

SILVER STATE-ISTICS



Nevada Chapter of American Statistical Association

Nevada Chapter Newsletter

Vol. 17 No. 5

Dear Nevada ASA Members and Friends,

Read on for information on what our Chapter has been up to.
Please keep an eye on our website for more details on upcoming events.

https://community.amstat.org/nevadachapter/home

Contents:

Fall Virtual Event	1
Elections	1
The President's Corner	1
Data Science Education Opportunities in	
Nevada	2
Invited Talks	2
Tutorial on Data Pipelines	4
Research Presentations	4
Position at Touro University	5
UNLV MS in Data Analytics	5
Joining NV-ASA!	5
NV-ASA Officers & Others	5

* * *

🗷 🛮 Fall Virtual Event 🗷 📑

Our **Annual Meeting** was held on-line October 3, at the end of the virtual *State of Data Science* event. Most of this Newsletter is devoted to discussions of the event. ≥

Elections

Election results for terms beginning January 1, 2021 were announced during our Annual Meeting. Chris Tong was elected to fill the remaining year of a two-year term as President, and Charles Davis was elected to fill the remaining year of a two-year term as Secretary. Sean Breckling was elected to his first term as Southern Vice President, and Alejandra Livingston was elected to another two-year term as Treasurer. Thanks to all those

participating in the Executive Committee and the election process.

 ∞

The President's Corner Christopher Tong

I would like to thank my fellow board members, especially outgoing Southern Vice President Shar Chaudhry, and the speakers and attendees of our virtual State of Data Science (SDS) and Annual Meeting on Oct.

3. The organizer (Shar) and speakers are all volunteers, and we owe them a debt of gratitude for donating their time and labors. At our annual business meeting we also honored David Thiel, founder of our K-12 Poster Competition, with a lifetime chapter membership. We also awarded a certificate of appreciation to Tia Maria Price, who took on the lead role after Mr. Thiel stepped down until her retirement in 2020. We are grateful that volunteers continue to step forward to contribute to this effort, with Beth Harris taking the lead in 2021.

This fall also featured a chapter officer election, with the two positions that were vacant in the last election (and thus filled by appointment) actually contested this time. I won an extremely close election as President, and would like to reach out to members whose vote I failed to earn. Please let me know what I should have done better in 2020, and how we can do better in 2021. Even better, your active involvement in the chapter will help advance our mutual goals.

I would also like to reflect on the SDS session on "State of Data Science Education Opportunities in Nevada". We had representatives of eight academic degree programs from NSHE institutions, several of which are new or still in the pipeline. Nevada State College deserves our congratulations for founding the first data science bachelor's degree programs in the state, which began in 2020. As I alluded to in my opening remarks of the meeting, the supply of data science talent from within the state does not meet the demand, and many Nevada companies hire data scientists from out of state. Meanwhile, the demand from students for

education in statistics and data science is also high, as illustrated by the UNLV Math Sciences department reporting that the fraction of their students interested in these areas exceeds the fraction of faculty teaching in them. Thus, the expanding set of data science, statistics, and analytics degree choices available to students in Nevada is timely and most welcome. Unfortunately, NSHE has suffered tens of millions of dollars in budget cuts this fiscal year, economic fallout from the COVID-19 pandemic. NV-ASA can help by fostering connections between these programs and entities in the private, public, and nonprofit sectors that could employ our graduating students.

Our chapter can be a natural home for data science in Nevada. We can be that home if we embrace the reality on the ground that many data professionals come from outside mathematics and statistics. As one who has worked and published in physics, statistics, and machine learning, I am convinced that each of these communities (and other data professions) has much to teach each other. No one discipline has a monopoly on good data habits or is free from bad ones.

Data Science Education Opportunities in Nevada

After the NV-ASA President's welcome we heard eight brief presentations on data science education opportunities in NV. Here are summaries; look for these presentations on our website.

We heard first from Tim Malacarne of **Nevada State College** about its recently approved Data Science

Program, being a BA and BS in Data Science. This new program is now searching for an assistant professor and working to expand links with local business and organizations.

