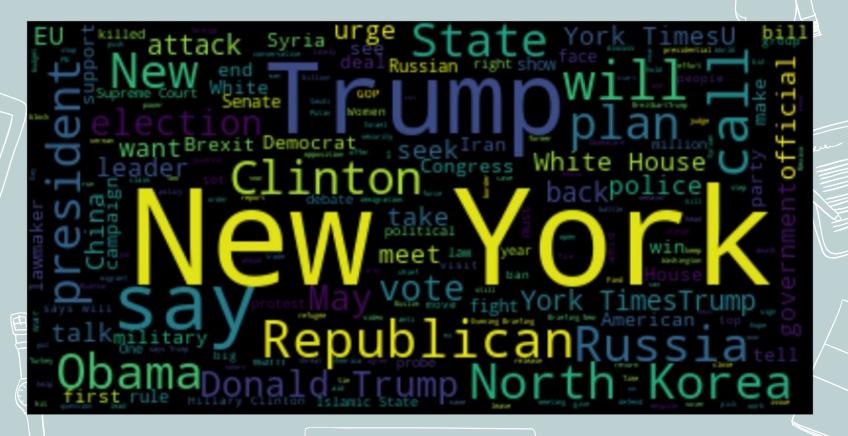




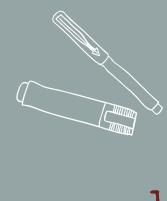




WORD CLOUD

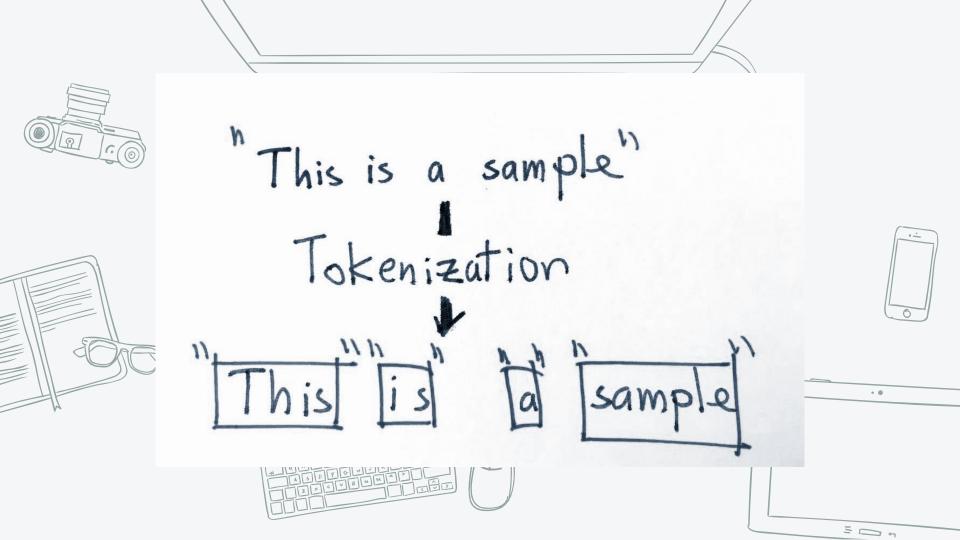


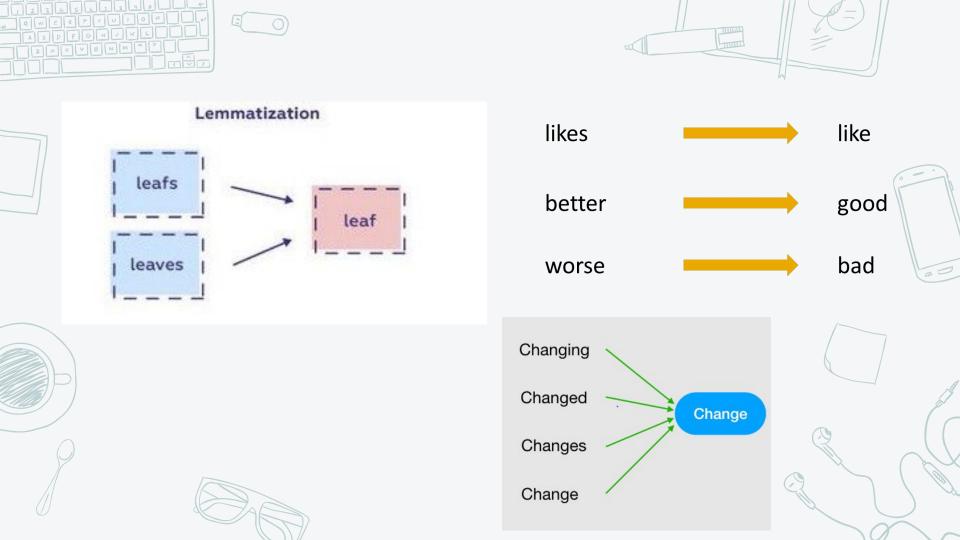
ASDF		Title	Text	Label	
	0	LAW ENFORCEMENT ON HIGH ALERT Following Threats Against Cops And Whites On 9-11By #BlackLivesMatter	No comment is expected from Barack Obama Members of the #FYF911 or #FukYoFlag and #BlackLivesMatter	1	(0)
	1	UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MOST CHARLOTTE RIOTERS WERE "PEACEFUL" PROTESTERS…In Her	Now, most of the demonstrators gathered last night were exercising their constitutional and protect	1	
	2	Bobby Jindal, raised Hindu, uses story of Christian conversion to woo evangelicals for potential 201	A dozen politically active pastors came here for a private dinner Friday night to hear a conversion	0	7
	3	SATAN 2: Russia unvelis an image of its terrifying new 'SUPERNUKE' – Western world takes notice	The RS-28 Sarmat missile, dubbed Satan 2, will replace the SS-18 Flies at 4.3 miles (7km) per sec an	1	
	4	Billionaire Odebrecht in Brazil scandal released to house arrest	RIO DE JANEIRO/SAO PAULO (Reuters) - Billionaire Marcelo Odebrecht, the highest-profile executive im	O	(
					~



1. ПОДГОТОВКА ИСХОДНОГО ТЕКСТА

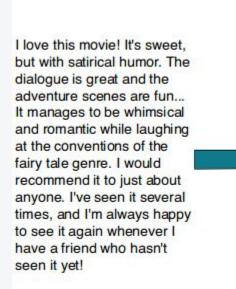
- 1. устранение пунктуации и цифр
- 2. word tokenization разбиение на список слованглийского языка
- 3. lemmatization приведение слов к её нормальной форме
- 4. избавление от stopwords

