Christopher J Caldarella

Data Scientist

Chapel Hill, NC | 720.689.5439 | christopher.j.caldarella@gmail.com | GitHub | LinkedIn | Website

PROFESSIONAL SUMMARY

Experienced Data Scientist and Software Engineer with 10 years of expertise in developing and implementing data-driven solutions for various industries. Proven track record in data analysis, machine learning, and software development. Adept at transforming complex datasets into actionable insights and efficient systems. Passionate about solving challenging problems and driving innovation.

TECHNICAL SKILLS

- **Programming & Scripting**: Python, C#, Ruby, JavaScript, HTML, CSS
- **Data Science and Machine Learning**: Apache Spark, Databricks, NumPy, Pandas, Scikit-Learn, OpenCV, NLTK, SpaCy, LangChain, Matplotlib, Seaborn, Plotly, Azure OpenAI, ChatGPT
- **Databases & Query Languages:** SQL, MySQL, TSQL, PostgreSQL, Relational Databases
- **Cloud & DevOps**: AWS, Azure, Azure DevOps, GitHub Actions, GCP, CI/CD
- **Tools and Libraries**: LangChain, FastAPI, Flask, Streamlit, PyTest, Git, GitHub, Power BI, Docker, Selenium, Beautiful Soup, Virtual Environments
- **Automation & Scripting:** Bash Shell, PowerShell, OOP, Scripting, Batch files
- **OS Proficiency:** Windows, Linux (Ubuntu, Debian), Raspberry Pi OS (Raspbian)
- **MS Office Suite:** Excel, Word, PowerPoint, Outlook
- **Collaboration & Documentation:** Markdown, SharePoint, PowerApps
- **Data Science Core Skills:** Data Analysis, Data Cleaning, EDA, Data Aggregation, Feature Engineering, Predictive Modeling, ETL, Data Pipelines
- **Data Science Specializations:** Regression, Classification, Neural Networks, CNN, NLP, Forecasting, Clustering, Generative AI, RAG (Retrieval-Augmented Generation), Computer Vision
- **Data Science Advanced Techniques:** Transfer Learning, Bayesian Statistics, Probability

PROFESSIONAL EXPERIENCE

Data Scientist Consultant | *Insight Enterprises Inc* | Remote | Oct 2021 – Present

- Delivered advanced data solutions for various clients, transforming and analyzing large datasets to optimize business processes.
- Developed and deployed machine learning models and data pipelines using tools like Python, OpenCV, TensorFlow, AzureML, and GCP Pub/Sub.

  Key Projects:
  - **Government Fire Employees Association**: Assisted with transferring sensitive data related to EMS and Fire incidents from a local government association to Azure using PySpark and building pipelines for medallion architecture in Azure Synapse.
  - **Toy Company USA**: Created data ingestion and transformation pipelines for Shopify API data integration with Salesforce and Ping Identity.
  - **Airplane Sensors USA**: Engineered features and developed an algorithm with 80% accuracy for hydraulic system leakage detection.
  - **Drainpipe Systems USA**: Developed a computer vision model for object detection and report generation using AWS Rekognition and SageMaker.
  - **Supply Chain Company International**: Created a custom Python library for data scientists using the AzureML Python SDK and integrated with VS Code via a Flask API.
  - **Generative AI Solution (Internal)**: Built a web-based Generative AI tool using Azure ChatGPT, LangChain, FastAPI, and Streamlit.
  - **AI Teleport (Internal)**: Developed a metadata interpretation tool using NLTK and SpaCy.
  - **Project Ambition (Internal)**: Assisted in ETL processes for a FinTech team using Databricks.

Staging Specialist | *NCR - Northeast Depot* | Concord, NH | Jul 2014–Mar 2021

- Led hardware and software upgrades, developed Python tools, and managed on-time equipment delivery to numerous clients.
- Created customer "packets" using Kanban workflow, built and maintained corporate images, and streamlined system setup and delivery processes.

Key Achievements:

- Developed tools in Python to assist with migration to new software APIs for on-time delivery.
- Created various automation scripts in Python and C#, enhancing operational efficiency and productivity.
- Managed project delivery details, including shipping, delivery, installation dates, and staging, ensuring timely delivery to over 100 clients.

EDUCATION AND TRAINING

Data Science Immersive | *General Assembly* | *Remote* | *Mar 2021 – Jun 2021*

- Completed 480+ hours of intensive training in data research, aggregation, engineering, and analysis using Python and libraries such as TensorFlow, Keras, and scikit-learn.

PROJECTS

- **Face and Object Tracking**: Utilized OpenCV and Deep Learning to track objects using a Raspberry Pi.
- **Reddit Post Classification**: Applied logistic regression and Naive Bayes models to predict subreddit comments with 71.08% accuracy.
- **Electricity Usage Prediction**: Forecasted electrical consumption in North Carolina using historical data and predictive analytics.
- **Hotdog/Not-Hotdog Classification**: Deployed an Online app for classifying images using various neural net models and transfer learning.
- **Air Quality Analysis**: Conducted exploratory analysis of air quality data in Cleveland, Ohio.

EDUCATION AND CERTIFICATIONS

Rutgers School of Engineering, New Brunswick, NJ
Bachelor of Science, Electrical and Computer Engineering
with a focus in Computer Science and Mathematics

Certificates:

- **Databricks Certified Data Engineer Associate | *Databricks* |** Remote **|** 2024
- **Azure AI Fundamentals | *Microsoft* |** Remote **|** 2023
- **Azure Fundamentals | *Microsoft* |** Remote **|** 2021
- **Data Science Immersive | *General Assembly* |** Remote **|** 2021
- **ScrumMaster | *Scrum Alliance* |** Denver, CO **|** 2020
- **Scrum Product Owner | *Scrum Alliance* |** Denver, CO **|** 2020