

پروژه دسته‌بندی اخبار با یادگیری ماشین

هدف پروژه

هدف از این پروژه، طراحی یک سیستم ساده برای تشخیص دسته‌بندی خبر از روی تیتر آن است. این سیستم از یادگیری ماشین استفاده می‌کند تا بتواند تیترهای جدید را به دسته‌هایی مانند «ورزشی»، «جهان»، «اقتصاد» و... نسبت دهد.

کتابخانه‌های مورد استفاده

```
import pandas as pd
import numpy as np
import joblib
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score
import streamlit as st
```

توضیح

- pandas برای خواندن داده‌ها
- numpy برای عملیات عددی

- joblib برای ذخیره مدل
- sklearn برای یادگیری ماشین
- streamlit برای رابط گرافیکی وب

فایل داده‌ها

- فایل news_sample.csv شامل ستون‌های:
 - o title تیتر خبر
 - o Category دسته‌بندی خبر
 - o

مراحل اجرای پروژه

1. خواندن داده‌ها و پیش‌پردازش

```
df = pd.read_csv("news_sample.csv")
df = df.dropna(subset=['title', 'category'])
```

2. تقسیم داده‌ها به آموزش و تست

```
x = df['title']
y = df['category']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

3. بردارسازی متن با TF-IDF

```
vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

4. آموزش مدل

```
model = LogisticRegression(max_iter=1000)
model.fit(X_train_tfidf, y_train)
```

5. ارزیابی مدل

```
y_pred = model.predict(X_test_tfidf)
print("دقت مدل:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

6. ذخیره مدل و بردار ساز

```
joblib.dump(model, 'model.pkl')
joblib.dump(vectorizer, 'vectorizer.pkl')
```

ساخت رابط کاربری با Streamlit

فایلی با نام news_classifier.py ساخته شد و کد زیر در آن قرار گرفت:

```
import joblib
import streamlit as st
import pandas as pd

st.set_page_config(page_title="تسخیص دسته خبر", layout="centered")

st.title("📰 تسخیص دسته خبر")

model = joblib.load("model.pkl")
vectorizer = joblib.load("vectorizer.pkl")

st.write("📄 چند نمونه از داده‌ها")
df = pd.read_csv("news_sample.csv")
if 'category' in df.columns:
    st.write(df[['title', 'category']].head())

text_input = st.text_area("لطفاً تیتر خبر خود را وارد کنید")

if st.button("پیش‌بینی دسته خبر"):
    if text_input.strip() != "":
        X = vectorizer.transform([text_input])
        prediction = model.predict(X)[0]
        st.success(f"🔍 دسته‌بندی خبر: **{prediction}**")
    else:
        st.warning("لطفاً یک متن وارد کنید")
```

نحوه اجرا

برای اجرای رابط کاربری داخل ترمینال در پوشه پروژه کد زیر را وارد می کنیم.

```
streamlit run news_classifier.py
```

فایل های نهایی پروژه

دیتاست اخبار	news_sample.csv
مدل آموزش دیده	model.pkl
پرداز ساز TF-IDF	vectorizer.pkl
رابط گرافیکی با Streamlit	news_classifier.py