ACCT2700—Forensic Analytics
Spring 2024

Instructor: Prof. Daniel Taylor (dtayl@wharton.upenn.edu)
Office hours: by appointment

Teaching Assistants: Ravi Arya (raviarya@wharton.upenn.edu)
Dawson Goldsmith (dawsong@wharton.upenn.edu)
Andrew Tong (andytong@wharton.upenn.edu)
Office hours: by appointment

"ACCT270" must be included in the subject of your emails to get a response

Physical Classroom: JMHH380 Computer Lab
Virtual Classroom: see link and password on Canvas

Course Description

Recent trends in Big Data and predictive analytics are revolutionizing the way sophisticated investors, stock exchanges, broker/dealers, regulators, and auditors interact with financial data. This course teaches students the hands-on skills necessary to manipulate large-scale financial databases and build predictive models and algorithms.

The course will cover three applications of forensic analytics to publicly-traded companies: (i) forecasting future earnings, (ii) predicting accounting fraud, and (iii) detecting insider trading. The course will draw on cutting-edge academic research in each area; introduce students to the basic SQL coding skills necessary to manipulate Big Data and conduct meaningful analyses; and leverage the datasets and computing power of Wharton Research Data Services.

The course is organized as a hybrid of a traditional seminar course and a computer science course. The first few classes of each unit will cover the conceptual material and source material related to each topic. The later classes in each unit will teach the programming skills needed to manipulate the respective datasets and run detection algorithms.

Curriculum:

The course can be used to satisfy any of the following Wharton undergrad curriculum requirements: (a) Business Breadth requirement; (b) Technology Innovation and Analytics requirement; (c) Analytics concentration; or (d) Accounting concentration.

Prerequisites:

Basic knowledge of financial statements and linear regression is required. Previous coding experience with Structured Query Language (SQL) is not required. The course is designed for students with no previous coding experience, and will provide students with basic SQL coding skills. Formal prerequisites: ACCT101 and STAT102 (or STAT431)—**students will need knowledge of multiple regression, t-statistics, R-squared**.

Course Overview:

In ACCT101, you learned how to prepare financial statements. In STAT102, you learned how to run a regression. Perhaps you've taken a computer programming class. In addition, you have been learning how to write, present, and work collaboratively with your peers. This class is going to push you to combine these skills—to extract meaningful insights from Big Data and communicate your insights to others.

The course is divided into three "units." Each unit covers a different topic, and a different application of forensic analytics. For each unit, students will complete and a team project that entails computer programming, statistical analysis, and technical writing. Each unit includes multiple sessions designed to orient you to the datasets we will study, and provide a basic programming knowledge necessary to manipulate the dataset and conduct basic analyses. The datasets we will use in the class are very large——you will not be able to use Excel.

Three units doesn't seem like that much, but the final team project in each unit—developing statistical models that forecast earnings more accurately than Wall Street analysts (unit #1), improving existing fraud predictions models (unit #2), and designing algorithms to detect insider trading (unit #3)—represent a major accomplishment.

Because later material builds on prior classes, you will find that the entire process will go by smoothly if you take the weekly readings and assignments seriously. **The first unit (forecasting earnings) emphasizes programming, the latter two units (accounting fraud and insider trading) emphasize concepts.**

This class is treated as a bridge to the real-world of business:

(1) The material will NOT be spoon fed to you. It is your responsibility to identify gaps in your knowledge and take actions to remedy those gaps. Learning is about knowing what you know, knowing what you don't know, and knowing what questions to ask.

(2) The class will entail significant hands-on learning outside of class. You will need to invest time to watch online videos and material on your own. The amount you get out of this class will be directly proportional to the amount of time you spend investing in the class.

(3) I expect you to be professional, come prepared for class, be a good team member, and participate in class.

Assignments

Your course grade will be determined as follows:

Individual Component	11 Individual Assignments (6% each)	66%
Group Component	3 Team Projects (8% each)	24%
	1 Team Presentation	10%

Individual Assignments: These short assignments will take the form of a short homework of one page or less. In some cases, they will test your knowledge of the concepts presented in the pre-recorded lectures. In other cases, they will test your programming ability and your understanding of estimated models.

