ACCT270: Predictive Analytics using Financial Disclosures

Classroom: JMHH 380
Instructor: Prof. Daniel Taylor
Office: SHDH 1312
Office hours: by appointment
Email: dtayl@wharton.upenn.edu

Teaching Assistants:
Bradford Lynch (lynchb@wharton.upenn.edu) Section 01
Michael Pearson (mipear@wharton.upenn.edu) Section 02
Jason Liebmann (liejason@wharton.upenn.edu) Section 03

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

Course Description

Annual reports, financial statements, and other mandatory disclosures contain large amounts of financial data that provide the foundation for a variety of forward-looking decisions. Recent trends in Big Data and predictive analytics are revolutionizing the way businesses and investors extract meaningful insights from these disclosures. This course teaches students the hands-on skills necessary to manipulate large-scale financial databases and build predictive models useful for strategic and investment decisions.

The course will cover three applications of predictive analytics in this setting: (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, estimate prediction models, and backtest algorithms. Given the widespread use of financial disclosures and the growing use of Big Data, this course is valuable for students pursuing a broad variety of careers ranging from investment bankers and equity analysts to consultants and business managers.

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 regression and hypothesis testing.

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. Throughout your time at Wharton 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 Big Data and predictive analytics: (1) forecasting future earnings + backtesting prediction models; (2) predicting accounting fraud; and (3) detecting insider trading. For each unit, students will complete and present 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 analyze these datasets in 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. Some of the academic articles might seem intimidating at first, but it is important to go to the source material. Don't worry if you do not understand everything. Learning is about knowing what you know, knowing what you don't know, and knowing what questions to ask. It is your responsibility to identify gaps in your knowledge and take actions to remedy those gaps. **The class will entail significant hands-on learning outside of class. At times you may feel like you are drinking from a fire hose. Stop and ask questions. Ask your peers. Ask me. Rely on your group. Pick up as much knowledge as you can.**

My goal is to help you become more informed about the practice of predictive analytics as it relates to financial data as used in the real-world. If you find yourself stuck on an individual or team assignment, don't hesitate to reach out to the TAs or myself.

Assignments

Your course grade will be determined as follows:

Individual Component	Class Participation	15%
	3 Individual Assignments / "Analyst reports" (10% each)	30%
Group Component	3 Team Projects / "Team white papers" (15% each)	45%
	1 Team Presentation	10%

Class Participation: Everyone starts with a 0% participation. Participation (questions) will be recorded on a team basis. Your team must ask questions in order to receive credit for participation, attendance alone is not sufficient. At the end of the semester, each team member will anonymously indicate what percent of team participation each member should earn (0-100%). *Individual class participation score = Team participation score * Average Allocation Your Team Assigns You.*[1] Attendance is necessary to do well. In addition to discussing the required reading, many classes will feature in-class programming exercises in which you will work in small groups.

Individual Assignments: These assignments will take the form of a brief "analyst report" on a given topic. Each individual assignment will ask you to collect some data and conduct rudimentary analyses. All assignments are provided on the first day of class.

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 team project for each unit. On the first day of each unit, each team will submit a written "Team white paper" on the prior unit (that includes the code used to conduct the analysis). For the final unit, teams will present their paper to the class. For the final presentation, I expect you to dress as if you were presenting your analysis in a formal business environment (i.e., wearing formal business attire). All team members will receive the same score on team assignments, and scores will be assigned on an absolute basis not a relative basis (i.e., teams are not competing against each other). **Attendance is required of all students in all presentations. Failure to attend any presentation will result in a zero for class participation.**

All three team projects will be provided on the first day of 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!

[1] E.g., Team X has three members, Rick, Stan, and Steve. Team X has 90% team participation. Each member anonymously gives 100% to Steve. Steve's individual participation is 90% = 90% * 100%. Stan gives himself 100%, Rick and Steve each give him 70%. Stan's individual participation is 72* = 90% * 80%.

We will work on the programming aspects of the projects together in class. I will teach you the core programming steps needed to execute the project in class as we move along.

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 a Forbes July 2017 report, 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 and is also available for purchase at Penn Computer Connection. That being said, you may use whatever SQL-compliant software you like (e.g., R is a popular alternative).[2]

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, and assignments on each topic. There is no textbook for this course. Instead, I will provide a set of notes and readings. 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.

---

[2] A word of caution about using R in the course. R is excellent for statistical analysis, however, it is not well-suited for the big datasets we will be using. R loads the entire dataset into RAM at once, other programs (including SAS) read file sections on demand. As such, the computer's RAM limits the size of the file that R can open. This does not mean it is not possible to use R for this class, only that it is more cumbersome to manipulate large data files. e.g., if you only have 16GB RAM you will not be able to easily view a 20GB file using R.

Topic and Assignment Schedule

This course is a hybrid of a traditional seminar course and a computer science course. All classes will meet in JMHH 380 computer lab. Programming classes, indicated below with *, will teach the programming aspect of the material 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.

