|  |  |
| --- | --- |
|  | Charles F. Dolan School of Business  IS 550 – Business Analytics & Big Data Mgmt., Course Syllabus, Fall 2018 |

# PROFESSOR INFORMATION

|  |  |  |
| --- | --- | --- |
| **Name**: Dr. Jie Tao | **Email**: [jtao@fairfield.edu](mailto:jtao@fairfield.edu) | **Phone**: (203) 254-4000 x 2232 |
| **Office**: DSB 2102 | **Office Hours**: By appointment | |

# COURSE INFORMATION

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Number:** IS 550 | **Credit Hours**: 3 | **Title:** Business Analytics and Big Data Management | |
| **Course Description**:  In this course, we will focus on big data analytics, from a balanced theoretical and practical perspective. We will start with the *scope*, *role*, and *principles* of big data in today’s business context, followed by two important pillars in big data analytics, namely *Parallel Computing* and *Deep Learning*. These two pillars represent the latest development in the business analytics/data science domain.  In addition to the understanding of aforementioned two pillars, students are supposed to complete a comprehensive project, with either a research or a practical flavor of the project. The comprehensive project is used to test the capabilities and understanding of all the knowledge in the domain – with a focus in any of the business domains (i.e. finance, accounting, marketing, etc.). | | | |
| **Prerequisites:** IS540, Working Python Programming Knowledge | | | **Course Duration**: Sep., 6th – Dec., 13th |
| **Course Delivery Method**: Meeting and Online (Blackboard: <http://fairfield.edu/blackboard>) | | | |

# COURSE MATERIALS

**Required Textbook:** None.

**Course Materials:** Major materials used in this course include academic articles, videos, lab tutorials, and other online resources. They will be distributed via Blackboard Course Management System under “course contents”. Additional Material will be released over websites such as Github, Kaggle, or Databricks.

**Technical Requirements:** We will use Apache Spark on a public cloud (<http://www.databricks.com/)> for the Spark part of the course. Students should also have a working IPython/Jupyter environment (i.e. via the Anaconda distributions). In addition, we will use the Nvidia Deep Learning Institute (DLI) lab training materials – which are distributed via Amazon Web Services and the Nvidia training platform. Detailed instructions will be provided as the course progresses.

**Reference Books:** The following books serve as good reference material for topics covered in this course:

* NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, by Pramod Sadalage, Martin Fowler, 1st edition, ISBN: 978-0-32-182662-6.
* Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, by EMC Education Services, 1st edition, ISBN: 978-1-11-887613-8.

# LEARNING OBJECTIVES

Upon completion of this course, the students should be able to:

* **LO 1.** Develop a broad understanding of the term “*big data analytics*”, and the techniques, technologies, and tools it encompasses;
* **LO 2.** Gain an appreciation via industrial applications and real-world scenarios regarding big data issues – *volume*, *velocity*, *variety*, and *veracity*;
* **LO 3.** Gain an overview of parallel computing platforms (e.g. Spark), and its various use cases;
* **LO 4.** Gain an overview of the deep learning technology, and its various applications;
* **LO 5.** Understand some latest and important applications of business analytics, namely text mining and computer vision – and apply the working knowledge in research/analytical projects.

# INSTRUCTIONAL METHODOLOGIES

Primary instructional methodologies would include: lecture notes; presentation slides; Lab work; supplementary reading materials; e-mail; and discussion board.

# GRADING POLICIES

This Course is graded based on a point system (with a total of **1,000** points), with multiple deliverables. The deliverables and associated points are shown below. I reserve the right to change following policies when deemed needed.

|  |  |
| --- | --- |
| **Deliverables** | **Point Value** |
| Labs (8, each worth 50 points) | 8 x 50 = **400** points |
| Course Project (2 Presentations & 1 Final Report) | **2\*100 + 300 = 500** points |
| Participation and Discussions | **100** points |
| Total | **1,000** points |

**Grading Scale** – *Percentages and letter grades are provided for your convenience, as this Cl (noted above) is based on a point system*.

