FE-582: Foundations of Financial Data Science
Spring 2021

Instructor:	Dragos Bozdog Office: Babbio 429A Email: dbozdog@stevens.edu Phone: (201) 216-3527
Teaching Assistant:	Zhiyang Deng Email: zdeng10@stevens.edu
Canvas:	https://sit.instructure.com/courses/42379
Schedule:	FE-582-A and FE-582-WS Tuesday (6:30pm-8:10pm) Hanlon Lab 1 (Babbio 4th floor) and Online (Zoom)
Office Hours:	TBD
Description:	This course will provide an overview of issues and trends in data quality, data storage, data scrubbing, and data flows. Topics will include data abstractions and integration, data management issues with collection, warehousing, preprocessing and querying, similarity and distances, clustering methods, classification methods, text mining, and time series. Case studies will be presented in support of the theoretical concepts. Furthermore, the Hadoop based programming framework for big data issues will be introduced along with any governance and policy issues. These concepts will be applied to areas such as real estate, social media and social networks, and capital markets financial data. A one credit Hanlon lab course, FE-513: Practical Aspects of Database Design is co-requisite to this course in order to facilitate learning of the practical side of data management.
Objective:	This course provides the theoretical and practical foundation for Financial Analytics. The main objective is to enable the students to create data-driven decision in the financial industry based on data science tools and analytics methods. The goal is to be enable effective business operations, enhanced customer services and product offerings, improved risk analysis, and risk management.
Co-Requisite	FE 513 – Practical Aspects of Database Design

Textbooks:

No single textbook covers all the topics. Several references will be used and supplementary notes will be provided whenever appropriate.

General References:

1. Charu C. Aggarwal, *Data Classification: Algorithms and Applications.* CRC Press, *2015.* (ISBN: 978-1-4665-8674-1)
2. Charu C. Aggarwal, *Data Mining.* Springer, 2015. (ISBN: 978-3-319-14141-8)
3. Deborah Nolan and Duncan T. Lang, *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*, CRC Press, 2015. (ISBN: 978-1-4822-3481-7)
4. Norman Matloff, *The Art of R Programming,* No Starch Press, 2011. (ISBN: 978-1-59327-384-2)
5. Cathy O'Neil and Rachel Schutt, *Data Science,* O'Reilly, 2014. (ISBN: 978-1-449-35865-5)

Outcomes:

After taking this course, the students will be able to:

1. Have a working knowledge of the issues of data quality, data storage, data scrubbing, data flows, and their potential solutions.
2. Understand and design various schemas needed for the representation of financial data.
3. Tackle problems dealing with data management issues such as collection, warehousing, preprocessing and querying.
4. Will get a primer on database management as well as advantages and disadvantages from the attached lab course FE 513.
5. Have a working understanding of all the databases available for them through the Hanlon lab.
6. Apply the newly acquired data management and database skills to financial data from the capital markets, social media, and the financial services sector.

Grading:

Assignments 60%
Project 40%

Graduate Student Code of Academic Integrity:

All Stevens, graduate students promise to be fully truthful and avoid dishonesty, fraud, misrepresentation, and deceit of any type in relation to their academic work. A student's submission of work for academic credit indicates that the work is the student's own. All outside assistance must be acknowledged. Any student who violates this code or who knowingly assists another student in violating this code shall be subject to discipline.

All graduate students are bound to the Graduate Student Code of Academic Integrity by enrollment in graduate coursework at Stevens. It is the responsibility of each graduate student to understand and adhere to the Graduate Student Code of Academic Integrity. More information including types of violations, the process for handling perceived violations, and types of sanctions can be found at www.stevens.edu/provost/graduate-academics .

Learning Accommodations:

Stevens Institute of Technology is dedicated to providing appropriate accommodations to students with documented disabilities. The Office of Disability Services (ODS) works with undergraduate and graduate students with learning disabilities, attention deficit-hyperactivity disorders, physical disabilities, sensory impairments, psychiatric disorders, and other such disabilities in order to help students achieve their academic and personal potential. They facilitate equal access to the educational programs and opportunities offered at Stevens and coordinate reasonable accommodations for eligible students. These services are designed to encourage independence and self-advocacy with support from the ODS staff. The ODS staff will facilitate the provision of accommodations on a case-by-case basis.

For more information about Disability Services and the process to receive accommodations, visit https://www.stevens.edu/office-disability-services . If you have any questions please contact: Phillip Gehman, the Director of Disability Services Coordinator at Stevens Institute of Technology at pgehman@stevens.edu or by phone 201-216-3748.

