3-21-2021

6th Analytics Without Borders Conference Program

Bentley University

Bryant University

Tufts University

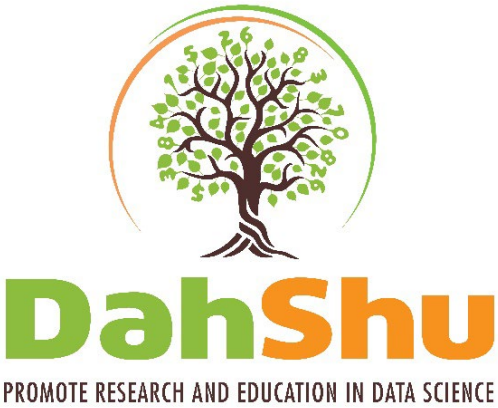
University of Massachusetts Lowell

BENTLEY UNIVERSITY Bryant UNIVERSITY Tufts UNIVERSITY UMASS LOWELL MANNING SCHOOL OF BUSINESS

ANALYTICS WITHOUT BORDERS

2021 Analytics without Borders Conference

Co-sponsored by



2021 Analytics without Borders Conference Program

Hosted by: Center for Analytics and Data Science-Data Analytics Research Team (CADS-DART) of Bentley University

Host Contact: Mingfei Li, Professor of Mathematical Sciences, Bentley University (mli@bentley.edu)

Location: Virtual conference for this year.

(ZOOM link and instruction will be sent to all registered participants by emails)

Date: March 26th (Friday), 2021

Summary of Schedules

The 6th Analytics Without Borders Conference Schedule

Start	End	Main Track	Track 1 (Research talks)	Track 2 (Student research competition talks)
9:20	9:30	Opening announcement and instructions		
9:30	9:45	Welcome speech by Dr. Donna Maria Blancero, Provost of Bentley University		
9:45	10:35	Keynote 1		
10:40	11:30	Career discussion panel for Graduate students	1A (Chair: Clifton Chow)	2A (Chair: Noah Giansiracusa)
11:35	12:25	Major talk	1B (Chair: Noah Giansiracusa)	2B (Chair: Michael McGuirk)
12:25	***13:25***	***Lunch break***		
13:30	14:20	Career discussion panel undergraduate students	1C (Chair: Maria Skaletsky)	2C (Chair: Jiaying Weng)
14:25	15:15	Discussion panel for Data Analytics and Data Science education	1D (Chair: Clifton Chow)	2D (Chair: Briana Trifiro)
15:20	16:10	Keynote 2		
16:15	16:20	Closing by the next host: Dr. Sandra Richtermeyer, Dean of Manning School of Business, UMass Lowell; Dr. Asil Oztekin, Associate Professor, Operations & Information Systems, Manning School of Business, UMass Lowell		

All time are Eastern Time Zone

Main Track

9:20am-9:30am: Opening announcement and instructions

9:30am-9:45am: **Welcoming remarks by Dr. Donna Maria Blancero, Provost of Bentley University,**
9:45am-10:35am: **Keynote 1**

How AI is changing the financial service industry? by Ren Zhang, PhD, Chief Data Scientist, BMO Financial Group

Introducer: Dominique Haughton, Bentley University

10:40am-11:30am: **Graduate Career Discussion Panel (Moderator: Mingfei Li)**
Panelists:

Rebecca Qu, Marketing Analytics & Operations Manager, Tufin
Christopher Wishon, Lead Operations Research Engineer, The MITRE Corporation
Joe Dery, Managing Director of Decision Sciences, Global Business Operations, Dell Technologies
Anthony Chen, Product Analyst, Data Science, Google
Kevin Xie, Senior Data Analyst, Cohere Health

11:35am-12:25pm: **Major Talk**
Essential Data Science and Analytics Areas from Industry Perspectives

by Victor S.Y. Lo, PhD, Senior Vice President of Data Science & Artificial Intelligence, Workplace Investing, Fidelity Investments

12:25pm-1:25pm: Lunch Break

1:30pm-2:20pm: **Undergraduate Career Discussion Panel (Moderator: Mihaela Predescu)**
Panelists:

August G Roesener, Lead Operations Research Engineer, The MITRE Corporation
Joe Dery, Managing Director of Decision Sciences, Global Business Operations, Dell Technologies
Eric Kerstens, Software Development Engineer, Expedia Group

2:25pm-3:15pm: **Discussion Panel on Data analytics, Business Analytics and Data Science Education (Moderator: Mihaela Predescu)**
Panelists:

Elena Naumova, Professor and Chair, Division of the Nutrition Epidemiology and Data Science, Friedman School of Nutrition Science and Policy, Tufts University

Joe Blitzstein, Professor of the Practice in Statistics, Director of Undergraduate Studies

Mingfei Li, Professor of Mathematical Sciences, Bentley University

3:20pm- 4:10pm**: Keynote 2**

Analytics in Action – Overcoming the Customer Knowledge Barrier by Nathaniel Lin, PhD, President, Analytics Consult, LLC

4:10pm- 4:15pm: **Announcement Student Research Competition Award** by the Award Committee

4:15pm- 4:20pm: **Closing Speech by**

Dr. Sandra Richtermeyer, Dean of Manning School of Business, University of Massachusetts Lowell.

Dr. Asil Oztekin, Associate Professor, Operations & Information Systems, Manning School of Business, University of Massachusetts Lowell.

