Assignment 1: Decision Trees

Procedure:
The data contains the time-series percentage increase in COVID-19 cases worldwide. The attributes are date, confirmed cases, recovered cases, number of deaths and increase rate of confirmed cases. Target attribute is the rate of increase in confirmed cases.

The data, as it was continuous, is made discrete as follows

- Month of the date
- log(base10) of the confirmed cases, recovered cases and number of deaths
- A few ranges of percentages for the target value

Then the **ID3 algorithm** applied on the data yielded the decision tree.
For the pruning, **chi-square statistic** with the chi-square-threshold of 1.65 is implemented.
The maximum depth reached is 4, so the decision tree is the same as when the depth is more than 4

Results:
Decision Tree with the maximum depth 4 yielded an accuracy of **91.30%**
Average of accuracy of 10 random splits with

- Depth 4: Accuracy = 80.87%
- Depth 3: Accuracy = 79.56%
- Depth 2: Accuracy = 80%
- Depth 1: Accuracy = 65%
- Depth 0: Accuracy = 55.65%