John Munro

Last update on June 6, 2024

jbmunro4@gmail.com • (813) 546-0723 • www.linkedin.com/in/jbmunro4
575 Brownwood Ave SE • Atlanta, GA • 30316

Summary

Data nerd with strong development skills, a proven history of designing and building creative data products, an intuition for data architecture, and strong leadership capabilities.

I enjoy an intensely challenging and creative career in various forms of Information Science that deepen my existing knowledge of Machine Learning, Data Analysis, Artificial Intelligence, Systems Design, and Cognitive Science while broadening my world view. I want my views to be challenged on a regular basis and to always be learning new things.

I have a passion for understanding the nature of Intelligence and strive to formalize or leverage aspects of Human Intelligence in everything that I build, whether from an AI and Automation perspective or from the perspective of Product Design and Usability.

Experience

AMP Recover — Atlanta, GA

Principle Software Engineer *Apr '19 – Feb '24*

Sole engineer for the first few years before hiring 2 part time engineers.

Upon joining, had to play archeologist with a live product that had not been touched by the previous engineer for a year and came with zero documentation or deployment instructions.

Implemented innumerable bug fixes and new features across backend, web, and mobile.

Reimplemented our aws infrastructure and deployment code using ansible and capistrano.

Upgraded our Ionic mobile app to use Capacitor instead of Cordova.

Built new mobile app for fitness evaluations using Asensei's motion tracking and coaching SDK.

Met directly with customers to understand needs and discuss designs.

Environment: Small healthtech startup with the mission to facilitate world-class orthopedic rehabilitation outcomes through individually tailored recovery protocols and deep patient engagement. Technologies included Ruby on Rails, Angular, the Ionic Framework, Capacitor, Cordova, iOS, Android, Amazon Web Services, and Ansible.

Hopper — Remote

Senior Software Developer *Sep '21 – Oct '23*

Designed and implemented high throughput search algorithms and infrastructure for Hopper's Virtual Interlining (VI) products.

Reduced VI's compute spend by 10x through widescale refactoring of our online algorithms.

Designed, built, and maintained VI's ETL infrastructure for offline elements of our online algorithms.

Improved price and duration beat rate and VI market share by implementing novel algorithmic features.

Rigorously validated new features before and after release.

Prototyped a Random Forest model that dramatically improved price prediction accuracy.

Participated in on-call rotations.

Environment: Hopper is an Online Travel Agency which employs the Single Threaded Ownership (STO) model across its 300+ person engineering teams. My small team was responsible for shopping the currently available airfare and returning optimized ticket combinations to provide customers with alternatives to the single-ticket options provided by airlines directly. Technologies included Scala, Google Cloud Services, Kubernetes (k8s), Python, and Airflow. Machine Learning and Statistical Methods included Exploratory Data Analysis, Time Series Analysis & Forecasting, and Random Forests.

Rented.com — Atlanta, GA

Software Developer & Data Scientist *Jul '18 – Apr '19*

Sole researcher, developer, and architect of early-stage prototypes for automated or machine-assisted rental property appraisals. Modeling tasks included forecasting total monthly and yearly revenue of short-term rental units as well as evaluating the performance of property managers themselves. Models were based on the AirDNA global dataset, the industry standard for rental property evaluation.

Prototypes were built in Python using Elasticsearch as the primary data store and real-time statistical aggregation engine.

As part of the software development team, I was responsible for the full stack implementation of various small features and improvements to our internal and customer facing financial tools.

Environment: First data scientist at a small fintech startup with the mission to provide loans to property managers of short-term rental units. My role was at the intersection of the software development and product teams where our mission was to support internal finance and customer success representatives with rental property appraisals through accurate revenue forecasts and risk assessments. Technologies included Python, Elasticsearch, Ruby on Rails, ClojureScript, Postgres, SciPy, NumPy, Scikit-Learn, and Matplotlib. Machine Learning and Statistical Methods included Exploratory Data Analysis, Time Series Analysis & Forecasting, Linear Regression, a variety of Clustering techniques, etc.