Parallel sessions Track 1 at breakroom 1

10:40am-11:30am: **Track 1A (Chair: Clifton Chow, Bentley University)**

Multicriteria Decision Frontiers for Prescription Anomaly Detection Over Time by Babak Zafari, Babson University

Class Balancing in Customer Segments Classification Using Support Vector Machine Rule Extraction and Ensemble Learning by Suncica Rogic, University of Montenegro, Montenegro

11:30am-12:25pm: **Track 1B (Chair: Noah Giansiracus, Bentley University)**

Blockchain 101 - Understanding the current NFT market, By Kevin Mentzer and Max Paul, Bryant University

Understanding Covid-19 Pandemic: Insights from Social Media Discussion, by Suhong Li, Bryant University

1:30pm-2:20pm **Track 1C (Chair Maria Skaletsky, Bentley University)**

The association of the occurrence of Alzheimer's disease with Atorvastatin and Angiotensin converting enzyme inhibitor by Ying Wang, Bentley University

A Machine Learning Approach to Study the Business Analytics and Data Science Education and Job Market during Pandemic by Rui Zhuo, Bentley University

Do You See What I See? Examining the relationship between political attitudes and the perception of COVID-19 data visualizations by Jonathan Ericson, Bentley University

2:25pm-3:15pm: **Track 1D (Chair Clifton Chow, Bentley University)**

Prolonging Retirement Coverage With The Target Volatility Investment Strategy by Zefeng Bai, Bentley University

The Effects of COVID-19 on Air Quality in Boston, MA by Alicia T. Lamere, Bryant University

Parallel sessions Track 2 at Breakout room 2

10:40- 11:30: **Student research talk session 2A (Session chair: Noah Giansiracusa)**

1. **10:40 – 10:45: Macroeconomy forecasting with dimension reduction and time series modeling** by Taoyanran Sun, Bentley University
2. **10:45 – 10:50: The Impact of Innovation in Tourism on Montenegro's Export Competitiveness** by Sanja Janketic, University of Montenegro
3. **10:50 – 10:55: A Health Analytical Approach to Predict Strokes** by Amar Singh, University of Massachusetts Lowell
4. **10:55 – 11:00: Predictive Analytics Prudential Life Insurance** by Katherine Welch, University of Massachusetts Lowell
5. **11:00 – 11:05: Using Machine Learning Technique to Analyze the Pattern of Customers Interactions on an eCommerce Website to Generate Revenue** by Monita Chea, Meymey Lov, University of Massachusetts Lowell
6. **11:05 – 11:10: Credit Risk for Personal Loans** by Hayden Goldsmith, Justin Varghese, Jacob Barnes, University of Massachusetts Lowell
7. **11:10 – 11:15: A Regression Analysis Study of World Happiness** by Stephen Basler, University of Massachusetts Lowell
8. **11:15 – 11:20: Using Data Analytics to Predict Consumer Credit Card Behavior** by Miguel Espinoza, Cassie Burke, University of Massachusetts Lowell
9. **11:20 – 11:25: A Predictive Analytic Approach to Identify Warning Signs of Corporate Bankruptcy** by Ian Goulet, University of Massachusetts Lowell
10. **11:25 – 11:30: Halloween Candy Power Ranking** by Kayla Rowles and Jeurys Santiago, University of Massachusetts Lowell

11:35- 12:25: **Student research talk session 2B (Session chair: Michael McGuirk)**

1. **11:35 – 11:40: Instrumental vs. Expressive: A Study of Voter Behavior Models Through the Lens of Identity in the 2016 Presidential Election** by Kaitlyn Fales, Bryant University
2. **11:40 – 11:45: Change Detection in Terrorism Events and Other Self Exciting Processes** by Moinak Bhaduri and Anuja Das, Bentley University
3. **11:45 – 11:50: At a Loss for Words: Analysis of Racism and Hate Speech in Soccer Based on Fan's Expressions on Social Media:** Cem Kullukcu, Babak Zafari, Babson College
4. **11:50 – 11:55: Change Point Analysis of Hurricane Activity** by Taoyanran (Alice) Sun and Yunting (Nancy) Sun, Bentley University
5. **11:55 – 12:00: Productizing Analytics Functionality: e-Commerce Platforms Businesses Empower Retailers** by Randy Miller, Babson College
6. **12:00 – 12:05: Effects of Race and Income on COVID Immunization and Death Rates** by Collin Fabian, Bentley University
7. **12:05 – 12:20: How Do We Value Happiness: A Multiple Linear Regression Model Determining the Key Factors of The World Happiness Score** by Nicole Wilson and Salma Dakiri, University of Massachusetts Lowell

1:30 - 3:15: **Student research talk session 2C-2D**

Oral Presentation:

1. **1:30 -- 1:45: Predicting Graduate Admissions** by William Hegarty, Nicholas Walls, Michael Mahlebjian, Uzair Khalid, University of Massachusetts Lowell
2. **1:45 -- 2:00: What is the Key to Happiness? A linear Regression Model and Analysis** by Maddy Roop and Drasti Patel, University of Massachusetts Lowell
3. **2:00 – 2:15:World Happiness Report Regression Analysis** by Mansi Thakkar, Lauren Foley, Isabelle Frost, Jessica Stevens, University of Massachusetts Lowell
4. **2:15 – 2:30: World happiness Report** by Nancy Chac, Nishtha Patel, Harsh Pandey, University of Massachusetts Lowell
5. **2:30 – 2:45: Under Pressure: A Case Study of the Effects of External Pressure on MLB Players using Twitter Sentiment Analysis** by Jonathan Huntley, Bryant University
6. **2:45 – 3:00: Improve the accuracy of Pneumonia Detection from Chest X-ray using Transfer Learning** by Son Nguyen and Nathan Angell, Bryant University
7. **3:00 – 3:15: Predicting Worldwide Box Office Revenue Using Machine Learning Models** by Cam Germaine and Son Nguyen, Bryant University

Presentation Abstracts

Morning Keynote Presentation

How AI is changing the financial service industry?

Speaker: Ren Zhang, PhD, Chief Data Scientist, BMO Financial Group

Abstract:

With a combination of an explosion of data, technological advancements and improved maturity of several promising techniques, artificial intelligence has demonstrated the potential to transform the entire financial services industry. It helps to provide relevant customer experiences, enhanced operational efficiencies, prudent credit risk decisions, proactive fraud detection and optimal price assignments. Though the potential impact is big, the transformation through AI comes with its challenges. It ranges from talent, data, technology, organization awareness, and regulatory support.

