

Ben Bobby

+1(510)461-6737 | benbobyabraham@gmail.com | github.com/benbobyabraham | linkedin.com/in/ben-boby | SF Bay Area

DATA SCIENTIST

Highly skilled and passionate Data Scientist with expertise in engineering metrics, data analysis, and visualization. Seeking a position with a successful, multi-national organization to contribute to improving software quality and influencing processes across the company. Excited to work in a vibrant community of engaged engineers and experienced data scientists.

KEY SKILLS:

Programming Languages and Libraries: Python3, JavaScript, Java, BASH/CMD scripting

Data Science: Data Analysis (NumPy, Pandas), Data Visualization (Matplotlib, Seaborn, SciPy), Machine Learning (scikit-learn, statsmodels), Deep Learning (Keras, PyTorch, Tensor Flow), Computer Vision (OpenCV), Data Processing (PySpark, Apache Storm), ML Algorithms (Statistical Methods, Linear Regression, Classification, Clustering, Ensemble Methods, Decision Trees, SVMs, K-Nearest Neighbors, K-means Clustering, NLP), Model Deployment (Flask, Django, Docker)

Full Stack Development: Django Rest framework, React, Redux, Flask,

Database: MySQL, PostgreSQL, NoSQL (MongoDB), Redis, Hive, HBase

DevOps Tools: Docker, Kubernetes, Jenkins, Ansible, Splunk, GitHub Actions, Grafana, Prometheus, Git, ELK

Cloud Services: AWS (S3, EC2, Lambda, ECS, Kinesis, RDS, DynamoDB), Azure Databricks, Digital Ocean Cloud, GCP

Additional Skills: Big Data (Hadoop, Spark, Kafka, Airflow), NLP Libraries (NLTK, SpaCy), Feature Engineering, Time Series Analysis, Reinforcement Learning, Neural Network Architectures (CNN, RNN, LSTM), Version Control Systems, Data Visualization (Power BI, Tableau, Plotly), Statistical Analysis, Hypothesis Testing, Microsoft Azure, Automated Machine Learning (AutoML) Frameworks (H2O.ai, Auto-sklearn, TPOT)

WORK EXPERIENCE

Senior Data Scientist, NTT DATA Inc. - CDTs Division

May 2022 - Dec 2023

- Worked on the NTS project, a big data ticketing system for Victoria, Australia, the backend for Myki cards.
- Worked on data analytics for Santander banking data, and derived metrics with an automated system.
- Managed and processed large amounts of data, handling over 2 million files from multiple IoT ticketing devices across the state.
- Led the team responsible for data management, including ingestion, cleaning, processing, and storing of processed data in the data warehouse.
- Implemented data backup systems with up to 2 years of data retention across multiple environments.
- Utilized Hortonworks Big Data architecture and worked with technologies such as Apache Spark, Apache NiFi, Apache Storm, Kafka data streaming, HBase database, and Hive data warehouse.
- Implemented Grafana and Prometheus for monitoring and performance analysis.
- Analyzed data incoming from station terminals to identify trends and patterns.
- Generated actionable metrics to drive data-driven decision-making.
- Created a dashboard for the team to monitor and analyze data.
- Deployed the dashboard on the test environment for team usage.
- Championed and supported the use of metrics across the organization, contributing to software quality improvements and implemented machine learning systems for managing high traffic and for high availability.
- Configured and analyzed data using Python, Tableau, Power BI and other relevant tools.

Data Scientist Faculty, Tech I.S. Coding School

May 2021 – Nov 2021

- Created course curriculum for students, focusing on data analysis, data processing with PySpark, and decision trees using Tableau.
- Trained junior data scientists hired as tutors, enhancing their tutoring skills and knowledge.
- Conducted data analysis of internal student data to assess success rates and identify actionable insights.

- Implemented changes based on insights, resulting in a 20% increase in incoming students.
- Performed A/B testing on the signup webpage, leading to UI improvements and improved lead generation numbers.
- Collaborated with application teams to enable data-driven controlled experiments.
- Conducted data analysis and applied data mining techniques to develop actionable conclusions.
- Built tools for data logging and automated data tasks, increasing efficiency in data scientist duties.
- Supported students in Kaggle competitions, supervised machine learning projects, and developed a roadmap for their data science competitive programming journey.

Data Scientist, Thriwe.com (previously Alliances Galore)

Aug 2020 - Mar 2021

- Acquired leads for the marketing process through data mining techniques.
- Built an ETL pipeline and data aggregation system using Scrapy in Python and MongoDB database.
- Scraped over half a million listings from various websites daily.
- Processed and cleaned the acquired data using Pandas and NumPy libraries.
- Created analysis visualizations using Matplotlib and Seaborn to extract insights from the data.
- Developed an automated pipeline for data acquisition, processing, and visualization.
- Created a machine learning model for predicting churn, comparing it to existing data to improve accuracy.
- Reduced workload on the sales team by automating tasks and increasing efficiency.
- Increased the number of tie-ups made by leveraging data-driven insights.
- Developed an automation tool to tag and classify images sent as proof by field executives.
- Implemented Convolutional Neural Network (CNN) models and employed computer vision techniques to perform accurate image classification.
- Deployed the system on AWS EC2 and S3 for cloud-based scalability.
- Created REST API endpoints using Flask for easy access to the data.

Technical Support & Database Administrator, K.C. Jacob & Co.

Feb 2019-June 2020

- Served as a database administrator, responsible for managing the storage of corporate documentation.
- Created an automated system for handling document submissions and filing processes using Python automation.
- Automated the process from submission to filing, reducing the processing time for new consignments from 3 days to under 5 hours.
- Achieved significant reduction in workforce requirement and tasks for the team.
- Conducted analysis on historical data to identify hotspots of delay and improve productivity.
- Utilized data analysis techniques to identify patterns and areas for improvement.
- Implemented measures to optimize productivity and streamline processes based on data insights.

Research Engineer, Vingyan Innovations Pvt. Ltd.

Aug 2018 - Jan 2019

- Developed computer vision and image processing applications for robotics projects.
- Created IoT solutions, including electronics and PCB design.
- Designed products related to robotics and IoT.
- Assisted senior management in making key business decisions.
- Developed courseware for students, which became the company's main product.
- Increased revenue by 80% through the successful launch and promotion of the courseware.
- Conducted data analysis of IoT solutions using OpenHAB platform.

EDUCATION

Golden Gate University, San Francisco

Master of Science, Business Analytics

Hindustan University, Chennai

Bachelor of Technology, Mechatronics and Robotics Engineering

IIT-B University, Bangalore

Advanced Certificate Program in Data Science