COURSE NUMBER: CSC 591/CSC 791
COURSE TITLE: **Algorithms for Data Guided Business Intelligence**
DESCRIPTION: Algorithmic design principles and best practices underlying data guided Business Intelligence (BI) will be taught through a set of hands-on use cases. Analytic pipelines for solving BI problems will be introduced from the end-to-end, practical guide (i.e., cookbook) perspective. These pipelines will be implemented through a series of mini-projects covering recommender systems, sentiment analytics, online advertisement, cybercrime and online fraud detection, Internet of Things analytics, social media analytics, web logs analytics, and supply chain analytics. The space of algorithms will include but will not be limited to deep learning, information fusion from dynamic heterogeneous and attributed graphs, and causal network inference. Tutorials and projects that teach students how to handle Big Data issues will utilize Apache Spark on top of lambda architectures. (3 credit hours)

PREREQUISITES:

- CSC 522 -- Automated Learning and Data Analysis (or its equivalent)
- Python Programming Language
- Undergraduate level knowledge of probability, statistics, and linear algebra.

Otherwise, consent of the instructor is required.

INSTRUCTOR:

- Prof. Nagiza F. Samatova
- Office: 2-272 EB 2, on Centennial Campus.
- Phone: 919-513-7575 (work) and 865-566-5586 (cell)
- Office hours are by appointment.
- Skype: nagiza.samatova1

COURSE STRUCTURE: The instructor will teach all the course lectures. Student learning activities will involve:

- ***Tutorials***: Apache Spark, R programming, Python Programming, Word2Vec+Doc2Vec or Tensorflow.
- ***Mini-Projects***: A github portfolio with 10 mini-projects for solving BI use cases from end-to-end perspective (team discussions but individual submissions).
- ***Lecture Quizzes:*** Weekly lecture material quizzes testing understanding of and critical reasoning about core concepts (individual).
- ***Reflection on Research Papers:*** For each topic (10 total), a reflection on 2-3 papers discussing pros and cons of the algorithms, innovations and contributions to the data science domain, and future research extensions (team submissions).
- ***Algorithmic Mini-homeworks:*** For each topic (10 total) (individual); randomly selected problems will be graded but solutions to all the problems will be provided.
- ***Capstone Project:*** Team-initiated capstone project solving a data-driven BI problem, including GitHub portfolio, demo, and presentation (in class/via WebEx during class time)
- ***Final Comprehensive Exam:*** A comprehensive individual home-take final exam, *in lieu of the Capstone Project*.

COURSE OBJECTIVES & LEARNING OUTCOMES (see details in Course Schedule):
By the end of the course, students will have gained knowledge of different algorithms and methodologies for (a) design and analysis of data science experiments and model diagnostics, (b) generalized linear models, (c) causal inference and reasoning under uncertainty, (d) deep learning, (e) pattern discovery in heterogeneous, dynamic, and multi-attribute graphs, (f) graph embedding, (g) information and model fusion, (h) internet of things analytics, and (i) dealing with Big Data using Apache Spark; they will be able to critically analyze the pros and cons of applying these techniques in different business intelligence contexts (recommender systems, sentiment analytics, fraud detection, market analysis, online advertising, social blogging, multi-modal searching), and will be aware of the broader set of applications that require such techniques. Finally, students will be able to conceptualize and design efficient and effective end-to-end algorithmic solutions for different business intelligence and decision making problems.

- to become comfortable with rigorous mathematical and statistical approaches to (a)-(h) categories of problems;
- to get an overview of and experience with the most prevalent algorithm design techniques: latent context modeling, embeddings, information fusion, extreme value theory for modeling rare events, causal reasoning, etc.;
- to acquire the ability to analyze the performance using a variety of metrics and from the end-to-end perspective in the context of real-world application requirements and constraints; and
- to be exposed to business intelligence problem domains in which algorithm design and analysis ((a)-(i)) have practical applications.

