HKUST BUSINESS SCHOOL 香港科大商學院

~~BIG~~ DATA ANALYTICS
ISOM 5270
SYLLABUS – Fall 2025

Course Name	~~Big Data~~ (Predictive) Analytics
Course Code	ISOM 5270
Course Schedule and Classroom	**L1:** Friday, 2:00 PM to 5:20 PM (Classroom: LSK-2003) **L2:** Monday, 7:00 PM to 10:20 PM (Classroom: LSK-2003) **L4:** Tuesday, 7:00 PM to 10:20 PM (Classroom: LSK-2003) **L5:** Friday, 9:30 AM to 12:50 PM (Classroom: LSK-2001)
Instructor	Prof. Carlos Fernández-Loría**,** ISOM Office: LSK5054A Email: imcarlos@ust.hk Office hours (book in advance to guarantee a spot): - Mon, Tue and Wed 3:00 PM to 5:00 PM
Teaching Assistant	Ray Pang Office: LSK 4065 Email: imncpang@ust.hk Office hours: - Thu 3:00 PM to 5:00 PM - Fri 10:00 AM to 12:00 PM

1. Promise of the Course

Data is reshaping the world. It is used not just to describe the past but to predict—to use what we know to figure out what we don't and make better decisions.

Streaming platforms use data to predict what we will enjoy and decide which movies to show us. Banks predict the risk of default and decide whether to approve a loan. Doctors predict the likelihood of disease and decide on early interventions. Election campaigns predict how we might vote and decide how to target their messages.

Today, this includes not only traditional predictive models but also newer systems such as generative AI. Though they differ in how we build and interact with them, both can be understood through the same predictive modeling lens.

The fact is that predictive modeling has become one of the defining forces of our time. This course will help you understand that force and use it wisely. By the end, you will know how to apply predictive modeling to make better business decisions.

To deliver on that promise, this course is built around helping you grow as a learner, thinker, and communicator. You will learn frameworks for reasoning about predictive modeling and practice them through case studies, discussions, and hands-on work with data and predictive models. In the process, you will develop the abilities that make these tools valuable in the real world.

Specifically, you will learn to:

- **Frame problems.** Break down business challenges into predictive tasks and organize the development of a solution into key components.
- **Reason about data needs.** Identify the data required to build and use predictive models for decision making, and the assumptions behind those choices.
- **Build models.** Use code to create predictive models from structured data and understand the principles behind how they are built.
- **Make data-driven decisions.** Apply predictive models to guide and improve decision making.
- **Evaluate solutions.** Judge how well a solution addresses the business problem and under what assumptions it works.
- **Compare prediction paradigms.** Recognize similarities and differences between traditional predictive models and generative AI, and reason about how these matter when designing and evaluating decision systems.
- **Communicate effectively.** Explain your reasoning in writing and discussion so that others can understand, evaluate, and act on it.
- **Learn independently.** Use resources (e.g., the textbook) to acquire new knowledge and solve unfamiliar problems.

2. What We Will Do to Realize the Promise

This course is not about memorizing answers; it's about learning how to think through and solve business problems with predictive modeling. You'll do that mainly by learning independently and applying what you've learned to assignments and case studies. I'll provide materials, guidance, and advice on how to do this effectively, but most of the conceptual understanding and coding practice will be up to you. I won't lecture much.

That doesn't mean you'll be on your own. I'll be there to guide and support you. My goal is to help you grow into someone who can tackle messy problems, defend your reasoning, and keep learning on your own.

Before class: You'll work on an assignment—usually a case study—and turn it in the night before class. These assignments aren't about being "right"; they're about practicing and learning by doing. Often, you'll need to pick up new concepts and try them out, sometimes by coding and testing models. I'll read your work to check how you're progressing and use it to shape our discussions.

During class: We'll use your work to drive the conversation. Together we'll compare approaches, clarify misconceptions, and connect what you did to broader concepts. The point isn't for me to hand you the "right" solution, but to help you practice reasoning, weighing alternatives, and drawing connections to fundamental concepts.

