Welcome to COS 781

Data Mining
People

- Course Coordinator and Lecturer
  - Will van Heerden
    - IT 5–38
    - wvheerden@cs.up.ac.za
    - E-mail communication preferable
    - Unavailable in weeks that Mr Riekert is lecturing
People

- Co-lecturer for the module
  - Marius Riekert
    - IT 5–43
    - mriekert@cs.up.ac.za
    - E-mail communication preferable
    - Unavailable in weeks that Mr Van Heerden is lecturing
What is this course about?

- **What it is:**
  - Data mining and exploratory data analysis
    - Searching for hidden meaning in data
    - A focus on AI-based techniques
  - A computer science subject
    - Focussed on algorithms
    - Doesn’t necessarily require you to write code
    - Requires written assignments (nothing too strange…)
    - Requires empirical analysis (experiments…)
    - Requires report writing (not as easy as it sounds…)
What is this course about?

- What it is:
  - Problems that data mining can be applied to:
    - Fraud detection
    - Analyzing the stock market
    - Medical analysis
    - Cross-marketing
    - Designer drugs
    - Homeland security
    - Crime prediction
What is this course about?

- What it is:
  - Data mining theory
    - Background
    - Data preparation
    - Exploratory data analysis
    - Data clustering
    - Self-organising maps for data mining
    - Induction and decision trees
    - Evolutionary computation for data mining
    - Swarm intelligence for data mining
    - Maybe some other topics (perhaps with guest lecturers)
What is this course about?

- What it is:
  - Two written assignments
    - Cover a section of work we’ve done in class
    - Some questions on that work to test your insight
    - Some questions on stuff that’s related but that we didn’t discuss in class (you’ll have to do research on these)
  - Some practical data mining exercises
    - Experiments
    - Report writing
  - Some presentations by you…
What is this course about?

- What it isn’t:
  - A detailed introduction to the relevant AI techniques
    - If you didn’t do (or didn’t enjoy) the first semester AI course, don’t worry too much
    - We’re not interested in the deep theory of these methods
    - As long as you know the basics of neural networks, EC, PSOs, and (maybe) ACOs, you’ll be fine
    - However, you have to be willing to put in effort to catch up on work you haven’t studied before
  - A databases or data warehousing course
  - A tutorial on the tools that are available
Why Data Mining?

- The world is increasingly data driven
  - More than 4.9 million English articles on Wikipedia
  - 1.44 billion users on Facebook, all generating data

- An extreme shortage of data mining experts
  - Particularly in South Africa
  - Internationally too

- Data mining is still a young field
  - Much future potential
  - Exposure to front-line research
Material

- **Study guide**
  - On the course website
  - Read it!

- **Study material**
  - No prescribed texts
  - Our own slides and notes
    - On the website as the course progresses
  - *Any additional material we give you…*
    - Articles, surveys, reports, etc.
  - Google and the library…
Assessment

- Four assignments
  - Two written assignments
  - Two reports on a real-world data set
    - Academic writing!
      - Spelling, grammar & format are very important
      - Write clearly, give complete background
      - Assume the reader has no prior knowledge of AI or data mining (explain everything that you use)
    - Proper experimental procedure!
      - 30 simulations, statistical comparisons
      - Justify your parameter settings
Assessment

- **Presentation**
  - A topic in data mining
    - Something that wasn’t covered in class
    - Pick something new, interesting and relevant
      - Big data
      - Unstructured data mining

- **Assessment**
  - Both lecturers evaluate
  - Averaged between the two marks

- Final lecture of the semester
Assessment

- Semester mark
  - Assignments (85%)
  - Presentation (15%)
Assessment

- Exam
  - Three-day take-home exam
    - It is three days long to give you time to consider your answers carefully, and to think
    - Not because we want three continuous days of writing (use the mark allocation as a guide)
  - Further details to be announced...
  - All the material we cover in class, as well as some things you will have to research
Assessment

- Final mark
  - Semester Mark (50%)
  - Examination (50%)
Assessment

- Additional details
  - All students have automatic exam entrance
  - You have to write the exam
  - To pass: final mark $\geq 50\%$
  - Distinction: pass with final mark $\geq 75\%$
  - No sick exam or supplementary
A final word…

- Course website
  - www.cs.up.ac.za/courses/COS781
  - Downloads available here
  - Check announcements regularly

- Plagiarism
  - You know what it is by now
  - UP is very serious about this
  - Don’t do it!

- Don’t underestimate the course

- Class representative…