|  |  |  |
| --- | --- | --- |
| **Total Points** | **Percentage** | **Letter Grade** |
| **930 – 1,000** | **93- 100** | **A** |
| **900 - 929** | **90 – 92.9** | **A-** |
| **870 - 899** | **87 – 89.9** | **B+** |
| **830 - 869** | **83 – 86.9** | **B** |
| **800 – 829** | **80– 82.9** | **B-** |
| **770 – 799** | **77 – 79.9** | **C+** |
| **730 – 769** | **73 – 76.9** | **C** |
| **700 - 729** | **70 – 72.9** | **C-** |
| **599 and below** | **59.9 and below** | **F** |

**Labs:**

8 labs have been set up for you to get hands-on experiences with big data analytics tools (please refer to course outlines – Page 5 for details). These tutorials are set up due to the hands-on nature of this course. The tutorials will cover all important aspects of this course, including parallel computing (Spark), deep learning and its applications (text analytics, computer vision, etc.). *Submission guidelines*: for some of labs, you will need to take screenshots of important steps in the lab, then copy and submit them in a MS Word document; for the rest of the labs, you should submit your IPython/Jupyter notebooks. You will have until the next meeting to submit your lab work (i.e. your first submission should be on 9/20 – by 6PM EST) – late submissions will not be graded.

**Course Project:**

There will be one (1) *comprehensive* and *individual* course project – which purpose is to assess your understanding of the topics and technologies discussed in this course. You have to select one (1) of two options below for your course project:

1. **Practical/Analytical project**: you will conduct a real-world like data analytics project – from business understanding to evaluation/deployment of your model. You need to embed at least one (1) topic/technology discussed in this course in your course project; or
2. **Theoretical/Research project**: you will conduct a literature review/case study project – where you will need to intensively analyze research articles/real world cases using at least one (1) topic/technology discussed in this course (some sample topics may include (but not limit to): deep learning/deep neural network; text mining/text analytics/topic modeling/sentiment analysis; computer vision/image processing.

General guidelines and grading rubrics can be found in Appendix A of this syllabus.

Despite which option you are selecting for the course project, there are several required deliverables in this course:

1. *Project Proposal*: specifying the project scope, expected outcome, responsibilities, and a possible time line of the project – this is supposed to be delivered in the form of presentations (100 points) – please note that the proposal needs to be approved by the instructor before moving on with your project;
2. *Progress Review/Milestone*: to be submitted in the pre-defined date (see Section 5 of this document for details) for the purpose of communicating progress and issues within the project (required, 0 points);
3. *Final Presentation:* entailing the progress and achievements in the project, reviewing methods/techniques used in the project, and communicate with the instructor/client – the presentation will be completed meeting (100 points);
4. *Final Report/Deliverable:* depending on the selection of the project options, the final deliverable can be a written report (for theoretical/research project) or an analytical report/pipeline (for practical/analytical project) – the deliverable will be used as the final assessment of the project, and this course (300 points).

**Deliverables Formatting Guidelines:** The report for each assignment should be **typed** in the form of a **single** Microsoft Word document. Note that hand-written deliverable reports will **NOT** be graded. The template for assignment report can be found in the course website (under “course contents”) – it is **REQUIRED** to use the template for your assignment reports where applicable. Diagrams used in each report need to be created using graphical tool (i.e. Microsoft Visio). Hand-drawn diagrams will **NOT** be graded. You should make use of them by submitting draft versions of the deliverables to me for feedbacks and comments – although they are not mandatory. Important dates for these deliverables are illustrated in Page 5 of this document.

