Marketing Models

Instructor: Professor (Dr.) Anirban Mukherjee (am5422@columbia.edu)
TA: Yegor Tkachenko (ytkachenko21@gsb.columbia.edu)

Course Aim and Objective

The course aims to introduce modern big data analysis, choice modelling, and machine learning methods—jointly referred to as marketing analytics—to conduct marketing research and develop market intelligence.

Traditional marketing methods were designed for small datasets where econometric efficiency was the key concern. In the modern “big data” landscape of varied, large, and evolving datasets: computational scalability, reliability, and reproducibility are key concerns. The course aims to describe and introduce cutting-edge toolkits designed for estimating/training empirical marketing models on large-scale marketing data.

The course is targeted to students interested in a wide variety of roles that require marketing data analysis. Quantitative/computational marketing data analysis is a fast-growing function in many industries and firms ranging from brands (e.g., Nestle and Unilever) to retailers (e.g., Amazon and Walmart) to traditional and social media platforms (e.g., Google and NYTimes) to advertising and media agencies (e.g., WPP and Omnicom). Quantitative/computational marketing data analysis is also crucial to management consulting (e.g., BCG Gamma) and equity research/trading (e.g., Point72). Furthermore, the course seeks to provide a solid foundation in data analysis for doctoral programs in quantitative/computational marketing.

To emphasize the development of applied, practical, and relevant knowledge through applications of tools and methods, the course will adopt a code-as-you-learn approach. The assignments and the project will require (extensive) coding in R. The TA and I jointly have many decades of experience in coding: We will work with you, both in-class and out-of-class, to make sure you have the technical knowledge needed to complete the assignments and the project.
## Lesson Plan

<table>
<thead>
<tr>
<th>Week</th>
<th>Materials</th>
<th>Readings/Notes</th>
</tr>
</thead>
</table>
| 1    | Introduction and Overview  
- Marketing research  
- Course administration  
- Refresher tutorial on R | Tutorial on R. |
| 2    | Descriptive MR: Strategic Marketing  
- Introduction to dplyr  
- Introduction to ggplot  
- Introduction to ggvis | Assignment 1 released. |
| 3    | Descriptive MR: Consumer Analytics  
- Introduction to data.table |  |
| 4    | Marketing Mix Modeling  
- Marketing mix modeling  
- Endogeneity and instrumental variables  
- Bayesian Regression Models Using Stan. |  |
| 5-8  | Modeling Customer Behavior  
- Logistic classification  
- Finite mixture model  
- Hidden Markov model  
- Support vector machines  
- Decision trees, CART, random forest | Assignment 1 due; Assignment 2 released. |
| 9-10 | Customer Listening  
- Feedforward network  
- Convolutional network  
- Recurrent network  
- Attention and transformer networks  
- Introduction to Tensorflow  
- Introduction to Keras | Assignment 2 due; Assignment 3 released. |
| 11   | Natural Language Processing/Understanding  
- Latent Dirichlet allocation  
- Sentiment classification | Assignment 3 due; Assignment 4 released. |
| 12   | Group Presentation | Assignment 4 due.  
Group project report due. |
| TBA  | |  |

Please note: the lesson plan is tentative and subject to change.
Individual assignment #1 (Marketing Mix Modeling): 15%
Individual assignment #2 (Customer Analytics—Spotify): 20%
Individual assignment #3 (Social Media Analytics—Threadless): 10%
Individual assignment #4 (Customer Listening—IMDB): 10%
Group presentation: 10%
Group project report: 20%
Class attendance and participation: 15%

Individual Assignments

Each assignment will begin in-class and need to be completed after-class. The assignment is described in a R notebook that includes instructions on how to download data and conduct preliminary analyses and lists assignment questions. You will answer the assignment questions in the R notebook (including both your code and results), which you will then submit for grading.

The four assignments are:

1. **Marketing mix modelling**: Marketing mix modelling is the bread-and-butter of the media industry (advertisers, brands, and publishers). Marketing mix modelling seeks to establish the causal relationship between the marketing activities of the firm (pricing, advertising, etc.) and various outcomes (market shares, profits, etc.). In this assignment, you will develop summary measures to describe the industry landscape and then learn to use various empirical models to measure and optimize the effectiveness of the marketing mix.

