

# **HSE Data Science Hack**

Команда: InternationalLiders



# RoadMap

Постановка  
цели

Предобработка  
данных

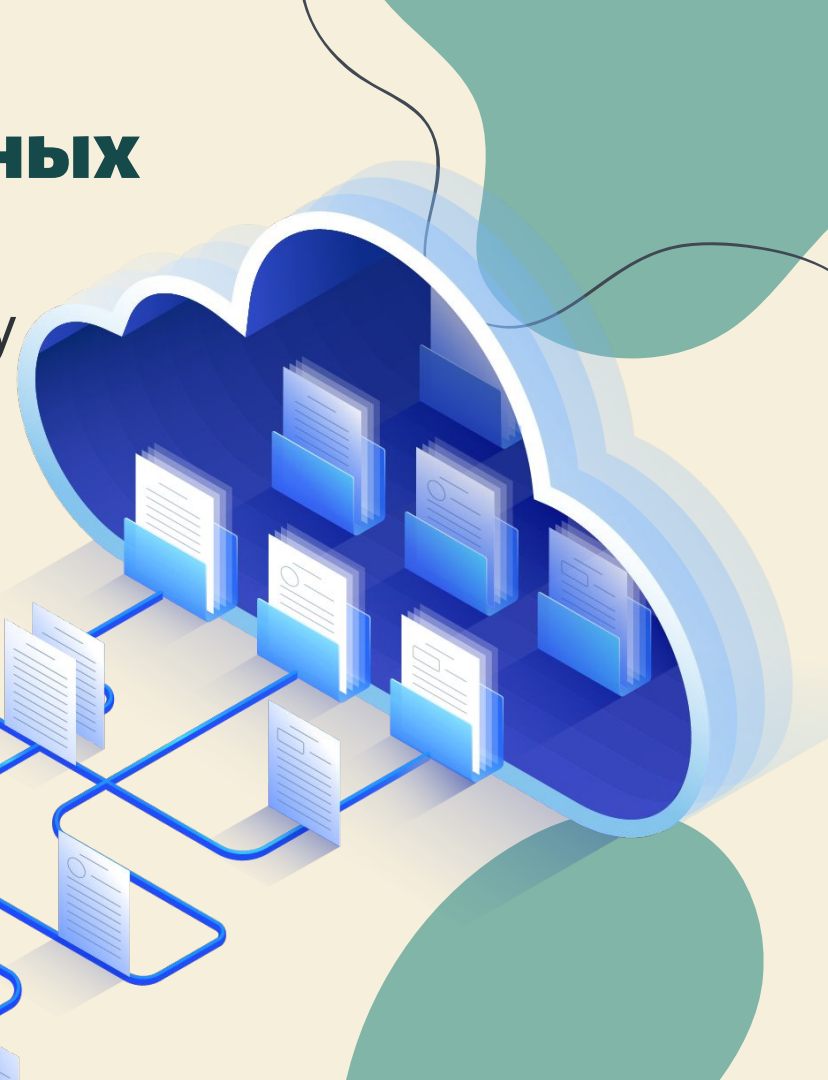
Выбор модели для  
sentiment

Выбор модели для  
category

# **Постановка задачи**

# Предобработка данных

- 1) приведение к нижнему регистру
- 2) удаление знаков препинания
- 3) удаление стоп-слов (кроме не)
- 4) провели лемматизацию



# Sentimental

Модель	Accuracy	ROC
GaussianNB	0.64	0.77
MLPClassifier	0.88	0.93
MNB	0.83	0.93
Log	0.86	0.94

# Category

Decision tree classifier