

LUKA NEGOITA, PHD

Portfolio: lukaneg.github.io

lukanegoita@gmail.com 

github.com/LukaNeg 

www.linkedin.com/in/negoita 

>> DATA SCIENCE | RESEARCH | MACHINE LEARNING

MOTIVATION With 14 years of data science experience, I excel in leading projects and delivering actionable insights that drive decisions and business growth. My strengths include creative problem-solving, clear communication, and aligning data solutions with business goals.

In other words, I love solving puzzles and am passionate about developing innovative, technically sound, and commercially viable solutions—all while clearly communicating across technical and general audiences.

My goal is to use these skills to contribute towards a culture of data-driven research and development, especially in organizations with a vision for social, health, or environmental impact. I thrive in teams that value curiosity, clear communication, and a culture of mutual growth.

SKILLS & TOOLS

Programming: R, Python, SQL, dbt, git, ggplot2, pandas, dplyr, Shiny, dbt, Javascript, HTML/CSS

Machine Learning: Linear and Logistic Regression, Time-series (longitudinal, multi-modal, etc), Mixed-models, LLM RAG and prompt engineering, Bayesian inference, Random Forest, KNN, k-means, NLP, Causal Impact, Gradient Boosting, Decision analysis

Other Skills: Data Visualization, Data Cleaning, Scientific publishing, Experimental design, Dashboards, Slide presentations, Google Cloud Services

Senior Data Scientist (May 2022 - present) — [Nutrisense, Inc](#)

Nutrisense is a Series-A health-tech startup that uses continuous glucose monitoring to help individuals improve their health. I led multiple cross-functional projects that leveraged advanced data science techniques to drive product development and strategic business decisions. **Below are four selected projects that highlight some of my skills and experience:**

1. Diabetes Remission Outcomes: Our Marketing team needed to quantify how many members were achieving diabetes remission, but we lacked the lab results to assess this. I developed an innovative method for calculating fasting glucose, a key diagnostic metric, by pulling sleep data from integrations and using glucose readings just before waking. After cleaning and engineering the required data, I validated the method against existing lab tests, confirming its reliability. I then built a longitudinal mixed-effect logistic model to evaluate member outcomes, demonstrating that 37% of highly engaged members achieved remission within six months. I wrote a whitepaper on these results, which is now being used by the Marketing team for promotional and educational purposes.

2. Competitor Outcome Comparison for Investors: As the Executive team prepared for investor meetings, they needed a way to show if our diabetes remission outcomes outperformed competitors. The challenge was that we needed an apples-to-apples comparison, but each study used different methods and metrics. I conducted an extensive literature review, extracting key summary statistics from relevant publications. Using these statistics, I simulated competitor datasets and used a Bayesian framework with a logistic model to generate probability distributions of the estimated outcomes in each study. I helped account for differences in study methods by introducing additional error into simulated data and using results from other publications to apply informed priors. The final result showed us a 93% probability that our members' outcomes were better than a major competitor's. I visualized the results as an infographic and presented them to our CEO, who used it to support investor pitches.

3. Semantic Search for Nutritionists: Our Nutritionist coaches rely on a large collection of articles and resources to help answer members questions. However, the sheer volume of articles made it time-consuming for Nutritionists to find relevant information. Recognizing this inefficiency, I took the initiative to develop an MVP solution leveraging APIs to OpenAI's large language models. I engineered the data pipeline, built a vector database of text embeddings from a subset of our articles, engineered and tuned the prompts, and deployed a web app that allowed Nutritionists to perform semantic searches, quickly locating relevant answers to their queries. After presenting this MVP to the CTO, the Engineering team built a final version of the app—now a critical tool that Nutritionists use daily to enhance member support.

