Huy Nguyen, PhD

Anaheim, CA

7148602017 huynguyen2406@gmail.com in/huy-nguyen-smart-worker github.com/chemcoder-2020

Data Scientist

Highly motivated and detail-oriented Data Scientist with over 3 years of experience in delivering enterprise-level data products in a software engineering environment. Skilled in transforming complex data into actionable insights and driving innovation through advanced analytical techniques.

- **Computer Vision & Deep Learning**: Developed advanced computer vision applications utilizing **scikit-image** and **Meta SAM models** to enhance semiconductor process metrology, ensuring adherence to industry standards and accelerating inspection cycles.
- **Data Querying & Execution**: Spearheaded the integration of **Pyodide** for efficient script execution in an internal data querying system, optimizing workflows and enabling real-time data analysis for improved decision-making.
- **Data Engineering & Analysis**: Architected end-to-end data solutions in **Databricks** with **PySpark**, creating robust data pipelines that enhanced data accessibility and reliability for multiple teams engaged in analytics and reporting.

Work experience

Lam Research 09/2020 – 02/2024

Data Scientist

- **AI-assisted Computer Vision**: Developed an AI-driven computer-vision engine using Meta SAM and scikit-image, achieving over 70% adoption among process engineers by demonstrating superior segmentation accuracy and efficiency compared to legacy tools and improving confidence in the tool transitioning process.
- **Plasma-Etching Recipe Development**: Leveraged innovative physical modeling and machine learning techniques to create new etch recipes, resulting in a reduction of metal contamination levels to 1e10 atoms/cm^2, exceeding customer specifications by 10x.
- **Data Pipeline Development**: Facilitated seamless data integration and analysis by assisting users in developing scripts through the **Pyodide** interface to extract and plot data for custom analyses. This contributed to a streamlined decision-making workflow across teams.
- **Business Intelligence Dashboard**: Designed a BI dashboard that reduced report creation efforts from days to just a few hours while increasing accuracy. The dashboard provides a comprehensive view of outreach and development efforts, enabling informed decision-making and outcomes analysis.
- **Recognition**: Featured in corporate reports presented to the Corporate VP of the Semiverse Solution team, highlighting significant contributions to improving semiconductor processes and internal data workflows.

Data Science Projects

End-To-End Process Digital Twin – Python | Azure

A set of data products that absorb data from equipment and process measurements to create a digital representation which feeds into the process data modeling and decision-making engines.

Equipment Data Query and Transformation – Pandas | Databricks | PySpark | SQL

Delivered a data transformation procedure to extract, transform, and load equipment/metrology data. Employed Pandas and Azure Databricks interactivity widgets to introduce low-code applications. Created data summary tables tailored to process engineers' requirements.

Computer Vision Metrology on SEM Image Data – OpenCV | Scikit-Image

Developed a computer vision plugin that measure critical dimensions of holes and pillars and implemented proprietary characteristic metrics in the measurement.

Chemical Process Data Science – Scikit-learn | Visualization | Simulation | Statistics

Generated recipes for process optimization and path-finding applications. Examples include:

- Modeled high-aspect ratio (HAR) conductor etch process with physical mechanistic steps.
- Root-cause analysis of metal contamination issue upon wafer entry.

Process Health Checker – Object-Oriented Programming | Pandas | Scikit-learn

Delivered a fault-finding application that identifies failed processes based on prescribed tolerances from engineers, providing a tool to standardize a baseline for R&D data quality.

Business Intelligence Dashboard – Streamlit | Plotly | JIRA | SQL | PowerPoint

Maintained a Bitbucket project of a visualization dashboard to mine internal JIRA data (stored in SQLite relational database) for statistical reports of user adoption and operational cost.

Education

PhD in Physical Chemistry — 08/2015 – 05/2020

University of Illinois at Urbana-Champaign

Researcher with over 200 citations on spectroscopic and electronic properties of nanostructures. Constructed a data pipeline of electron microscope data collection which includes creating A/D circuit boards, writing firmware (C++) and signal processing codes (Python), and publishing results.

BS in Physical Chemistry — 08/2013 – 05/2015

University of California at Berkeley

Researcher. Processed signals from streak cameras with Python and MATLAB.

Skills

Python, Machine Learning, Data Wrangling (ETL), Data Visualization (Azure Databricks, Power BI), Statistical Analysis, Pandas, PySpark, Computer Vision (OpenCV, Scikit-image), SQL, Predictive Modeling, Team Collaboration, Problem-Solving, Communication, Adaptability, Critical Thinking.