Noah Dcruz

● Dcruz.n@northeastern.edu ● (339) 215-9580 ● Medford, MA ● LinkedIn ● Github

EDUCATION

Northeastern University, Boston, MA
Candidate for Master of Science, Applied Mathematics, GPA: 3.72 — Dec 2024

Bentley University, Waltham, MA
Bachelor of Science, Finance, Concentration in Corporate Finance — May 2021

SKILLS and LEADERSHIP

- Programming/Software: Advanced Python (NumPy, Pandas, Matplotlib, Scikit-learn, SciPy, TensorFlow, PyTorch, Spark), Advanced R, MATLAB, Intermediate SQL, Advanced VBA, Tableau, Power BI, SPSS, Advanced Excel, SAP, Bloomberg Terminal
- Advanced Math: Statistical methods, Differential Equations, Optimization Techniques, Stochastic Processes, Numerical Analysis, Systems Modeling

PROFESSIONAL EXPERIENCES

ConcertAI, Cambridge, MA — May - Aug 2024
Biostatistics Intern

- Authored a comprehensive Statistical Analysis Plan to guide the evaluation of comorbidity impacts using the Charlson Comorbidity Index, including survival analysis and data source comparisons to ensure actionable insights in clinical research
- Collaborated cross-functionally with scientists to drive data-driven outcomes in precision medicine projects
- Automated data integration processes using SQL in Amazon Redshift, streamlining querying and analysis of large datasets, which reduced manual data handling time by 30% and improved report generation times
- Developed visualizations in R to illustrate survival probabilities stratified by Charlson Comorbidity Index score categories, enabling clear insights into patient survival differences
- Performed Cox Proportional Hazards modeling in R on large-scale patient datasets, assessing the impact of various patient characteristics on survival outcomes, which contributed to actionable insights in oncology research

Keros Therapeutics, Lexington, MA — July - Dec 2023
Data Modeling & Pharmacometrics Co-Op (full-time)

- Processed and cleaned 2GB of pharmacological data, enhancing data analysis efficiency by 15% through data wrangling
- Collaborated with 5+ departments to define clinical pharmacology questions, speeding up project timelines by 20%
- Reviewed 100+ scientific articles to strengthen our modeling approaches with the latest pharmacometrics insights
- Led M&S analysis by feature engineering/dimensionality reduction, to identify factors for drug efficacy and safety
- Simulated dose modifications with cross-validation and hyperparameter tuning, achieving a 10% improvement in outcomes
- Communicated complex modeling insights to stakeholders through data storytelling, strengthening 3 key drug projects
- Improved clinical trial predictive accuracy by 12%, using time series and anomaly detection in pharmacometrics

COURSEWORKS AND PROJECTS

Machine Learning and Statistical Learning Theory — Fall 2022

- Analyzed genomic data using CRISPR and expression analysis to classify cancers, applying LDA, QDA, LR, SVM, and Random Forests. Achieved precision up to 78% with LDA, 68% with LR, and a notable 98% in specific cases, highlighting model accuracy in cancer classification
- Studied Linear/Logistic Regression, Decision Trees, Random Forests, Naive Bayes, XGBoost, LightGBM, Neural Networks
- Identified five key genes via Random Forests feature analysis, informing CRISPR KO strategies. Presented findings comparing model efficacy to benchmarks, emphasizing the project's impact on cancer genomics

Probability — Fall 2022

- Applied transition matrix and goodness of fit test in Python, simulating 100 data points; affirmed Markov Chain model's accuracy in predicting Carbon Monoxide state transitions
- Modeled Carbon Monoxide air quality time series data using Markov Chains, after data cleansing and state discretization. Achieved an empirical distribution closely mirroring the stationary distribution, validating the Markov Chain model's fit

Collecting, Storing, and Analyzing Data — Spring 2024

- Developing expertise in large-scale data repository creation using traditional and NoSQL solutions for enhanced analytics
- Acquiring hands-on experience in data management and predictive modeling with R, SQL, and the CRISP-DM framework