

# Программа предсказания тем фильма

Клиент:

```
import streamlit as st
import requests

st.set_page_config(page_title="Предсказание тем статей")

st.title("Предсказание кластера статей")
input_text = st.text_area("Введите описание статьи", height=200)

if st.button("Предсказать"):
    if not input_text.strip():
        st.warning("Пожалуйста, введите текст статьи")
    else:
        try:
            response = requests.post(
                url="http://127.0.0.1:8000/predict",
                json={"text": input_text}
            )
            response.raise_for_status()
            result = response.json()

            st.subheader("Предсказанный кластер")
            st.success(result['cluster'])

            st.subheader("Вероятности всех кластеров")
            st.text(result['probabilities'])

        except requests.exceptions.RequestException as e:
            st.error(f"Ошибка при запросе к серверу: {e}")
        except Exception as e:
            st.error(f"Произошла ошибка: {e}")
```

Сервер:

[illegible]

```
def fun_punctuation_text(text):
    text = text.lower()
    text = ''.join([ch for ch in text if ch not in string.punctuation])
    text = ''.join([i if not i.isdigit() else ' ' for i in text])
    text = ''.join([i if i.isalpha() else ' ' for i in text])
    text = re.sub(pattern: r'\s+', repl: ' ', text, flags=re.I)
    st = ')\xa0'
    text = ''.join([ch if ch not in st else ' ' for ch in text])
    return text
```

1 usage

```
def fun_lemmatizing_text(text):
    tokens = word_tokenize(text)
    res = []
    morph = pymorphy3.MorphAnalyzer(lang='ru')
    for word in tokens:
        p = morph.parse(word)[0]
        res.append(p.normal_form)
    return " ".join(res)
```

1 usage

```
def fun_tokenize(text):
    t = word_tokenize(text)
    tokens = [token for token in t if token not in russian_stopwords]
    return " ".join(tokens)
```

1 usage

```
def fun_pred_text(text):
    text = fun_punctuation_text(text)
    text = fun_lemmatizing_text(text)
    text = fun_tokenize(text)
    return text
```

```

def predict_cluster(text):
    cluster_description = {
        0: "0 - Наука, магия, судьбы людей",
        1: "1 - Выживание и сложный выбор",
        2: "2 - Приключения, преследования, злодеи",
        3: "3 - Поединки, путешествия",
        4: "4 - История, война, судьба человека",
        5: "5 - Космос, магия, комедия",
        6: "6 - Трагедии, война, фашизм"
    }
    processed_text = fun_pred_text(text)
    text_vectorized = vectorizer.transform([processed_text])
    probabilities = model.predict_proba(text_vectorized)[0]
    sorted_clusters = sorted(
        [(cluster_description[i], prob) for i, prob in enumerate(probabilities)],
        key=lambda x: -x[1]
    )
    main_cluster = sorted_clusters[0][0]
    probabilities_str = "\n".join([f"{name}: {prob:.2%}" for name, prob in sorted_clusters])
    return main_cluster, probabilities_str

1 usage
class Item(BaseModel):
    text: str

@app.post("/predict")
def post_pred_text(item: Item):
    try:
        main_cluster, probabilities = predict_cluster(item.text)
        return {
            'cluster': main_cluster,
            'probabilities': probabilities
        }
    except Exception as e:
        raise HTTPException(status_code=500, detail=str(e))

```

## Результат:

### Предсказание кластера статей

Введите описание статьи

В оккупированной нацистами Польше актерская труппа оказывается вовлечена в игры одного из солдат с немецким шпионом, цель которых - раскрыть последнего...

Предсказать

### Предсказанный кластер

6 - Трагедии, война, фашизм

### Вероятности всех кластеров

- 6 - Трагедии, война, фашизм: 40.00%
- 0 - Наука, магия, судьбы людей: 20.00%
- 4 - История, война, судьба человека: 20.00%
- 5 - Космос, магия, комедия: 20.00%
- 1 - Выживание и сложный выбор: 0.00%
- 2 - Приключения, преследования, злодеи: 0.00%
- 3 - Поединки, путешествия: 0.00%

# Предсказание кластера статей

Введите описание статьи

Один бродяга решил отдохнуть за государственный счет и, симулируя безумие, попал в психиатрическую лечебницу. Здесь он решил ввести свои правила жизни пациентов, что вошло в конфликт с официальными порядками в больнице. Этот конфликт привел к тому, что его стали лечить от излишней независимости и непоседливости, в результате чего он действительно стал сумасшедшим. Теперь ему не надо притворяться, чтобы находится здесь.

Предсказать

## Предсказанный кластер

5 - Космос, магия, комедия

## Вероятности всех кластеров

- 5 - Космос, магия, комедия: 40.00%
- 2 - Приключения, преследования, злодеи: 20.00%
- 4 - История, война, судьба человека: 20.00%
- 6 - Трагедии, война, фашизм: 20.00%
- 0 - Наука, магия, судьбы людей: 0.00%
- 1 - Выживание и сложный выбор: 0.00%
- 3 - Поединки, путешествия: 0.00%