



Algorithms, Big Data and Online Marketplaces

Welcome to the course!



Our class

The purpose of this class is very straightforward: we want to bring algorithms, big data and online marketplaces together to address the following questions:

How to collect data?

How to use analytics and algorithms to analyze data?

How to guide firms' business operations using the insights obtained from data?





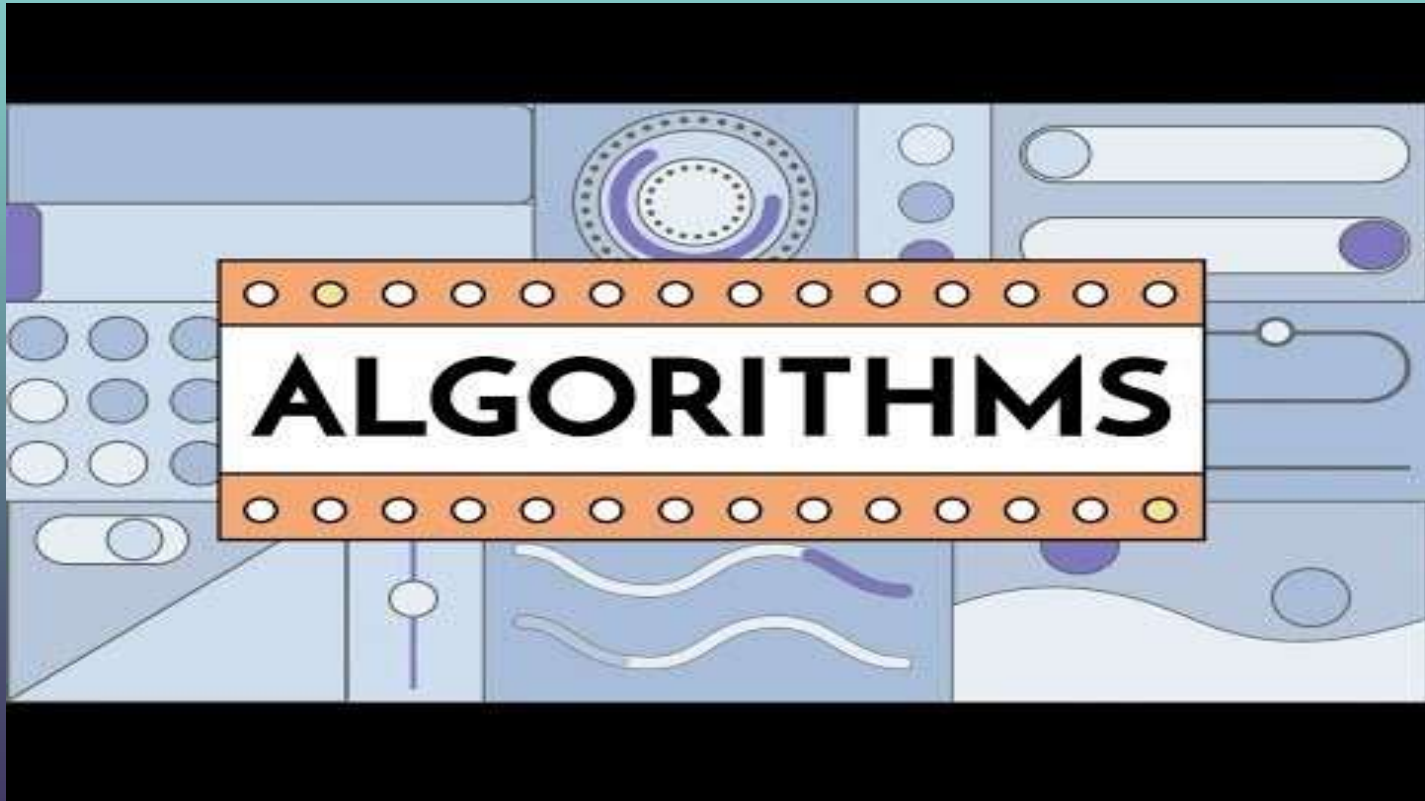
Questions

What is an algorithm?

Do you know any algorithms?



Questions



What is...

An algorithm?

Xi Li

What is an algorithm?

19/11/2023, 11:30:27



ChatGPT (Tokens used 97)

An algorithm is a set of instructions or steps that a computer program follows to solve a problem or perform a task. Algorithms are used in many areas of computing, including search engines, data analysis, and artificial intelligence. They are designed to be efficient and effective, and can be written in different programming languages depending on the needs of the application.


19/11/2023, 11:30:27

This is perhaps the only B-school class with an “algorithm” in its title!



