

홍수 발생 가능성 예측 보고서

학번 : 2318071

이름 : 손준화

Git Hub : <https://github.com/SONJUNHWA/FinalExams/blob/main/FinalExams.py>

1. 안전 관련 머신러닝 모델 개발 요약

- 데이터 수집 : 강수량, 기온, 습도, 강수 시간, 지형 등 다양한 환경 데이터를 수집한다.
- 데이터 전처리 : 누락된 데이터 처리, 변환 및 정규화 작업을 수행하여 머신러닝 모델의 입력으로 사용한다.
- 모델 선정 : 로지스틱 회귀(Logistic Regression), 랜덤포레스트(Random Forest), 서포트 벡터 머신(Support Vector Machine, SVM) 등 다양한 알고리즘을 고려한다.
- 성능 평가 : 모델의 성능은 정확도, F1 점수, AUC-ROC 등 여러 지표를 통해 평가한다.

2. 개발 목적

a. 머신러닝 모델 활용 대상

- 재난 예방 및 관리 기관
- 정부의 자연재해 대응 부서
- 지역 사회 안전 센터 및 응급 관리 조직

b. 개발의 의의

- 예방적 안전 관리 : 홍수 지역에 대한 경고 시스템으로 활용하여 인명 피해와 재산 손실을 줄이는 데 기여할 수 있다.
- 의사결정 지원 : 홍수 발생 예측을 통해 효율적인 자원 배분 및 긴급 대피 결정을 지원한다.
- 기후 데이터 활용 : 큰 강우 이벤트에 대한 분석을 통해 변화하는 기후에서의 위험 인자를 평가한다.

3. 데이터 수집 및 전처리

- 데이터 세트 :
 - 독립 변수 :
 1. Rainfall(강수량) : 일정시간 동안의 강수량
 2. Temperature(온도) : 홍수 발생과 관련된 기온 변동
 3. Humidity(습도) : 지도 위성 데이터에서 측정된 습도.
 4. Soil Moisture(강수 지속 시간) : 토양 내 수분 비율로, 홍수 위험에 크게 영향을 미침
 5. Rain Duration(강수 지속 시간) : 일정 시간 동안의 강수 패턴
 - 종속 변수 :
 - *Flood Occurrence(홍수 발생 여부) : 1(홍수 발생), 0(홍수 미발생)

```
import pandas as pd

# 데이터 로드
data = pd.read_csv('flood_data.csv')

# 결측치 처리 및 필요 열 선택
data.dropna(inplace=True)
data = data[['Rainfall', 'Temperature', 'Humidity', 'SoilMoisture', 'RainDuration', 'Flood']]
```

4. 머신 러닝 모델 개발

- 여러 머신러닝 알고리즘을 선택하여 모델을 정의하고 훈련한다. 본 프로젝트에서는 로지스틱 회귀, 랜덤포레스트, svm 을 사용한다.

```

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, f1_score, roc_auc_score

# 데이터 로드
data = pd.read_csv('flood_data.csv')

# 데이터 전처리
data.dropna(inplace=True) # 결측치 제거
X = data[['Rainfall', 'Temperature', 'Humidity', 'SoilMoisture', 'RainDuration']]
y = data['Flood']

# 훈련 세트와 테스트 세트로 나누기
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 데이터 정규화
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# 모델 초기화
models = {
    "Logistic Regression": LogisticRegression(),
    "Random Forest": RandomForestClassifier(),
    "Support Vector Machine": SVC(probability=True)
}

# 모델 훈련 및 평가
results = {}
for name, model in models.items():
    model.fit(X_train, y_train) # 모델 훈련
    preds = model.predict(X_test) # 예측
    accuracy = accuracy_score(y_test, preds)
    f1 = f1_score(y_test, preds)
    auc = roc_auc_score(y_test, model.predict_proba(X_test)[:, 1])
    results[name] = {'Accuracy': accuracy, 'F1 Score': f1, 'AUC': auc}

# 결과 출력
for model_name, metrics in results.items():
    print(f"{model_name} - Accuracy: {metrics['Accuracy']:.2f}, F1 Score: {metrics['F1 Score']:.2f}, AUC: {metrics['AUC']:.2f}")

```

5. 개발 결과

a. 성능 지표에 따른 머신러닝 성능 평가

- 모델 성능은 다음과 같다:
 - 정확도(Accuracy) : 전체 샘플 중에서 예측이 올바른 비율
 - F1 Score : 정밀도와 재현율의 조화 평균
 - AUC-ROC : 모델이 양성 클래스를 얼마나 잘 구분하는지를 나타내는 곡선
 -

b. 결과 해석

- 랜덤 포레스트 모델이 가장 높은 성능을 보이며, svm 또한 예측 성능을 나타냈다.

6. 결론

이번 프로젝트를 통해 홍수 발생을 예측하기 위한 머신러닝 모델을 성공적으로 개발하였으며, 랜덤 포레스트 모델이 특히 높은 정확도와 안정성을 보였다. 이러한 예측 시스템은 재난 예방 및 긴급 대응을 위한 중요한 도구로 활용될 수 있다.