



Rensselaer

SYLLABUS

Data Analytics BCBP 4960/ ITWS 4600/ITWS 6600/ MATP 4450/ CSCI 4960/
MGMT 4962/ MGMT 6962

3 Credit Hours
Fall 2022

Room Location(s):

Section1: In-person class location: VCC South - Time: 10:00am -11:50am ET

Section2: In-person class location: CII 3206 - Time: 2:00pm – 3:50pm ET

Course Website: <https://tw.rpi.edu/classes/data-analytics-fall-2022>

Prerequisites or Other Requirements:

Knowledge in Data Structures (CSCI 1200), Database Systems (CSCI 4380) preferred.

For cross-listed courses: Students may not receive credit for both the 4000 level and 6000 level versions of this course.

INSTRUCTOR

Instructor Name: Thilanka Munasinghe

Office Location: Lally 315

Tel. No: +1-857-998-8767

Email Address: munast@rpi.edu

Instructor Office Hours: Tuesdays/Fridays 12:30pm ET – 1:30pm ET or by
appointment via email

TEACHING ASSISTANT

Teaching Assistant: Shivam Sonawane

TA Office Location: Lally 215

TA Office Hours: 2 pm – 4 pm ET on Wednesdays OR by Appointment

TA(s) Email Address: sonaws@rpi.edu

COURSE DESCRIPTION:

Introduce students to relevant methods to recognize and apply quantitative algorithms, techniques and interpretation To develop students' strategic thinking skills, combined with a solid technical foundation in data and model-driven decision-making. Develop ability to apply critical and analytical methods to formulate and solve science, engineering, medical, and business problems Students will examine real-world examples using modern cyberinfrastructure to place statistical and data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science.

By the end of the course, students can effectively communicate analytic findings to non-specialists Data and Information analytics extends analysis (descriptive and predictive models to obtain knowledge from data) by using insight from analyses to recommend action or to guide and communicate decision-making. Thus, analytics is not so much concerned with individual analyses or analysis steps, but with an entire methodology.

The world at-large is confronted with increasingly larger and complex sets of structured/unstructured information; from sensors, instruments, and generated by computer simulations; data is "hidden" in websites, application servers, social networks and on mobile devices. As a nation, assimilating information across disparate domains (e.g., intelligence, economics, science) has the potential to provide improved capabilities for decision makers. In commerce and industry, analytics-driven enterprises are becoming mainstream. Yet, there is a shortfall in the key education skills needed to meet the growing needs.

Traditional enterprises are moving toward analytics-driven approaches for core business functions. In the government and corporations, cybersecurity problems are prevalent. The investment in advanced analytics capabilities could potentially be more broadly leveraged today and greater than any prior government investments in computing. Emphasis is now placed on disruptive data and information sources on the Web and Internet: using Web Science and informatics to explore social networks, platform competition, the "long tail" and economic or resource impacts of the search for new findings. Key topics include: advanced statistical computing theory, multivariate analysis, and application of computer science courses such as data mining and machine learning and change detection by uncovering unexpected patterns in data.

Introduce students to relevant methods to recognize and apply quantitative algorithms, techniques and interpretation To develop students' strategic thinking skills, combined with a solid technical foundation in data and model-driven decision-making. Develop ability to apply critical and analytical methods to formulate and solve science, engineering, medical, and business problems Students will examine real-world examples using modern cyberinfrastructure to place statistical and data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science. By the end of the course, students can effectively communicate analytic findings to non-specialists

REQUIRED TEXT(S):

Refer to Reading/ Assignment/ Reference list for each week on the course webpage:

<https://tw.rpi.edu/classes/data-analytics-fall-2022>

Following Reference material (available through RPI library - RCS login required):

- [Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die \(online\) \(RECOMMENDED\)](#)
- [Big data analytics: turning big data into big money](#)
- [Big Data Analytics: Turning Big Data into Big Money \(online\)](#)
- [Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph \(online\)](#)
- [Big Data Analytics with R and Hadoop \(online\)](#)
- [R for Everyone: Advanced Analytics and Graphics \(online\)](#)
- Introduction to Statistical Learning with R - 7th Edition

COURSE GOALS/OBJECTIVES

- Introduce students to relevant methods to recognize and apply quantitative algorithms, techniques, and interpretation
- To develop students' strategic thinking skills, combined with a solid technical foundation in data and model-driven decision-making.
- Develop the ability to apply critical and analytical methods to formulate and solve science, engineering, medical, and business problems
- Students will examine real-world examples using modern cyberinfrastructure to place statistical and data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science.
- By the end of the course, students can effectively communicate analytic findings to non-specialists