Anna Panorska gave us an overview of statistics and data science programs in UNR's Department of Mathematics and Statistics. These include BS and BA programs in Math with Stat specialization, and MS and PhD programs in Statistics and Data Science. Anna described the educational goals of the PhD program, and then listed the current jobs (or professional positions) of 10 undergraduate and MS recipients, which are as diverse as professional positions at Disney and Amazon and NV State Biostatistician.

Zhijian Wu, chair of the UNLV Department of Mathematical Sciences, described their programs, including MS in Math Sciences with concentration in Applied Statistics with either thesis or comprehensive exam options as well as a PhD in Math Sciences.

Daniel Wright told us about possibilities in Assessment and Quantitative Analysis Education at UNLV. A PhD stream in Educational Psychology exists already, and an MSc is being developed. Contact him at daniel.wright@unlv.edu for further information.

Matthew Strickland gave an overview of the MS program in **Biostatistics at UNR**, describing four ways in which biostatistics differs from other areas of statistics and data science, emphasizing types of applications involved and the related skills taught: focusing on health and disease; causal rather than predictive inference; using health information systems; and statistical consulting with health professionals.

Chad Cross gave a brief history of **Epidemiology and Biostatistics in the UNLV School of Public Health**, which in July 2020 became its own department, offering MPH and PhD degrees. After describing the School's infrastructure and current quantitative training opportunities, he listed various professional labels describing their graduates and various fields of work, including biostatistician, statistician, epidemiologist, and data analyst in academia, government, and the healthcare industry.

Kal Joshi, MSBA Program Director, described **the Online MS in Business Analytics at UNR**. His presentation described the curriculum, including a preprogram free and introductory "Skills Boot Camp". He concluded with two extended student testimonials.

Finally, Michael Lee of the UNLV Lee Business School told us about their MS in Applied Economics and Data Intelligence, MS in Management Information Systems, and certificate in Management Information Systems, along with a new MS program in Cybersecurity and a proposed interdisciplinary MS program in data analytics.

 \leq

Invited Talks

፞ቝፙዺዀቝዹዺዀቝቝፙዺዀቝዺፙዀቝፙዺዀቝፙ

ଌୄ୷୷ୡୄ୵ୡ୷ୡୡ୵ୡ୷ୡୡୡ୷୷୷ୡୡୡ୷୷ୡୡୡ୷୷ୡୡ

David Christoff is Director of Football Analytics for the Las Vegas Raiders. His presentation is titled *Controlled Chaos: The History of Data in the NFL*. David walked us through a history of data acquisition and utilization in the NFL. The earliest source he mentioned was *A Course in Football for Players and Coaches* (1912) by Glenn Scobey "Pop" Warner; this early period is labeled "Prehistory".

The Early NFL Period went from 1920-1960. In the early days statistics were available in local newspapers, with tallying expanding in time. By the end of that

period more hardware became involved, in field-tosideline communication and study of game films.

He calls the 1960s and 1970s the "first wave", with NFL partnering with Elias Sports Bureau and the development of computer-aided player draft analysis. David cites the article "Extra Point Strategy in Football" (Richard Porter, *The American Statistician*, December 1967).

The 1980s and 1990s are called the "GSIS Era", for Game Statistics and Information Systems. 1988 brought *The Hidden Game of Football* (Bob Carroll, Pete Palmer, and John Thorn), the first "football analytics" book; the concepts in it are used today with minor modification.

Data acquisition methods continue to evolve, to include RFID plotting of player movement and machine-learning development of predictive algorithms.

Menglin Cao is Senior VP and Head of AI NLP Model Development CoE, Wells Fargo. Her title is Developing AI/ML Models in a Regulated Environment – Current Challenges and Path Forward. Her presentation has three main sections. The first is Motivation, describing twelve roles ("advancements") in which AI/ML is involved in her industry and why these are coming under heightened scrutiny and increasing regulatory requirements. The second presents four categories of current challenges for AI/ML practitioners, and the third suggests paths forward for these practitioners in a regulated environment.

