

# Susanna Morin, MSc

Dallas, TX | 408-606-0840 | [susanna.morin@gmail.com](mailto:susanna.morin@gmail.com) | [LinkedIn](#) | [GitHub](#)

## PROFESSIONAL SUMMARY

---

I use statistics and predictive analysis to find meaningful insights in population health data and present those insights to non-technical audiences in a simple way using graphs and dashboards. I have experience engineering fast and accurate machine-learning models in Python and R, using state-of-the-art approaches to determine the highest performing model. I pride myself with strong analytical skills, attention to detail and ability to communicate with a team effectively.

## SKILLS

---

**PROGRAMMING/LANGUAGES** C++, Python, R, SQL, Microsoft Suite

**DATA ANALYSIS TOOLS** Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, PowerBI

**DATA STORYTELLING** Data Wrangling, Statistical Analysis, Predictive Analysis, Empirical Trends, Data Visualization

**MACHINE LEARNING** Linear and Logistic Regression, Principal Components Analysis, Feature Extraction, Model Selection

## EDUCATION

---

### University of California

*Master of Science: Biomedical Informatics 2022*

*Bachelor of Arts: Computer Science, Bioinformatics 2021*

## PROFESSIONAL EXPERIENCE

---

### Gainwell Technologies | November 2022 – Present

#### Senior Healthcare Data Scientist

- Built a survival analysis model using the cox proportional hazards regression to investigate the association between the survival time in the Severe and Persistently Mentally Ill (SPMI) patients and predictor variables of interest with the potential to inform policy in healthcare
- Maintained and created tables for downstream analysis using cloud-based Electronic Health Records (EHRs), claims data and population health data for predictive analytics
- Conducted a total cost of care time series analysis on healthcare provider utilization trends
- Developed a model that predicted average cost savings for attributed and enrolled members that received care coordination compared to those not enrolled in the program on a yearly basis

### UCSF Biomedical Informatics | July 2021 – November 2022

#### Clinical Data Scientist

- Conducted data validation in collaboration with Nima Aghaeepour Laboratory at Stanford for a journal publication that
  - Focused research efforts on patient phenotyping and trajectory prediction in neonatal health and morbidity
  - Queried patient information from clinical databases using SQL; Analyzed relationships between women's health factors and offspring health outcomes in pre-term labor using statistical hypothesis testing
- Generalized a CNN model measuring knee osteoarthritis and improved performance by changing the biomarker from bone shape to cartilage thickness

### UC Santa Cruz Genomics Institute | July 2019 – April 2021

#### Bioinformatics Engineer

- Developed a genotyper using the Markov Chain Monte Carlo probabilistic model that supports standard variant calling formats; Improved accuracy and performance of genotyper using the Min-Cut algorithm to break out of sampling bottlenecks maximizing mixing efficiency; Established evaluation methods that compare accuracy metrics against gold-standard datasets

### Gladstone Institute | 2018; 2020

#### Data Science Analyst | July 2020 – November 2020

- Investigated how single-cell RNAseq data and single-cell ATACseq data from mouse hearts correlated with each other across drug treatment and disease states to successfully predict enhancer activation due to heart stress
- Built a support vector machine (SVM) model to predict enhancer activation (single cell ATAC-seq) from expression data (single-cell RNA-seq); Presented research findings at the Cold Spring Harbor's Biological Data Science Conference 2020

### Bioinformatics Analyst, Cardiovascular Disease | June 2018 – August 2018

- Researched gene expression in the interventricular septum of the heart via RNA in situ hybridization that assisted in narrowing down candidate genes; Visualized signals in RNA molecules using fluorescent microscopes; Utilized bioinformatics clustering tools such as Seurat to analyze RNA-seq experiments that provided information on gene expression in mouse heart tissue