Traffic Volume Estimation - Project Documentation

README.md

Traffic Volume Estimation using Machine Learning

This project predicts hourly traffic volume using weather, time, and holiday data. It uses a linear regression model as a baseline and can be extended with more advanced ML models.

Dataset

- Metro Interstate Traffic Volume
- Source: [UCI ML Repository](https://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume)

Getting Started

Requirements

```bash
pip install -r requirements.txt
```

Run Script

```bash
python traffic_volume_main.py
```

Run Jupyter Notebook

```bash
jupyter notebook notebooks/traffic_volume_prediction.ipynb
```

Project Structure

- `src/`: Source code for preprocessing and model training

Traffic Volume Estimation - Project Documentation

- `data/`: Dataset CSV file
- `notebooks/`: Jupyter Notebook version of the pipeline

requirements.txt

```
pandas
numpy
scikit-learn
matplotlib
seaborn
jupyter
```

src/preprocess.py

```
import pandas as pd

def load_and_preprocess(filepath):
    df = pd.read_csv(filepath)
    df['date_time'] = pd.to_datetime(df['date_time'])
    df['hour'] = df['date_time'].dt.hour
    df['dayofweek'] = df['date_time'].dt.dayofweek

    df = df.drop(['date_time', 'weather_description'], axis=1)
    df = pd.get_dummies(df, columns=['weather_main', 'holiday'], drop_first=True)

    X = df.drop('traffic_volume', axis=1)
    y = df['traffic_volume']

    return X, y
```

src/model.py

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

def train_model(X, y):
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

    model = LinearRegression()
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    mse = mean_squared_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)

    return model, mse, r2, y_test, y_pred
```

traffic_volume_main.py

```
from src.preprocess import load_and_preprocess
from src.model import train_model
import matplotlib.pyplot as plt
import seaborn as sns

# Load and preprocess data
X, y = load_and_preprocess('data/Metro_Interstate_Traffic_Volume.csv')

# Train model
model, mse, r2, y_test, y_pred = train_model(X, y)

print(f"Mean Squared Error: {mse}")
print(f"R^2 Score: {r2}")
```

```
# Plot actual vs predicted
plt.figure(figsize=(10,5))
sns.scatterplot(x=y_test, y=y_pred)
plt.xlabel("Actual Traffic Volume")
plt.ylabel("Predicted Traffic Volume")
plt.title("Actual vs Predicted Traffic Volume")
plt.grid(True)
plt.show()
```