STUDENT LEARNING OUTCOMES

1. Students to demonstrate knowledge of relevant analytic methods, and to recognize and apply quantitative algorithms, techniques and interpret results
2. Students to demonstrate strategic thinking skills, combined with a solid technical foundation in data and model-driven decision-making.
3. Students to develop the ability to apply critical and analytical methods to formulate and solve science, engineering, medical, and business problems
4. Students will examine real-world examples to place data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science.
5. Students must effectively communicate analytic findings to non-specialists.
6. [6000 level]: Students must develop and demonstrate an ability to apply appropriate analytic techniques under conditions of uncertainty, be able to build optimization models that incorporate random parameters: static stochastic optimization, two-stage optimization with recourse, chance-constrained optimization, and sequential decision making

COURSE ASSESSMENT MEASURES

Assessment	Due Date	Learning Outcome #s
Homework	Each homework assignment due date listed in the assignment.	1, 2, 3, 4, 5, 6
Project	Last day of the classes.	1, 2, 3, 4, 5, 6
Presentation	Last day of the classes.	5

4000 Level	Points	6000 Level	Points
Assignment 1: Review an existing case study for Data Analytics	5%	Assignment 1: Review an existing case study for Data Analytics	5%
Assignment 2: Data Analytics Lab2 submission	10%	Assignment 2: Data Analytics Lab2 submission	5%
Assignment 3: Distribution analysis and comparison of distributions, visual analysis, statistical model fitting and testing	15%	Assignment 3: Distribution analysis and comparison of distributions, visual analysis, statistical model fitting and testing	12%
		Assignment 3: 6000 Level: Filter the distributions you explored in Question 1 in the Assignment 3 using one or more of the other variables that are in the dataset	3%
Assignment 4: Pattern, trend, relations: model development and evaluation on provided datasets	15%	Assignment 4: Pattern, trend, relations: model development and evaluation on provided datasets	12%

		Assignment 4: 6000 Level: Draw conclusions from this study – about the model type and suitability/ deficiencies. Describe what worked and why/ why not	3%
Assignment 5: Term project proposal and presentation.	5%	Assignment 5: Term project proposal and presentation. <u>6000 level students need to use minimum of two different datasets for the project</u> (multiple datasets required for 6000-level students)	10%
Assignment 6: Term Project - exploring the relationships and distributions within your data to lead to optimized predictive models. Start with a hypothesis, claim, or questions. Think of one or more ways to construct model(s) ¹ , find or collect the necessary data, and do both preliminary analysis, detailed modeling, validation, summary (interpretation) and (if any) resulting decisions.	25%	Assignment 6: Term Project - exploring the relationships and distributions within your data to lead to optimized predictive models. Start with a hypothesis, claim, or questions. Think of one or more ways to construct model(s) ¹ , find or collect the necessary data, and do both preliminary analysis, detailed modeling, validation, summary (interpretation) and (if any) resulting decisions.	20%
Electronic version of the Poster and presentation	5%	Assignment 6: <u>6000-level students must develop at least four different types of models</u> , not just change the number of variables for a given model type. Describe how you determined which datasets you used in this project, the criteria, source, data and information-types in detail, associated documentation and any other supporting materials.	5%
		Electronic version of the poster and presentation	5%
Assignment 7 - Predictive and Prescriptive data analytics, Model Development, Validation, Optimization and Tuning – Develop 2 or more models and	20%	Assignment 7 - Predictive and Prescriptive data analytics, Model Development, Validation, Optimization and Tuning - validate predictive models	14%

validate predictive models (minimum 2 models)		Assignment 7 – 6000 level students need to develop 3 or more models (minimum 3)	6%
--	--	---	----

Total	100%	Total	100%
--------------	-------------	--------------	-------------

Students taking a 6000 level course, regardless of student status (i.e., Undergraduate or Graduate) must satisfy the learning outcomes at the 6000 level if they expect to receive graduate credit for the course.

GRADING CRITERIA

7 Assignments adding up to the total score of 100%.

Note: courses for credit at the 6000 level cannot receive “D/D+/D-” grades

Grade – letter scale:

93% + is an A;
 90%-92% is an A-
 87%-89% is a B+
 83%-86% is a B
 80%-82% is a B-
 77%-79% is a C+
 73%-76% is a C
 70%-72% is a C-
 65%-69% is a D+
 60%-64% is a D
 0%-59% is a F.