Speaker's Bio:

Ren is the Chief Data Scientist for BMO Financial, leading the Enterprise AI CoE, in charge of defining the Enterprise Data Science and Intelligent Process Automation strategy, driving adoption of AI and Machine Learning capabilities to automate processes and deliver better predictive decisions that will help the businesses drive accelerated revenue, cost productivity and customer outcomes.

Ren has over 15 years' experience as a senior AI leader within various financial organizations. Prior to BMO Financial, she worked at Prudential Financial, where she was Vice-President and Head of Data Science. Prior to Prudential, Ren was Executive Director of Data Science and Innovation at Commonwealth Bank of Australia. And at American Express, she held progressively senior leader roles, ranging from credit risk management, to loyalty analytics, to fraud risk strategy, and to risk capabilities, with her last being Vice-President, Risk and Information Management of Enterprise Growth. Ren holds a PhD in Statistics from The Wharton School at the University of Pennsylvania.

Afternoon Keynote Presentation:

Analytics in Action – Overcoming the Customer Knowledge Barrier

Speaker: Nathaniel Lin, PhD, President, Analytics Consult, LLC

Abstract:
As consumers purchasing more online, the challenges of knowing the customers are becoming ever more taunting. This is especially true with the reliance on intermediaries such as Google search engines and social media advertising platforms such as Facebook and Instagram, in reaching their customers. As a result, the intimate customer knowledge the business owners require to better serve their customers are increasingly fragmented and often beyond their reach. This talk describing an attempt to use analytics to bridge the customer knowledge gap and thus allowing the business owners to leverage such intimate customer knowledge to more cost effectively serve their customers!

Speaker's Bio:

Nathaniel Lin is an experienced leader in marketing and business analytics across various industries worldwide and has over 20 years of front-line leadership experience applying actionable advanced analytics strategies to the world's largest companies in technology, finance, automotive, telecommunications, retail, and many other businesses. These include IBM, Fidelity Investments, OgilvyOne, and Aspen (now Epsilon) Marketing Analytics. He is currently the President of Analytics Consult and was the Chief Data Scientist at the NFPA. Dr. Lin has been leading efforts to develop leading edge Big Data Analytics technology and knowledge assets to deliver unparalleled values to Chinese consumers and U.S. clients. Nathaniel enjoys being a teacher, author, and hands-on practitioner and business executive in the application of advanced data analytics in a wide variety of businesses. He is a frequently invited speaker in analytics events and had advised over 150 CEOs in the U.S. and abroad on analytics and Big Data issues. He wase also invited by WWW2010 as one of the four expert panelists (together with the heads of Google Analytics, eBay Analytics, and Web Analytics Association) on the Future of Predictive Analytics. As a recognized analytics expert, he has partnered with Professor Tom Davenport to benchmark analytics competencies of major corporations across different industries.

To fulfill his passion in developing future analytics leaders, he has taught Strategic CRM and Advanced Business Analytics to MBA and PhD students at the Georgia Tech College of Management, Boston College Carroll School of Management, Bentley U, Northeastern University and Hult International Business School. Nathaniel has a PhD in Engineering from

Birmingham University (UK) and an MBA from MIT Sloan School of Management. He has also published over 10 peer-reviewed papers in leading academic journals and the author of "Applied Business Analytics" published by the Financial Times/Pearson Press in 2015 focusing in bridging the gap between analytics and business insights and strategies. The book has been translated and published in Chinese in 2016.

Major Talk

Essential Data Science and Analytics Areas from Industry Perspectives
Speaker: Victor S.Y. Lo, PhD, Senior Vice President of Data Science & Artificial Intelligence, Workplace Investing, Fidelity Investments

Abstract

Data science and analytics have become more popular than ever in this age of Big Data and technological development. The academic literature covers many important analytics areas from the research perspectives. In this talk, I will focus on the most essential data science and analytics areas from the industry points of view. These practical areas are important to meet today's business and industry demand and key to career development for students and professionals. The discussion can also be helpful for researchers interested in industry applications.

Speaker's Bio:

Victor S.Y. Lo is a seasoned Big Data, Marketing, Risk, and Finance leader with over 26 years of extensive consulting and corporate experience employing data-driven solutions in a wide variety of business areas, including Customer Relationship Management, Market Research, Advertising Strategy, Risk Management, Financial Econometrics, Insurance Analytics, Product Development, Healthcare Analytics, Operations Management, Transportation, and Human Resources. He is actively engaged with causal inference and is a pioneer of Uplift/True-lift modeling, a key subfield of data science.

Victor has managed teams of quantitative analysts in multiple organizations. He is currently SVP of Data Science & AI, Workplace Investing at Fidelity Investments. Previously he managed advanced analytics/data science teams in Personal Investing, Corporate Treasury, Managerial

Finance, and Healthcare and Total Well-being at Fidelity Investments. Prior to Fidelity, he was VP and Manager of Modeling and Analysis at FleetBoston Financial (now Bank of America), and Senior Associate at Mercer Management Consulting (now Oliver Wyman).

For academic services, Victor has been a visiting research fellow and corporate executive-in-residence at Bentley University. He is also an elected board member of the National Institute of Statistical Sciences (NISS), where he provides guidance to the board and general education to the statistics community. Additionally, he has been serving on the steering committee of the Boston Chapter of the Institute for Operations Research and the Management Sciences (INFORMS) and on the editorial board for two academic journals. Victor earned a master's degree in Operational Research and a PhD in Statistics and was a Postdoctoral Fellow in Management Science. He has co-authored a graduate-level econometrics book and published numerous articles in Data Mining, Marketing, Statistics, and Management Science literature, and is co-authoring a graduate-level data science textbook titled "Cause-and-Effect Business Analytics."