2. **Customer analytics**: Customer analytics is an important and fast-growing type of marketing data analysis. As recently as half a decade ago, firms mainly relied on intuition and qualitative marketing research when deciding on pricing and inventory. With improvements in technology, it is now common for firms to collect and analyze large-scale data describing customer behavior, thereby taking a quantitative approach to profitable customer relationship management. In this assignment, you will study data from Spotify. You will learn to estimate/train the most common and important empirical choice models (classifiers) used in marketing practice to describe consumer behavior.

3. **Social media marketing**: With the rise of social media and the growth of the internet, large-scale customer interactions (e.g., Pinterest, Twitter, and Instagram) have become both a vital source of information for marketing mix design (e.g., product design) and a crucial medium for marketing interventions.
(e.g., social media advertising and influencer marketing). In this assignment, you will examine data from Threadless, a pioneering crowdsourcing company. The data describes both users’ voting and purchase behavior. In the assignment, you will learn to use various machine learning algorithms, including neural network models, to predict customer purchase behavior from customers’ stated preferences.

4. Customer listening: Until recently, most marketing data analysis dealt with numerical data (e.g., prices and market shares). Social media, however, involves many media (e.g. audial, textual, and visual media). In this assignment, you will analyze textual movie reviews from the Internet Movie Database. Specifically, you will learn to use various deterministic and probabilistic machine learning methods for both sentiment classification and topic learning.

**Group project and presentation**

The group project will be completed in groups of 4 or 5 (assigned by the instructor after the course has begun). Each group will use any selection of the methods described during the course to complete a specific task. You will be graded on both application and originality, with bonus points to the group with the best performing solution. Details to follow.

**Class attendance and participation**

To do well in the course, it is important that you arrive on-time, prepared, rested, and happy. Your participation in each session will be graded on a 0 to 1.5 scale per session, where a 0 corresponds to an unexcused absence and a 1.5 corresponds to perfect participation (thoughtful comments that build on the discussion, electronic devices used only for classroom learning). Each point corresponds to 1% of the final grade. Therefore, students with a perfect record of excellent class participation will earn up to a bonus of 3% towards the final grade.
Instructor Biography

Professor (Dr.) Anirban Mukherjee is Visiting Associate Research Scholar at the Chazen Institute, Columbia Graduate School of Business (Columbia University); Visiting Fellow at the Emerging Markets Institute (INSEAD); Fellow at the Institute on Asian Consumer Insight (Nanyang Technological University); Honorary Visiting Professor at Chaoyang University of Technology; and Affiliated Faculty at the Science of Networks in Communities (SONIC) research group (Northwestern University).

Professor Mukherjee has over 15 years of experience in researching and teaching quantitative and computational marketing. He has developed and taught courses at on marketing strategy, marketing analytics, and marketing research at Columbia University, Cornell University, INSEAD, and Singapore Management University. His research has been published by elite research journals (e.g., Journal of Marketing Research, Journal of Retailing, and Management Science), featured in the popular press (e.g. Forbes and Business Times), and received numerous awards. He has been invited to give research presentations at more than 50 universities in North America, Europe, Asia, and Australia, including Cornell University, INSEAD, Northwestern University, University of California-Berkeley, and Washington University.

Professor Mukherjee holds a B.Sc. in Electrical and Computer Engineering (2003), a M.Sc. (2008) in Quantitative Marketing, and a Ph.D. in Quantitative Marketing (2009), all from Cornell University. He studied at The Doon School, Dehra Dun (353 KA, 1999).

Teaching Assistant

Yegor Tkachenko is a 4th-year PhD student at Columbia Business School, where he studies applied machine learning. His primary interest is in designing robust, scalable autonomous intelligent systems and agents capable of optimal action to advance specified business objectives. He holds an MS in Operations Research (MS&E) from Stanford University (2016) and a BS in International Relations and Business from Taras Shevchenko University (2012). While at Stanford, he worked as a research assistant at Stanford Intelligent Systems Laboratory (SISL) of Aeronautics and Astronautics Department, under Prof. Mykel J. Kochenderfer, doing research on deep reinforcement learning. Prior to that, in 2013-2014, he spent a year working for the BlackRock Global Client Insights team.