1. **Course number and name:**
   COE 594-05 Big Data and Analytics

2. **Credits and contact hours**
   3 – 0 – 3

3. **Prerequisites and/or co-requisites**
   Prerequisites: consent of instructor
   Competencies: Exposure to statistical methods and programming.

4. **Name and Contact Information of Instructor:**

   Dr. Imran Zualkernan

5. **Course Description (Catalog Information)**
   Covers the end-to-end process for big data analytics including Extract-Transform-and-Load (ETL), descriptive analytics, supervised and un-supervised learning methods, deep learning, and big data storage and clustering eco-systems. Includes model evaluation techniques.

6. **Textbook, title, author and year:**

   Primary:

   Readings, excerpts from book chapters, and notes.

   Secondary:

   Ian Goodfellow and Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press, 2016.

   Tan, Steinbach, Kumar, *Introduction to data mining*, Pearson New International Edition. First edition, Essex Pearson, 2014.

   Hadley Wickham, Garrett Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, O'Reilly Media, 2016. (http://r4ds.had.co.nz/)

   François Chollet with J. J. Allaire, *Deep Learning with R*, Manning Publications, January 2018.

   Foster Provost and Tom Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media; 1 edition (July 27, 2013).

   Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics: The New Science of Winning*, Harvard Business Review Press; 1 edition (March 6, 2007).

7. **Learning Outcomes**

This course requires student to demonstrate the following:

1. Identify and describe the various big data analytic processes.
2. Explain the components and alternatives of a big data eco-system.
3. Conduct complex data manipulation and visualization tasks on large data sets.
4. Design ETL techniques for complex and large data sets.
5. Apply and evaluate supervised learning methods large data sets
6. Apply and evaluate unsupervised learning methods for large data sets
7. Build, evaluate and optimize deep learning models for big data

8. Teaching and Learning Methodologies:

Different teaching and learning methods will be adopted in class to help students achieve the course's learning outcomes and to encourage student participation, creativity, and interaction with each other. Methods include formal lectures, class discussion, case studies, and a term project.

9. Course Topics and Schedule:

Topic	Weeks
Introduction to Big Data and Analytics – Case Studies	1
Introduction to Data Analysis and Visualization	1
Big Data Storage - Hadoop, Hive, RDFS and Spark	2
Extract Transform and Load (ETL)	2
Supervised Learning (Logistic regression, SVM, Bayes, Decision Trees)	2
Unsupervised Learning (K-Means, Gaussian Mixtures)	1
Evaluating Supervised and Unsupervised Learning Models	1
Deep Learning (CNN, RNN)	2
Explanation tools for Deep Learning Models	1
Optimization for Deep Learning Models	1
Evaluation	2
Total:	**16**