Prolonging Retirement Coverage With The Target Volatility Investment Strategy

Author(s) Zefeng Bai, PhD Candidate of Business Analytics, Bentley University

Abstract
Motivated by the recent market turbulence triggered by the COVID-19 pandemic, we use the popular target volatility investment strategy as an extra layer of protection to several conventional retirement portfolios in the pension decumulation context. We use the Monte Carlo simulation approach to evaluate the performance of retirement portfolios under different stochastic environment. Over a thirty-year hypothetical retirement span, we find that the retirement portfolio attached with the target volatility mechanism shows significantly higher sustainability compared to the conventional retirement portfolio. We conclude that the target investment strategy could be a suitable option for investors looking for a higher level of stability in retirement coverage after retired. Our results are robust with a consideration of transaction cost.

Understanding Covid-19 Pandemic: Insights from Social Media Discussion

Author(s) Suhong Li, Professor of Information Systems and Analytics, Bryant University

Abstract
World Health Organization announced the Covid-19 outbreak as a pandemic on March 11 2020. Starting from Wuhan China, this virus has infected and killed millions of people all around the world. The purpose of this study is to understand the Covid-19 pandemic through discussion on social media platforms. This study is based on more than 720 million tweets from over 49 million users who tweeted about this topic between March and December, 2020. The results show that Covid-19 topics have been widely discussed in various languages and countries. The top 5 tweeted languages are English, Spanish, Portuguese, Indonesian and French. The top 5 countries based on the number of users tweeting about Covid-19 are the US, the UK, India, Brazil and Spain. Based on hashtags, Covid-19 discussion can be classified into four topics: prevention measures, news media, location, and policy/impact. The results showed that tweets

with the prevention measure hashtags such as #socialdistancing, #stayhome, #wearamask and #vaccine received overall positive rating, while all hashtags in the news media category (#foxnews, #smartnews #fakenews, #breaking, and #news) received negative rating, reflecting people's pessimistic reaction to Covid-19 news. In addition, the findings of this study show that the months or countries with more cases/deaths are associated with a higher number of tweets and a more negative sentiment. These findings may indicate the number of cases/deaths are positively related to the popularity of Covid-19 topics, and negatively related to the overall sentiment rating of these topics.

Keywords: Covid-19, Sentiment Analysis, Twitter Analytics

The Effects of COVID-19 on Air Quality in Boston, MA

Author(s) Christopher Ethier, Alicia T. Lamere*, Dan McNally
Speaker's Affiliation: Bryant University

Abstract
During the year 2020, COVID-19 spread throughout the globe, drastically impacting our daily behavior. With societies under lockdown, many places of work shut down as well, potentially reducing levels of pollution. This study attempts to determine whether the lockdown of the United States of America in the spring of 2020 caused a measurable reduction in the pollution levels in the areas surrounding Boston, Massachusetts. Several air quality metrics and pollutant levels, along with mobility data for residents, were examined. Not only were significant reductions in pollutant levels observed in 2020 compared to previous years, but strong correlations were also observed between pollutant levels and mobility data, confirming the relationship between changes in resident behavior and air quality.

Keywords: covid-19, correlation analysis, pollution, association

Multicriteria Decision Frontiers for Prescription Anomaly Detection Over Time

Author(s) Babak Zafari* , Tahir Ekin, Fabrizio Ruggeri,
Speaker's Affiliation: Babson University

Abstract
Health care prescription fraud and abuse result in major financial losses and adverse health effects. The growing budget deficits of health insurance programs and recent opioid crisis in United States have accelerated the use of health care fraud analytical methods to address opioid abuse and fraud. Unsupervised methods such as clustering and anomaly detection could help the health care auditors evaluate the billing patterns when embedded into rule-based frameworks. These decision models can aid policy makers in proactively detecting potential suspicious activities. This manuscript proposes an unsupervised temporal learning based decision frontier model using the real world Medicare Part D prescription data collected over five years. First, temporal proba-bilistic groups of drugs are retrieved using a structural topic model with covariates. Next, we construct combined concentration curves, Gini measures, and opioid scores considering the weighted impact of temporal observations for prescription patterns, in addition to

the Gini values for the cost. The novel decision frontier utilizes this output and enables health care practitioners to assess the trade-offs among different criteria and identify audit leads to detect abusive prescribers.

Keywords: multivariate anomaly detection; Medicare Part D; topic model; prescription patterns

Blockchain 101 - Understanding the current NFT market

Author(s) Kevin Mentzer and Max Paul, Associate Professor of Information System and Analytics, Bryant University

Abstract
In this session we will explore the current non-fungible token market which has gained significant interest in the past few months with crypto-assets selling for upwards of $6.6M and the trading card market being disrupted with new entrant Dapper Labs with their collaboration with the National Basketball Association and the product NBA Topshot. We'll cover the crypto-asset market in general including cryptocurrencies, tokens, and NFTs and discuss how these assets are gaining prominence in the various financial markets.

Keywords: Blockchain NFT Crypto-assets DeFi

Class Balancing in Customer Segments Classification Using Support Vector Machine Rule Extraction and Ensemble Learning

Author(s) Suncica Rogic* Ljiljana Kascelan
Speaker's Affiliation: University of Montenegro, Faculty of Economics

Abstract
An objective and data-based market segmentation is a precondition for efficient targeting in direct marketing campaigns. The role of customer segments classification in direct marketing is to predict the segment of most valuable customers who is likely to respond to a campaign based on previous purchasing behavior. A good-performing predictive model can significantly increase revenue, but also, reduce unnecessary marketing campaign costs. As this segment of customers is generally the smallest, most classification methods lead to misclassification of the minor class. To overcome this problem, this paper proposes a class balancing approach based on Support Vector Machine-Rule Extraction (SVM-RE) and ensemble learning. Additionally, this approach allows for rule extraction, which can describe and explain different customer segments. Using a customer base from a company's direct marketing campaigns, the proposed approach is compared to other data balancing methods in terms of overall prediction accuracy, recall and precision for the minor class, as well as profitability of the campaign. It was found that the method performs better than other compared class balancing methods in terms of all mentioned criteria. Finally, the results confirm the superiority of the ensemble SVM method as a preprocessor, which effectively balances data in the process of customer segments classification.

