Surani Matharaarachchi, Ph.D.

Curriculum Vitae

66 Chancellors Cir, Department of Statistics, University of Manitoba, R3T 2N2.

I +1 431 338 5077 • # matharas@myumanitoba.ca • ð [surani-matharaarachchi](https://www.linkedin.com/in/surani-matharaarachchi)  [yhoDCisAAAAJ](https://scholar.google.com/citations?user=yhoDCisAAAAJ) • D [Surani-Matharaarachchi](https://www.researchgate.net/profile/Surani-Matharaarachchi) • Citizenship status: Canadian

Professional Summary........................................................................................................

With a strong foundation in machine learning, natural language processing, and statistical methodologies, I bring experience in advanced model development, feature engineering, and deployment. My research emphasizes handling data imbalance and optimizing model performance, including extensive work on synthetic data generation techniques such as Deep-ExtSMOTE, which integrates autoencoders for high-dimensional data classification. I am passionate about leveraging innovations in LLMs to enhance AI-powered personal assistant capabilities.

Research Interests.............................................................................................................

Machine Learning, Large Language Models (LLM), NLP, Knowledge Representation, Model Optimization, Statistical Learning, Deep Learning Techniques, Data Imbalance, Bayesian Methods

Education.......................................................................................................................

Doctor of Philosophy (PhD) Candidate, 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, & Dr. Mike Domaratzki, PhD.

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.](https://mspace.lib.umanitoba.ca/items/3a64f151-8deb-4a0b-9e95-846ed23551c3)

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)

Professional and Research Experience....................................................................................

Research Assistant

○

*Department of Statistics, University of Manitoba Sep 2019 - Present*

* Applied association rule mining to detect Long COVID symptom patterns in over 1M social media records using Python’s NLP toolkit NLTK and LLMs, uncovering key insights into symptom co-occurrence and progression.
* Identified Long COVID patients in Manitoba by extracting symptoms from health data using NLP, machine learning, and resampling, achieving 0.95 sensitivity, 0.81 specificity, and a 0.94 AUC. This approach revealed 24.7% of cases as Long COVID—14 times the previously known cases. Collaborated with MCHP using their secure RAS platform for data analysis.
* Developed 4 advanced SMOTE algorithms with self-developed code, using proximity-based, probabilistic, and Bayesian weighting mechanisms to address abnormal instances in minority classes. Applied in both simulated and real-world data, these methods improved representational accuracy for minority classes in highly imbalanced datasets.
* Achieved a 25% improvement in average F1 Score with new Deep-ExtSMOTE by integrating autoencoders, Bayesian resampling, and tailored embeddings for continuous and categorical data in TensorFlow and Keras. This approach advanced class imbalance handling in high-dimensional data, with all computations were done on the Digital Research Alliance of Canada cluster.
* Demonstrated strong independent research skills and effective time management by balancing complex algorithm development, data analysis, and publication efforts alongside full-time professional responsibilities, ensuring consistent progress and timely completion of all project milestones.

Data Scientist

○

*Department of Education and Early Childhood Learning, Government of Manitoba Feb 2024 - Present*

* Led the development of three public education data dashboards on the government website as part of a key government initiative, along with internal apps, using Power BI Desktop and Power BI Server to deliver streamlined, data-driven insights supporting decision-making in the education sector.
* Extracted, cleaned, and processed data from the Education Information System (EIS), an integrated database on Oracle and MSSQL, using SQL for ETL tasks and R and Python for advanced analysis and reporting.
* Conducted data analyses and fulfilled data requests using R, delivering customized insights to stakeholders and generating comprehensive provincial test reports to support policy review and data-driven decision-making in education.
* Developing a Unified Predictive Model for High School Graduation Risk in Manitoba using machine learning to analyze longitudinal student data, enabling early identification of at-risk students from Grades 9 to 12 and supporting targeted interventions.

Data Science Leader in Training (LTP)

○

*Government of Manitoba Dec 2022–Feb 2024*

* Collaborated with multiple departments and agencies to support diverse projects, sharing expertise and aligning data-driven solutions with specific departmental needs. Served as an R support partner in the Data Science Practicum Program, contributing technical skills to enhance project outcomes.
* Engaged in a project to develop predictive models for property assessments using XGBoost, performing data cleaning, pre-processing, feature engineering, hyper-parameter tuning, and model fitting to achieve over 90% accuracy, ultimately reducing manual workload by approximately 75%.
* Evaluated the impact of COVID-19 on high school education outcomes, using regression models to quantify effects and identify the most vulnerable student groups through analysis of performance trends and demographic factors.
* Investigated the effects of heat waves on health-related illnesses using data from MCHP and Environment Canada, to identify patterns and correlations between extreme heat events using logistic regression models. These findings support public health responses and preventative measures for vulnerable populations.