# COURSE OUTLINE

The following course outline is tentative and subject to change when and where deem needed.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Week | Lecture | LO. | Meeting Lab \* | Project Milestones |
| W1 1 | **Orientation to course**   * Intro to Big Data Analytics * Big Data Analytics Life Cycle | **1, 2** | **Lab 1:**   * **Setup Spark on Databricks** * **Python Review** | **Course Project Kick-off** |
| We 2 | **Analytical Framework:**   * Spark I | **3** | **Lab 2:**   * **Spark Basics/Structured Streaming** |  |
| W 3 | **Analytical Framework cont’d**:   * Spark II | **3** | **Lab 3:**   * **Spark for Data Scientists** |  |
| 4 | **Big Data Visualization:**   * Spark III | **3** |  | **Project Proposal Presentation** |
| 5 | **Big Data Application: 1**   * Text Analytics/NLP | **5** | **Lab 4**   * **Text Analytics Python Labs** |  |
| 6 | **Big Data Application: 2**   * Text Analytics/NLP | **4** | **Lab 5**   * **Text Analytics Python Labs** | **Project Coaching Session I - TBA** |
| 7 | **Business Analytics Development:**   * Deep learning part I | **4** | **Lab 6:**   * **Deep Learning Intro** | **Milestone Report Due** |
| 8 | **Business Analytics Development:**  Deep learning part II | **4, 5** | **Lab 7 :**   * **Computer Vision labs** |  |
| 9 | **Business Analytics Development:**  Deep learning part III | **4** | **Lab 7 cont’d :**   * **Computer Vision labs** |  |
| 10 | * **Deep Learning on Computer Vision** | **1, 5** |  |  |
| 11 | **Other Deep Learning Applications-1** | **4,5** |  | **Project Coaching Session II - TBA** |
| 12 | **Other Deep Learning Applications-2** | **4,5** | **Lab 8 :**   * **Twitter Data** |  |
| 13 | **Final Chapter:**  (Big) Data Analytics career path | **1,4,5** |  |  |
| 14 | **Final Project** |  |  | **Meeting Final Presentation**  **Final Report Due: Dec, 13th** |

# Appendix A: Project Grading Rubrics

Following rubrics are tentative and subject to change when and where deem needed. Please not the letter grades provided are only for your convenience.

Option I: Practical/Analytical Project

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Content** | | | | **Presentation** | | | |
|  | **A/A-** | **B+/B/B-** | **C+/C/C-** | **F** | **A/A-** | **B+/B/B-** | **C+/C/C-** | **F** |
| **Proposal Presentation/ Final Presentation** | 1. Defines the business problem in the project clearly; 2. Discuss the data collection method/process clearly; 3. Presents the steps in the project clearly – at least 4 in CRISP-DM; 4. Defines the timeframe of the project, and each step in it clearly; 5. Define the final outcome of the project clearly. | 1. The business problem is defined; 2. Process of data collection is discussed; 3. Some steps in the project are presented; 4. Timeline is defined – but some steps are missing; 5. The final outcome of the project is mentioned. | 1. The business problem is not clearly defined; 2. Data collection is not clearly discussed; 3. Crucial steps in completing the project are missing; 4. The time frame is not clear; 5. The final outcome of the project is ambiguously defined. | 1. The business problem is not defined; 2. Data collection steps not discussed; 3. Subsequent steps in project not discussed – or a majority of steps missing; 4. Timeframe is not defined – or it is not feasible; 5. Final outcome is not defined – or it is not feasible. | 1. Presentation Slides are designed professionally and clearly; 2. Presentation is attractive and interesting; 3. Presentation slides are embedded with diagrams; 4. Presentation slides are created without typo; 5. Presentation is interactive and generates discussions. | 1. Presentation Slides are self-explanatory; 2. Presentation is informative; 3. Diagrams in slides are not illustrative; 4. Typo are scarce in the presentation slides; 5. Presentation is somewhat interactive and generates discussions. | 1. Presentation Slides are logical; 2. Presentation is informative but missing some crucial parts; 3. Diagrams are not used when needed; 4. Some typos can be found on slides; 5. Presentation is a little interactive. | 1. Presentation Slides are poorly designed; 2. Presentation is not informative; 3. Typos are popular in presentation slides; 4. Presentation is not interactive and generates no discussions. |
| **Final Report** | 1. Discusses all important steps in the project; 2. Explains the final outcome in the project clearly; 3. Techniques/methods used in project well explained; 4. Results are well presented and interpreted; 5. Coverage of the data mining process/pipeline is clear. | 1. Discusses mostly important steps in the project; 2. Provides some explanation toward the final outcome; 3. Techniques/methods used in project are explained; 4. Results are presented, and some interpretation provided; 5. Phases in the data mining project are discussed. | 1. Some important steps in project are not discussed; 2. The final outcome is not explained, or does not match the proposal; 3. Techniques/methods are not explained; 4. Results are not clear or missing crucial parts; 5. Important phase in the data mining pipeline is not discussed. | 1. Some important steps in project are not discussed; 2. The final outcome is not explained, or does not match the proposal; 3. Techniques/methods are not explained; 4. Results are not presented; 5. The analytical pipeline is flawed. | 1. Writing style is professional and clear; 2. Writing is attractive and interesting; 3. Report is embedded with diagrams; 4. Report are written without typo; 5. Report is well indexed and versioned – proper citation style followed. | 1. Academic Writing style is followed; 2. Writing is grammatically correct; 3. Report is well organized; 4. Report are written without a few typos; 5. Report is indexed – and consistent citation style used. | 1. Academic Writing style is somewhat followed; 2. Writing has scarce grammar errors; 3. Organization of report is acceptable; 4. Report are written without some typos; 5. Report is not well indexed – and citation style is not consistent. | 1. Academic Writing style is not followed; 2. Writing has some grammar errors; 3. Organization of report is illogical; 4. Report are written without some typos; 5. Report is not indexed – and citation style is not consistent. |