1/16 Class #1. Course Introduction

Discuss Course Structure and Organization
Discuss IT solution and Cloud-based Big Data
Cover the concept of relational databases and database keys
Cover the structure of basic financial datasets (CRSP, Compustat, IBES)
Submit Bios

1/21* Class #2. Introduction to Relational Databases, SAS, and SQL

Introduce SAS concepts related to DATA step and PROC command
Introduce SQL concepts related to CREATE TABLE and JOIN statements
Introduce lag operator

>>>>>>>>>> UNIT #1: FORECASTING FUTURE EARNINGS <<<<<<<<<<

1/23 Class #3. Forecasting Earnings: Concepts

Review accounting concepts related to earnings
Beaver (1968), "The Information Content of Annual Earnings Announcements" *Journal of Accounting Research*
Ball and Brown (1968), "An Empirical Evaluation of Accounting Income Numbers" *Journal of Accounting Research*

DUE: Select and name "Analyst Teams"
[in Canvas use "People" > "Analyst Teams"]

1/28 Class #4. Forecasting Earnings: Concepts

Review accounting concepts related to earnings
Doyle, Lundholm, and Soliman (2006), "Extreme Future Stock Returns Following Earnings Surprises" *Journal of Accounting Research*.
Taylor (2011), "Post-Earnings Announcement Drift and Related Anomalies" in *Handbook of Equity Market Anomalies*, Ch 4: 91-115. Wiley Publishing.

1/30 Class #5. Forecasting Earnings: Concepts

Review accounting concepts related to accrual accounting
Review the difference between accruals and cash flows
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*

2/4 Class #6. Forecasting Earnings: Concepts

Review DuPont Analysis

Bradshaw, Richardson, Sloan (2001) "Do Analyst and Auditors Use Information in Accruals? *Journal of Accounting Research*.

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

DUE: Individual Assignment #1

2/6* Class #7. Forecasting Earnings: Programming

Extracting data on past performance and earnings
Measuring earnings expectations and earnings surprise
Correlating deviations from expectations with future performance

2/11* Class #8. Forecasting Earnings: Programming

Decomposing earnings into cash and accrual components
Estimate persistence of cash and accrual components

2/13* Class #9. Forecasting Earnings: Programming

Replicate existing prediction models
How to Backtest Forecasts I

2/18* Class #10. Forecasting Earnings: Programming

How to Backtest Forecasts II
Horse-race prediction models

2/20* Class #11 Forecasting Earnings: Programming

Improving existing predictions models

2/25* Class #12 Forecasting Earnings: Programming

Improving existing predictions models

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

2/27 Class #13. Accounting Fraud: Concepts

Famous cases of earnings management
Cover readings:

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*

Lee, Ingram, and Howard (1999) "The Difference between Earnings and Cash Flow as an Indicator of Financial Reporting Fraud" *Contemporary Accounting Research*

Dechow, Ge, Larson, and Sloan (2011), "Predicting Material Accounting Misstatements" *Contemporary Accounting Research*

DUE: Group Assignment #1

3/3 Class #14. Accounting Fraud: Concepts

Cover readings:

Beneish (1999), "Detection of Earnings Manipulation" *Financial Analysts Journal*

Beneish, Lee, and Nichols (2011), "To Catch a Thief: Can Forensic Accounting Help Predict Stock Returns?" working paper

Price, Sharp, and Wood (2011) "Detecting and Predicting Accounting Irregularities: A Comparison of Commercial and Academic Risk Measures" *Accounting Horizons*

3/5 Class #15. Accounting Fraud: Do We Know It As It Happens? — GUEST LECTURE

Matt Jacques, Chief Accountant, SEC Division of Enforcement

----- SPRING BREAK -----

3/17 Class #16. Accounting Fraud: Data

Introduction to AAER data and Audit Analytics Data

Replicate existing fraud prediction algorithms

DUE: Individual Assignment #2

3/19* Class #17. Accounting Fraud: Programming

Incorporate additional data in existing algorithms

3/24* Class #18. Accounting Fraud: Programming

Incorporate additional data in existing algorithms

3/26* Class #19. Accounting Fraud: Programming

Backtest new algorithms

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

3/31 Class #20. Insider Trading: Concepts — GUEST LECTURE

Jonathan Vogan, Quantitative Research Analyst, SEC

DUE: Group Assignment #2

4/2 Class #21. Insider Trading: Concepts

Cover readings:

Jagolinzer, Larcker, and Taylor (2011) "Corporate governance and the information content of insider trades" *Journal of Accounting Research*
Cohen, Malloy, Pomorski (2012) "Decoding Inside Information" *Journal of Finance*
Arif, Kepler, Schroeder, and Taylor (2018) "Audit Process, Private Information, and Insider Trading" working paper.

4/7* Class #22. Insider Trading: Data
Introduction to disaggregated trading data
Aggregating trade to the person-level, company-level

DUE: Individual Assignment #3

4/9* Class #23. Insider Trading: Programming
Replicating existing insider trading detection algorithms

4/14* Class #24. Insider Trading: Programming
Incorporate additional data in existing algorithms
Backtest new algorithms

4/16* Class #25. Insider Trading: Programming
Incorporate additional data in existing algorithms
Backtest new algorithms

4/21 Class #26 Final Team Project Presentations ("Detecting Insider Trading")

4/23 Class #27 Final Team Project Presentations ("Detecting Insider Trading")

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

Professor Daniel Taylor's research focuses on opportunistic/fraudulent reporting and insider trading. His research appears in leading academic journals; has been featured in the popular media, including such outlets as the *Wall Street Journal*, the *New York Times*, and the *Economist*; has been cited in final rulings by the U.S. Securities and Exchange Commission; and has informed investigations by multiple enforcement agencies, including the FBI and DoJ. He received a PhD in Business from Stanford University, a MA in Economics from Duke University, and a BS in Economics from the University of Delaware.