CHATHURANI SENEVIRATHNA
Orlando | senevirathna.c.v@gmail.com | https://www.linkedin.com/in/chathurani-senevirathna/
407.633.8631 | https://github.com/Csenevirathna

>> DATA SCIENCE | MACHINE LEARNING

Analytical, quality-oriented, recent Ph.D. university graduate with a proven background in researching and analyzing a wide range of data. Skilled in Python, Pandas, NumPy, Matplotlib, Scikit-Learn, and Machine Learning algorithms from DSI data science professional certificate course in addition to holding a Ph.D. in Industrial Engineering, MSc, and BSc in Mathematics. Passionate about solving business problems using Data Science & Machine Learning.

SKILLS AND TOOLS

Programming: Python (Base, Pandas, NumPy, Matplotlib, Scikit-Learn), SQL, SAS, NetLogo

Machine Learning: Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis

Other: Statistics, SPSS, GitHub, Data Visualization, MS Office, Tableau, Jupyter Notebook, AWS, Modelling and Simulations

WORK EXPERIENCE

University of Central Florida | Orlando, FL

Postdoctoral Scholar - Complex Adaptive Systems Laboratory | *2022 – Present*

- Used AWS Databrew (data preparation tool), AWS Glue Crawler (database), AWS S3 to extract, transform, and load social media data (Reddit, 4-chan; 100 million rows) and produced SQL queried dataset using ASW Athena to lead research focus on understanding the influence pathways across social media actors.

Graduate Student Researcher |*2017 -2022*

- Created a new network model, analyzed data quantitatively (Python, JS-Divergence, Spearman's Correlation, ANOVA) and developed data visualization using matplotlib and seaborn to understand how online social network users' actions influence each other, using Twitter and GitHub activity data of hundreds of thousands of users, resulting a peer-reviewed journal paper (more than 2000 reads in mdpi.com and received citations).

Graduate Research Assistant - Complex Adaptive Systems Laboratory |*2017-2021*

- Developed an agent based model to study the spread and evolution of information (NetLogo) on social media (GitHub, Twitter, Reddit, and YouTube).
- Led, coordinated, and oversaw a 5-member team from 2 national institutions to execute the SocialSim project funded by DARPA, achieving a 2nd place finish and securing $6.2 million in funding.
- Manipulated structured and unstructured online social network data with millions of users and analyzed data quantitatively (descriptive statistics, correlation research) to understand the effect of information overload of social media users on information spread, resulting in two publications.
- Performed Spearman's Correlation test on google trend data and GDELT (Global Database of Events, Language and Tone) data of the Venezuelan presidential crisis (Python) to understand the association between google search and publication of news articles to develop an information diffusion model.
- Created and implemented novel mathematical models to calculate online social networks users' information processing capacity and probability of performing action upon receiving messages from their connections.

Graduate Teaching Assistant (GTA) - Department of Mathematics |*2014 -2017*

- Assisted 3 faculty members in teaching Ordinary Differential Equations, Calculus I, College Algebra, Pre-Calculus, Trigonometry for undergraduates.

SELECTED DATA SCIENCE / MACHINE LEARNING PROJECS

- **Predicting Customer Loyalty Score Using ML**

  Built predictive models using Linear Regression, Decision Tree, and Random Forest methods in Python that estimated customer loyalty scores for customers that couldn't tag. Used R2 and adjusted R2 metrics to choose the best performing model to predict the loyalty score. **|** *Inferential statistics, Regression, Customer loyality, Python*

- **You Are What You Eat" Customer Segmentation**

  Used k-means clustering on grocery transaction data to split out customers into distinct "shopper types" that could be used to better understand customers over time, and to target customers more accurately with relevant content & promotions. **|** *Unsupervised learning, Clustering, Customer segmentation, Python*

- **Enhancing Targeting Accuracy Using ML**

  Used Logistic Regression to understand the probability of customers signing up for a delivery club to mail more targeted selected customers to reduce cost. **|** *Inferential Statistics, Propensity estimation, Classification, Customer targeting, Pyhton*

- **Understanding the Alcohol Products Relationships Using ML**

  Used Association Rule Learning to understand the relationships of alcohol products to understand which products could be put together to promote bundle products and reorganize the shelf for customer satisfaction. Used Apriori Algorithm. **|** *Unsupervised learning, Association rule learning, Python*

- **Quantifying Sales Uplift**

  Used Causal Impact Analysis on customer data to understand and quantify the uplift on sales for customers who joined a free delivery club, over and above what they would have spent had the club not come into existence. **|** *Causal impact analysis, Python*

- **Assessing Campaign Performance Using Chi-Square Test for Independence**

  Used Chi-Square test for independence on grocery transaction data to understand the significance of the difference in signup rate of the customers for a free delivery club, based on the quality of the mail they received during the campaign to optimize the ROI from future campaigns. **|** *Inferential statistics, AB testing, Hypothesis testing, Chi-square, Python*

EDUCATION

University of Central Florida | Orlando, FL, USA
Doctor of Philosophy in Industrial Engineering | *05/2022*
Master of Science in Mathematics | *05/2017*
University of Peradeniya | Sri Lanka
Bachelor of Science in Mathematics | *09/2011*

COURSES AND CERTIFICATES

DSI Data Science Professional Certification

Actionable Learnings: ● Extracting & manipulating data using SQL. ● Application of statistical concepts such as hypothesis tests for measuring the effect of AB Tests. ● Utilizing GitHub for version control, and collaboration. ● Using Python for data analysis, manipulation & visualization. ● Applying data preparation steps for ML including missing values, categorical variable encoding, outliers, feature scaling, feature selection & model validation. ● Applying Machine Learning algorithms for regression, classification, clustering, association rule learning, and causal impact analysis for measuring the impact of an event over time. ● Machine Learning pipelines to streamline the ML pre-processing & modelling phase. ● Deployment of a ML pipeline onto a live website using Streamlit. ● Using Tableau to create powerful Data Visualizations. ● Turning business problems into Data Science solutions.