The Advancements listed are Fraud, Personalization, Ops Risk, Anomalies, Targeting, Insights, Decisions, Experience, Conversations and Q/A, Text/Docs Clustering and Classification, Sentiment Analysis, and Text and Speech Recognition. For example, Anomalies involves examining large databases for records that may be erroneous or mis-categorized (as one would expect), Decisions involves making automated decisions more rapidly and consistently, and Targeting involves using ML models to target marketing messages to better focus on consumers' financial needs at the right time and place. Federal SR11-7 guidance (dating from April 2011) is to be applied to all banking organizations supervised by the Federal Reserve; it deals with model risk management. Compliance becomes more challenging with the rapid pace of AI/ML model evolution.

Current challenges are grouped into four categories. One involves a need for adequate training data in unstructured situations, coupled with considerations for handling data that might contain personal identifiable information. Another challenge is the computational resources involved with ML and deep learning processes, along with an information security mindset. Another is the ongoing change from individual model developers to

team model development, with its challenges in communication. Extending the latter are the challenges in taking a model once developed from development to implementation stages. Specific needs associated with these challenges include the following: developing ways to track, reproduce, and document experiments; obtaining accurate model comparisons across experiments and team members; and managing changing datasets throughout a project and storing model artifacts in useful ways.

Menglin suggested several strategies for addressing these challenges. Some involve establishing consistent, repeatable workflow practices with respect to dataset creation and labeling, exploratory data analysis, model and error analysis, and explain-ability analysis. Also involved is training the data scientists in using and adhering to these workflows, as well as developing or procuring software that allows these things to be done.

Jackie Zuker is Director of Data Products with EMPLOYERS Insurance. Her title is *Canary in the Coal Mine: How to Leverage AI as an Early Warning System*. Jackie described a methodology with four steps: assembling a relevant database; preparing and performing feature extraction on the database; doing "anomaly detection"; and passing the results through a personalizable "recommendation engine". The latter can be adjusted to meet the needs and desires of individual users.

The anomaly detection procedure she presented is Prophet, originally developed by Facebook; see "Forecasting at Scale" (Sean Taylor and Benjamin Letham, *The American Statistician*, 2018). It uses seasonal decomposition and estimation with Bayesian modeling in its forecasting; anomalies are data points exceeding the confidence/prediction limits developed by the model.

The next question is "What anomalies do I want to see?". This is where the recommendation engine comes in. This collects feedback from users and logs their responses, using them to personalize user preferences to develop a way of determining which anomalies (alerts) a user would be likely to want to see first today.

A major benefit for this approach is that users can spend less time identifying problems and more time solving them with this proactive data acquisition strategy.

Ari Kaplan is Director of AI Evangelism and Strategy with DataRobot. His title is *How Companies are Transforming into AI-Based Organizations*. His opening statement is "The intelligence revolution has begun; every person, every company, and every nation will be impacted by unprecedented disruption." A November 2018 quote from the *Harvard Business*

Review is "The [AI] winners may take all and late adopters may never catch up." He portrays this as the fourth stage of the Industrial Revolution, prior stages being (mechanical), (electrical & mass production), and (computers, internet, & automation).

Ari listed 18 areas in which AI successes have been made; a few are financial services, gaming, health care, insurance, retail, and sports, some of which were discussed by the other speakers. His first example was in baseball. Changes attributed to AI include increased strikeout percentages, fewer innings pitched by starting pitchers, more defensive shifts, and (perhaps consequently) longer games. He showed us examples of the use of very detailed on-field data acquisition providing information for analysis of baseball, basketball, and football patterns.

He stated that only one in three AI projects succeed. 96% of projects experience difficulties with model development and 90% cite obstacles in deploying models once developed. Only 1% monitor the performance of their AI and ML assets once in production. He outlined a 12 aspects of the process of developing an effective model, grouped into 4 categories: (1-2) knowing the business and the data; (3-8) having the analytical skills needed to deal with the data and write the code needed; (9-10) understanding the underlying math and algorithm basics; and (11-12) having the practical experience and ability to interpret and explain models. And that's where DataRobot comes in. DataRobot provides automated machine learning tools intended to remove items (3-12) above from the "AI Bottleneck".