Team Projects: These projects will test: (1) your conceptual understanding of the material; (2) your ability to manipulate and analyze big data; (3) your ability to convey your thoughts formally in presentations and technical writing; and (4) your ability to work together in a team with diverse personalities—effectively all of the talents needed to have a successful career. There will be one written team project per unit, and a final team presentation at the end of class where **teams will have a maximum of 10 minutes to present their final project to the class.**

You should budget your time and not wait until the last minute to get started. These projects are each significant accomplishments and will require significant investment from every team member. For example, it is not easy to forecast earnings more accurately than Wall Street analysts (e.g., those at Goldman Sachs), but it can be done and should make for a good story to tell recruiters!

For presentations, I expect you to dress as if you were presenting your analysis in a formal business environment (i.e., wearing formal business attire). Attendance is required for all students in all presentations. Presentations are scheduled for the final two days of class. Unexcused absence for any of the presentations will result in a 10 point reduction to your final grade.

Participation: Notice that class participation is not an input into the above grading scheme. Class participation is purely extra credit, up to an extra 10% of your grade (extra 10 pts). Your participation will be entirely at the discretion of the professor and the TAs. You will be evaluated on the extent to which you ask questions of the instructor, guest speakers, and participate in class discussion. **There is no extra credit for attendance. Other than presentations, attendance is not required. You are an adult and can manage your own time and priorities.**

The Importance of a Good Team

On the first day of class I will ask students to submit a brief biography and assess their strengths and weaknesses in terms of creative thinking, programming skill, statistical analysis, writing, and presentation. I will also ask students to provide estimated weekly availability for team work. For

example, if your schedule is constrained such that only Friday, Saturday, Sunday are free for team work, you should probably seek out students with similar availability. This information will then be distributed to all students so that you may efficiently match with other students to build the best team. Team selection is due by the third class. Maximum team size = 3

I recommend you seek out students whose skills *complement* your own. For example, if you are an excellent computer programmer, but not fond of public speaking or writing, you should think about pairing with students who excel at writing and presenting (but who might not be the best programmers). Ideally, each member brings a different strength to the team and each members learn from other's talents.

This course is a not an individual or team competition. You are not being graded relative to your peers. My goal is for you to learn, and push the frontiers of what you are capable of—both individually and working as a team. Effectively, to mimic what it is like to work as part of a small, dedicated team of analysts charged with a specific task (e.g., the real world!).

Grading disputes

All grading disputes must be appealed within one week following the return of the assignment. To have an individual assignment or team project re-graded, please submit a written description of your disagreement. For team projects, every member must sign the re-grade request. I reserve the right to review the entire assignment/project; thus your grade could go up *or* down.

Course Materials & Software

The datasets we will use in this class will usually range in size from 1GB to 30GB. To handle data of this size, we require both a programming language and software that is Taylor-made for large datasets.

SQL-programming language

The course will provide students with an introduction to Structured Query Language (SQL). SQL is ubiquitous in data science, and is at the heart of all relational databases including Amazon Redshift, Oracle, MySQL, PostgreSQL, and Microsoft Access. Beyond managing data, SQL can be invoked in many popular packages including SAS, R, and Python—making it a "must know" for anyone interested in analyzing large datasets. According to Forbes, SQL is the second-most in-demand skill in data analysis (the most in-demand skill was "critical thinking").

SAS-software

SQL is a programming language not a software interface. Thus, we need to use software to execute SQL commands. I will provide in-class instructions for SQL within SAS software. SAS is very widely-used in industry (specifically in banking and finance): 94% of the Fortune 100 use SAS. Importantly, all of the data vendors for this course provide their databases in native-SAS format.

Why? Due to the nature of the software, SAS is faster than competing software at loading and manipulating Big Data. This is especially important for this course, as we will be working with the data in class (where waiting 10 minutes for a 10GB file to load is not practical). SAS is available for you to use remotely with Amazon WorkSpaces.

Amazon WorkSpaces

All of the data and software used in the class is stored in the cloud. Wharton Computing Labs has setup a virtual desktop that will allow you to "remote connect" to access the files and software from anywhere with internet access. You can then use the software and access the files as if you were physically in the computer lab. Software and installation instructions are available on Canvas.

Canvas

All course materials are located on the course Canvas site. This site includes all readings, cases, PowerPoint slides, pre-recorded lectures, and assignments on each topic. There is no textbook for this course. Instead, I will provide a set of notes, readings, and pre-recorded lectures. With regard to programming, it is up to you to take notes and ask questions when we cover in class programming exercises. For those of you not familiar with databases or SQL, you may need to undertake some serious self-study. There are excellent tutorials available online.