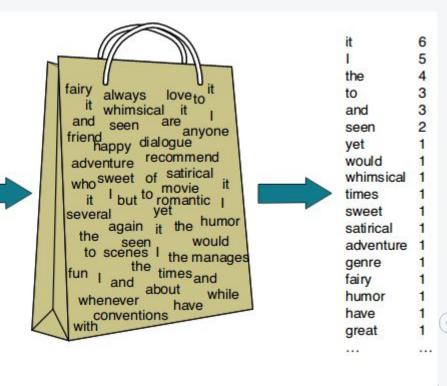


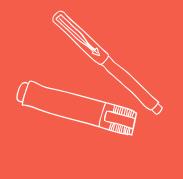












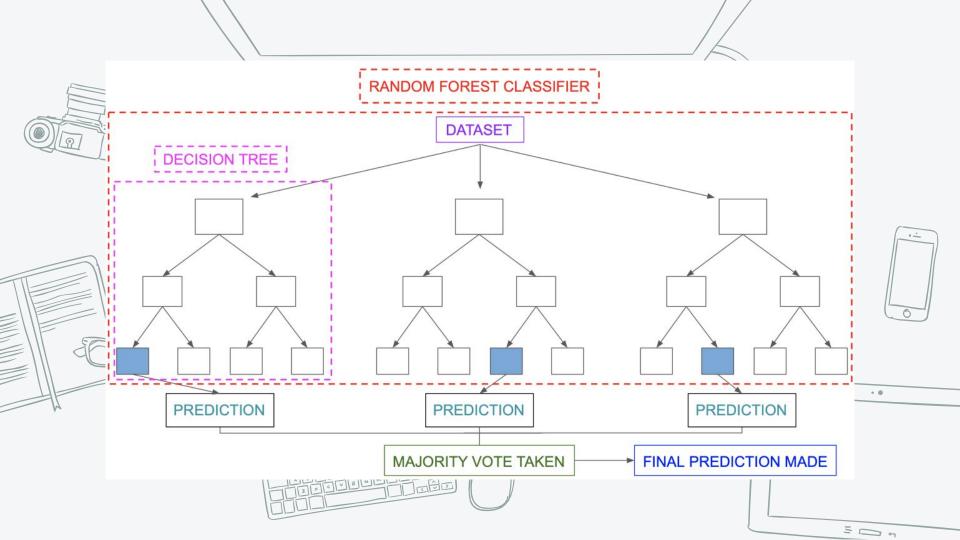
2. RANDOM FOREST CLASSIFIER

это множество решающих деревьев. В задаче регрессии их ответы усредняются, в задаче классификации принимается решение голосованием по большинству



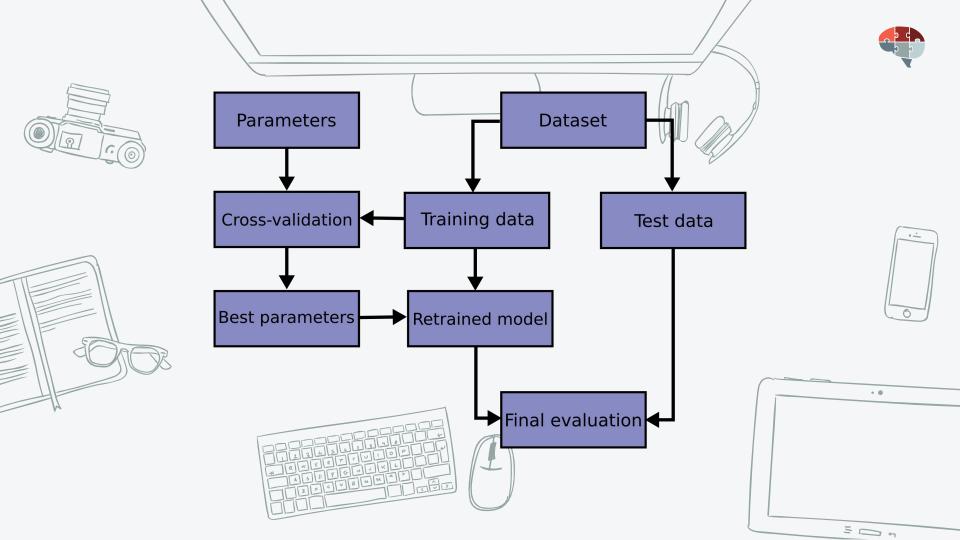






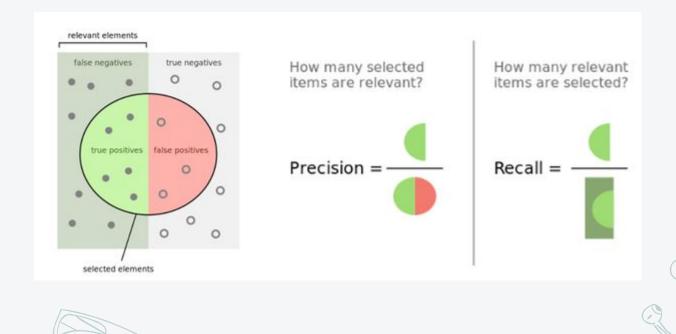














96%



Precision in real news

94%

Accuracy

















- sklearn.ensemble.RandomForestClassifier
- 3.3. Metrics and scoring: quantifying the quality of predictions