After class: You'll have the option to write and share notes on what we covered—the problem, the reasoning, the key concepts, and the conclusions. The idea is to encourage you to make sense of the material in your own words, while also creating a resource your classmates can learn from. Quality matters more than frequency; what counts is making contributions that are genuinely useful to others.

Later in the semester: After Class 4, you'll also take on milestone case studies alongside the pre-class work. The pre-class assignments are about exploring new concepts; the milestone cases are about testing your mastery of concepts you should already know. They'll require you to code predictive models, test them, and make decisions from your results.

3. How We Will Work Together

The promise of this course can only be realized if we all do our part. Learning is joint work: I'll bring structure, guidance, and support; you'll bring effort, preparation, and presence. How much you engage is your choice, and your learning will reflect that. But coming to class also requires your commitment to contribute to a learning environment that supports not only you, but your classmates as well.

From me:

- **Make class worth your time.** I'll do everything I can to ensure our classes are engaging, useful, and aligned with the promise of the course.
- **Be accountable.** You can ask "WHYSHIC?" ("Why should I care?") at any time. I commit to explaining how what we're doing connects to the promise of the course and to your growth as a thinker, learner, and communicator.
- **Be available.** I'll respond to forum posts and emails within 24 hours. Please use the forum first—it helps me avoid answering the same question multiple times and lets everyone benefit from the answer. There will also be office hours every weekday (with me or the TA); you can book in advance to guarantee your spot.

From you:

- **Come prepared.** Work on the assignments before class so we have something real to discuss together.
- **Be present.** Stay engaged and don't do unrelated activities during class. Laptops, tablets, and phones should stay closed unless you're using them to look up something directly relevant to what we're discussing. Otherwise, they're just distractions that make it harder to focus and have good discussions.
- **Contribute to others.** Support your peers (share notes, do pre-class work together) and enrich discussions by engaging with the problem, asking good questions, and building on each other's ideas.

If you can't commit to this at a given moment, it's perfectly fine to step out and return when you're ready.

4. The Plan

Class	Date	Topics (subject to changes)	Deliverables
1	L1: Oct 31 L3: Oct 27 L4: Oct 28 L5: Oct 31	Frame decision problems as prediction tasks Introduction to machine learning (ML) Data science process	- Read the syllabus - Pre-class work #1
2	L1: Nov 07 L3: Nov 03 L4: Nov 04 L5: Nov 07	Key data choices: Instance, target, features Learning from data: common modeling approaches, loss functions, model representations.	- Pre-class work #2
3	L1: Nov 14 L3: Nov 10 L4: Nov 11 L5: Nov 14	Model selection: dealing with model complexity and overfitting, out-of-sample testing, hyper-parameter tuning	- Pre-class work #3
4	L1: Nov 21 L3: Nov 17 L4: Nov 18 L5: Nov 21	Expected value framework Alternative evaluation measures (AUC, Lift)	- Pre-class work #4 - Team formation.
5	L1: Nov 28 L3: Nov 24 L4: Nov 25 L5: Nov 28	Causal analytics A/B Testing	- Milestone case #1 - Pre-class work #5
6	L1: Dec 05 L3: Dec 01 L4: Dec 02 L5: Dec 05	Unstructured data: transactions, text, images Deep learning	- Milestone case #2 - Pre-class work #6
7	L1: Dec 12 L3: Dec 08 L4: Dec 09 L5: Dec 12	Large Language Models Gen AI vs “Good Old-Fashioned ML” Wrap-Up	- Milestone case #3 - Pre-class work #7
8	Available Dec 13-21	Final Exam (take-home, individual, time constrained)	

5. Tools for the Journey

To work on the assignments, you'll often need to pick up new data science concepts or code to test your ideas. How you do that is up to you: YouTube, Google, ChatGPT, and other AI tools are all fair game. What matters is that you use the tools available to learn what you need in order to solve the problems in front of you.

