

Серверная часть

```
from fastapi import FastAPI
import pickle
from pydantic import BaseModel
import pandas as pd
import numpy as np
import string
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import pymorphy3

app = FastAPI()

with open('model_lr.pkl', 'rb') as file:
    model = pickle.load(file)

with open('vectorizer_h.pkl', 'rb') as file:
    vectorizer = pickle.load(file)

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(r"\s+", ' ', text, flags=re.I)
    text = re.sub('[a-z]', "", text, flags=re.I)
    st = '> \xa0'
    text = ".join([ch if ch not in st else ' ' for ch in text])
    return text

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

def fun_tokenize(text):
    nltk.download('stopwords')
    russian_stopwords = nltk.corpus.stopwords.words("russian")
    russian_stopwords.extend(['т.д.', 'т', 'д',
'это', 'который', 'свой', 'своём', 'всем', 'всё', 'её', 'оба', 'ещё', 'должный', 'фильм',
'кинокартина', 'сцена', 'режиссер', 'актер', 'мин', 'кадр', 'роль'])
    t = word_tokenize(text)
    tokens = [token for token in t if token not in russian_stopwords]
    text = " ".join(tokens)
```

```

    return text

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):
    text_vectorized = vectorizer.transform([fun_pred_text(text)])
    probabilities = model.predict_proba(text_vectorized)[0]

    mapping = {
        0: 'Криминально-социальные драмы',
        1: 'Семейно-философские драмы',
        2: 'Городские социальные истории',
        3: 'Приключенческие/детективные сюжеты',
        4: 'Военно-исторические драмы',
        5: 'Историко-культурные нарративы'
    }
    probabilities_list = []
    for cluster_num, prob in enumerate(probabilities):
        cluster_name = mapping[cluster_num]
        probabilities_list.append({
            "cluster_name": cluster_name,
            "probability": float(prob)
        })

    return probabilities_list

class Item(BaseModel):
    text: str

@app.post("/predict")
def post_predict_text(item: Item):
    return {'probabilities': predict_cluster(item.text)}

```

Клиентская часть

```

import streamlit as st
import requests
import os

os.environ['HTTP_PROXY'] = "
os.environ['HTTPS_PROXY'] = "

st.title("Предсказание кластера фильма")
input_text = st.text_area("Введите текст", height=200)

if st.button("Предсказать"):
    if input_text == "":

```

```

st.write("Введите текст")
else:
    data = {
        "text": input_text
    }
    url = 'http://127.0.0.1:8000/predict'
    response = requests.post(url, json=data)
    result = response.json()
    probabilities = result.get("probabilities", [])

    st.markdown("""
        ### Вероятности тем
        """)

    for prob in probabilities:
        st.write(f"{prob['cluster_name']}: {prob['probability']}")

```

Криминально-социальные драмы:

Предсказание кластера фильма

Введите текст

Восьмилетний мальчик Коул постоянно видит мертвых людей и разговаривает с ними. Ему пытается помочь детский психолог Малькольм Кроу. Хэйли Джозел Осмент, которому в момент съемок было 9 лет, получил роль во многом благодаря тому, что был единственным мальчиком, пришедшим на кастинг в костюме и галстук. Фраза "Я вижу мертвых людей" считается одной из самых популярных фраз в истории кинематографа и самой пародируемой в различных фильмах. Восьмилетний мальчик Коул постоянно видит мертвых людей и разговаривает с ними. Ему пытается помочь детский психолог Малькольм Кроу. Хэйли Джозел Осмент, которому в момент съемок было 9 лет, получил роль во многом благодаря тому, что был

Предсказать

Вероятности тем ⇄

Криминально-социальные драмы: 0.2637263326478187

Семейно-философские драмы: 0.1409178625151367

Городские социальные истории: 0.22696920243568142

Приключенческие/детективные сюжеты: 0.09433332031215351

Военно-исторические драмы: 0.04568131214264885

Историко-культурные нарративы: 0.2283719699465608

Семейно-философские драмы:

Предсказание кластера фильма

Введите текст

должный очистить окрестность от повстанец и девочка и военный пытаться найти свой путь к свобода или победа и у весь время на исходенеподадек от свой дом офелия находить лабиринт и она проводник становиться фея в центр лабиринт жить фавн готовый рассказать офелия и она предназначение девочка принцесса но чтобы найти свой королевство она нужно пройти три страшный испытание, испания 1944 год история маленький девочка офелия жестокий отчим военный офицер очистить окрестность повстанец девочка военный пытаться найти путь свобода победа весь время исходенеподадек дом офелия находить лабиринт проводник становиться фея центр лабиринт жить фавн готовый рассказать офелия предназначение девочка принцесса найти королевство нужно пройти страшный испытание

