

CIS 84000 - Doctoral Seminar: Data Analytics Prof. Yuanfeng Cai Fall 2022	
Instructor: Yuanfeng Cai	Office: VC Room 11-231 Phone: (646)-312-3152
E-Mail: Yuanfeng.Cai@baruch.cuny.edu	Office Hours: By appointment
COURSE SYLLABUS	

Course Objectives and Description: This course provides both the foundation and knowledge of analytics research. It examines the research directions and methodologies in the area of data mining and analytics. Students are expected to gain a thorough understanding of both traditional and new data mining methods, learn their applications in various problem domains, and know the challenges in these research areas.

This is a PhD level seminar course. The course is structured around extensive reading and discussion of the research literature on the analytics field. Active in-class and after-class participation is essential. Students are expected to prepare assigned readings carefully and participate in class discussions actively.

Required Readings: Each week, students are expected to read journal articles related to the specific topic. Please see schedule for the full list of articles in each topic.

Grading Policy: Course grades will be based upon the following weights. No late submission is accepted.

Course participation	20%
Dissertation report	30%
Term Paper	50%

Course Participation: Each student is expected to read all papers listed in the weekly readings before class and be ready to discuss all of them. Each student is expected to summarize one paper in a reflection report and present it in class. Other students are expected to be discussants and participate in the critical evaluations.

Dissertation report: Each student is expected to find one recent IS doctoral dissertation paper in data analytics area. The student should analyze the dissertation and present the findings in one written report and one oral presentation. The report is expected to contain: 1) A summary of the dissertation; 2) An analysis of the dissertation; 3) A possible extension of the dissertation.

Term paper: Throughout the semester, each student should develop a research proposal in the data analytics area. Students could choose a research question they are interested, prepare the literature review, write about how they would study that problem and design the research methods. The proposal should also include the discussion about future work and the analysis of potential limitations. Each student is expected to present the proposal and submit a written paper in the last week.

Tentative Weekly Schedule

The following is a tentative plan of the topics and the corresponding readings for each week, subject to revision as opportunities arise.

Week 1 (08/26): Introduction & Data Acquisition

Chen, H., Chiang, R.H.L., & Storey V.C. (2012) Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36 (4), pp. 1165-1188

Shi, Z., Rui, H., & Whinston, A. B. (2014) Content sharing in a social broadcasting environment: Evidence from twitter. *MIS Quarterly*, 38(1), pp. 123–142

Fraginière, E., Gondzio, J., Sarkissian, R., & Vial, J. (2000) A Structure-Exploiting Tool in Algebraic Modeling Languages. *Management Science*, 46(8), pp.1145-1158.

Yin, D., Bond, S., & Zhang, H (2021) Anger in consumer reviews: Unhelpful but persuasive? *MIS Quarterly*, 45(3), pp.1059-1086.

Week 2 (09/09): Classification and regression I

Shmueli, G. & Koppius, O.R. (2013) Predictive analytics in information systems research, *MIS Quarterly*, 35(3), pp. 553-572

Jiang, Z., Mookerjee, V., & Sarkar, S. (2005) Lying on the Web: Implications for Expert Systems Redesign. *Information Systems Research*, 16(2), pp. 131-148.

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003) Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation. *Management Science*, 49(3). pp. 312-329

Zhang, W. & Ram, S., (2020) A Comprehensive Analysis of Triggers and Risk Factors for Asthma Based on Machine Learning and Large Heterogeneous Data Sources. *MIS Quarterly*, 44(1), pp.305-349.

Week 3 (09/16): Classification and regression II

Bapna, R., Goes, P., Wei, K., & Zhang, Z. (2011) A Finite Mixture Logit Model to Segment and Predict Electronic Payments System Adoption. *Information Systems Research*, 22(1). pp. 118-133.

Yang, C.S., Wei, C.P. & Chiang, Y.S. (2014) Exploiting technological indicators for effective technology merger and acquisition (M&A) predictions, *Decision Sciences*, 45(1), pp. 147-174.

Cui, G., Wong, M., & Wan, X. (2012) Cost-Sensitive Learning via Priority Sampling to Improve the Return on Marketing and CRM Investment. *Journal of Management Information Systems*, 29, pp. 341–374.

Atish P.S., & Jerrold H.M. (2004) Evaluating and tuning predictive data mining models using receiver operating characteristic curves. *Journal of Management Information Systems*, 21, pp. 249-280

Week 4 (09/23): Performance evaluation and challenges

Lin, M., Lucas Jr, H., & Shmueli, G. (2013) Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. *Information Systems Research*, 24(4). pp. 906-917.

Yang, M., Adomavicius, G., Burtch, G., & Ren, Y (2018) Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining. *Information Systems Research*, 29(1). pp. 4-24.

Qiao, M., & Huang, K. (2021) Correcting Misclassification Bias in Regression Models with Variables Generated via Data Mining. *Information Systems Research*, 32(2). pp. 462-480

Simester, D., Timoshenko, A., & Zoumpoulis, S. (2020) Targeting Prospective Customers: Robustness of Machine-Learning Methods to Typical Data Challenges. *Management Science* 66(6). pp. 2495-2522

Week 5 (09/30): Unsupervised learning

Wei, C., Chiang, R., & Wu, C. (2006) Accommodating individual preferences in the categorization of documents: a personalized clustering approach. *Journal of Management Information Systems*, 23 (2), pp. 173-201

Bapna, R., Goes, P., Gupta, A., & Jin, Y. (2004) User Heterogeneity and its Impact on Electronic Auction Market Design: An Empirical Exploration. *MIS Quarterly*, 28(1), pp. 21-43.