Keywords: direct marketing, customer classification, class imbalance, SVM-Rule Extraction, ensemble

The association of the occurrence of Alzheimer's disease with Atorvastatin and Angiotensin converting enzyme inhibitor

Author(s) Ying Wang*, Mingfei Li, Lewis E Kazis, Weiming Xia,
Speaker's Affiliation: Bentley University

Abstract
Variables accelerating the onset of Alzheimer's disease (AD) during the pre-symptomatic stage are not well established. Retrospective studies have revealed the beneficial effects of long-term use of angiotensin converting enzyme inhibitor (ACEI) and statin medications on delaying occurrence clinical manifestation of AD. However, controversial results exist for individual drugs under each medication class. The presented study compared the risks of developing AD among users under three individual drugs (Lisinopril, Atorvastatin, and Simvastatin). We utilized Cox proportional hazard models with propensity score balancing to calculate the hazard ratios (HR) of taking individual medications on the occurrence of dementia with AD, based on the International Statistical Classification of Diseases (ICD) 9 or 10 codes. Patients with Atorvastatin (adjusted HR: 0.697, CI (0.514, 0.944)) or Lisinopril (adjusted HR: 0.928, CI (0.884 0.973)) were significantly less associated with AD onset and enjoyed a prolonged pre-symptomatic survival time to AD onset, compared to those with Simvastatin. In terms of age at target medication initiation, Lisinopril users were associated with less risk of AD onset among patients who start medication at an early age (≤ 65 yr); Atorvastatin users were associated with less risk of AD onset among patients who start medication at a late age (> 65 yr).
Keywords Alzheimer's disease, Anti Hypertension drugs.

A Machine Learning Approach to Study the Business Analytics and Data Science Education and Job Market during Pandemic

Author(s) Mingfei Li. Rui Zhuo*, Heyao Wang
Speaker's Affiliation: Bentley University

The recent technology revolution in data and AI motivates data related education in higher education, including in business schools. Many analytics or data science programs were established in recent 5 years and many schools are considering opening their own analytics or data science majors/minors in the near future. However, with the outbreak of the pandemic, the job market of data scientists/data analyst/business analyst related positions is undergoing a big change in America. This on-going study collected all the data analytics, business analytics and data science programs for both undergrad and graduate levels and job ads across all states in May 2020 from Indeed and aimed to answer the questions of what skills are the mostly demanded in certain area for certain industry and if the programs match local market better than the national market for certain industry.

Student Lightning Talks+e-posters Session:

Macroeconomy forecasting with dimension reduction and time series modeling
Author(s) Taoyanran Sun
Speaker's Affiliation: Bentley University
Abstract
How to regulate an economy's growth and stability through different economic components is a major study topic among macroeconomists. To make better predictions and investigate the connections between high-dimensional covariates, in this research, two latest dimension reduction algorithms and time series forecasting were applied to the FRED dataset. The method was also tested by the simulation process.
Keywords Dimension Reduction; Time Series

The Impact of Innovation in Tourism on Montenegro's Export Competitiveness
Author(s) Sanja Janketic
Speaker's Affiliation: University of Montenegro
Abstract
The author decided to study The Impact of Innovation in Tourism on Montenegro's Export Competitiveness, because tourism makes up as much as a third of the Montenegrin economy. In that sense, it is necessary to constantly innovate in this sector, which is so dependent on all external influences, and an example of that is the situation caused by the pandemic of virus COVID-19. Since Montenegro is in the transition process for a long time now, the concept and the process of innovation is something that needs to be clearly understood in both in the scientific and economic sectors.
In order to master the concept of innovation, the author gave numerous examples of innovation in the European Union, as well as some of the innovation in tourism in Montenegro. Slovenia's experience in innovation in tourism was significant, especially when it comes to the use of social networks in tourism. The author compared numerous financial indicators of Montenegro, countries in its region, as well as countries that are at the top of the European innovation list, such as Sweden and Denmark.
After listing all the adopted strategies in the field of innovation and tourism in Europe and Montenegro, the author prepared a questionnaire for her master's thesis, based on the CIS questionnaires for measuring innovation, adapted to the Montenegrin tourism sector.
Based on the answers from the Questionnaire, given by hotels, tourist organizations and travel agencies in Montenegro, through the IBM SPSS program (T-test of independent variables and ANOVA test), tabular and graphical presentation of research results was made, which presented the situation in Montenegrin tourism sector in terms of innovation and in terms of the use of social networks. The questionnaire contained as many as 80 questions, and the answers were presented graphically and tabularly throughout the paper.
Keywords Innovation, Innovation in tourism in Montenegro, Innovation in tourism in European Union, Sustainable tourism, Competitiveness, Social networks in tourism

A Health Analytical Approach to Predict Strokes
Author(s) Amar Singh
Speaker's Affiliation: University of Massachusetts Lowell
Abstract

Objective: To use characteristics of patients as variables in data mining approaches to determine stroke prediction. According to the World Health Organization (WHO), strokes are the second leading cause of death globally, responsible for 11% of total deaths. As such, it is essential for healthcare facilities to understand patterns within stroke patients to provide better care to patients and preventing death. This paper uses data analytics to analyze such patterns in patients who have and have not experienced a stroke to determine those at higher. The dataset used in this study predicts whether a patient is likely to get a stroke based on input parameters such as gender, age, various diseases, and smoking status. Three different methods were used to perform the data analytic approach: decision tree classifier, logistic regression, and K neighbor classifier. A subset of the original train data is taken using the filtering method for machine learning and data visualization purposes. Training and testing datasets were established using both random 66%-33% split and 10-fold cross-validation.
Keywords data mining; variables; parameters; predictive analytics; modeling; classification; decision tree; machine learning

Predictive Analytics Prudential Life Insurance
Author(s) Kimberly Young, Andrew Ackerman, Jenkin Pang, Katherine Welch*, Muskan Kaw
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
The goal is to develop a predictive model that accurately classifies the 8 risk categories of an application.
The results will help Prudential better understand the predictive power of the data points in the existing assessment, enabling them to significantly streamline the process.
Keywords Life Insurance, Predictive Analytics