One example Ari details is Kroger's use of personalized marketing. If you have a Kroger/Smith discount card, all of your purchases are tracked through time series, and you receive personalized discount coupons at checkout and personalized coupons and recipes by mail, aligned to your shopping patterns. Other retail examples he gives involve sales predictions and related inventory modeling, goals being minimizing customer dis-satisfaction and mark-downs.

He ended by suggesting a path toward getting started, with 100 real-life use cases of AI in business to learn from. The link is datarobot.com/yourpath. λ

Tutorial on Data Pipelines

Jonathan Sims, data scientist with NOW Insurance, presented a tutorial titled *Pipelines for Data Science*. The motivation slide gives the following quote from *The DataOps Cookbook: Methodologies and Tools that Reduce Analytics Cycle Time While Improving Quality*

(Chris Bergh, Gil Benghiat, and Eran Stroud): "Automate and Orchestrate – Data Scientists spend 75% of their time doing data engineering, It's about time that data professionals took a page from DevOps. Automate workflow and the development of new analytics. Orchestrate the end-to-end data pipeline so we stop sucking the life out of data scientists. A single data engineer should be able to support ten data analysts and scientists, who in turn should be supporting 100 business professionals. An automated pipeline can get you there."

The aims of the tutorial are to gain experience in deploying AWS (Amazon Web Services) infrastructure from code (using Pulumi), using an automated testing framework (Great Expectations), scheduling and managing many pipelines (Airflow), and deploying cheap cloud infrastructure (using Python script). Prerequisite skills/experience include Python, Git, building and running Docker containers, and basic skills in modeling, data wrangling, and analysis.

Details and materials may be found at https://github.com/etothexipi/sdsnv2020.

Φ

Research Presentations

Blake Hament of UNLV gave us a presentation titled Data for Robots - Working with Exotic Modalities like Images, Point Clouds, and Hyperspectral Signatures. His abstract is the following. "Data Science is the foundation of Machine Learning and today's Artificial Intelligence. To design intelligent robots and automated systems, we must imbue them with the same rules of logic and probability that dictate our own decisions. This presentation covers several projects from the UNLV Drones and Autonomous Systems Lab (DASL), with a focus on using exotic data inputs like images, point clouds, and hyperspectral signatures. A range of popular Python tools and best practices are introduced, with code examples from PyTorch, Sci-Kit Learn, and Pandas."

Anjan Mandal of UNLV's presentation is *Bayesian Directional Parameter Estimation Under Continuity* (joint work with Kaushik Ghosh). The goal of this research is to develop methods to be used in neuroimaging for brain research. Magnetic Resonance Imaging (MRI) is a common non-invasive technique not involving radiation; diffusion MRI uses water diffusion to probe the anatomy of biological tissues. Water tends to diffuse faster along fibers; the anisotropy (directional patterns) in MRI data can give insight into tissue structure. Neuron axions with similar destinations form big bundles called "white matter fiber tracts". Diffusion MRI can reveal details of their anatomy: location, size,

shape, and connections. Clinical applications can involve detecting brain abnormalities and differentiating types of tumors and growth orientations.

Anjan described goals and techniques in the current state of the art. Motivation for his current research comes from limitations. He identified these: the method handles fiber crossings poorly; being an in-vivo technique, it is not possible to compare fitted models with "ground truth"; standard models use 64 gradient directions, while a fully parameterized model can have at least 9 estimable parameters; calculations can be time consuming.

His solution is the SPACE (Spatial Prior for Angular Continuity Estimation). He described this Bayesian model in detail. Fitting is by Monte Carlo Simulation. He concluded by reminding us of the challenges involved in model validation when model-fitting must use in-vivo data, and gave some suggestions.

This presentation will be available on our website.