Brightbloom (formerly eatsa) — San Francisco, CA

Senior Data Engineer & Lead Data Scientist — *Mar '16 – May '17*

Data Science Lead on Menu Personalization and Dynamic Pricing. Collaborated closely with Product, Design, Engineering, the Food and Operations teams, the Consumer Science team, Analytics, and Executive Leadership to design and implement the initial technical solutions for Menu Recommendations and Frequency-Based Discounts. Designed various pricing models in collaboration with Finance. Designed various Menu Recommendation Algorithms in collaboration with the Science Team, Food Team, and Engineering. Designed and implemented Cibo, a production micro-service built on top of Django to compute and serve personalized menus and prices to our customers.

Technical Lead on Project Blackbird, a very high priority initiative led by the CEO, to measure and improve operational efficiency. Rapidly designed and implemented a custom Inventory Management Solution after arduously working through complex operational requirements, accounting requirements, executive reporting requirements, and intensely dirty data. Generated product requirements and evaluated inventory management solutions, workforce management solutions, and Enterprise Resource Planning solutions (ERP's). Solely responsible for the vast majority of technical implementation for a highly customized ERP integration. This included eatsa's first production micro-service which was built in Python on top of the serverless framework.

Significantly improved accuracy and removed volatility in the customer-facing Order Delivery ETA provided to customers once they've purchased food through the mobile apps.

Collaborated tightly with the analytics team to rapidly stand up the initial analytics infrastructure in preparation for Business Intelligence and advanced Machine Learning applications. Provided a centralized data warehouse to correlate and analyze information across a mix of proprietary and third party data sources across the engineering, marketing, product, and science teams. Worked closely with executive leadership and the analytics team to provide frequent board updates and ad-hoc analysis.

Environment: First Data Engineer and Data Scientist at this diverse, early stage fast food startup focused on hardware and software automation toward the goal of quick, healthy, and affordable food. Technologies included Ruby on Rails, Python, iOS, Redshift, DynamoDB, Redis, PostgreSQL, Looker, Tableau, Django, SciPy, NumPy, Scikit-Learn, and Matplotlib. Machine Learning and Statistical Methods included Bayesian Inference, Collaborative Filtering, Exploratory Data Analysis, Queuing Theory, Time Series Analysis & Forecasting, Linear Regression, and Graphical Models.

Anomali (formerly ThreatStream) — Redwood City, CA

Principle Data Scientist, Labs Team — *May '15 – Mar '16*

Individual contributor with a high level of autonomy and a wide range of responsibilities across the entire company. My tasks frequently required Core Product Engineering, Threat Intelligence Research, Exploratory Data Analysis, Data Visualization, Business Intelligence (BI) Metric Design, End to End Product Design, and User Experience (UX) Design.

Designed and engineered the next iteration of Anomali's Machine Learning-based Domain and IP Reputation engine, the core product's central use case. This new, highly distributable micro-service included very significant improvements in model accuracy, response time, and throughput. It was designed to auto-tune the underlying Machine Learning model and report thorough goodness metrics for regular, semi-automated updates as feature distributions, data sources, and code change.

Worked on targeted research projects with security analysts, vetted new data sources, designed gamification mechanisms for our online communities, and contributed to the sales team by integrating Anomali's many data sources to give a 360° view of customers as they move through the sales process and make use of Anomali's SaaS security platform.

Environment: Intermediary between Labs (Threat Research Team) and the core product engineering teams. Technologies included Python, SciPy, NumPy, Scikit-Learn, Matplotlib, Django, PostgreSQL, Splunk, the ELK Stack (Elasticsearch, LogStash, Kibana), Looker, HTML, CSS, JavaScript, jQuery, Angular, R, and GGPlot2. Machine Learning and Statistical Methods included Random Forests (Decision Forests), k-NN's, DBScan, SVM's, Time Series Analysis, Time Series Forecasting, Feature Selection, Stratified Sampling, and Simple Random Sampling.