Examples of Algorithms

The Netflix Prize: In 2006, Netflix launched a competition called the Netflix Prize, offering a \$1 million prize to anyone who could improve their movie recommendation algorithm by 10%. This attracted thousands of participants from around the world, including data scientists, mathematicians, and computer scientists. It took several years, but eventually, a team named “BellKor's Pragmatic Chaos” won the prize in 2009, improving Netflix's recommendation algorithm by 10.06%







Who should take this course?

MSc Students who are

- Interested in marketing in the digital age.

- Interested in data analysis and programming in general.

- Interested in joining a big-tech firm or start a career in the Internet industry.





Social Listening

Through the years, colleagues, students, and TAs share students' online chatter with me. Now, let me share some with you.



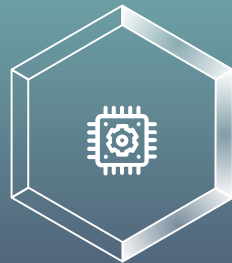
This course adopts methodologies from



**Computer
Science**



Statistics



Economics




Leveraging your competitive advantage

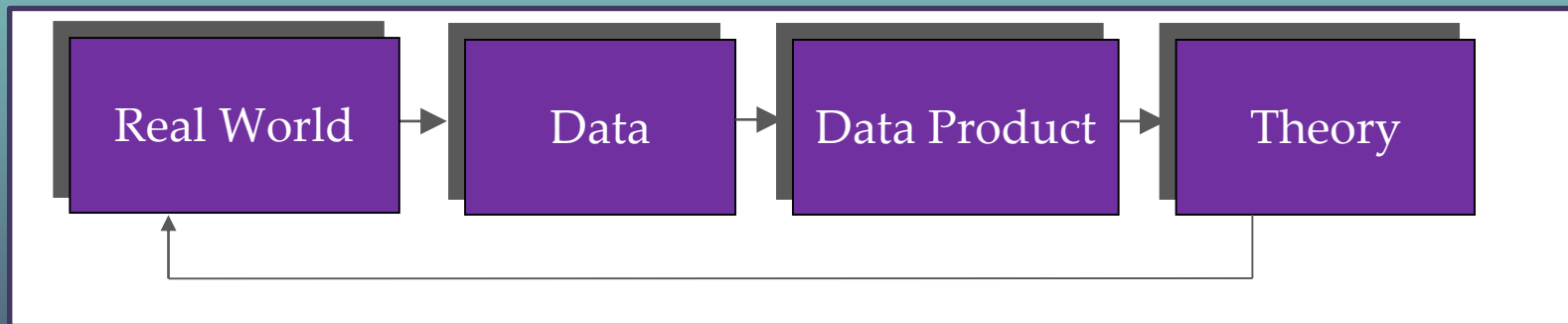
Compared to traditional marketers, you know how to program and how to analyze data.

Compared to statisticians and computer scientists, you understand consumers better, and know how to apply results to business settings.

Compared to economists, you not only know the theory, but also know how to apply the theory to solve real-world marketing problems, and test the theory using real-world data.



The LOOP





It seems that our class motto is

“Talk is cheap, show me the data!”



We emphasize on data, data, and data

Recall that the program offers you an **MSc** degree in Marketing, not an **MA** degree.

Why? Because you are learning quantitative methods, not just how to make presentations and talk about business ideas.



HKU
BUSINESS
SCHOOL
港大經管學院

Master of Science in Marketing
理科碩士(市場營銷學)

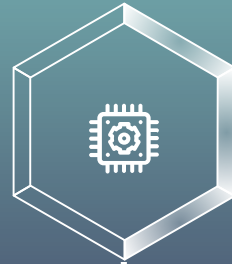
We solve problems for



Firms



Platforms



Individuals

Do I really want to take the course?

I don't know. It depends. But let me offer you some advice.

Take the course if

You are interested in data analysis, and you are considering to become to data scientist in the future.

You want to understand how the online marketplaces work.

You want to explore rigorous research methodologies from different areas.

Don't take the course if

You hate data analysis or programming.

You already know the materials to be covered in the class.

You want to pick an easy course to fulfill your credit requirement.



Additional Course Information






The Instructor

Xi Li, Professor of Marketing
Director, Asia Case Research Centre
Associate Director, Institute for Digital Economy and Innovation

PhD in Management, University of Toronto.
M.Phil. in Operations Research, HKUST.
B.E. in Computer Science, Tsinghua University.

Research interests: Algorithms, big data and online marketplaces.
Email: xili@hku.hk
Office Hour: by appointment





Our Teaching Assistant

Ray Song

Can be reached via email: songray@hku.hk



My Research

We observed an interesting phenomenon: In a crowdfunding campaign, two very similar products often have very different prices. See the following examples. We addressed this puzzle.

