README

Data Science Project on Data Science Salaries

Nicholas Nagy

2023-04-04

Employment Type Predictor using Decision Tree Classification

This project uses a Decision Tree Classifier from the `sklearn` library to predict the employment type based on salary and remote work ratio. The model is trained and tested on a dataset of salaries (`ds_salaries.csv`).

Prerequisites

The project is implemented in Python. You need to have the following Python libraries installed:

- pandas
- pyarrow (not required but suggested)
- sklearn

You can install these packages using pip:

```
pip install pandas pyarrow sklearn
```

Usage

Run the `ClassificationModel.py` script:

```
python ClassificationModel.py
```

Project Structure

The project consists of a single Python script `ClassificationModel.py` which includes:

- A `ClassificationModel` class that encapsulates the Decision Tree Classifier model from `sklearn`. It includes methods for training the model, making predictions, and evaluating the model's performance.
- A `Model` function that splits the data into training and testing sets, trains the model, makes predictions, and evaluates the model's performance.
- The script reads a CSV file named `ds_salaries.csv` into a pandas DataFrame. The target variable is set to the `employment_type` column of the DataFrame, and the features are set to the `salary_in_usd` and `remote_ratio` columns.
- The script runs the `Model` function a specified number of times (`NUM_OF_TESTS`), each time adding the accuracy score to a running total. Finally, the average accuracy score is calculated and printed to the console.

Authors

- Nicholasnagy-ux