ATTENDANCE POLICY

Either in-person OR Virtual (online) participation is required during both lectures and labs sessions, attendance will be taken each day/session. Student is responsible to sign the attendance sheet either on paper or on an electronic form before end of the class lecture or lab session.

ACADEMIC INTEGRITY

Student-teacher relationships are built on trust. For example, students must trust that teachers have made appropriate decisions about the structure and content of the courses

they teach, and teachers must trust that the assignments that students turn in are their own. Acts that violate this trust undermine the educational process.

The Rensselaer Handbook of Student Rights and Responsibilities and the Graduate Student Supplement (For 6000 level and above courses) define various forms of Academic Dishonesty and you should make yourself familiar with these. In this class, all assignments that are turned in for a grade must represent the student's own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration. Submitting the same work for multiple assignments/courses/labs without prior consent from the instructor(s) not allowed. Submission of any assignment that is in violation of this policy will result in (1) an academic (grade) penalty and (2) reporting to Associate Dean of Academic Affairs and either the Dean of Students (for Undergraduates) or the Dean of Graduate Education (for Graduate students).

In this course, the academic penalty for a first offense is zero grade for the relevant portion of the grade. A second offense will result in failure of the course.

If you have any questions concerning this policy before submitting an assignment, please ask for clarification.

ACADEMIC ACCOMMODATIONS

Rensselaer Polytechnic Institute strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on a disability, please let me know immediately so that we can discuss your options.

To establish reasonable accommodations, please register with The Office of Disability Services for Students (<mailto:dss@rpi.edu>; 518-276-8197; 4226 Academy Hall). After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion."

COURSE CALENDAR

Detailed course calendar with each week's reading material is available on the course webpage at: <https://tw.rpi.edu/classes/data-analytics-fall-2022>

General outline of the course calendar:

Group 1 - Intro/ Setup

- Week 1 (Aug. 30/ Sept. 02): Introduction to Course, Case Studies, and Preview of Course Material/Refresher on basic statistics.
- Assignment 1
- Week 2 (Sept. 06: No classes—Follows Monday's schedule/ Sept. 09): Introduction/ refresher on basic statistics continue / Starting with Data and Information Resources, Role of Hypothesis, Synthesis and Model Choices, R/ RStudio introduction and Intro to Labs.

- Week 3 (Sept. 13: / Sept. 16): Introduction to Analytic Methods, Types of Data Mining for Analytics, Data filtering, hypothesis exploration, visual analysis, model consideration and assessment (lab)
- (Lab) Assignment 2

Group 2 - Patterns, relations, descriptive analytics

- Week 4 (Sept. 20 / Sept. 23): Weighted kNN, Clustering, early decision trees, Exercises for linear regression, kNN and K-means (lab), trees, plotting
- Assignment 3
- Week 5 (Sept. 27/ Sept. 30): Interpreting: Regression, Weighted kNN, Clustering, and Bayesian Inference, Exercises for clustering, plotting, bayesian inference (lab)
- Assignment 4
- Assignment 5
- Week 6 (Oct. 04/ Oct. 07): Assignment 5 presentations (Tuesday and Friday)
- Assignment 6
- Week 7 (Oct. 11 / Oct. 14): Lab weighted kNN, decision trees, random forest

Group 3 - Predictive Analytics

- Week 8 (Oct. 18/ Oct. 21): Cross-Validation Trees, Dimension Reduction and Multi-Dimensional Scaling
- Week 10 (Oct. 25/ Oct. 28): Spring Break: No Classes
- Week 10 (Nov. 01/ Nov. 04): Support Vector Machines, Lab for Trees, DR, MDS, SVM
- Week 11 (Nov. 08/ Nov. 11): Factor Analysis, Factor Analysis lab
- Week 12 (Nov. 15/ Nov. 18): Interpreting PCA, MDS, DR, and FA , Boosting, Bootstrapping, Bagging, Boosting, Bootstrapping, Bagging (lab)
- Assignment 7

Group 4 - Evaluating and validating, prescriptive analytics

- Week 13 (Nov. 22/ Nov. 25: No Classes – Thanksgiving Break): Cross-validation, Revisiting Regression - local methods, Lab - Cross-validation, Regression - local methods and continue on project and assignment work
- Week 14 (Nov. 29/ Dec. 02): Local Regression ctd, Mixed Models, Optimizing, Iterating, (Fischer Linear Discriminant)
- Week 15 (Dec. 06/ Dec. 09): Prior Lab Review, Hierarchical Linear and Mixed Models, Latent Class Mixed Models, Lab, Assignment 7 due
- Week 16: (TBD): Final Project and Poster Due

COVID-19 code of conduct

All students must comply with all health and safety protocols specified by the Institute under the [Return-to-Campus plan](#) available at the [Rensselaer COVID-19 website](#). Appropriate action will be taken against those who do not comply fully with these protocols”.

