• Project by loading and preprocessing The dataset.

INTRODUCTION

Data Preprocessing in Data Mining

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task.

Some common steps in data preprocessing include:

Data preprocessing is an important step in the data mining process that involves cleaning and transforming raw data to make it suitable for analysis. Some common steps in data preprocessing include:

Abstract

Artificial intelligence (AI) is a leading technology of the current age of the Fourth Industrial Revolution (Industry 4.0 or 4IR), with the capability of incorporating human behavior and intelligence into machines or systems. Thus, AI-based modeling is the key to build automated, intelligent, and smart systems according to today's needs. To solve real-world issues, various types of AI such as analytical, functional, interactive, textual, and visual AI can be applied to enhance the intelligence and capabilities of an application. However, developing an effective AI model is a challenging task due to the dynamic nature and variation in real-world problems and data. In this paper, we present a comprehensive view on "AI-based Modeling" with the principles and capabilities of potential AI techniques that can play an important role in developing intelligent and smart systems in various real-world application areas including business, finance, healthcare, agriculture, smart cities, cybersecurity and many more. We also emphasize and highlight the research issues within the scope of our study. Overall, the goal of this paper is to provide a broad overview of AI-based modeling that can be used as a reference guide by academics and industry people as well as decision-makers in various real-world scenarios and application domains

AI Based Prediction and Database (lodingand preprocessing dataset)(source c

Creating an AI-based prediction system with dataset loading and preprocessing involves several steps. I'll provide a high-level overview and some sample Python code to get you started.

Data Collection:

Gather the dataset you want to work with. Ensure it's in a format that's easy to work with, such as CSV, Excel, or a database.

Neural Networks and Deep Learning

Deep learning (DL) [80] is known as another popular AI technique, which is based on artificial neural networks (ANN). Nowadays, DL has become a hot topic in the computing world due to its layer-wise learning capability from data. Multiple hidden layers, including input and output layers, make up a typical deep neural network. Figure 4 shows a general structure of a deep neural network (hiddenlayer=N

and N ≥

2) comparing with a shallow network (hiddenlayer=1

). DL techniques can be divided into three major categories, highlighted in our earlier paper Sarker et al. [80]. These are as below:

Fig. 4

figure 4

Data Loading:

Load the dataset into your Python environment. You can use libraries like Pandas to do this. Here's a basic example:

python

Copy code

```
import pandas as pd

# Load the dataset
data = pd.read_csv('your_dataset.csv')
```

Data Preprocessing:

Prepare the data for training your AI model. This typically includes handling missing values, encoding categorical variables, and scaling numerical features. Here's a simple example:

python

Copy code

```
from sklearn.preprocessing import StandardScaler

# Handle missing values
data = data.dropna()

# Encode categorical variables (if any)
data = pd.get_dummies(data, columns=['categorical_column'])

# Scale numerical features
scaler = StandardScaler()
data['numerical_column'] = scaler.fit_transform(data[['numerical_column']])
```

Split Data into Training and Testing Sets:

Divide your dataset into a training set and a testing set to evaluate your model's performance.

python

Copy code

```
from sklearn.model_selection import train_test_split

X = data.drop('target_column', axis=1)
y = data['target_column']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Build and Train Your Model:

Choose an AI model (e.g., a machine learning algorithm or a neural network) and train it on the training data. Below is a simplified example using scikit-learn's RandomForestClassifier:

python

Copy code

```
from sklearn.ensemble import RandomForestClassifier

model = RandomForestClassifier(n_estimators=100, random_state=42)
```

model.fit(X_train, y_train)

Make Predictions:

Use your trained model to make predictions on the testing data:

python

Copy code

y_pred = model.predict(X_test)

Evaluate Model Performance:

Assess how well your model performs by using appropriate metrics (e.g., accuracy, precision, recall, or custom metrics). Scikit-learn provides useful functions for this:

python

Copy code

from sklearn.metrics import accuracy_score, classification_report

accuracy = accuracy_score(y_test, y_pred)

print("Accuracy: ", accuracy)

print(classification_report(y_test, y_pred))

Remember that this is a simplified example. Depending on your specific use case, you may need more advanced data preprocessing, hyperparameter tuning, and potentially deep learning frameworks like TensorFlow or PyTorch. Additionally, you can use SQL or NoSQL databases to store and retrieve your dataset if it's large or distributed.

Future Aspect and Research Issues

Artificial intelligence is influencing the future of almost every sector and every person on the planet. AI has acted as the driving force behind developing technologies for industrial automation, medical applications, agriculture, IoT applications, cybersecurity services, etc. summarized in “Future Aspect and Research Issues”, and it will continue to do so for the foreseeable future. This interdisciplinary science comes with numerous advancements and approaches that are possible with the help of deep learning, machine learning algorithms, knowledge-base expert systems, natural language processing, visual recognition, etc. discussed briefly in “Potential AI techniques”. Thus, by taking into account the capabilities of AI technologies, we illustrate three essential terms, mentioned in “Introduction” within the scope of our study. These are

PRESENTED BY

S.PRAVEENKUMAR

V.JAYAPRAKASH

S.ABISHEAK

M.JAYASEELAN