Предсказать

Вероятности тем ⇄

Криминально-социальные драмы: 0.08425355832486446

Семейно-философские драмы: 0.25153840688415646

Городские социальные истории: 0.22077695334423733

Приключенческие/детективные сюжеты: 0.1519615361066464

Военно-исторические драмы: 0.04090637292533703

Историко-культурные нарративы: 0.2505631724147584

Городские социальные истории:

Предсказание кластера фильма

Введите текст

исследователи учены по разным областейный момент выбрать не самый удачный
разыгаться гроза да ещё и один из сотрудник пытаться украсть образец эмбрион выключить
система безопасность динозавр на свобода они инстинкт говорить сам за себя оставаться на
остров опасно но и сбежать с он не такто просто,учёный смочь возродить доисторический
динозавр использовать клетка днк след кровь динозавр необычно прибыльно открыть остров
неповторимый сафари директор будущее парк юрский период приглашать группа
исследователейучёный разный областейный момент выбрать самый удачный разыгаться
гроза сотрудник пытаться украсть образец эмбрион выключить система безопасность
динозавр свобода инстинкт говорить оставаться остров опасно сбежать такто просто

Предсказать

Вероятности тем

Криминально-социальные драмы: 0.1240050680011932

Семейно-философские драмы: 0.20716385776085283

Городские социальные истории: 0.4066081136320384

Приключенческие/детективные сюжеты: 0.14246167199547757

Военно-исторические драмы: 0.04312365269629796

Историко-культурные нарративы: 0.07663763591414002

Приключенческие/детективные сюжеты:

Предсказание кластера фильма

Введите текст

золотой жиле в окрестностях сьерра мадре вложив в дело все свои сбережения отправляются на поиски золотого песка найти золото оказывается гораздо проще чем не потерять человеческий облик при его виде и вернуться домой живыми, два бедный американец прознать от престарелый старатель о золотой жила в окрестность сьерра мадр вложить в дело всё свой сбережение отправляться на поиск золотой песок найти золото оказывается гораздо простой чем не потерять человеческий облик при он вид и вернуться домой живой, бедный американец прознать престарелый старатель золотой жила окрестность сьерра мадр вложить дело сбережение отправляться поиск золотой песок найти золото оказывается гораздо простой потерять человеческий облик вид вернуться домой живой

Предсказать

Вероятности тем ↔

Криминально-социальные драмы: 0.0410856386074955

Семейно-философские драмы: 0.04273741171117259

Городские социальные истории: 0.05484839380291938

Приключенческие/детективные сюжеты: 0.7990479953498617

Военно-исторические драмы: 0.019420942413418164

Историко-культурные нарративы: 0.042859618115132586

Военно-исторические драмы

Предсказание кластера фильма

Введите текст

Клинт Иствуд играет ветерана Корейской войны, противостоящего группировке азиатских иммигрантов в пригородах Детройта.",клинт иствуд играет ветерана корейской войны противостоящего группировке азиатских иммигрантов в пригородах детройта,клинт иствуд играть ветеран корейский война противостоящий группировка азиатский иммигрант в пригород детройт,клинт иствуд играть ветеран корейский война противостоящий группировка азиатский иммигрант пригород детройт

Предсказать

Вероятности тем

Криминально-социальные драмы: 0.13037228055025007

Семейно-философские драмы: 0.16327252512412496

Городские социальные истории: 0.15775663637299883

Приключенческие/детективные сюжеты: 0.14711113189969163

Военно-исторические драмы: 0.19421225144967838

Историко-культурные нарративы: 0.20727517460325606

Предсказание кластера фильма

Введите текст

Продолжение истории о Питере Паркере - юноше, случайно ставшим Человеком-Пауком и его подруге ЭмДжей, которым снова придется столкнуться лицом к лицу с коварными врагами.",продолжение истории о питере паркере юноше случайно ставшим человекомпауком и его подруге эмджей которым снова придется столкнуться лицом к лицу с коварными врагами,продолжение история о питер паркер юноша случайно стать человекомпаукома и он подруга эмджей который снова придется столкнуться лицо к лицу с коварный враг,продолжение история питер паркер юноша случайно стать человекомпаукома подруга эмджей снова придется столкнуться лицо лицо коварный враг

Предсказать

Вероятности тем

Криминально-социальные драмы: 0.07699083974853217

Семейно-философские драмы: 0.13452373826846875

Городские социальные истории: 0.09827942802641626

Приключенческие/детективные сюжеты: 0.0937905430231152

Военно-исторические драмы: 0.0393564565375114

Историко-культурные нарративы: 0.5570589943959563