Endgame (acquired by Elastic) — San Francisco, CA

Lead Data Scientist, Hiring Manager — *Jun '14 – May '15*

Built a diverse team of product-focused Data Scientists in an effort to create various analytical models for a variety of Cyber Security applications. Implemented cross-functional collaboration to evangelize our capabilities and learn more about potential use cases across the company. Initiated a culture of healthy peer review along with multiple successful initiatives to improve the team's security domain knowledge.

Designed and contributed to a Data Science Platform that supports analytic R&D in both Batch and Streaming environments at scale. This platform serviced a new product aimed at giving System and Security Administrators visibility into cloud based production environments to identify anomalous behavior indicative of security threats, misconfigurations, or inappropriate resource allocation. Our focus was on Multidimensional Anomaly Detection using a combination of Supervised and Unsupervised Machine Learning and Graph-Theoretic approaches to perform Offline (Batch) Clustering and Online (Real Time Streaming) Classification or Regression.

Environment: Small, distributed team working toward Endgame's first commercial product. Technologies included Spark, Scala, PySpark, Storm, Kafka, Redis, Java, Python, NumPy, SciPy, Scikit-Learn, Matplotlib, Pandas, R, GGPlot2, d3.js, HTML, JavaScript, CSS, jQuery, Angular, React, PostgreSQL, Elasticsearch, Hadoop, and Cassandra. Machine Learning and Statistical techniques were largely adapted for our specific use cases, inspired by traditional algorithms such as Random Forests (Decision Forests), SVM, k-NN, k-Medoids, and Time Series Forecasting.

Washington, DC

Senior Backend Developer, Tech Lead — *Oct '12 – Jun '14*

Researched, designed, and implemented a full stack product to identify trends in large, High Dimensional Data sets using Unsupervised Machine Learning methods to aid in data exploration. Iterated closely with the customer to define their needs and determine appropriate solutions which provide a significant advantage over their old methods and redefine how they think about their specific problems. As Tech Lead, I hired, onboarded, and led a small team of developers to support the product going forward.

Environment: Small, customer facing development team within an early stage startup focused on building crucial products for a very high profile government use case. Technologies included Python, Django, R, GGPlot2, NumPy, SciPy, Scikit-Learn, Matplotlib, Pandas, d3.js, HTML, JavaScript, CSS, jQuery, Angular, PostgreSQL, Elasticsearch, Redis, Hadoop, and Pig. Machine Learning and Statistical Techniques included DBScan, k-NN, k-Medoids, Random Forests (Decision Forests), Cluster Analysis, Simple Random Sampling, Stratified Sampling, Principle Component Analysis (PCA), Feature Selection, Random Forests, etc.

Atlanta, GA

Malware Analysis Engineer (Data Scientist) — *Oct '11 – Oct '12*

Designed, built, patented, and presented Clairvoyant Squirrel, a novel real time classification system which utilizes Random Forests and other Machine Learning techniques over Big Data-scale network sensor output to score and classify the maliciousness of domain names.

Researched and implemented a wide variety of AI, Machine Learning and Statistical Models to classify and categorize malicious characteristics of malware executables, TCP packet headers, malware domain names at various points in the DNS hierarchy, and Botnet Command and Control (C&C or C2) communication patterns.

Frequently performed Exploratory Data Analysis, Data Visualization, Statistical Sampling, and ETL using a wide variety of tools, languages, and frameworks. Techniques included Supervised and Unsupervised Machine Learning, Classification, Regression, Cluster Analysis, Natural Language Processing (NLP), Time Series Analysis, Time Series Forecasting, and Dimensionality Reduction.

