

INFO411/INFO911

Data Mining and Knowledge Discovery

Instructions:

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 (illustrative Figures and Tables and explanations)
- Discussion (discovered knowledge?)

Task: Modeling Songs to Predict Release Year

Background:

The popularity of music genres and music styles is changing over time. This raises the question on whether it is possible to predict the release year of a song from its audio features.

Definition of the task:

Download the dataset available at:

<http://archive.ics.uci.edu/ml/datasets/YearPredictionMSD#>

You are to investigate in the difficulty of the task, investigate suitable classification methods for the task, and offer comparative results along with an analysis of the results.

Requirements:

1. Present a general description of the dataset and present the general properties of the dataset.
2. Describe the approach and methods that you used for this task. Discuss the strengths and limitation of these methods.
3. Present and explain your answers to the objectives of this project.
4. Offer a discussion on (the suitability of) alternative approaches that may have been considered for this task.
5. Summarize: What new and interesting things did you discover while working on this project?