COURSE SCHEDULE (DRAFT):

See a separate Excel file with the Schedule

COURSE ASSIGNMENTS:

A. Diagnostic test:
- **Due Jan 10**
- Topics will be provided ahead of time
- Moodle quiz
- 2 hours to complete
- Mandatory

B. Topic-specific Assignments
- 10 topics in total
- see topic schedule (draft)

B.1: Lecture material quiz (591/791)
- Time estimate: 4 hrs/wk
  - lecture time (2.5 hrs)
  - quiz (1.5 hrs)
- Moodle quiz
- Core concepts
- What-if scenarios
- Pros and cons
- Questions to ask on the topic
- **Due every Sunday** on topics covered that week

B.2: Reflection on research papers
- Time estimate: 2 hrs/wk
- One submission per Team
- 2 pages (max.)
- 2-3 assigned papers
- Elevator speech: Underlying Intuition and Main Idea (591/791)
- Pros of the algorithms (591/791)
- Cons of the algorithms (591/791)
- Answer specific questions (591/791)
- Major contributions to CS field (CSC 791 students only)
- Algorithmic Innovations (CSC 791 students only)
- Future research extensions (CSC 791 students only)

B.3: Mini-project with BI use cases (591/791)
- Time estimate: 5 hrs/wk
- Using Python and/or R
- End-to-end solution
- GitHub portfolio with all 10 use cases
- Team-based but individual portfolio submission

B.4: Algorithmic mini-homeworks
- Time estimate: 2 hrs/wk
- **Problem sets might be different for 791 and 591 students**
- Random problems are graded
- Solutions are provided

C. Tutorials

- Apache Spark: **Due Jan 7**
- R programming: **Due Jan 7**
- Python programming: **Due Jan 7**
- Word2Vec & Doc2Vec: **TBD**

D. Capstone Project

- Time estimate: 20 hrs/student

- ***BI Use Case: End-2-end*** (591 students)
  GitHub portfolio
  Demo
  Presentation:

- ***Survey/Method Paper*** (791 students)
  Proposal:
  Future Research Directions:
  Paper:

D. Final Exam

- Time estimate: 20 hrs
- In lieu of Capstone project
- Comprehensive, hometake
- Three business days to complete
- **Due date: TBD based on NCSU calendar**

E. Participation on Piazza

COURSE LOGISTICS:

- ***Communication***:
  - All questions/issues/suggestions must be posted on piazza only
  - No emails will be answered; private posts are ok
  - Except for personal issues or assignment clarifications, TAs will wait 24 hrs before helping with questions to give students sufficient time to interact and earn participation points.

- ***Assignments***:
  - Individual unless stated otherwise.
  - Each assignment due date is on the Calendar available on the Moodle course page. Typically, the due date is on Sunday. But two business days delay without penalty is allowed. If you did not click the "submit" button, but uploaded your files and saved them in what is called a "draft" in Moodle; then we will still be able to view and grade the drafts.
  - Electronic submission via Moodle is required.
  - In some cases, only part of a homework (randomly chosen) will be graded. The solutions for the homeworks will be posted after the due date.
  - For all the assignments you are required to submit your own work (unless explicitly stated as a team-based assignment). However, you may discuss your approach to any assignment (except for the exams) with your classmates. In this case, you must provide

the names of the students you discussed your approaches with. If you fail to mention your collaborators and similarity is detected there will be an automatic 20 percent deduction.

- ***Late Assignments****:*
  - Please *do not ask us* permission to submit anything late (or give excuses for late submissions). We have implemented the following policies to avoid these issues altogether.
  - Two mini-homeworks and two mini-projects may be submitted late without penalty: one homework and one project from the set of assignments before Spring Break and one homework and one project from the set of assignments before Exam week.
  - If the students choose to submit the assignments late then the teaching staff will not be responsible for handling student questions in a timely manner; their focus will be primarily on the currently due assignments.
  - No late submissions for the Lecture Assignments, Capstone Project, and the Final Exam unless within the purview of the university policies.

- ***Attendance***: not required
- ***Video lectures***: available to all students
- ***Team Assignments***:
  - Teams: up to 3 students
  - Teams are self-assembled
  - Teams can vary from assignment-to-assignment
- ***Office hours***: <u>By Appointment Only; first point of contact is the TA(s); if the issue is not resolved then contact the instructor</u> (unless it is a personal issue)
- ***Regrading***: No regrading requests will be handled until the week of April 18. This way the teaching staff will spend time on assisting the students and not discussing possible mistakes in grading.