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 billing address data using Python NLP tools and data from Python REST services with the NRCAN API to extract geo-location information, reducing the need for manual address verification by 60%.

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 by predicting infections, recoveries, and deaths using time series predictive and SEIR models, processing and cleansing data from publicly available sources, and integrating Python REST services to streamline data analysis and reporting.

Industry Experience...........................................................................................................

Freelance Data Science Recruitment Consultant

○

*Self-Employed Dec 2023*

* Provided freelance consulting services to Callia Inc. in Winnipeg, Canada, using expertise to offer feedback on the technical evaluation of data scientist candidates, conducting comprehensive test reviews, and collaborating with hiring managers for informed hiring decisions.

Data Scientist

○

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

* Led the data integration for successfully implementing a dashboard, an analytics tool, and a mobile app with the OLAP Druid database, achieving a 60% reduction in data processing time, optimizing database architecture, developing data models, and ensuring data integrity at every stage as part of a new initiative.
* Provided structured layouts and design specifications for each tool to the designer and collaborated closely with software engineers and architects to ensure seamless integration and alignment with project goals.
* Extracted data from MongoDB and Elasticsearch databases to perform daily reporting, using Kafka and Flink to streamline data processing workflows.
* Led project coordination with client stakeholders and cross-functional teams, facilitating clear communication, aligning project goals, managing timelines, and ensuring deliverables met client expectations and quality standards.

Data & Report Analyst

○

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

* Led the backend development of interactive regression and time series widgets within a new dashboard, using Python to automate models for predicting future sales and trends. Integrated data from multiple sources to provide robust analytical insights and improve forecasting accuracy by 30%.
* Implemented data warehouse and ETL processes with Pygrametl, managing and querying databases such as PostgreSQL, MySQL, Microsoft SQL Server, and BigQuery to enable efficient data storage, retrieval, and analysis.
* Conducted unit testing and actively participated in the Quality Assurance process, identifying and resolving defects, collaborating with QA teams to improve test coverage, and ensuring the reliability and performance of application features before deployment.
* Worked in an agile environment, collaborating closely with backend and frontend engineers, QA specialists, and project managers. Engaged in sprint planning, daily stand-ups, and retrospectives to ensure alignment, address challenges, and deliver high-quality features on schedule.

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. https://doi.org/10.1016/j.mlwa.2024.100597.
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.](https://doi.org/10.7717/peerj-cs.1081)

1. 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.](https://doi.org/10.2196/37984)
2. 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.](https://doi.org/10.1016/j.sleepe.2022.100041)
3. 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.](https://doi.org/10.23889/ijpds.v7i3.1924)
4. 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).

1. 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.](https://doi.org/10.24433/CO.6033651.v1)

Manuscripts Under Review

1. 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.
2. 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.
3. 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.

In Preparation

1. Matharaarachchi S. and Muthukumarana S. (2024). “NBA Player Performance Evaluation using SMOTE Based Machine Learning models.”

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.”

1. 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. ‘CANSSI Show Case 2024’. Lightening talk title: “Long COVID Prediction in Manitoba Using Clinical Notes Data: A Machine Learning Approach.”
2. ‘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”
3. ‘CANSSI Show Case 2023’. Lightening talk title: “Long COVID Prediction in Manitoba Using Clinical Notes

Data: A Machine Learning Approach.”

1. ‘Data to Action Day 2023’, organized by the Data Science Program, Government of Manitoba. Lightning Presentation title: “Machine Learning in Government.”
2. Statistical Society of Canada (SSC) Annual Meeting 2022. Abstract presentation title: “Discovering long COVID symptom patterns: Association rule mining in social media tweets.”
3. 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.”

1. 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, TeachingTrack 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-Cˆot´e, 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 1000 - Basic Statistical Analysis I
* STAT 2000 - Basic Statistical Analysis II
* STAT 4150 (Level II) - Bayesian Analysis and Computing (with R & Python)
* Exam Invigilator, Grader and TA at the Statistical Help Center

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