

INFO411: Data Mining and Knowledge Discovery

Project 12

Instructions:

This task is a real-world data mining problem. You are required to prepare a set of presentation slides which must include (1) the full name and student number of each student in the group, the contribution (in percent) of each group member, (2) your proposed data mining approach and methodology; (3) the strengths and weaknesses of your proposed approach; (4) the performance measures that can evaluate your data mining results; (5) the results and a brief discussion. Below is the recommended structure of your slides:

- Introduction (define the problem and the goal)
- Methods (propose approaches, and discuss their strengths and weaknesses)
- Results (Figures and tables of data analysis)
- Discussion (discovered knowledge from data mining)

Task: Sea Temperature and Salinity

Background:

Acidification moorings are co-located at three National Reference Stations: the Yongala NRS in Queensland (replaced in September 2013 after Tropical Cyclone Yasi); the Maria Island NRS in Tasmania; and the Kangaroo Island NRS in South Australia (removed in June 2013, to be redeployed in April/May 2014). These moorings provide key observations to help understand and address the problem of increasing ocean acidification in Australian waters. The Yongala and Maria Island acidification moorings are located to characterise changes down the east coast of Australia and the influence of the East Australian Current on CO₂ uptake and acidification from the Great Barrier Reef to the Southern Ocean. The Kangaroo Island mooring monitors the deeper waters upwelled on the South Australian shelf which are expected to have higher CO₂

The observations include CO₂ concentrations in air and water, surface water temperature and salinity.

<https://imos.aodn.org.au/>

Note that some information such as latitude and longitude are not very useful for this data mining task.

Requirements:

1. Explore the relationships between CO₂ concentrations in air and water, surface water temperature, and salinity.
2. Also, investigate any trends over time and any differences between locations.
3. Present relevant visualisations of the data, which help to illustrate the relationships, trends and differences found in the previous items.
4. Provide the performance evaluation of any fitted models, including details of cross-validation or splitting into training, validation and/or testing sets.
5. Present your interpretations and conclusions.