That said, I'll provide support for two key materials:

- **Textbook: *Data Science for Business* (Provost & Fawcett).** This will be our main reference for core concepts. I'll often point to chapters and sections that are helpful to review to complete assignments. I encourage you to get a copy before our first class (a digital version is available through the HKUST library).

- **Software: Python.** This is the coding tool I'll support, and all code examples will be in Python. If you already know another programming language, you're free to use it, but I won't provide support for it. Don't worry if you haven't coded before; the level required here is not very advanced, and many students start from scratch and do well. That said, this is not a Python course. I won't use class time to teach programming basics, but I'll provide resources, answer questions, and help you troubleshoot. You're also encouraged to use generative AI (e.g., ChatGPT, DeepSeek) to assist you. In short: I'll support you all the way, but learning to code is your responsibility.

Getting Started with Python

It is OK if you've never coded before. You don't need to become an expert programmer for this course; you just need to become comfortable enough to run and experiment with code. Here are two simple ways to get started:

- **Install Anaconda.** This is free software that gives you everything you need to run Python on your computer.

- **Use Google Colab.** A free, browser-based option that works like Google Docs but for code—no installation needed.

Then, practice with Pandas, a Python tool for working with tables of data (like Excel, but more powerful). This YouTube playlist is a good place to start:

https://www.youtube.com/watch?v=ZyhVh-qRZPA&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS

You don't need to memorize everything or watch all videos. Just get comfortable with basic tasks like looking at rows, columns, and summaries of data.

Later in the course, I'll give you Python code that walks through how to build and use predictive models. You'll then adapt that code to work on the assignments, which is where most of the learning happens.

6. How We Will Track Progress

Grades are part of the course, but not the point—your learning is. I've set up things to help you learn by doing, lower stress where I can, and reward meaningful contributions.

Pre-class work (20%): You'll submit a pre-class assignment the night before each class. These are graded on effort, not correctness. The goal is to learn new concepts by applying them to problems. You can use any resources you like (forum, classmates, generative AI) but your submission must reflect *your own* reasoning. I'll read these before class and may ask you to share your approaches, so be ready to explain your thinking. Late work loses 10% if submitted before class; after that, it won't be accepted.

Quizzes (10%): At the end of each class, there will be a short closed-book quiz. These are cumulative, so anything we've covered so far is fair game (including that day's material). The goal is to keep you on track and reinforce key concepts. To reduce stress, I'll drop your three lowest scores. Because of this policy, there are no retakes; missed quizzes are covered by the drops.

Class contributions (20%): This is about how you make the class better for everyone.

- **In class discussions**: Quality matters much more than quantity. I look for comments that move us forward—clarifying reasoning, addressing unresolved questions, or connecting back to key concepts. I'll keep track of standout contributions, and in the quizzes, you'll have the chance to tell me about your contributions so I don't miss them.

- **Through notes**: You can write up and share short notes on a class—covering the problem, the reasoning, the key concepts, and the conclusions. There's no hard deadline for these, but earlier notes are more valuable (and rewarded more) since they're more useful to classmates. You don't need to do this every class; what matters is that when you do, your notes are clear and genuinely helpful.

Milestone cases (21%): After Class 4, you'll submit three milestone case studies (one per week, 7% each) to demonstrate mastery of concepts you've already learned. You'll code predictive models, test them, and use the results to make decisions. You can work on these alone or in pairs (confirm your group by Class 4; otherwise, I'll assign partners if available). Late work loses 10%, or 50% if over a week late.

Final Exam (29%): The exam will be take-home, available Dec 13–21. You can start whenever it works for you, but once you begin you'll have a fixed amount of time (still TBD). It's open-book: you can use your notes, the internet, and tools like ChatGPT. But it must be your own work; you are not allowed to work on it with anyone else.

Re-grading policy: If you think there's been a grading mistake:

1. Email the TA first to clarify.
2. If it's still unresolved, email both the TA and me within 3 days of receiving your grade, with a short explanation. Late requests will not be accepted.
3. I'll re-grade the whole submission, which means your score could go up or down.