Using Machine Learning Technique to Analyze the Pattern of Customers Interactions on an eCommerce Website to Generate Revenue
Author(s) Monita Chea, Meymey Lov
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
The e-Commerce business is playing an essential role in the retail industry. The growth of this industry benefits all businesses, especially e-commerce companies. Due to the change in shoppers' behavior and the advancement of technology, more and more retail shoppers hop online to search for products to fulfill their needs. In this age of technology, data analysis is leveraged by many of these companies to make sure that their marketing campaigns and policies are built to target the right customer and generate sales revenue. For this project, an open-sourced dataset on online shoppers purchasing intention is used to build different analytic models and test the accuracy rate.
Keywords Data analytics, online shopper, customer behavior, machine learning algorithms

Credit Risk for Personal Loans
Author(s) Hayden Goldsmith, Justin Varghese, Jacob Barnes
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
Banking is a competitive market and banks need to have the competitive advantage over their competitors. If banks are able to mitigate losses by assessing potential credit risk with people

who apply for loans, they would be able to reduce costs. The goal of this study is to use machine learning and data mining software to analyze the credit worthiness of borrowers. The dataset being utilized is from Credit One Bank, based in Las Vegas, Nevada. Attributes such as checking and savings account balances, years at employment, and number of dependents will be used to determine whether or not an individual will default on a loan. Further, we will utilize the purpose of credit taken, property, and job commitment to assess credit risk. The proposed models that will be used are decision trees, and Support Vector Machines (SVM). 10-fold cross validation will be used to test the dataset. By using Weka, the models will be generated and tested to determine which model would yield the highest quality results.
Keywords Credit Risk, Credit Worthiness, Data Mining, Decision Trees, Weka

A Regression Analysis Study of World Happiness
Author(s) Stephen Basler*, Mikayla Corbeil, Abby McLaughlin
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
We ran a regression model with data from a world happiness report based on 2019. The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

Using Data Analytics to Predict Consumer Credit Card Behavior
Author(s) Britney Danh, Cassie Burke*, Miguel Espinoza*, Molly Houlihan, Ruby Lin
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
The usage of credit cards increases more and more every year, and is now the most popular purchasing method in the United States. As more people possess one or more cards, it becomes increasingly important to understand why consumers make decisions. Consumer decisions on their credit cards can drastically affect their finances and future. Using predictive analytics, we hope to find information on what decisions certain types of people will make in their finances, and why. We want to analyze how different people with different educational backgrounds, marital statuses, and incomes view credit cards and how they go about making decisions. We also hope to understand why consumers would decide to close their cards. Not everyone is fully informed about how credit cards work and are fearful of the financial struggles they can inflict. This information can lead to smarter decisions being made by consumers and bankers. We hope to gather a greater understanding of the influences, such as income, marital status, and education level, and how they impact the financial decisions people make.

A Predictive Analytic Approach to Identify Warning Signs of Corporate Bankruptcy
Author(s) Ian Goulet*, Kevin Manning, Peter Christakis, Thomas Condon
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
When looking at corporate bankruptcy, profits and cash flows are the factors that are most often discussed and looked at as reasons why a company went bankrupt. This report aims to identify

the underlying causes for companies declaring bankruptcy beyond profits and cash flows. This will give more detailed insight into the factors affecting corporate bankruptcy and could highlight early identifiers of bankruptcy. These early warning signs could be used to shape where investors choose to invest and give companies the ability to react to potential concerns in a timelier manner. Artificial neural networks were used to identify the relationship between key business metrics and the success or failure of the specific company. Training and testing datasets were established with 10-fold cross-validation to limit any potential bias in the datasets as well as increase the accuracy of the results from testing.

Halloween Candy Power Ranking
Author(s) Kayla Rowles and Jeurys Santiago
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
Analyzing which halloween candy is the best and highest ranking based on different attributes of the candies.

Instrumental vs. Expressive: A Study of Voter Behavior Models Through the Lens of Identity in the 2016 Presidential Election
Author(s) Kaitlyn Fales
Speaker's Affiliation: Bryant University
Abstract
Studying voter behavior through the lens of identity is central to making sense of the 2016 presidential election. The traditional models for explaining voter behavior are rational choice and behavioralism. The former is grounded in instrumental partisanship and a voter's issue positions, with the latter grounded in an expressive, psychological attachment to partisanship. More recent, social identity theory related models discuss voter behavior through group belonging and the partisan mega-identity (Mason 2018). My analysis used the ANES 2016 Time Series Study. To measure a voter's issue positions, I created a new Identity Index alongside the expansion of an established Issue Index. To measure the more expressive aspects of voter behavior, the demographic identities of race, age, class, and gender were used, as well as feeling thermometer measures. My analysis showed that the Identity Index was a better proxy for determining voter behavior in 2016 than the Issue Index. The demographic variables, except for gender, were also significant in both models tested. Through the feeling thermometers, certain groups could be identified as having more influence in voter behavior than others. MANOVA analysis demonstrated that the feeling thermometer scores also vary based on different demographic interactions. The findings of this research demonstrate that the consideration of identity in the rational choice and behavioralism models is crucial to explain voter behavior in 2016.
Keywords: voter behavior, identity, rational choice, behavioralism, 2016 election.

Change Detection in Terrorism Events and Other Self Exciting Processes
Author(s) Anuja Das*, Moinak Bhaduri
Speaker's Affiliation: Bentley University
Abstract
For decades, one of the most critical threats to global security has been domestic and international terrorism. It is a threat that, far from subsiding, seems to be growing at a terrifying rate in both quantity and violence. The increasingly socio- and ethno-political nature of terror

factions lends itself to the progressively complex nature of terror attacks that seemingly strike when you are at your most vulnerable. But are these destructive events truly unpredictable?