LIFE ON THE LINE



A unique book of portraits connecting and celebrating life along the Arctic Circle.

Created by:
Cristian Barnett

201 backers pledged £12,134 to help bring this project to life.

Last updated September 28, 2014

£30

THE BOOK. A signed copy of LIFE ON THE LINE, your name included as a supporter in the book and an invitation to the launch.

£150

THE SPECIAL BOOK. A collector's edition of LIFE ON THE LINE (limited to just sixty-six copies), real cloth and slipcase, signed with a personal dedication by Cristian Barnett, invitations to events, plus special thanks in the book and online.

My Research



We are familiar with HK's local online marketplace, HKTVMall. Interestingly, at HKTVMall, consumers can observe the historical prices of products that they are interested in, a feature not available at other major online sellers.

We ask: Why is HKTVMall willing to disclose the historical prices? What happens when consumers can observe it?



Course Website

In addition to the official course website provided by the university, we will also use a semi-official course website. You can find the latest materials and updates on this course website.

https://ximarketing.github.io/_pages/teaching/

Password for ABOM: 7025

Please do not share the course content with others.





Textbook? No.
Real Data? Yes.






Course Overview

Data analytics with R.

You can use Excel or SPSS, but they are too simple and cannot handle complex data analysis projects.

R is a free software that is commonly used for statistical analysis.

It is not only useful for digital marketing, but also useful for other purposes such as machine learning and optimization.






Course Overview

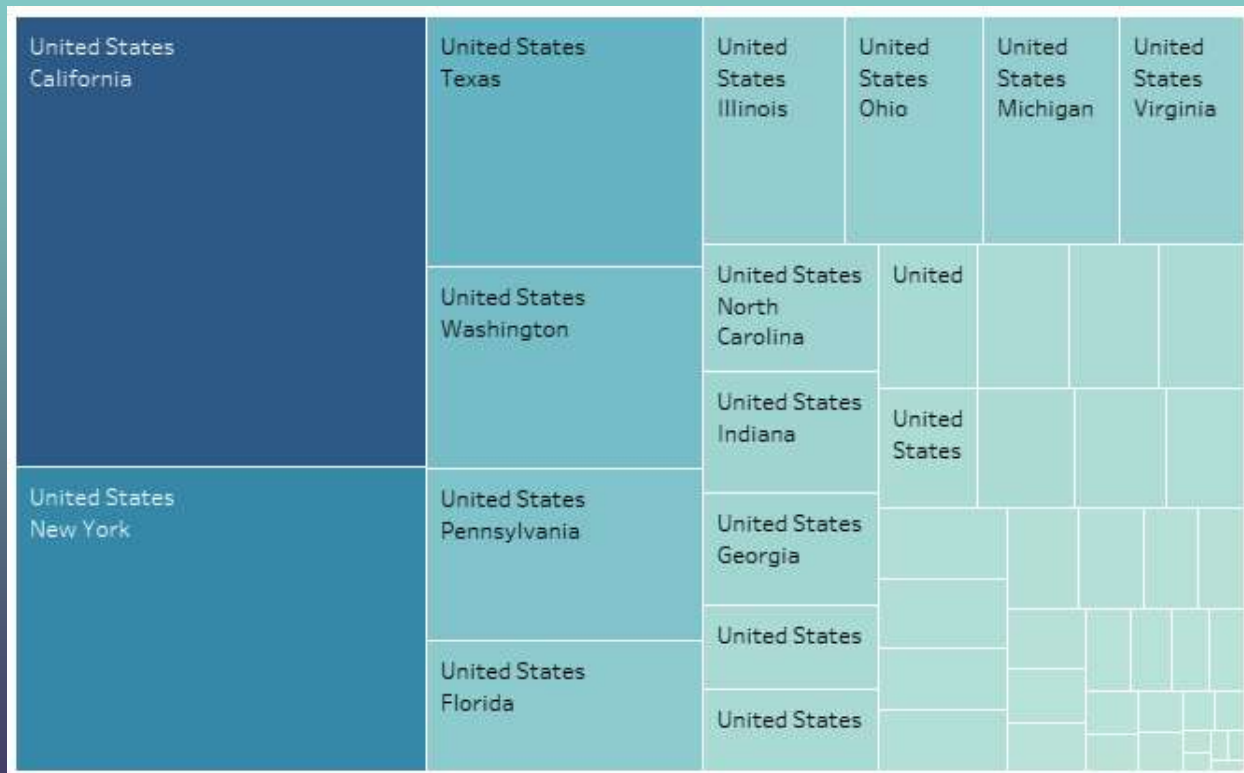
Data visualization with Tableau.

Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs.

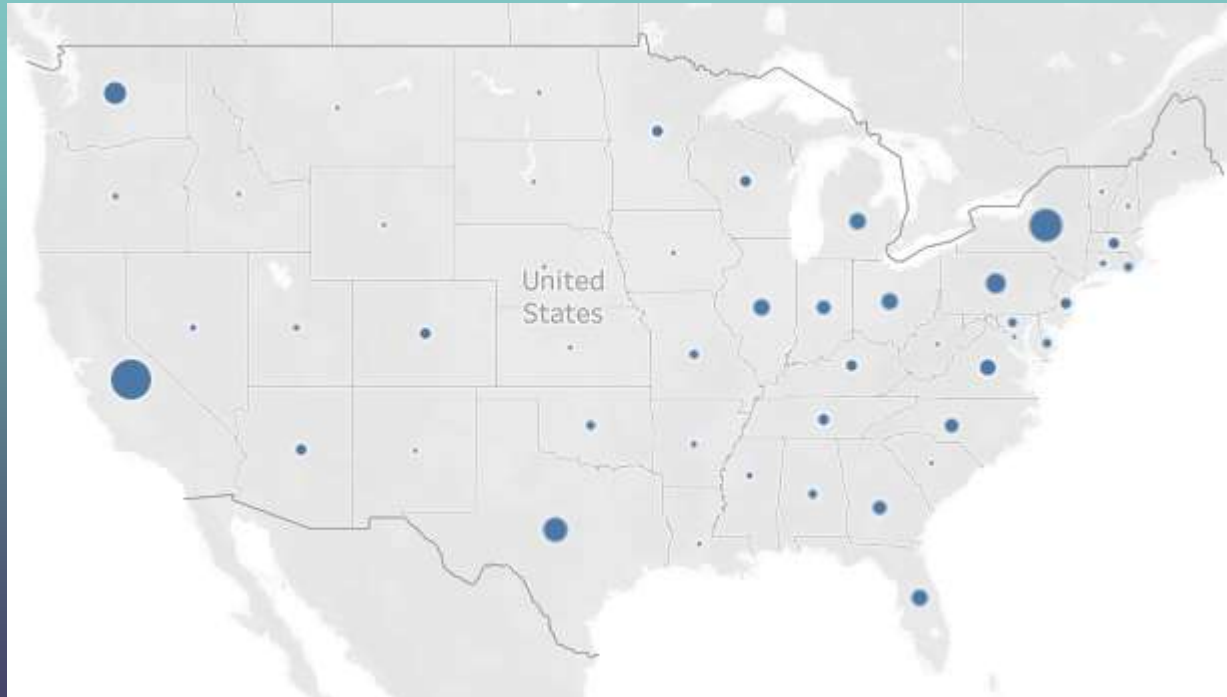
When you want to show your results to others, the best way to use figures --- *“A picture is worth a thousand words”*.



Tableau



Tableau






Course Overview

Web Scraping with R.

Data scraping allow you to scrawl information from websites, e.g., online marketplaces. For example, you can collect product information from HKTV Mall, weather information from Hong Kong observatory, and tweets from Twitter.

Warning: Do not expect that you can become an expert in data scraping within one class. This is simply an introduction, and you need more practice yourself.






Course Overview

Web Scraping with R.

Why choosing R? R is powerful and is widely used for data analysis.

Python is also a great choice (and maybe a better choice for machine learning tasks). However, it is not that friendly for beginners --- you need to take a more specialized course to learn it.



Course Overview

This may be the only web scraping class in a business school (I don't know the answer, though).

We will see how to gather information from HKU faculty webpage:






Course Overview

This may be the only web scraping class in a business school (I don't know the answer, through).

Note: We are only going to cover the simplest techniques for web scraping. If you want to learn something more complex (e.g., how to login to your Moodle and download your materials automatically), you still need take some additional courses.



Course Overview

2021 Nobel Prize in Economics



Joshua D. Angrist and Guido W. Imbens

“for their methodological contributions to the analysis of causal relationships”




Course Overview

Causality

What is the fundamental difference between economics/marketing and statistics/machine learning? It is causality.

In statistics and machine learning, we ask if X **predicts** Y. But now, our question is, does X **cause** Y?

We will talk about when and how to draw conclusions on causality.





Course Overview

Logistic Regression

You should be already familiar with linear regression, the simplest statistical model for predicting.

But linear regression only works for certain dependent variables, and it works poorly with binary dependent variables.

