

# INFO411/INFO911

## Data Mining and Knowledge Discovery

### Project 5

#### 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) a description of the task, (3) your proposed data mining approach and methodology; (4) the strengths and weaknesses of your proposed approach; (5) the performance measures that can evaluate your data mining results; (6) the results a brief discussion and a conclusion.

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/or tables of data analysis)
- Discussion (discovered knowledge)

#### Task: Global surface Temperature Analysis and Prediction

##### **Background:**

The analysis and prediction of global surface temperatures play an important role in many areas such as atmospheric weather prediction and climate change analysis just to name a few. The National Oceanic and Atmospheric Administration (NOAA) releases global weather observation data on a regular basis. Those measurements are obtained by either earth bound sensors or from observations from space.

You are to work with the NOAA global surface temperature anomaly dataset which can be retrieved from the following link:

<http://www.esrl.noaa.gov/psd/data/gridded/data.noaaglobaltemp.html>

The dataset contains monthly temperature anomalies and statistics from 1880 to present. The data is spatially gridded in cells of size 5° x 5° (longitude x latitude). The dataset is in netCDF format and can be converted to text via the command

```
ncdump -b c air.mon.anom.nc >air.mon.anom.nc.dump
```

The required software is available from: <ftp://ftp.unidata.ucar.edu/pub/netcdf/>

##### **Definition of the task:**

Use the temperature anomaly dataset (air.mon.anom.nc) to detect patterns (i.e. seasonal patterns), detect anomalies, and to make a prediction on temperature anomalies in the future. Note that the latter is a time series learning problem.

1. Visualize your results.
2. Provide a qualitative analysis of your results.
3. Compare your results with those published by Thomas M. Smith in the paper "Improvements to NOAA's historical merged land-ocean surface temperatures analysis", 2007.

##### **Requirements:**

1. Present a general description of the dataset and present a the general properties of the dataset.
2. Deploy data mining methods suitable for the analysis of the dataset.
3. Deploy data mining methods suitable for predicting future temperature anomalies.
4. Summarize: What new and interesting things did you discover while working on this project?