Topic and Assignment Schedule

This course is a hybrid of a traditional seminar course and a computer science course. Programming classes will teach basic coding relevant to the unit, and will cover a variety of in-class programming tasks. These classes will be hands-on sessions: students will write code and work with data in real-time.

1/22 Class #1. Course Introduction

Discuss Course Structure and Organization
Submit student bios

1/24 Class #2. Concepts—Stock Prices and Earnings

Earnings
Ball and Brown (1968), "An Empirical Evaluation of Accounting Income Numbers" *Journal of Accounting Research*
Doyle, Lundholm, and Soliman (2006), "Extreme Future Stock Returns Following Earnings Surprises" *Journal of Accounting Research*.

1/29 Class #3. Concepts— Stock Prices and Earnings

Accruals
Sloan (1996), "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *The Accounting Review*.
Richardson, Sloan, Soliman, and Tuna (2005) Accrual reliability, earnings persistence, and stock prices. *Journal of Accounting and Economics*
Bradshaw, Richardson, Sloan (2001) "Do Analyst and Auditors Use Information in Accruals? *Journal of Accounting Research*.

Ratio Analysis
Soliman (2008) "The Use of DuPont analysis by Market Participants" *The Accounting Review*.

Select "Analyst Teams"
[in Canvas use "People" > "Analyst Teams"]

1/31 Class #4. Coding—Intro to SAS
Introduce SAS interface
Introduce concepts related to DATA step and PROC command
Individual Assignment Due—Forecasting Concepts

2/5 Class #5. Coding—Intro to SQL
Introduce SQL concepts related to CREATE TABLE and JOIN statements
Introduce lag operator
Replicate Sloan

2/7 Class #6. Coding—Forecasting Earnings
Introduce analyst forecast data
Test for predictability in analyst forecast errors
Individual Assignment Due—SAS Basics

2/12 Class #7. Coding—Backtesting a Quant Strategy
Introduce CRSP return data
Introduce portfolio construction and return calculation
Individual Assignment Due—Demonstrate ability to Estimate Naïve Model

2/14 Class #7. Lab—Forecasting Earnings

2/19 Class #8. Lab—Forecasting Earnings
Individual Assignment Due—Demonstrate ability to Compute Portfolio Returns

2/26 Class #10. Lab—Backtesting

2/28 Class #11. Lab—Backtesting
UNIT #1 PROJECT DUE

>>>>>>>>>>>>>>>> UNIT #2: DETECTING ACCOUNTING FRAUD <<<<<<<<<<<<<<<<

3/11 Class #12. Concepts—Accounting Fraud
Dechow, Hutton, and Sloan (1996) "Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC." *Contemporary Accounting Research*. On Canvas.
Dechow, Ge, Larson, and Sloan (2011), "Predicting Material Accounting Misstatements" *Contemporary Accounting Research*. On Canvas.
Lee, Ingram, and Howard (1999) "The Difference between Earnings and Cash Flow as an Indicator of Financial Reporting Fraud" *Contemporary Accounting Research* On Canvas.
Beneish, Lee, and Nichols (2011), "To Catch a Thief: Can Forensic Accounting Help Predict Stock Returns?" working paper. On Canvas.

3/13 Class #13. Coding—Fraud Prediction
Introduction to AAER data
Replicate existing fraud prediction algorithms
Individual Assignment Due—Fraud Concepts

3/18 Class #14. TBD
Individual Assignment Due—Case Studies on Fraud I

3/20 Class #15. TBD
Individual Assignment Due—Case Studies on Fraud II

3/25 Class #16. Views of a Professional Fraud Hunter (guest lecture, virtual, 12pm ET)
9AM Individual Assignment Due (Madoff and the SEC)

3/27 Class #17. Activist Short-Selling (guest lecture, virtual, 10:15am ET)
9AM Individual Assignment Due—Case Studies on Fraud III

4/1 Class #18. Prosecuting Mega Frauds (guest lecture, in person)
9AM Individual Assignment Due—SDNY, Nikola, and FTX

4/3 Class #19. Anatomy of an Internal Fraud Investigation (guest lecture, in person)
In-Class Simulation.

