HackerRank Data Scientist Hiring Test

01h : 01m
to test end

1

☆ Predict Life Expectancy

Governments, research institutes, and organizations like the United Nations and the World Bank try to understand the relationship between the life expectancy of a country or a geographical area and socioeconomic factors. Such analysis is valuable in deciding economic and social policies. Can you construct a reliable model that predicts the life expectancy of an area (country, region, group of countries) using socioeconomic variables and identify how different features influence that?

Files:

train_data.csv	Data used for training along with target variable
test_data.csv	Data on which predictions are to be made
submission_example.csv	Example of the submission containing the real index and random predictions.

Every row of the train_data or the test_data represents socioeconomic variables of a geographical area. That area could be a country, a group of countries, a region or a big country's provision.

Schema

Goal:

For every row in the test data, you must predict the value of the life expectancy. The predictions must be saved in a .csv file with the name 'submission.csv'.

The CSV file must have two columns.

- The first column must be the index of the test set
- The second column must have the predicted value of every corresponding index value.

Evaluation Metric:

The metric used for evaluating the performance of the predictive model will be the mean absolute error of the predictions from the ground truth (the real values of the life expectancy for every row in the test set).

```
Mean Absolute Error = the average absolute difference between y_true and
y_pred where y_true refers to real true values (ground truth) and y_pred
refers to predicted values.
```

Deliverables:

- submission.csv
- Well commented Jupyter notebook

Your notebook should contain your solution, visualizations, and thought process, including the top features that go into the model. If required, please generate new features. Make appropriate plots, annotate the notebook with markdowns, and explain the necessary inferences. A person should be able to read your notebook and understand the steps are you taking and the reasoning behind them.

Reset Project Fullscreen

HackerRank Data Scientist Hiring Test

01h : 01m
to test end

/

Name	Last Modified
Questions.ipynb	14 minutes ago
submission_exam...	29 minutes ago
test.csv	29 minutes ago
train.csv	29 minutes ago

Code

Data Science Challe

[11]:
```
# If you'd like to install packages
# This will ensure your notebook ha

#import sys
#!{sys.executable} -m pip install
```

[12]:
```
#Libraries
import pandas as pd
import numpy as np
pd.set_option("display.max_columns"
```

Data Description

Column
surface_area
agricultural_land

0 1 Python 3 | Idle

Submit answer & continue

You can change your submission later.

HackerRank Data Scientist Hiring Test

01h : 01m
to test end

1