Change point detection is a field that already contains a robust volume of literature and research. However, the detection of change points within self-exciting events remains relatively unexplored. Our research tackles this niche and develops a new class of statistics within it. Self-exciting events, like terrorism, have multiple layers of intensity owing to the increased probability of retaliation that stems from a single attack. We identify changes in the underlying intensities of these point processes by reversing the flow of time. By implementing our statistics in various simulations, we have demonstrated a marked increase in accuracy of detection over those of our competitors and are able to identify potential change points within the immigrant and offspring regimes without requiring overly complex models. We take the study a step further and identify similarities among terror categories by exploring change point proximities through the Hausdorff metric. Bootstrap interval estimates are also provided.

At a Loss for Words: Analysis of Racism and Hate Speech in Soccer Based on Fan's Expressions on Social Media

Author(s) Cem Kullukcu*, Babak Zafari
Speaker's Affiliation: Babson College

Abstract
Discrimination, especially in the form of racism, is a very prominent topic in modern-day soccer. Along with the advancement of technology, it has recently become easier for soccer fans to express their views through a plethora of social platforms, including, but not limited to, social media, fan boards, and discussion websites. As a result, technology has added a new dimension through which racism can be transmitted. Although anti-racism soccer organisations have treatment programs to combat such ambiguity and racism online, there is a widespread belief that these efforts are not sufficient. Data analytics methods such as natural language processing models could help both the soccer associations and social media firms identify the level of online abuse and provide guidance to the most vulnerable groups. Therefore, given the contemporary implications of online racist abuse, this paper examines the terminology and diction used in soccer-related tweets during the 2018/19 English Premier League (the highest tier of soccer in the United Kingdom) season. Specifically, the paper explores whether any difference exists within tweets aimed at players who are black, Asian, and minority ethnic (BAME), and those who are white. First, the tweets aimed at the top-40 most followed BAME and white players during the 2018/19 EPL season are extracted, and sentiment analysis is performed to see if the sentiments expressed by English football fans differ towards players from different backgrounds – with particular attention to differences in racial abuse and hate speech. Next, key themes are identified using a structural topic model with racial ethnicity defined as the covariate. This paper then uses the output from the sentiment analysis and structural topic model to test the initial hypothesis and present findings that will provide the reader with a stronger understanding of online abuse in soccer.

Change Point Analysis of Hurricane Activity

Author(s) Taoyanran(Alice) Sun and Yunting(Nancy) Sun

Speaker's Affiliation: Bentley University
Abstract
When a hurricane strikes a coastal area, it brings numerous hazards that seriously affect people's lives. Regrettably, recent decades are witnessing an increased frequency of strong hurricanes. Established research validates the threatening rate and offers possible speculations regarding the cause. A void exists, however, in locating near-exact points in time around which the frequencies jump significantly. We popularize and examine the applicability of a recent sequential testing approach to pin-point locations of such structural breaks. Deviations from a null assumption of stationarity (in favor of a self-exciting class) will be checked through a novel statistic constructed through time-reversal. We will demonstrate reliable classification and estimation power, in addition to the controlling of false alarms, through extensive simulations, and will characterize intensities under which our proposal outperforms its traditional competitors. Brownian-bridge-based goodness-of-fit tests will confirm model justifiability. Clustering of oceanic basins through metrics such as the Hausdorff will offer crucial insights on how the North Atlantic's evolution closely mimics the South Indian's or the South Pacific's. We explain how change-point-induced partitions can be exploited to offer forecasts such as seventeen probable hurricanes and tropical storms in the North Atlantic or thirty in the West Pacific next year.
Keywords Hurricane; Change Point; Point Process; Forecasting; Cluster.

Productizing Analytics Functionality: e-Commerce Platforms Businesses Empower Retailers
Author(s) Randy Miller
Speaker's Affiliation: Babson College
Abstract
While the long-term behavioral changes catalyzed by the COVID-19 public health pandemic are difficult to predict, macroeconomic retail industry trends are intensifying, and more stores are being forced to reinvent their operating models to integrate online distribution. This study will review the various analytics functionalities that online retail platforms like Amazon, Shopify, eBay and others provide to sellers to support their online growth.
By benchmarking the various functionalities that e-Commerce stores provide to sellers to build their stores, this presentation seeks to understand the modern online retailer. In reviewing user interfaces, revenue models, and tools we aim to determine best practices for selling products and services in the online environment. Specific attention will be paid to the analytics tools which provide sellers insight into website visitors, visualizations of their online store traffic, and analysis of store's transactions.

Effects of Race and Income on COVID Immunization and Death Rates
Author(s) Collin Fabian
Speaker's Affiliation: Bentley University
Abstract
In this analysis I will be looking at the potential differences in COVID immunization and death rates between race and income in the United States. For data regarding COVID immunizations and deaths, I will be using the CDC database; specifically a dataset that tracks total COVID deaths and immunizations by county, as well as the proportions of deaths by race. For income data, I will gather median household income data, by county, from the National Census database.

My overall goal is to identify if certain racial groups or income classes within the United States receive adverse treatment for COVID compared to others.
Keywords COVID-19, COVID, racial inequality, income inequality

Student Oral Presentation Session:

How Do We Value Happiness: A Multiple Linear Regression Model Determining the Key Factors of The World Happiness Score
Author(s) Nicole Wilson and Salma Dakiri
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
The World Happiness Report is released every year on Kaggle.com, it consists of data taken from surveys around the world. Data is taken from 155 countries around the world, each survey question asks people to rank how much each factor such as GDP, social support, healthy life expectancy, freedom to make decisions, generosity, and perception of control, affect their happiness. The scale used is known as the Cantril ladder, which asks people to rank each factor's role in their life visualizing a 0-10 scale as a ladder, with 10 being the best and 0 the worst Results are then averaged to get a ranking and score, the highest score determines the happiest countries, and the lowest determines the unhappiest.
This project will specifically find what factors have been the most influential to the world happiness score from 2015 to 2020. Using regression analysis we will analyze the data provided and remove x variables until we can draw a conclusion about the most influential factors that contribute to happiness. This will ultimately lead to obtaining and confirming the top most influential factors to the happiness score, and ones that could be eliminated from the model as they are not significant.
Keywords World Happiness, Regression, Variable, Correlation

Predicting Graduate Admissions
Author(s) William Hegarty, Nicholas Walls, Michael Mahlebjian, Uzair Khalid
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
Our project that we are researching highlights multiple variables that are needed in order to get accepted into certain graduate programs. Our data set contains several parameters that are considered important throughout the application process for masters programs. The data set uses UCLA's guidelines to accept applicants into their graduate programs.
Keywords Graduate admissions, multiple regression, predictive analysis

What is the Key to Happiness? A linear Regression Model and Analysis
Author(s) Maddy Roop and Drasti Patel
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
We are using a dataset from Kaggle.com on the World Happiness Report. We are using a multiple linear regression model to find the key factors that effect happiness worldwide. The variables include GDP per capita, health, and freedom, among others.