4/8 Class #20. Lab—Fraud Prediction
UNIT #2 PROJECTS DUE 4/8

>>>>>>>>>>>>>>>> UNIT #3: DETECTING INSIDER TRADING <<<<<<<<<<<<<<<<

4/10 Class #21. Concepts—Insider Trading
Classical/Misappropriation
SEC Case Against Capital One Analyst
SEC Case Against Harmon's VP Tax
Controversies
Shadow Trading Case; NFT case; Crypto case;
Political Intelligence (followup)

4/15 Class #22. Concepts—Corporate Insiders
Holding Foreign Insiders Accountable
10B5-1 Plans
SEC Case Against EDGAR Hackers

4/17 Class #23. Coding—Statistical Detection

4/22 Class #24. Lab—Insider Trading
Individual Assignment Due—Insider Trading Concepts

4/24 Class #25. Lab—Insider Trading

4/29 Class #26. Final Team Project Presentations ("Detecting Insider Trading")
Attendance required, or 10pt reduction

5/1 Class #27. Final Team Project Presentations ("Detecting Insider Trading")
Attendance required, or 10pt reduction
UNIT #3 WRITTEN PROJECTS DUE

Should you take this course? Is this course relevant? Thoughts on the material and syllabus from professionals:

"FinTech is revolutionizing asset management and investment decisions. The blend of accounting, finance, and computer science that this course offers is ideally suited for preparing students for a variety of careers on Wall Street, including quantitative investing. This course and associated skills would clearly distinguish your undergraduate students from much of the competition---especially if they were seeking a job at a quant fund."

- former Global Head of Equity Research, Blackrock

"The biggest difference between equity investing today and a few decades ago is the sheer volume and velocity of the financial information that must to be processed quickly. As a result, big data analytics and quantitative investment techniques rank high on the strategic agenda of most equity fund managers today. Of particular shortage are analysts who can combine data analytic skills with a deep understanding of the financial information that can be extracted from corporate reports. This is the type of investment professional our business schools should be training more of today."

- former Head of US Equity Research; Co-head of N. America Active Equity Strategies, Global Head of Active Equity Research, Blackrock; Co-founder and General Partner, Nipun Capital

"I've seen astrophysicists who can value complex derivatives without any knowledge of finance or accounting, computer scientists who can code up stock selection algorithms without any understanding of valuation, and accounting and finance majors who are clueless about data. The key to success is to be able to analyze markets through multiple different lenses. Know finance and accounting: this will give you an edge over math-types. But also know how to design and backtest forecasting models using Big Data: this will give you an edge over MBA-types. This class offers an outstanding blend of these skills, and is exactly the type of class students need to take if they want to get the best jobs with the best firms."

- former Vice President, Citadel Investment Group

"The Big Data revolution is here. To be successful in today's environment, students need to have business skills, need to be able to extract insights from data, and need to know how to communicate those insights to others. Most individuals have one of those three talents. Those individuals who can do two or more are rare and very highly valued. The class looks challenging, but if you can succeed, you'll be able to do all three. Students who like to code will benefit from the business applications; and students who like business will benefit from learning how to write code. The hands-on nature of the class looks truly amazing. I wish I could have taken it as an undergrad."

- former Exotic Derivatives & CDO Trader, JP Morgan

"The Wall Street Journal recently proclaimed that the Quants Rule Wall Street Now. My own experience confirms that this is no overstatement. But quantitative skills alone are not enough. They need to be integrated with good investment intuition and a detailed knowledge of the structure and limitations of the underlying data. This course should give you the right combination of skills to help you on your way to becoming a successful quant."

- former Managing Director, Blackrock

ABOUT PROFESSOR TAYLOR

Daniel Taylor is the Arthur Andersen Chaired Professor at The Wharton School, and is director of the Wharton Forensic Analytics Lab. He is an award-winning researcher and teacher with extensive expertise on corporate disclosures, insider trading, and fraud prediction. He has published extensively on these topics in leading academic journals; led seminars at dozens of top business schools across the globe; and won numerous academic and industry awards.

Prof. Taylor's research is relevant to a variety of practitioners and regulators seeking to understand, detect, and deter illict activity in capital markets. His research frequently appears in the business media, and has been cited in rules and regulations promulgated by the SEC. Most recently, his research on insider trading was the impetus behind the SEC's proposed rule changes to 10B5-1 trading plans; changes to Form 144 filings; the Holding Foreign Insiders Accountable Act; and multiple investigations by the SEC, DoJ, and FBI. He received his bachelor's degree from University of Delaware, his master's from Duke University, and his PhD from Stanford University.