4. Lead Conversion and Churn Prediction: Our Marketing and Product teams needed insights into which leads were likely to convert into paying customers and which of those customers were at risk of churn. I led the development of two predictive models—one for lead conversion and another for churn prediction. I worked closely with stakeholders to ensure the models provided actionable results aligned with their goals. I pulled all relevant data sources, including over 100 features for each model, and used cross-validation to select features, tune hyper-parameters, and select the final models (LightGBM for both). I also used Shapley values to identify the factors associated with conversion and churn, making the non-intuitive discovering that the characteristics of good leads also often align with early churn. I shared results from both models with stakeholders and collaborated with others in my team to deploy the lead model in production for Marketing to optimize paid marketing campaigns.

Additional Experience @ Nutrisense

I led the product data team in developing predictive models, causal inference analyses, and dashboards using R, Python, and SQL. I collaborated across teams to improve user experience and user health outcomes. I built and maintained data pipelines using dbt, ensuring clean, reliable data for product scaling. I authored two research whitepapers, presented a conference poster, and regularly communicate technical results to non-technical stakeholders, promoting a data-driven culture across the company.

Founder and Lead Data Science Instructor (May 2020 - present) — [R for Ecology School](#)

- I aimed to make data science more accessible to biologists worldwide, so I built an online school dedicated to teaching biologists how to use R in a way that reduces the stress burden of learning how to code.
- The business was profitable from the start, providing scholarships for anyone in need and hiring three assistants to help run the school.
- Through this process I designed and created the online courses and numerous resources via [my blog](#), [YouTube](#), and [Twitter feed](#).
- The school has now taught over 1000 students from more than 150 countries.

Ecologist and Biostatistician (Oct 2017 - Nov 2021) — Charles Darwin Foundation

- I led research studies and analyzed multi-modal environmental data using Bayesian, regression, model selection, and time-series analyses to improve ecosystem restoration efforts in the Galapagos Islands.
- As part of this I also developed models optimizing cost-benefits in restoration, which demonstrated the potential to reduce costs by 34% while maintaining successful restoration.
- Just prior to the lockdowns of 2020, I built a simulation model that I used to accurately predict the arrival of COVID-19 to the Galapagos Islands. I presented this to local officials which likely prompted the process of closing the islands to visitation and thereby reducing the death toll. [View report here.](#)
- I designed data visualizations, infographics, and technical reports for communicating results to stakeholders and general public using ggplot2, plotly, Rmarkdown, and other visualization tools.
- As part of this, I built online dashboards with R-Shiny for visualizing and analyzing financial and ecological data.
- Finally, wrote and published two first-author peer-reviewed publications as a result of my analyses and results.

Founder and CEO of a Mental Health Startup (Sept 2015 - Jan 2018) — HabitU Lab

- I led a development team to build an app designed to help people improve their mental health by combining a habit-tracking system with behavioral psychology.
- The app is still online and [available to use here.](#)
- I bootstrapped the tech startup and successfully acquired seed funding after competing in numerous business competitions and participating in a startup incubation program
- I did all of this while simultaneously working towards my PhD in Biology
- This work led to my award as Student Entrepreneur of the Year 2017 by Whitman School of Management, Syracuse University

Research Scientist (Nov 2010 - May 2020) — Syracuse University and College of the Atlantic

- I led four major research programs at two universities, executing all design, data collection, and analyses
- I wrote and hosted ten successful grants, fellowships, and crowdfunds totalling over \$175,000 USD
- I authored 9 peer-reviewed publications and numerous conference presentations and technical reports
- This work included the successful defense of my PhD thesis which won one of the top biology dissertation awards in 2018.

Senior Research Analyst (Dec 2017 - Mar 2018) — The Research Foundation at SUNY

- I conducted statistical analyses on multi-modal environmental data from studies in the Galapagos Islands
- I generated graphical and tabular summaries of analyses, producing narratives of the statistical methods and interpretations of results that were later used in scientific publications

EDUCATION **PhD (Quantitative Biology)**
2012 - 2018 — Syracuse University, New York

BA (Human Ecology)
2007 - 2011 — College of the Atlantic, Maine