

**COURSE NUMBER: CSC 591**

**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, and others. The space of algorithms and methodologies will include but will not be limited to deep learning, information fusion from dynamic heterogeneous and attributed graphs, and probabilistic reasoning under uncertainty, among others. Tutorials and projects that teach students how to handle Big Data issues will utilize Apache Spark. (3 credit hours)

**PREREQUISITES:**

- CSC 522 -- Automated Learning and Data Analysis (or its equivalent)
- Placement Tests (available on Moodle) on Python, Statistics, Linear Algebra, Machine Learning
- Otherwise, consent of the instructor is required.

**INSTRUCTOR**

**Dr. Nagiza Samatova** <samatova@csc.ncsu.edu>

My office is 2272 EB 2, on Centennial Campus.

I have an open-door policy if I am on campus. Otherwise, schedule via Piazza.

Cell: 865-566-5586

**TEACHING ASSISTANTS**

The best communication medium is Piazza, but meetings can be arranged if necessary.

Use Piazza to arrange meetings. Personal emails will be ignored.

- **Ronil Pancholia:** For questions related to Topics 1, 2, 5, 7
- **Yifan Zhao:** For questions related to Topics 3, 4, 6, 8
- **Jianxun Wang:** Grader

**COURSE OBJECTIVES & LEARNING OUTCOMES:** 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, and (h) 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)-(g) 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.

## **COMMUNICATION**

Please ask for help using the Piazza message board if you have trouble understanding the concepts or have questions about specific homework problems.

All communication to us should take place on Piazza (private or anonymous messages are possible). We will not be monitoring our email for any course related issues, we will only monitor the Piazza. Except for personal issues or assignment clarifications, TAs will wait 24 hours before helping with questions to give students sufficient time to interact.

## **EXAM**

There will be a **cumulative final exam**. The exam will be **home-take** and is open to course materials on Moodle but nothing else. To be more specific, anything besides lecture slides/videos, assignment solutions, notes, and piazza posts is not allowed. It is likely will be timed: the details will be posted later.

## **ASSIGNMENTS**

### **Submission:**

All assignments must be submitted thorough Moodle.

### **Assignment Schedule:**

All the assignments and due dates are summarized in the **Assignment Schedule** file on Moodle and Google Drive. The file might be updated: Piazza posts will be posted in case of the updates.

### **Late Submission Policy:**

Every assignment will have **an extra of TWO (2) business days** for late submission without penalty. Please, do not ask us for any extra permission to submit anything later (or give excuses for late submissions). Business days are weekdays excluding national holidays and official school breaks.

However, **THREE (3)** assignments may be submitted later than 2 business days without penalty under the following rules:

- All late submissions for assignments must be **submitted before the start date of the final exam**.

## Teamwork:

All assignments must be done **individually** unless stated otherwise. However, you may discuss your approach to any assignment (except for the exam) 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, then there will be an automatic fail for the assignment and a possible plagiarism case will be open with the university.

## GRADING

To reduce stress over grades and promote collaboration, we will adopt a grading system inspired by a number of US medical schools. The assignments and exams will be graded as a **binary pass or fail** (P/F). Typically, if you complete roughly 80% of the assignment or more, you will be given a pass, namely, you met the learning objectives for this topic; otherwise, a fail. The actual cut-off (e.g., 80%, 75%, 70%) might be adjusted by the course TAs based on the class performance, namely, the passing grade might be curved with 80% used in most cases. Ultimately, at the end of the course we will have to submit a letter grade. This letter grade will be determined as follows:

Grade	Ways to Achieve the Grade
A+	Fail no more than two (2) assignment and pass both the final exam and capstone.
A	Fail no more than three (3) assignments and pass both the final exam and capstone.
A-	Fail no more than four (4) assignments and pass both the final exam and capstone.
B+	Fail no more than five (5) assignments and pass both the final exam and capstone.
B	Fail no more than six (6) assignments and pass both the final exam and capstone.
B-	Fail no more than seven (7) assignments and pass both the final exam and capstone.
C+	Fail no more than eight (8) assignments and pass both the final exam and capstone.
C	Fail no more than nine (9) assignments and pass both the final exam and capstone.
C-	Fail no more than ten (10) assignments and pass both the final exam and capstone.
F	Fail more than eleven (11) assignments or fail both the final exam and capstone.

Feedbacks on Pass assignments will NOT be provided, and it will be your responsibility to look into the provided solution to figure out if/where you might have made any mistakes.

However, if you receive a Fail, we will provide feedback as to why you did not meet the learning objectives for that assignment.

**IMPORTANT: Failed Final Exam or Capstone (but not both):** If you FAIL either Capstone or Final exam but NOT both then you can still PASS the course. But in this case, your earned letter grade will be down-graded by the full letter: A → B+ or B → C+.

**IMPORTANT:** We will not entertain grade disputes until the end of the semester during the last week of classes. If you feel an assignment was graded incorrectly, please, make a note for your record and bring up the issue to the TA/grader at that time.

### **BONUS ASSIGNMENTS**

The students will be able to use bonus assignments to replace the assignments (not the exams) that they have a Fail (zero) score. One bonus assignment can substitute one Failed assignment. No more than TWO Bonus assignments will be allowed. BUT the use of the bonus assignments should NOT upgrade the student's grade by more than a letter grade. For example, if the student's grade without the bonus assignment is B then the student may choose to improve it to A- using bonus assignments but the student cannot improve to A or A+ using bonus assignments alone, as specified in the Table above.

### **STUDENTS WITH DISABILITIES**

Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the [Disability Services Office](#) at Suite 1900, Student Health Center, Campus Box 7509, 515-7653.

### **ACADEMIC INTEGRITY**

*All the policies outlined in the [Code of Student Conduct](#) (Section 8) apply. Please (scroll down) read this carefully and ask me if you have any questions.*

The essence of University policy is: ***Don't do anything that will give you an unfair advantage over other students.*** The Code of Student Conduct has a list of behaviors that are unacceptable.