Ghoshal, A., Menon, S., & Sarkar, S. (2015) Recommendations Using Information from Multiple Association Rules: A Probabilistic Approach. *Information Systems Research* 26(3), pp. 532-551.

Lu, Y., Gupta, A., Ketter, W., & van Heck, E. (2016) Exploring bidder heterogeneity in multichannel sequential B2B auctions. *MIS Quarterly*, 40(3), pp. 645–662

Week 6 (10/07): Network

Li, Z., Fang, X., & Sheng, O. (2017). A Survey of Link Recommendation for Social Networks: Methods, Theoretical Foundations, and Future Research Directions. *ACM transactions on management information systems*. 9(1), Article 1.

Hu, D., Zhao, J., Hua, Z., & Wong, M. (2012) Network-based modeling and analysis of systemic risk in banking systems. *MIS Quarterly*, 36(4), pp. 1269-1291.

Lee, Y., Hosanagar, K., & Tan, Y. (2015) Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. *Management Science* 61(9), pp. 2241-2258.

Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quarterly*, 36(4), pp. 1189–1216.

Week 7 (10/14): Text analytics I

Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation. *ACM transactions on management information systems*. 9(2), 5.

Lash, M., & Zhao, K. (2016). Early predictions of movie success: the who, what, and when of profitability. *Journal of Management Information Systems*, 33, pp. 874-903

Arazy, O., & Woo, C. (2007) Enhancing Information Retrieval Through Statistical Natural Language Processing: A Study of Collocation Indexing, *MIS Quarterly*, 31(3), pp. 525-546.

Wei, C., Hu, P., Tai, C., Huang, C & Yang, C. (2008) Managing Word Mismatch Problems in Information Retrieval: A Topic-Based Query Expansion Approach. *Journal of Management Information Systems*, 24(3), pp. 269-295

Week 8 (10/21): Text analytics II

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems*, 26(3), 12.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), pp. 217–248.

Ravichandran, T., & Deng, C. (2022). Effects of Managerial Response to Negative Reviews on Future Review Valence and Complaints, *Information Systems Research*, Articles in Advance

Lau, R., Liao, S., Wong, K., & Chiu, K. W. (2012). Web 2.0 environmental scanning and adaptive decision support for business mergers and acquisitions. *MIS Quarterly*, 36(4), pp.1239-1268

Week 9 (10/28): Recommender systems and online feedback mechanism

B. Pathak, R. Garfinkel, R. Gopal, R. Venkatesan, & F. Yin, (2010) Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2), pp. 159-188

Hu, N., Pavlou, P., & Zhang, J. (2017) On self-selection biases in online product reviews, *MIS Quarterly*, 41 (2), pp. 449-471

Duan, W., Gu, B., & Whinston, A.B. (2008) Do online reviews matter? – An empirical investigation of panel data. *Decision Support Systems*, 45 (4), pp. 1007-1016

Lu, X., Ba, S., Huang, L., & Feng, Y. (2013) Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Information Systems Research*, 24 (3), pp. 596-612

Week 10 (11/04): Temporal analysis

Li, X., Groot, M., & Bäck, T, (2021) Using forecasting to evaluate the impact of COVID-19 on passenger air transport demand. *Decision Sciences*. Article in Advances.

Gao, Y., Duan, W., & Rui, H (2022) Does Social Media Accelerate Product Recalls? Evidence from the Pharmaceutical Industry. *Information Systems Research*. Article in Advances.

Lee, G & Raghu, T (2014) Determinants of mobile apps' success: Evidence from the app store market, *Journal of Management Information Systems*, 31(2), pp. 133–170.

Yu, X., Liu, Y., Huang X., & An, A. (2012) Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain, *IEEE Transactions on Knowledge and Data Engineering*, 24(4), pp. 720-734.

Week 11 (11/11): Anomaly detection

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62(12), pp. 3393-3672.

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: a meta-learning framework for detecting financial fraud. *MIS Quarterly*, 36(4), pp.1293-1327.

Zhou, L., Burgoon, J., Twitchell, D., Qin, T., & Nunamaker Jr, J. (2004). A comparison of classification methods for predicting deception in computer mediated communication. *Journal of Management Information Systems*, 20(4), pp. 139–166

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7), pp. 1146-1160.

Week 12 (11/18): Artificial intelligence

Jain, H., Padmanabhan, B., Pavlou, P., & Raghu, T. (2021). Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society. *Information Systems Research*, 32(3), pp. 675–687.

Sturm, T., Gerlach, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organizational Learning. *MIS Quarterly*, 45(3), pp. 1581-1602.

Sniezek, J., Wilkins, D., Wadlington, P., & Baumann, M. (2014). Training for Crisis Decision-Making: Psychological Issues and Computer-Based Solutions. *Journal of Management Information Systems*, 18(4), pp. 147–168

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza J. (2021). Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. *Information Systems Research*, 32(3), pp. 713-735.

Weekly 13 (12/02): Dissertation report presentation and feedback

Weekly 14 (12/09): Proposal presentation and feedback