Suriyapriya Selvanathan

+1 412-954-8828 | sselvana@andrew.cmu.edu | Linkedin.com/Suriyapriya S | https://suriyapriya.github

EDUCATION

Carnegie Mellon University — **Pittsburgh, PA, May 2026**
Master of Science in Health Care Analytics and Information Technology — **QPA:** 3.82 / 4.00
Coursework: Machine Learning, Advanced AI, Applied Econometrics, Database Management, Natural Language Processing, Optimization, Decision and Risk Modeling, Statistics, R with Analytics, Healthcare Information Systems, Project Management

PSG College Of Technology — **Coimbatore, India, Apr 2023**
Bachelor of Engineering in Biomedical Engineering — **GPA:** 3.94 / 4.00
Coursework: AI, ML and Data Science, Statistics, Pattern Recognition and Neural Networks, Digital Signal Processing

EXPERIENCE

Highmark Health / Allegheny Health Network (AHN), Pittsburgh, PA — **Jan 2026 – Present**
Capstone Project Data Lead
- Spearheading a strategic initiative to model longitudinal cancer cost trajectories, utilizing Python and SQL to engineering data workflows that ingest complex administrative claims and clinical utilization data.
- Developing predictive risk algorithms and conducting cost-driver sensitivity analyses to identify financial variances, directly supporting Value-Based Care resource allocation.
- Translating technical model outputs into actionable insights for clinical leadership, ensuring data-driven strategies align with population health goals and potential cost-saving opportunities.

Carnegie Mellon University, Pittsburgh PA — **Aug 2025 – Present**
Research Assistant – Prof. Rema Padman
- Engineered a feature labeling scheme to process over 65,000 raw game telemetry events from a pediatric RCT , constructing chronological event sequences to map user trajectories.
- Developed a discrete, first-order Markov Chain model (7 states) to visualize stochastic gameplay patterns, applied the MDA framework and Poisson regression to identify a counterintuitive priming effect"influencing health behaviors.

Allegheny Health Network (AHN), Pittsburgh, PA — **Jun – Aug 2025**
Data Science Intern
- Designed a clinical data platform and decision support system to process and analyze Real-World Data (RWD) from over 15,000 health screenings, while managing project timelines, budget, and resources in MS Project and mapping data flows in Visio.
- Built a robust data pipeline to integrate diverse structured and unstructured data sources, creating longitudinal patient records essential for time-series analysis and research.
- Created modules for data quality validation and built interactive dashboards (Geopy, Folium, Tableau) to help clinicians flag high-risk patients, improving data curation efficiency by 83%.

Latent View Analytics, Tamil Nadu, India — **Jun 2023 – Jul 2024**
Data Analyst
- Analyzed longitudinal clinical data from over 2 million patient records, using survival analysis and feature importance from Machine Learning (XGBoost) models to identify key predictors of disease progression and treatment adherence.
- Engineered and deployed ensemble models(Gradient boosting) for real-time patient risk stratification, leading to a 25% improvement in accuracy of identifying high-risk candidates for preventative care interventions.

Sree Chitra Tirunal Institute for Medical Sciences & Research, Kerala, India — **May – Jul 2022**
Clinical Analyst Intern
- Spearheaded the full-stack development of a patient portal application (modeled after Epic's MyChart), creating a user-friendly interface for patients to access real-time health outcomes.
- Developed statistical models (regression, classification) to assess patterns in clinical data, stratify high-risk patients, and provide personalized insights within the application to support evidence-based decision-making.

PROJECTS AND RESEARCH

Diabetes Readmission Risk Assessment System — **Jan 2026 – Present**
- Built a clinical decision support system using Gradient Boosting to predict 30-day hospital readmissions from a longitudinal dataset of diabetic patients, achieving 94.7% accuracy
- Engineered features capturing temporal patterns and comorbidity interactions, addressed missing data via imputation, and handled class imbalance with SMOTEENN.
- Conducted sensitivity analysis to identify key predictors of readmission, providing actionable insights for care management protocols.

Optimal Placement of Safe Consumption Sites Using a Predict-Then-Optimize Framework — **Aug – Dec 2025**
- Developed and deployed 106 individual ARIMA time-series models to accurately forecast future fatal overdose events for every ZIP code in Allegheny County.
- Formulated a maximal covering location problem, implemented an optimization model using Gurobi to identify optimal location for a new site, increasing network coverage by 10 percentage points, covering additional 50 forecasted deaths.

SKILLS

Programming: Python (pandas, NumPy, scikit-learn, TensorFlow/Keras), SQL, NoSQL, MATLAB, R, C/C++, Stata
Data Science & ML: Time Series Analysis, Statistical Modeling, Supervised/Unsupervised Learning, NLP (Hugging Face, spaCy, LLMs), Causal Inference, Deep Learning (PyTorch, TensorFlow/Keras)
Data Engineering & Cloud: ETL, BigQuery, Snowflake, Databricks, Spark, Hadoop, Docker, Git, Cloud Platforms (GCP, AWS, Azure)
Visualization & BI: Streamlit , Tableau, Power BI, RShiny, Matplotlib, Seaborn, Excel (Pivot Tables, Solver, VLOOKUP)

LEADERSHIP & AWARDS

- Safety21 US DOT Transportation Fellow
- Super Analytics Case Competition- Finalist
- 2025 Patient Safety Fellowship Recipient
- Heinz College Merit Scholar
- CMU (Tartan) Student Ambassador
- PLUS(NGO) Mathematics Tutor