Students taking Lally School of Management classes are required to abide by the **COVID-19 code of conduct** below.

This code will apply to any class that meets fully or partially in an on-campus physical classroom for in-person instruction.

Violations: Refusal to comply with the COVID-19 code of conduct will be treated just as any classroom disruption, which will receive request for immediate compliance, failing which the student will be asked to leave the classroom. Any further noncompliance will result in the dismissal of the entire class. All Covid-19 related violations will be reported by the instructor to the Compliance Officer and the Dean of Students. A student found to be in violation of the code, or required repeated reminders for compliance, will be asked to participate in all classes remotely. This is to protect their health and safety as well as the health and safety of their classmates, instructor, and the university community.

Traffic Flow and Social Distancing: Students and faculty will respect the need for social distancing. They are required to follow the traffic flow arrows posted in all rooms and buildings, including bathrooms and common areas.

In-Class Seating: Students should sit in the appropriate designated seating in the classroom. Students are not allowed to move furniture or sit in seats not designated by the Institute.

Cleaning of Spaces: Students are encouraged to clean the surfaces of the chairs/tables/desks they occupy before they sit down and as they prepare to leave. Cleaning and sanitizing solutions will be provided the classroom.

Students who are ill, under quarantine for COVID-19, or suspect they are ill should not come to class. All faculty will make every reasonable effort to accommodate the student’s absence and will communicate that accommodation directly to the student. Students who need to report an illness should contact the [Student Health Center](#) via [email](#) or call 518-276-6287. For student seen off campus, a student may request an excused absence via www.bit.ly/rpiabsence with an uploaded doctor’s note that excuses them.

Rensselaer Polytechnic Institute
On- and Off-Campus Support Resources: Fall 2022
Remember, seeking help is a strength, not a weakness

ON-CAMPUS HEALTH & WELLNESS SUPPORT

Student Health Center*

Mon-Fri, 8:30 am – 5:00 pm EST

The mission of the Student Health Center (SHC) is to keep students healthy so that they may achieve their academic, personal, and athletic goals. The SHC provides confidential, accessible, cost-effective, current evidence-based treatment for acute and chronic physical health problems. At this time, appointments are being offered virtually (phone and video). Call 518-276-6287 to schedule an appointment, or schedule one through your Student Health portal. **There are no walk-in appointments available at the Student Health Center during this time.**

**information subject to change based on pandemic conditions*

Counseling Center*

Mon-Fri, 8:30 am – 5:00 pm EST (some weekday evening hours available by appointment)

The goal of the Counseling Center is to help students maximize their sense of well-being as well as their academic, personal, and social growth. Appointments are free and confidential, and in-person at the Counseling Center, 4th Floor of Academy Hall. Some WebEx and phone appointments will be offered as needed. Please contact the Counseling Center for this service. Appointments can be made by calling 518-276-6479 or email: counseling@rpi.edu Counseling Center staff are available in case of a crisis on evenings and weekends (call Public Safety at 518-276-6611 and ask to speak with the on-call counselor).

**information subject to change based on pandemic conditions*

Office of Health Promotion

Health promotion initiatives at Rensselaer are evidence-based and comprehensive efforts to improve health knowledge, behaviors, and skills of Rensselaer students. Health Educators provide campus programming on a variety of health topics, and are available for one-on-one consultations around issues including, but not limited to: sleep hygiene, mental health, sexual health, alcohol and other drugs, LGBTQIA+ topics, sexual assault prevention, and more. All appointments are free and confidential and take place via WebEx. To schedule an appointment, email: healthed@rpi.edu Follow us on social media for daily health tips and event information! Instagram: [rpi.studenthealth](https://www.instagram.com/rpi.studenthealth) | Twitter: [@RPIhealth](https://twitter.com/RPIhealth) | Facebook: RPI Student Health Services | Discord: <https://discord.gg/8DZJJ38zWj>

Disability Services for Students

The Office of Disability Services for Students (DSS) assists Rensselaer students with disabilities in gaining equal access to academic programs, extracurricular activities, and physical facilities on campus. DSS is the designated office at Rensselaer that obtains and files disability-related documentation, assesses for eligibility of services, and determines reasonable accommodations in consultation with students. Call 518-276-8197 or email dss@rpi.edu for more information.