Course Overview

HAP 780 Data Mining in Health Care

Janusz Wojtusiak, PhD
George Mason University
Fall 2016
“Two men were examining the output of the new computer in their department. After an hour or so of analyzing the data, one of them remarked: Do you realize it would take 400 men at least 250 years to make a mistake this big?”

-Nicholas Negroponte
Presentations may be Recorded

• I may record some lectures for use for future students
• The camera is facing screen/whiteboard/me
• Your voice may be recorded if you ask questions
  – If you do not wish to be on the recording, please let me know and help me identify moments of the presentation on which your voice is present.
Syllabus

• How to contact me
  – Email: jwojtusi@gmu.edu
  – Office hours: By appointment, before class, Wednesdays 1pm – 4pm
  – Office location: 108 Northeast Module
  – Please do not leave voicemail

  – Please include HAP 780 in subject of all emails
Syllabus

• How to contact you
  – GMU email
    • Check SPAM/JUNK folders
    • Remove old emails
    • Set-up forwarding
  – MyMason portal (blackboard)
Syllabus

• Recommended Prerequisites
  – HAP 700 Introduction to Health Informatics
    • Informatics background
  – HAP 709 Healthcare databases
    • Database skills
  – HAP 602 Statistics
    • Statistics/research skills are important
  – HAP 780 should be taken close to the end of the program
Course Objective

- Make your life miserable ...
In this class you will learn

• What is KDDM and what it does
• Which KDDM methods to apply
• How to apply KDDM methods
• How to interpret and apply results
• What is relation of KDDM to other disciplines

• Right combination of theory and practical skills
Your Part

• Participation
• Weekly assignments
• Weekly reading
• KDDM topic presentation
• Semester-long project
• Participation outside classroom
Grading

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>96+</td>
<td>A</td>
</tr>
<tr>
<td>90-95</td>
<td>A -</td>
</tr>
<tr>
<td>86-89</td>
<td>B +</td>
</tr>
<tr>
<td>80-85</td>
<td>B</td>
</tr>
<tr>
<td>75-79</td>
<td>B -</td>
</tr>
<tr>
<td>70-74</td>
<td>C</td>
</tr>
<tr>
<td>0-70</td>
<td>F</td>
</tr>
</tbody>
</table>
Reading


More Reading

- Data Mining and Knowledge Discovery (Springer)
- IEEE Transactions on Knowledge and Data Engineering
- Journal of American Medical Informatics Association (Oxford)
- Artificial Intelligence in Medicine Journal (Elsevier)
- Health Informatics Journal (Sage)
- Journal of Intelligent Information Systems (Springer)
- International Journal of Medical Informatics (Elsevier)
- Machine Learning (Springer)
- ... and many more ....
Presentation

• Prepare ~10 minute presentation on a selected topic related to data mining or healthcare data
• Every week 1-2 presentations
• Should describe some specific things and be interesting
  – Example: Find a DM-related article and present it

• Avoid two presentations about the same
• Avoid presentations about topics that will be covered in class (see weekly schedule in the syllabus)
• Topics and dates priority based on when you sign up
  – We will use doodle to set up presentations
Presentations

• Use topics from published papers. Here are some examples:
  – Exploring Generalized Association Rule Mining for Disease Co-Occurrences (AMIA 2012)
  – Leveraging Derived Data Elements in Data Analytic Models for Understanding and Predicting Hospital Readmissions (AMIA 2012)
  – The Analytic Information Warehouse (AIW): A platform for analytics using electronic health record data (JBI 2013)
Projects

- Most important component of the grade
- **Use some real data, terminologies, systems!**
- Analyze data to get new results
- Create new DM methods
- Test and compare methods
- Build classification or prediction models
- Detect outliers in data

- Focus on a specific group of patients, make sure the experiments are well designed (problem, cohort, ...).
Projects

• Most project reports should include:
  – Problem description
  – Data selection
  – Data pre-processing
  – Selection of DM methods
  – Application of DM methods
  – Analysis of results
  – Review of literature and related work
  – Conclusions
Projects

• Direct application of existing methods to existing datasets is not sufficient

• Projects must demonstrate significant efforts as well as understanding of problems

• Use dataset(s) that is available
Participation Outside Class

• It is not about membership (and paying fee)
• It is about going to a meeting, talking to people, doing some networking, participation
• If you present something (on a conference or workshop) it is even better
Membership/Participation Examples

- AAAI Fall Symposium Series
- HIMSS annual conference
- AMIA symposium
- HIMSS Capital area
- HIMSS VA
- Health Tech Net
- Academy health
- HIMSS Mentorship program!
Honor Code

• Do not cheat ....
• Do things yourself
• Do not plagiarize

• Do not copy-and-paste when doing assignments

• Details in the syllabus
Individuals with Disabilities

- Work with GMU disability office
- Details in the syllabus
Schedule

• Posted in the syllabus in Blackboard
• No class next week – labor day

• For all important dates, check academic calendar on GMU website.
Contact Again

• Janusz Wojtusiak, PhD
• jwojtusi@gmu.edu

• In emails, please make sure I know who you are.
• Please include HAP 780 in subject of all emails