INFO296A-2: thought Leaders n Data Science and Big Data Analytics Project and Seminar Report Instructions and Guidelines, v1.0

Objectives:

1. **A) Projects:** 1. Full blown product development for data science and analytics startup.
   2. Data Science and Analytics projects with real world data and context.
   3. Research version of 2 (above) with algorithm development and paper
   - Mentors will be from from industry, VCs and campus

   B**) Seminar Reports**: Thorough (Competitive) analysis of landscape of
   -i) Enterprise Analytics firms
   ii) Web Analytics firms
   iii) IoT firms
   iv) Various verticals, both high tech and otherwise

2. Students will form teams, typically 3-4 students, with a team lead/POC (with other substituting as required).
   - Each team: Please provide list of names, email addresses, cell # contacts, and the name of the team lead

3. Each team will submit either
   - A Project Report
     OR
   - A Seminar Report

4. Each team will need to define which team member contributed to which part of the report.
5. Each report would be 10-20 pages long, typically 15-20 pages, although focused and substantive content is far more important than length.

6. Projects: Could be for 2 & 3 units. The units will be used in calibrating effort and performance.
7. Project participation subject to review of CVs, project proposal, and instructor approval

8. Seminar Reports: Typically 2 units. Could be 3 units with approval from instructor.

9. Seminar cover a set of broad ranging industry talks, and provide perspectives on
   - Domain and business needs
   - Infrastructure needs
   - Analytics needs
   - State of the art

   8. Due dates:
   A) Seminar reports

- Seminar reports are due 3/14 (Seminar Report 1) and 5/2 (Seminar Report 2) ;
- Preferred: 3/13 (Seminar Report 1), 5/1 (Seminar Report 2)
- Drafts are due 3/7 (Draft Seminar Report 1) and 4/25 (Draft Seminar Report 2)

B) Project Reports

- Reports due Project report Phase 1 - 2/8 (extended to 2/22), Project report Phase 2- 3/7, Project report Phase 3 - 4/11, Project report Phase 4 - 5/2

i) Project Report – Phase 1: Verbal description of problem, identified data sources and gaps, tentative models
ii) Phase 2: Model of problem, business model, initial data analysis. Write down state variables, decision variables etc, as also dependent and independent variables, and differentiate. Explain the model and rationale for developing it.
iii) Phase 3: Mode development and analysis. Additional data analysis, Preliminary conclusions.
iv) Phase 4: Final model, analysis, results, insights etc.
v) Please provide and cite references

10.Project participation subject to review of CVs, project proposal, and instructor approval

11. Project types:
Either
i) Startup product development (with industry mentor)
OR
ii) Industry/real-world problem solution (with industry data and engagement)
OR

ii) industry/real-world problem solution and research/algorithm development and testing/validation (with industry data or other data)

OR

iV) industry/real-world problem solution with challenge or open source data

Themes:
- Enterprise Analytics:
Enterprise databases (DB)and Business Intelligence (BI)
- Web Analytics
Leading to Hadoop, Spark/Shark, Streaming + Analytics
- Internet of Things

Continuous sensing and proactive response
What is new and different about it?

Data Science & Analytics:
Components
Data collection, storage, and basic processing
Architecture and Infrastructure
Analytics
Domain
Business Needs
To solve real Big Data problems, need expertise in some or all of these areas
Need to form teams!

12. Seminar Report Format: Thorough (Competitive) analysis of landscape of
-i) Enterprise Analytics firms

ii) Web Analytics firms
iii) IoT firms
iv) Various verticals, both high tech and otherwise

- Identify the needs and gaps in each industry vertical or segment
- Identify the state-of-the-art from speakers, other sources (journals, magazines, newspapers, white papers, websites, research papers, products and services of firms, interviews etc.)
- Identify potential opportunities for short-, medium-, and long - term products and services, as also algorithms, infrastructure, business model and other enablers/solutions
- Do this by enterprise/web/IoT cut, as well as verticals e..g Healthcare,
- Identify challenges and away forward
- Explain your logic an d
- Use tables, graphs, charts, etc. , as needed, to succinctly communicate the main points
- Include references (separately, and in addition to the 20 pages)
- Note: Given the Convergence of Technologies in ICT (Information and Communication Technologies) and their prevalence in High Tech (and Low Tech), when comparing firms in a vertical or industry segment, choose firms or their sub-units very carefully.
- Please study Porter Model and Business Models

- Themes:
  - Enterprise Analytics:
  Enterprise databases (DB)and Business Intelligence (BI)
  - Web Analytics

Leading to Hadoop, Spark/Shark, Streaming + Analytics
- Internet of Things
Continuous sensing and proactive response
What is new and different about it?

Data Science & Analytics:
Components
Data collection, storage, and basic processing
Architecture and Infrastructure
Analytics
Domain
Business Needs
To solve real Big Data problems, need expertise in some or all of these areas
Need to form teams!

Format: Identify

i) Industry vertical and/or horizontal needs
ii) Current state of the art of product, services, algorithms
iii) Reasons for gap between need and available solution
iv) Solution/resolution: Product, service, (infrastructure, software), algorithms, UI,business model, culture etc. etc.
v) Integrate speakers 'inputs and contrast with your knowledge and reading.
vi) Provide landscape perspective, and growth over time, not a linear extrapolation.