Surani Matharaarachchi, PhD

Curriculum Vitae

66 Chancellors Cir, Department of Statistics, University of Manitoba, R3T 2N2.
+1 431 338 5077 • matharas@myumanitoba.ca • surani-matharaarachchi
yhoDCisAAAAJ • Surani-Matharaarachchi • Citizenship status: Canadian

I am applying for the Assistant Professor position to contribute to the advancement of Statistics and Data Science through both research and teaching. With a solid foundation in statistics and computer science and experience across academia, industry, and the public sector, I am dedicated to addressing data-driven challenges by developing novel methods, fostering student success, and contributing to the department's research and academic excellence.

Research Interests: Machine Learning, Statistical Learning, Classification, Data Imbalance, Feature Engineering, Algorithmic Approaches, Deep Learning Techniques, Bayesian Methods

Education

- **Doctor of Philosophy (PhD), Statistics**
  University of Manitoba, Canada *Sep 2021–Nov 2024*
  GPA: 4.13/4.5
  Thesis: New Developments for Addressing Class Imbalance Issue in Classification Tasks.
  Supervisors: Prof. Saman Muthukumarana, PhD, Department of Statistics, University of Manitoba & Dr. Mike Domaratzki, PhD, Department of Computer Science, Western University, Ontario.
- **Master of Sciences (MSc), Statistics**
  University of Manitoba, Canada *Sep 2019–Jun 2021*
  GPA: 4.33/4.5
  Thesis: Assessing feature selection methods and their performance in machine learning with imbalanced data.
  Advisors: Prof. Saman Muthukumarana, PhD & Dr. Mike Domaratzki, PhD
- **Bachelor of Sciences (BSc), Statistics (Special)**
  University of Sri Jayewardenepura, Colombo, Sri Lanka *Nov 2011–Dec 2015*
  GPA: 3.8/4.0 (First Class) - Honors
  Thesis: Study on Parliamentary General Electoral Systems in Sri Lanka.

Publications

Peer-Reviewed Publications

1. **Matharaarachchi S.**, Domaratzki M, Muthukumarana S. (2024). "Enhancing SMOTE for Imbalanced Data with Abnormal Minority Instances." Machine Learning with Applications.
2. **Matharaarachchi S.**, Domaratzki M., Muthukumarana S. (2022). "Minimizing features while maintaining performance in data classification problems." PeerJ Computer Science 8:e1081, `https://doi.org/10.7717/peerj-cs.1081`.

3. **Matharaarachchi S.**, Domaratzki M., Katz A., Muthukumarana S. (2022). "Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets." JMIR Form Res, https://doi.org/10.2196/37984.
4. **Matharaarachchi S.**, Domaratzki M., Marasinghe C., Muthukumarana S., and Tennakoon V. (2022). "Modeling and Feature Assessment of the Sleep Quality among Chronic Kidney Disease Patients." Sleep Epidemiology, https://doi.org/10.1016/j.sleepe.2022.100041.
5. Enns, J., Katz, A., Yogendran, M., Urquia, M., Muthukumarana S., **Matharaarachchi, S.**, Singer, A., Nickel, N., Star, L., Cavett, T., Keynan, Y., Lix, L. and Sanchez-Ramirez, D. (2022) "A population data-driven approach to identifying 'Long COVID' cases in support of diagnosis and treatment." International Journal of Population Data Science, 7(3), https://doi.org/10.23889/ijpds.v7i3.1924.
6. **Matharaarachchi, S.**, M. Domaratzki, and S. Muthukumarana (2021). "Assessing feature selection method performance with class imbalance data." Machine Learning with Applications, https://doi.org/10.1016/j.mlwa.2021.100170
   This paper was awarded with the Reproducibility Badge Initiative (RBI).
7. **Matharaarachchi, S.**, M. Domaratzki, and S. Muthukumarana (2021) Assessing Feature Selection Method Performance with Class Imbalance Data [Source Code]. https://doi.org/10.24433/CO.6033651.v1.

Manuscripts Under Review

8. **Matharaarachchi S.**, M. Domaratzki, A. Katz, S. Muthukumarana. (2024). "Long COVID Prediction in Manitoba Using Clinical Notes Data: A Machine Learning Approach." Intelligence-Based Medicine.
9. **Matharaarachchi S.**, M. Domaratzki, S. Muthukumarana. (2024). "Deep-ExtSMOTE: Integrating Autoencoders for Advanced Mitigation of Class Imbalance in High-Dimensional Data Classification." Journal of Data Science.
10. Katz A., O. Ekuma, J. E. Enns, T. Cavett, A. Singer, D. C. Sanchez-Ramirez, Y. Keynan, L. M. Lix, R. Walld, M. S. Yogendran, N. Nickel, M. L. Urquia, L. Star, K. Olafson, S. Logsetty, R. Spiwak, J. Waruk, **S. Matharaarachichi** (2024). "Identifying people with post-COVID condition using linked, population-based administrative health data from Manitoba, Canada: Prevalence and predictors in the COVID-positive population." BMJ Open.

