Libraries

Sklearn.ensembel

Sklear.datasets

Radnomforest classifier

Skleran.model selection

Presentation

Random forests are a popular supervised machine learning algorithm that can handle both regression and classification tasks. Below are some of the main characteristics of random forests:

* Random forests are for supervised machine learning, where there is a labeled target variable.
* Random forests can be used for solving regression (numeric target variable) and classification (categorical target variable) problems.
* Random forests are an ensemble method, meaning they combine predictions from other models.
* Each of the smaller models in the random forest ensemble is a decision tree

In a random forest classification, multiple decision trees are created using different random subsets of the data and features. Each decision tree is like an expert, providing its opinion on how to classify the data. Predictions are made by calculating the prediction for each decision tree and then taking the most popular result. (For regression, predictions use an averaging technique instead.)

Coding

print(dataset.describe())

dataset = dataset[dataset['year'] >= 1994]  # Remove cars older than 30 years

dataset = dataset[dataset['selling\_price'] > 0]  # Remove rows with zero price

# Remove rows with any null values

    cleaned\_data = data.dropna()

    print("\nData after removing null values:")

    print(cleaned\_data)

    # Remove duplicate rows

    cleaned\_data = cleaned\_data.drop\_duplicates()

    print("\nData after removing duplicate rows:")

    print(cleaned\_data)

    # Optionally, save the cleaned dataset to a new CSV file

    cleaned\_data.to\_csv(r"C:\Users\Msi\Downloads\cleaned\_student\_monitoring\_data.csv", index=False)

# Display first few rows

dataset.head()

# Check for missing values

dataset.isnull().sum()

# Check for duplicates

dataset.duplicated().sum()

دا الكود اللي هو فتح كلام انا بحط الاكسل ف فاريابل عشان اعدل عليه

COVIDdata = pd.read\_csv(r'D:\ML-project\covid\_worldwide.csv')

# Display first few rows

dataset.head()

# Check for missing values

dataset.isnull().sum()

# Check for duplicates

dataset.duplicated().sum()

print(dataset.describe())

ataset = dataset[dataset['year'] >= 1994]  # Remove cars older than 30 years

dataset = dataset[dataset['selling\_price'] > 0]  # Remove rows with zero price

cleaning

reading

df\_main=pd.read\_csv("digital\_marketing\_campaign\_dataset.csv")

top of ds

df.head()

summarize null rows

df.isnull().sum()

dublicated summerivze

df.duplicated().sum()

df.dtypes

print(dataset.describe())

dataset = dataset[dataset['year'] >= 1994]  # Remove cars older than 30 years

dataset = dataset[dataset['selling\_price'] > 0]  # Remove rows with zero price

print("Cleaned dataset saved successfully!")

# Rename columns to use consistent naming convention

COVIDdata.rename(columns={

    'Total Cases': 'total\_cases',

    'Total Deaths': 'total\_deaths',

    'Total Recovered': 'total\_recovered',

    'Active Cases': 'active\_cases',

    'Total Test': 'total\_test',

    'Population': 'population',

    'Country': 'country'

}, inplace=True)

# Plot histograms of all numeric columns

COVIDdata.hist(figsize=(10, 8))

plt.suptitle("Distribution of Numeric Columns")

plt.show()

https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/