World Happiness Report Regression Analysis

Author(s) Mansi Thakkar, Lauren Foley, Isabelle Frost, Jessica Stevens
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
Multiple Linear regression is a statistical tool used to calculate the value of a dependent variable from independent variables. In this model we use multiple regression to understand the relationship between world happiness by country, based on measurements of GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. World happiness rank is the dependent variable, or output. This means it depends on the independent variables, or inputs, of GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity and perceptions of corruption. By using multiple linear regression, we determine an equation for the data, which shows the strength of the relationship between the inputs and happiness as well as how much variability in happiness is explained by the equation. Through this model we hope to understand how accurately countries can be ranked by happiness according to these variable inputs, and how these methods may be altered to provide a more comprehensive and accurate depiction of world happiness and how it can be better measured.
Keywords Regression, happiness, variability

World happiness Report
Author(s) Nancy Chac, Nishtha Patel, Harsh Pandey
Speaker's Affiliation: University of Massachusetts Lowell
Abstract
The World Happiness Data is a dataset surveying global happiness, ranking all countries on their level of happiness and it is considered a benchmark for measuring global happiness. This dataset was retrieved from Kaggle.com, but the actual data are used from the Gallup World Poll. First published in 2012, this is its 2019 version, consisting of 156 countries. There are seven variables including: GDP per capita, Social support, Healthy life expectancy, freedom to make choices, generosity and perceptions of corruption. The dataset from Kaggle.com consists of different reports starting from 2015 up to 2019. The countries are measured by numerical scores, GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. Based on our research, we found that governmental policies and corruption are great factors towards people's happiness. One example is how the governments in Nordic countries provide social welfare and other benefits to its citizens. According to the World Happiness Report 2020, citizens are satisfied with the governments' policies which created a positive affect on how they live. However, the opposite occurred to residents from Latin America and Mexico due to experiencing inequality and corruption within the government. Based on the World Happiness Report 2019, citizens of these countries were fed up with their political leaders and poor policies given. We also found that East Asian countries perform well economically while poorly in happiness. Also the effect of income on happiness is the lowest in Thailand and Philippines while the highest is in South Korea and Taiwan.
Keywords Regression, happiness, predictive

Under Pressure: A Case Study of the Effects of External Pressure on MLB Players using Twitter Sentiment Analysis
Author(s) Jonathan Huntley
Speaker's Affiliation: Bryant University

Abstract
Performance under pressure and psychological momentum are well-documented topics in sports psychology, but most research focuses on “in-game” pressure. This study views pressure more broadly to examine how the external pressure of fans, quantified using the sentiment of tweets mentioning the players, can affect how MLB players perform.
Although external pressure is intangible, it can impact a player’s psyche and performance. This investigation focuses on players Chris Sale and David Price. A new process was developed leveraging the Vader package in Python that can generate tweet sentiment to compare to several performance metrics from Baseball Reference.
Results proved to be promising with correlation analysis pointing to some association between sentiment and performance. There was also an observed difference in how both players handled the pressure depending on whether they played for a small or large market team. An anecdotal study of the 2018 season showed even more interesting differences between Sale’s and Price’s performance and Twitter sentiment. Price’s performance and Twitter’s sentiment moved in a cyclical manner throughout the season whereas Sale’s results were much more consistent and less sensitive to change. Finally, a study focused on the impact of both pressure and past performance on future outings showed results consistent with past studies on the subject. For example, Sale was most likely to perform well under pressure if he preceded the start with a very good or bad outing rather than an average outing. Information like this could be useful for front offices and managers.
More analysis should be conducted to confirm and expand on the findings of this project. However, this case study can be used as a foundation for a new and innovative approach to player evaluation, ultimately complementing existing methods and informing decisions regarding otherwise intangible factors.
Keywords Baseball, Sports, Twitter, Psychology

Improve the accuracy of Pneumonia Detection from Chest X-ray using Transfer Learning.
Author(s) Son Nguyen and Nathan Angell*
Speaker’s Affiliation: Bryant University
Abstract
Over the past few years, deep neural networks and transfer learning have developed into powerful tools to solve computer vision tasks such as image recognition and object detection. In this work, we will study the benefits of using this transfer learning tool to detect Pneumonia from our chest X-ray dataset. This work will emphasize the power and convenience of transfer learning, for the goal of improving the accuracy of deep learning models. We will also study the effectiveness of the learning transferred from well-known pre-trained models such as MobileNet, Inception, Resnet, and EfficientNet in our Pneumonia detection chest -X-ray dataset.
Keywords Transfer Learning, MobileNet, deep neural networks

Predicting Worldwide Box Office Revenue Using Machine Learning Models
Author(s) Cam Germaine* and Son Nguyen
Speaker’s Affiliation: Bryant University
Abstract
Predicting Box Office Revenue is extremely important in the industry in selecting cast and crew, allocating the correct budget, and identifying variables that will maximize profit for movie

creators. Our work will study multiple machine learning models to predict revenue, along with identify important features. These models include decision tree, random forest, gradient boosting, support vector machine and neural networks. The performance of these models will be determined with a variety of evaluation metrics including mean squared errors and R-squared. This study will have an increasingly important role in the movie industry as the box office looks to rebound from the COVID-19 pandemic in the most efficient way possible.