Environment: Sole data scientist on a small research team focused on developing cutting edge security products within this early stage startup. Frequently used R, GGPlot2, Weka, Java, Python, NumPy, SciPy, Pandas, Matplotlib, Scikit-Learn, NLTK, Hadoop, Pig, Redis, Elasticsearch, MongoDB, RabitMQ, and PostgreSQL. Statistical and Machine Learning techniques included Random Forests (Decision

Forests), k-NN's, SVM's, Neural Networks, Naive Bayes, Decision Trees, k-Medoids, Linear and Logistic Regression, Spectral Clustering, DBScan, Principle Component Analysis (PCA), Latent Dirichlet Allocation (LDA), TF-IDF, Simple Random Sampling, and Stratified Sampling.

Autonomy (acquired by Hewlett-Packard) — Atlanta, GA

Technology Specialist — *Jan '11 – Oct '11*

Rapidly implemented Full Stack Proof of Concepts for targeted customer demos of Autonomy's Machine Learning-based enterprise search technology and vertically integrated solutions. Acted as intermediary between customers and engineering to communicate business requirements and debug solutions. Worked internationally on a high profile customer site to configure, deploy, and test new Autonomy products in a large production environment.

Environment: Very fast paced and high pressure sales environment utilizing proprietary backend and frontend components along with Perl, Python, HTML, JavaScript, and CSS. 80% travel and 20% remote.

Education

Georgia Institute of Technology — Atlanta, GA

Bachelor degree in Computer Science & Minor in Psychology — *2006 – 2010*

Specialized heavily in Artificial Intelligence, Cognitive Neuroscience and Computer Networking. Completed 3 years of Undergraduate Research focused on AI, Automated Story Telling, and Game Design with a thesis on computational models for automating machinima generation. Took courses in Machine Learning, Statistics, Experimental Design, Robotics, Computer Networking, Cognitive Psychology, and Neuropsychology.

Skills Overview

Data Science and Engineering Skills:

Machine Learning: Random Forests (Decision Forests), k-NN, SVM, Decision Trees, K-Means, K-Medoids, DBScan, Hierarchical Clustering, Self Organizing Maps, Neural Networks, Principle Components Analysis (PCA), Supervised Machine Learning (Classification, Regression), Unsupervised Machine Learning (Clustering), Graph Analytics (Network Analysis), Natural Language Processing (NLP)

Statistics: Information Gain Analysis, Linear Regression, Logistic Regression, Holtz-Winters Time Series Forecasting, Data Visualization, Statistical Hypothesis Testing, Statistical Distributions

Software Engineering: Test Driven Development (TDD), Production Software Development, Rapid Prototyping

Leadership Experience: Product Tech Lead, Big Data Software Architecture, Team Management and Hiring, Cross Functional Collaboration, Project Management, Project Planning

Frameworks, Tools, and Libraries:

Python: Django, GGPlot2, Matplotlib, NumPy, PySpark, Scikit-Learn, SciPy

Ruby: Capistrano, Carrierwave, Devise, Ruby on Rails, Searchkick, Sidekiq

JavaScript: Angular, d3.js, jQuery

Mobile: Android, Capacitor, Cordova, Ionic Framework, iOS

Java: Cordova, Spark, Storm, Weka

Scala: Spark

Databases and Big Data Cloud Technologies:

SQL and No-SQL Databases: Elasticsearch, MongoDB, PostgreSQL, Redis

Cloud Technologies: Amazon AWS, DynamoDB, ec2, Google Cloud, Hadoop, HDFS, Kafka, Kubernetes (k8s), Map Reduce, RabitMQ, Redshift, Spark, Storm, s3

Languages: Bash, C, C++, ClojureScript, CSS, HTML, Java, JavaScript, Matlab, Perl, Pig, Python, R, Ruby, Scala, SQL

Business Analyst Tools: Looker, Tableau

Operating Systems: Centos, iOS, Linux, MacOS, Ubuntu, Unix

Interests

Non-exhaustive and in alphabetical order: cooking, data visualization, eating, education, game design and development, robotics, snowboarding, UX and product design