

Проект 9: Распознавание эмоций в текстах

Описание задачи: Классификация эмоций (радость, грусть и т.д.) в твитах.

Стек технологий: Python 3.x, pandas, scikit-learn, nltk, tensorflow.

Начальный каркас кода:

Импорт

```
import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.svm import SVC

from sklearn.feature_extraction.text import CountVectorizer

from sklearn.metrics import accuracy_score
```

Шаг 1: Данные

```
data = pd.DataFrame({
    'text': ['I am happy', 'So sad today', 'Excited!', 'Depressed'],
    'emotion': ['joy', 'sadness', 'joy', 'sadness']
})
```

Шаг 2: Векторизация

```
vectorizer = CountVectorizer()

X = vectorizer.fit_transform(data['text'])

y = data['emotion']
```

Шаг 3: Разделение

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

Шаг 4: Обучение

```
model = SVC()  
model.fit(X_train, y_train)
```

Шаг 5: Оценка

```
predictions = model.predict(X_test)  
print("Accuracy:", accuracy_score(y_test, predictions))
```

TODO: Используйте BERT для лучшей точности

Направления для дальнейшего развития: Перейти к deep learning (BERT), интегрировать с Twitter API, визуализировать распределение эмоций.