Option II: Theoretical/Research Project

This website provides detailed instructions for doing literature review (<http://libguides.gatech.edu/researchprocess/litreview>).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Content** | | | | **Presentation** | | | |
|  | **A/A-** | **B+/B/B-** | **C+/C/C-** | **F** | **A/A-** | **B+/B/B-** | **C+/C/C-** | **F** |
| **Proposal Presentation/ Final Presentation** | 1. Defines the research problem in the project clearly; 2. Discuss the research database and information retrieval process; 3. Presents the steps in the project clearly; 4. Defines the timeframe of the project, and each step in it clearly; 5. Define the final outcome of the project clearly. | 1. The research problem is defined; 2. Process of information retrieval is discussed; 3. Some steps in the project are presented; 4. Timeline is defined – but some steps are missing; 5. The final outcome of the project is mentioned. | 1. The research problem is not clearly defined; 2. Data collection is not clearly discussed; 3. Crucial steps in completing the project are missing; 4. The time frame is not clear; 5. The final outcome of the project is ambiguously defined. | 1. The research problem is not defined; 2. Data collection steps not discussed; 3. Subsequent steps in project not discussed – or a majority of steps missing; 4. Timeframe is not defined – or it is not feasible; 5. Final outcome is not defined – or it is not feasible. | 1. Presentation Slides are designed professionally and clearly; 2. Presentation is attractive and interesting; 3. Presentation slides are embedded with diagrams; 4. Presentation slides are created without typo; 5. Presentation is interactive and generates discussions. | 1. Presentation Slides are self-explanatory; 2. Presentation is informative; 3. Diagrams in slides are not illustrative; 4. Typo are scarce in the presentation slides; 5. Presentation is somewhat interactive and generates discussions. | 1. Presentation Slides are logical; 2. Presentation is informative but missing some crucial parts; 3. Diagrams are not used when needed; 4. Some typos can be found on slides; 5. Presentation is a little interactive. | 1. Presentation Slides are poorly designed; 2. Presentation is not informative; 3. Typos are popular in presentation slides; 4. Presentation is not interactive and generates no discussions. |
| **Final Report** | 1. Discusses all important steps in the project; 2. Explains the final outcome in the project clearly; 3. Methodologies used in project well explained; 4. Results are well presented and interpreted. | 1. Discusses mostly important steps in the project; 2. Provides some explanation toward the final outcome; 3. Methodologies used in project are explained; 4. Results are presented, and some interpretation provided. | 1. Some important steps in project are not discussed; 2. The final outcome is not explained, or does not match the proposal; 3. Methodologies are not explained; 4. Results are not clear or missing crucial parts. | 1. Some important steps in project are not discussed; 2. The final outcome is not explained, or does not match the proposal; 3. Methodologies are not explained; 4. Results are not presented. | 1. Writing style is professional and clear; 2. Writing is attractive and interesting; 3. Report is embedded with diagrams; 4. Report are written without typo; 5. Report is well indexed and versioned – proper citation style followed. | 1. Academic Writing style is followed; 2. Writing is grammatically correct; 3. Report is well organized; 4. Report are written without a few typos; 5. Report is indexed – and consistent citation style used. | 1. Academic Writing style is somewhat followed; 2. Writing has scarce grammar errors; 3. Organization of report is acceptable; 4. Report are written without some typos; 5. Report is not well indexed – and citation style is not consistent. | 1. Academic Writing style is not followed; 2. Writing has some grammar errors; 3. Organization of report is illogical; 4. Report are written without some typos; 5. Report is not indexed – and citation style is not consistent. |