Disability Services Confidentiality Policy
Student Disability Files are kept separate from academic files and are stored in a secure location within the Office of Disability Services. The Family Educational Rights Privacy Act (FERPA, 20 U.S.C. 1232g; 34CFR, Part 99) regulates disclosure of disability documentation and records maintained by Stevens Disability Services. According to this act, prior written consent by the student is required before our Disability Services office may release disability documentation or records to anyone. An exception is made in unusual circumstances, such as the case of health and safety emergencies.

Inclusivity:

Name and Pronoun Usage
As this course includes group work and class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect. This includes the ability for all students to have their chosen gender pronoun(s) and chosen name affirmed. If the class roster does not align with your name and/or pronouns, please inform the instructor of the necessary changes.

Inclusion Statement
Stevens Institute of Technology believes that diversity and inclusiveness are essential to excellence in academic discourse and innovation. In this class, the perspective of people of all races, ethnicities, gender expressions and gender identities, religions, sexual orientations, disabilities, socioeconomic backgrounds, and nationalities will be respected and viewed as a resource and benefit throughout the semester. Suggestions to further diversify class materials and assignments are encouraged. If any course meetings conflict with your religious events, please do not hesitate to reach out to your instructor to make alternative arrangements.

You are expected to treat your instructor and all other participants in the course with courtesy and respect. Disrespectful conduct and harassing statements will not be tolerated and may result in disciplinary actions.

Mental Health Resources:

Part of being successful in the classroom involves a focus on your whole self, including your mental health. While you are at Stevens, there are many resources to promote and support mental health. The Office of Counseling and Psychological Services (CAPS) offers free and confidential services to all enrolled students who are struggling to cope with personal issues (e.g., difficulty adjusting to college or trouble managing stress) or psychological difficulties (e.g., anxiety and depression) and who can visit the office in person. CAPS is open from 9:00 am – 5:00 pm Mondays, Wednesdays, Thursdays and Fridays and from 9:00 am – 7:00 pm on Tuesdays during the Fall and Spring semesters; appointments are highly encouraged. For those students who cannot visit the Stevens campus for an in-person appointment, you can contact a local mental health care provider for an in-person appointment, or if you are enrolled in the Stevens Student Health Insurance, you may call Care Connect for 24/7 mental health support at 1-888-857-5462.

For further information please visit the CAPS webpage on Seeking Help Off-Campus.

Emergency Information:

In the event of an urgent or emergent concern about the safety of yourself or someone else in the Stevens community, please immediately call the Stevens Campus Police at 201-216-5105 or on their emergency line at 201-216-3911. These phone lines are staffed 24/7, year round. For students who do not reside near the campus and require emergency support, please contact your local emergency response providers at 911 or via your local police precinct. Other 24/7 national resources for students dealing with mental health crises include the National Suicide Prevention Lifeline (1-800-273-8255) and the Crisis Text Line (text "Home" to 741-741). If you are concerned about the wellbeing of another Stevens student, and the matter is not urgent or time sensitive, please email the CARE Team at care@stevens.edu. A member of the CARE Team will respond to your concern as soon as possible.

	Topic
Week 1	Introduction to Financial Data Science. Data Science Process. Sample Data Processing. The Basic Data Types. The Major Building Blocks: A Bird's Eye View. Introduction to R. Case Study: Exploratory Data Analysis (NYC Real Estate)
Week 2	Financial Data Quality Issues and Data Scrubbing. Data Preparation. Feature Extraction and Portability. Data Cleaning. Data Reduction and Transformation. Handling Missing Entries. Handling Incorrect and Inconsistent Entries. Scaling and Normalization. Data Reduction and Transformation. Sampling for Static Data and Data Streams. Dimensionality Reduction Intro.
Week 3	Web page retrieval, scrapping, regular expression extraction, basic statistical techniques to identify wrong data entries. Case Study: Data and Web Technologies. Linear Model. Piecewise linear model.
Week 4	Similarity and Distances. Impact of High Dimensionality. Generalized Minkovski Distance. Match-Based Similarity Computation. Impact of Data Distribution. ISOMAP. Impact of Local Data Distribution. Similarity on Categorical Data. Similarity on Mixed Quantitative and Categorical Data. Text Similarity Measures. Time Series Similarity Measures.
Week 5	Classification Methods.
Week 6	Tree-Based Methods.
Week 7	Clustering Methods.
Week 8	Financial Time Series Data. Using Decision Tree to Trade Stock. Building a Trading Strategy. Handling Time-Dependent Data in R. The Prediction Models.
Week 9	Mining Text Data. Document Preparation and Similarity Computation. Specialized Clustering Methods for Text. Probabilistic Algorithms. Co-Clustering. Topic Modeling.
Week 10	Case Study: Using Statistics to Identify Spam
Week 11	Blockchain. Cryptocurrencies.
Week 12	Outlier Detection.
Week 13	Hadoop. HDFS. MapReduce. Hive. Pig.
Week 14	Review and Catching up.
Finals Week	Final Project Presentations.