Logistic regression is introduced to deal with the issue.






Course Overview

Going beyond logistic regression

Logistic regression bears some similarities with some human tasks such as autonomous driving and digit recognition.

Indeed, our human brain also calculates logistic functions.

A fundamental machine learning algorithm, artificial neural network, is a generalization of the logistic regression we discussed.





Course Overview

Text Analysis

In the past, we focus merely on numerical data (e.g., sales, profit, purchases, price, time etc.).

However, today, most of the data take other forms. Many of them are text information.

This includes online reviews, product descriptions, Tweets, SMS messages, forum discussions, firm announcements etc.





Course Overview

Text Analysis

We are going to take some simple measures to extract meaningful information from text data.

Sentiment analysis: It classifies text based on sentiment polarization (positive vs. negative).

Latent Dirichlet Allocation (LDA): It classifies text based on the topic of the text.



Course Overview

Price Personalization

Do you know behavior-based pricing (杀熟)? Why does firm charge higher prices to old consumers and lower prices to new consumers? Is doing so profitable?

More generally, how should firms use consumer data to offer them personalized prices to improve profit?

Should public policymakers regulate price discrimination based on big-data technologies?

Course Overview

Price Personalization: Policy considerations

中华人民共和国个人信息保护法

(2021年8月20日第十三届全国人民代表大会常务委员会第三十次会议通过)

第二十四条 个人信息处理者利用个人信息进行自动化决策，应当保证决策的透明度和结果公平、公正，不得对个人在交易价格等交易条件上实行不合理的差别待遇。

通过自动化决策方式向个人进行信息推送、商业营销，应当同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式。

通过自动化决策方式作出对个人权益有重大影响的决定，个人有权要求个人信息处理者予以说明，并有权拒绝个人信息处理者仅通过自动化决策的方式作出决定。



Course Overview

Recommender Systems

Every time you visit Amazon, Taobao and YouTube, you always receive some recommendations from these platforms.

The recommendations are made based on your past behavior and characteristics of the products/services.

We will talk about how online platforms make personalized recommendations using big-data technologies.





Grading

50% Group work:

40%: Two data projects, 20% each.

10%: Providing an interesting observation from online (or even traditional) marketplaces (possibly with your own explanations; this is *optional*), and present it in class. You also need to join the discussion of other groups.

20% In-class participation:

Class attendance and participation in discussions. (TA will take notes).

30% Individual Assignment:

Data Collection and Analysis






Data Projects

In this course, we are introducing one (or two?) practice class on data analysis. This is one unique feature of our course.

You have already learned about R and Tableau. Perhaps you can also use many other tools such as SPSS, Excel, Python etc.

I will give you some real-world business data, and your goal is to study the data using the tools you know --- It's entirely up to you how you want to play with the data!





Data Projects

You are going to work with your teammates on the data projects. Then, each group will submit a report illustrating the findings from the data. **It will be graded, and it affects your final grade.**

We will start with the data analysis in the class, and you will finish all the analysis after the class.

The purpose of the data project is to help you understand the data analysis methodologies and gain a sense of real data that data scientists are analyzing everyday.





Overall Course Structure

We have 10 lectures over the entire module. Among them, we are going to have

- 1 introduction class (today)
 - 2 data workshops
 - 7 regular lectures
- 



Presentation

In the class, each group is required to provide an interesting observation from online (or even traditional) marketplaces, and invite other groups to discuss about it. You may also share your own thoughts and discuss about other groups' topics. This counts for 10 points toward your final grades.

I will give you examples later on.


The grading is based on the “interestingness” of your observation (5 points), as well as your discussion of other groups' topics (5 points).





Presentation

The presentations/discussions will be in the beginning of a few lectures, and the order of presentation will be drawn randomly after group formation is completed.





TODO List

If you decide to take this course, here is what you should do after today's class (Important!):

Form groups with your classmates.

Individual task: Install R on your laptop. We will be using it in the next class.

Bring your laptop with you for the next lecture!





Group Formation

Each group consists of 6~8 students.

You need to **choose a name** for your group, e.g., “Marketers”, “Fantastic”, “A Plus” ...

The deadline for submitting your group information is:
23:59 Dec 12 (Tuesday; one day after the add-drop period)

Let the TA know if you cannot find a group.





Let's Download and install R.

Your installation path must not contain any non-English characters. Otherwise, you will have troubles using it.

安装路径必须为纯英文，否则运行可能出错。





Next, let's download R-Studio.
It is also free.

Your installation path must not contain any non-English characters. Otherwise, you will have troubles using it.

安装路径必须为纯英文，否则运行可能出错。