Position at Touro University

ዀፙዺዀዀፙዺዀዀፙዺዀዀፙዺዀዀፙዺዀዀ

*ᢐ*ᢦᢐ*ᢎ*ᡐᢐᢦᡐᢐᢦᡐᢐᢦᡐᢐᢦᡐᢐᢐᡇᡐᢐᢐ

Data Scientist position at Touro University Nevada

This position will support data-related activities in Institutional Effectiveness (e.g., institutional research, assessment, planning and accreditation) and in research (e.g., biomedical, health-related and educational research projects). This is a job that offers a wealth of opportunities and challenges for a person with strong analytical and communication skills. The position posting is currently being approved by HR, and it should be posted soon. Please pass this along to any qualified people you know. Details will be available at https://tuncareers-touro.icims.com/jobs/searcht.

UNLV MS in Data Analytics

ቝ**ዺ**ጜቝቝዺጜዀቝኇዺዀቝኇዺዀቝኇዺዀቝኇዺ

୕

A MS Data Analytics degree has been approved, starting at UNLV in Fall 2021. This is an interdisciplinary program among the Colleges of Engineering, Hospitality, Liberal Arts, Sciences, the Lee Business School, and the School of Public Health. It will be housed in the Graduate College. The program is designed to appeal to and accept a wide variety of students from a range of academic disciplines. Core courses in technology and management will be offered to the College of Engineering and Lee Business School. Specialized classes in statistical analysis will be offered by the

Colleges of Hospitality, Liberal Arts, and Sciences, and the School of Public Health.

ୡ୭୕ୄ୷୶ୣ୴ୢ୵୷୶ୡୢ୵୷୶ୡ Joining NV-ASA! <mark>ୡୢୢ୰୵୷୶</mark>ୣ୰୰୷୶ୢ୰୵୷୶ୢ୵୷୶

Only a minority of the people who receive this newsletter are members of the Nevada Chapter of the American Statistical Association (NV-ASA). Dues are nominal. For full-time students at NV institutions, the cost is \$10 per year (\$2 for student members of the national ASA). Otherwise, it is \$20 per year (\$10 for members of the national ASA). One can become a Life Member for \$100. You can join NV-ASA when you renew your national ASA membership (or join for the first time); this can be done on-line at www.amstat.org/chapters.

Otherwise, whether a national ASA member or not, you can join through PayPal on our website or by contacting our Treasurer Alejandra Livingston. Any way you do it, please obtain an information form from our website, complete it, and send it to Alejandra at the address listed on the form.

Why should you join? NV-ASA events provide opportunities for networking and contact with other statisticians working in a wide variety of areas in Nevada. But in addition to that, a major reason is that your dues help support the outreach activities of the NV-ASA including the K-12 Poster Competition and Career Days. Our financial needs are not great, so long as we all pitch in our modest amounts.

EXEKBARKEKBARKEKBARKEKBARKEK NV-ASA Officers & Others EXEKBARKEKBARKEKBARKEKBARKEK

Voting officers are

President: Christopher Tong (2020-2021)
Past President: Alicia Chancellor Hansen

(2020-2021)

Northern Vice President: Glenn Waddell (2020-2021)

Southern Vice President: Sharang Chaudhry

(2019-2020)

Sean Breckling (2021-2022)

Secretary: Charles Davis (2020-2021)

Treasurer: Alejandra Livingston

(2019-2022)

Chapter Representative: Gayle Allenback (2020-2022)

Also involved are

Webmaster: Alicia Chancellor Hansen

Poster Competition Lead: Elizabeth Harris Newsletter Editor: Charles Davis Silver State-istics welcomes news items and letters from members and friends of the NV-ASA on matters of interest to the Chapter and the profession. Manuscript or items can be sent as a Microsoft Word document, PDF, or within an e-mail. Silver State-istics is published by the Nevada Chapter of the American Statistical Association.

All items appearing in *Silver State-istics* are from the individuals providing them, and are not intended in any way to represent positions or opinions of any employer, government agency, client, the NV Chapter of the ASA, or the National ASA.

© 2020 Nevada Chapter of American Statistical Association

For contact information, go to https://community.amstat.org/nevadachapter/home
Our address for regular mail is NV-ASA, PO Box 3311, Sparks, NV 89432-3311