Honors, Awards and Recognition

- University of Manitoba Graduate Fellowships (UMGF), (CAD 72,000) *2021–2025*
- Faculty of Graduate Studies Travel Award, (CAD 750) *2024*
- WNAR Student Paper Travel Award, (USD 500) *2024*
- Manitoba Centre for Health Policy (MCHP) Scholarship, (CAD 10,000) *2021–2022*
- 2nd-place-winning team of the Bison Transport Data Challenge, International Data Science NEXUS *2021*
- Third place winner of BIRS "Cut to the Chase" video competition, Math Science Career Fair *2021*
- 1st-place-winning team of the Bold Data Challenge, International Data Science NEXUS *2020*
- International Graduate Student Entrance Scholarship, (CAD 5,400) *2019*
- 1st-place-winning team at the Inter-University Statistics Quiz competition, University of Colombo *2012*

Conference Participation

Invited Presentations:

1. International Statistics Conference 2024 (ISC2024). Title: "Uncovering Symptoms and Predicting Long COVID Using Social Media Tweets and Clinical Notes Data: A Machine Learning Approach."
2. International Statistics Conference 2024 (ISC2024). Title: "Deep-ExtSMOTE: Integrating Autoencoders for Advanced Mitigation of Class Imbalance in High-Dimensional and Big Data Classification."
3. Three Minute Thesis (3MT®) 2024, Faculty of Graduate Studies, University of Manitoba. Title: " New Developments for Addressing Class Imbalance Issue in Classification Tasks"
4. 4^{th} International Conference on Future of Preventive Medicine & Public Health (Future of PMPH 2024). Title: "Machine Learning-based Identification of Long COVID Syndrome: Leveraging Encounter Notes Symptoms."
5. Departmental Seminar, Department of Statistics, University of Manitoba, 2021. Title: "Assessing Feature Selection Methods and Their Performance in High-Dimensional Classification Problems."

Contributed Presentations:

1. '2024 WNAR/IMS/Graybill Annual Meeting, Fort Collins, Colorado' - Student Paper Competition presentation title: "Novel Approaches to Mitigate Abnormal Instances in Imbalanced Datasets - for Improved Classification Performance."
2. 'CANSSI Show Case 2023'. Lightening talk title: "Long COVID Prediction in Manitoba Using Clinical Notes Data: A Machine Learning Approach."
3. 'Data to Action Day 2023', organized by the Data Science Program, Government of Manitoba. Lightning Presentation title: "Machine Learning in Government."
4. Statistical Society of Canada (SSC) Annual Meeting 2022. Abstract presentation title: "Discovering long COVID symptom patterns: Association rule mining in social media tweets."
5. Joint Statistical Meetings (JSM) 2021. Topic-Contributed Abstract presentation title: "Modeling and Inference with Feature Importance for Assessing the Quality of Sleep among Chronic Kidney Disease Patients."
6. Statistical Society of Canada (SSC) Annual Meeting 2021. Abstract presentation title: "Assessing Feature Selection Methods and their Performance in High-Dimensional Classification Problems."

Attended Conferences/Workshops:

1. 2024 CRA-WP Virtual Career Mentoring Workshop Series: Transitioning: Challenges and Strategies, Parenting and Work-Life Balance, Becoming an Outstanding Teacher and Supporting All Students, Teaching-Track Faculty Perspectives and Challenges
2. INFORMS Annual Meeting, Seattle, Washington, USA, October 2024.
3. 'Evidence to Action Day 2023', organized by Manitoba Centre for Health Policy.
4. Fundamentals of Causal Inference: With R, CANSSI Prairies Workshop Series in Data Science, University of Winnipeg, 2023.
5. 18^{th} Annual IPAC Leadership Summit 2023, organized by the Institute of Public Administration of Canada.
6. EMILI's Annual Agriculture Enlightened Conference, October 2022.

7. Data Science Pre-Conference workshop on Tools for Bayesian data science and probabilistic exploration by Prof. Alexandre Bouchard-Côté, 2019.

Teaching Experience

- **Sessional Instructor**
  Department of Statistics, University of Manitoba *Summer 2022*
  - STAT 1150 - Introduction to Basic Statistics and Computing (with R)

- **Teaching Assistant**
  Department of Statistics, University of Manitoba *Sep 2019 - Apr 2022*
  - STAT 2000 - Basic Statistical Analysis II (n = 9)
  - STAT 4150 (Senior Level TA) - Bayesian Analysis and Computing (with R & Python) (n = 3)
  - Exam Invigilator, Grader and TA at the Statistical Help Center

Professional and Research Experience

- **Data Scientist**
  Department of Education and Early Childhood Learning, Government of Manitoba *Feb 2024 - Present*
  - Key Contributor to a new government initiative, leading the creation of two education data dashboards using Power BI that enhance decision-making in the education sector.
  - Extracting and processing data from databases to meet various analytical needs.
  - Completing data requests and analyses using R, tailored to diverse stakeholders.
  - Creating detailed provincial test reports to support decision-making and policy review.

- **Data Science Leader in Training (LTP)**
  Government of Manitoba *Dec 2022–Feb 2024*
  - Collaborated with multiple departments/parties (Consumer Protection, Health, Municipal and Northern Relations, Manitoba Centre for Health Policy (MCHP)) on various projects.
  - Developed predictive models for property assessments using machine learning regression techniques, encompassing data cleaning, pre-processing, hyper-parameter tuning, and model fitting to ensure accurate and reliable outcomes.
  - Evaluated the impact of COVID-19 on education outcomes and heat waves on health-related illnesses.

- **STEP Student (Student Temporary Employment Program) - Data Science**
  Department of Consumer Protection and Government Services, Government of Manitoba *Jul 2022–Dec 2022*
  Supervisor: Anna Slavina, PhD, Director, Data Science Program
  - Processed, cleansed and verified the integrity of address data using Python NLP tools.
  - Integrated Python REST services using NRCAN API, handled geo-location data, created maps, and performed ad-hoc analyses.

- **Research Assistant**
  Department of Statistics, University of Manitoba *Sep 2019 - Nov 2024*
  - Conducted advanced research on class imbalance issues in high-dimensional data, focusing on novel resampling and feature selection techniques.

- Collaborated with the Manitoba Centre for Health Policy (MCHP), using their closed RAS platform to securely access and analyze healthcare data.

Data Analytics Graduate Student Intern - Data Science
City of Winnipeg, Manitoba *Nov 2020–Feb 2021*
Supervisor: Jennifer Bodnarchuk, PhD, Senior Data Scientist, Department of Innovation & Technology

- Assessed the spread of COVID-19 and predicted the number of infections, recoveries, and deaths using time series predictive and SEIR models.
- Processed, cleansed, and verified data integrity using publicly available sources.
- Integrated Python REST services to facilitate data analysis and reporting.

Industry Experience

Freelance Data Science Recruitment Consultant
Self-Employed *Dec 2023*

- Provided freelance consulting services to Callia Inc., Winnipeg, Canada.
- Participated in the technical evaluation of data scientist candidates, conducted comprehensive reviews of tests, and collaborated with hiring managers for informed decisions.

Data Scientist
nCinga Innovations (Pvt) Limited, Colombo 07, Sri Lanka *Sep 2018–Jul 2019*

- Enhanced data collection procedures to include information relevant to building analytic systems.
- Developed data modeling for Online Analytical Processing (OLAP).
- Led project coordination with client stakeholders and team.

Data & Report Analyst
Duo Software (Pvt) Limited, Colombo 02, Sri Lanka *Sep 2017–Sep 2018*

- Implemented data warehouse and ETL processes. Maintained and queried databases (PostgreSQL, MySQL, Microsoft SQL Server, Big Query, NoSQL databases).
- Conducted unit testing and participated in the Quality Assurance process.
- Performed statistical analysis for projects and design reports using Power BI and Stimulsoft.

Scholarly and Professional Activities and Affiliations:

University Service:

- Tenure Track Search Committee Member, Department of Statistics, University of Manitoba *2022*

Service to Profession as a Manuscript Peer Reviewer:

- Journal of Medical Artificial Intelligence - JMAI *2024*
- Digital Health: Sage Journals *2023*
- Journal of Informatics in Medicine Unlocked *2022*

Professional Memberships:

- Statistical Society of Canada (SSC) *2020-2025*

- INFORMS *2024-2025*
- The Western North American Region of The International Biometric Society (WNAR of IBS) *2024-2025*
- Statistics Graduate Students' Association (SGSA) *2019-2024*

Professional Development & Certifications:

- Respectful Workplace Policy Training - Government of Manitoba *2024*
- Anti-Racism: Understanding Ourselves & Our Systems - Government of Manitoba *2024*
- Our Shared Journey forward Truth & Reconciliation Training - Government of Manitoba *2024*
- Harnessing the Power of Data with Power BI (Authorized by Microsoft) – Coursera *2024*
- Diversity & Inclusion Policy Training - Government of Manitoba *2022*

Technical Profile

- **Programming**: Python, Pandas, Numpy, Scikit-learn, GeoPandas, R
- **Visualization**: Power BI, R Shiny
- **Machine Learning and Data Mining**: Classification, Feature Selection, SMOTE, Algorithmic Approaches, Data Imbalance
- **Deep Learning Techniques**: Autoencoders, Neural Networks, TensorFlow, Keras
- **Statistical Tools**: Minitab, SPSS, E-Views, OpenBugs, MATLAB, Maple, MS Excel
- **Databases & Query Languages**: MySQL, PostgreSQL, MS SQL, Druid, BigQuery, MongoDB, Elasticsearch
- **Development Tools**: SAP, PyCharm, Jupyter, Postman
- **Project Tools**: Agile, GIT, JIRA
- **Documentation**: MS Word, MS PowerPoint, LaTeX
- **Advanced Data Techniques**: Data Mining, NLP, Big Data, Data Cubes, OLAP

References

- References available on request