RISHABH SOLANKI

+1(508) 717-5407

rsolanki@umassd.edu ◇ LinkedIn ◇ rishabh01solanki.github.io

SUMMARY

As a Data Scientist with a strong background in Computational Physics, I specialize in utilizing machine learning methodologies for analyzing large and complex datasets. I am skilled in Python, Java, SQL, and High-Performance Computing (HPC), particularly in developing and optimizing algorithms for data-intensive tasks. With proficiency in using Linux, Git, and AWS services, I streamline workflows, deploy scalable machine learning solutions, and enhance operational efficiency. My unique blend of skills allows me to tackle machine learning challenges effectively, providing actionable insights that drive decision-making.

WORK EXPERIENCE

Data Scientist — January 2019 - May 2021
GKV Software Solutions — *Dehradun, India*

- Confronted with the challenge of inaccurate machine learning predictions, I applied Python with Numpy, Pandas, and Scikit-Learn for data cleaning and preprocessing. This not only enhanced the accuracy of our Naive Bayes and Random Forest models by 25%, but also significantly improved decision-making reliability, marking an important milestone in our machine learning journey.
- Grappling with time-intensive manual data extraction, I automated the workflow using Python and Pandas, reducing data preparation time by 40%. This allowed for efficient model development with TensorFlow, Keras, and XGBoost, and robust evaluation using libraries like Matplotlib and Seaborn, vastly enhancing overall productivity.
- Faced with scalability issues for machine learning models, I employed AWS services like SageMaker, EC2, S3, Lambda, and CloudFront for efficient training, deployment, and sharing of models. Coupled with Apache Spark for processing large datasets, this approach led to a 20% reduction in deployment time, underlining the value of cloud services in our workflow.
- To tackle stakeholder skepticism about machine learning models, I used interpretable decision tree models, backing their efficacy with clear data visualizations. I also utilized NLP and Deep Learning techniques on unstructured data for better model performance, significantly bolstering stakeholder trust, and facilitating broader acceptance of complex AI models. Throughout all projects, I applied Agile methodologies using Jira and Git, ensuring smooth collaboration, progress tracking, and version control.

Data Science Intern — June 2017 - September 2018
Instruments Research and Development Establishment — *Dehradun, India*

- Engineered a sophisticated data preprocessing pipeline using Java and MATLAB. This pipeline was geared towards real-time interpretation of substantial spectral and image datasets — a cornerstone in photonics research. Through this data science-infused approach, I expedited the extraction of key insights, substantially propelling our research endeavors.
- Fostered cross-functional collaboration to develop cutting-edge solutions for data modeling and predictive analysis. By implementing APIs and using SQL for data manipulation, I streamlined data extraction from multiple sources. This work significantly enhanced research productivity and paved the way for the adoption of more data-driven practices across the team.

PROJECTS

Data-Driven Modeling and Analysis of White Dwarf Mergers — September 2021 - May 2023
UMass Dartmouth — *MA, USA*

The research involved developing a new algorithm for analyzing white dwarf merger processes, focusing on reliability and stability improvements. Leveraging data science and mathematical techniques, the objective was to overcome limitations and refactor the previous algorithms, enhance the accuracy of predictions, and provide critical insights into astrophysical phenomena.

- Utilized Python programming expertise to optimize large-scale simulations of white dwarf mergers. The application of advanced data science techniques and refactored algorithms significantly increased our simulation speed and efficiency, leading to a 20% reduction in computation time. As a result, we could handle datasets 35% larger than before, yielding more accurate predictions and crucial research insights, such as the potential discovery of a new class of astrophysical objects.
- Addressing the issue of slow software release cycles, I streamlined code review, testing, and documentation processes. By adopting Test-Driven Development (TDD) and thorough code reviews, as well as utilizing Git for version control, we cut down our deployment time by 25%.

- Leveraged Linux command line tools on VSCode to integrate software components and automate routine tasks, further enhancing our operational efficiency. This approach led to a boost in productivity and fostered seamless collaboration across the team, demonstrating the value of an optimized workflow in a data-intensive research environment.

EnergyDash November 2022 - January 2023

Developed a detailed electricity usage analysis code ([available here](#)) in Python using Tensorflow and Pandas to track usage in hourly intervals and identify consumption patterns. It also suggests usage optimization strategies and includes monitoring of solar panel energy production.

Time-Series Supernovae Classifier December 2021 - August 2022

Utilized machine learning and numerical computing techniques in Python to successfully classify supernovae by reading FITS file data, extracting features and labels, and training and testing the classifier. This involved applying analytical skills to do system-level design and reliability testing.

Personal Website Design and Development June 2022 - May 2023

Created a comprehensive portfolio website ([available here](#)) to showcase personal projects and skills. Implemented using HTML, CSS, and JavaScript with emphasis on responsive design for optimal display across devices. Incorporated an interactive blog section to share insights and updates on my latest projects. Utilized Git for version control during development.

EDUCATION

University of Massachusetts Dartmouth, Master of Science in Physics 2023
GPA 4.0 Relevant Coursework: Linear Algebra, Computational Physics, Statistics, Applied Mathematics, and Data Structures. USA

University of Petroleum and Energy Studies, Bachelor of Science in Aerospace Engineering 2018
Minor in Math and Physics India

SKILLS

Programming Languages	Python, Java, SQL
Libraries	Numpy, Pandas, Scikit-Learn, Tensorflow, Keras, XGBoost
Machine Learning Techniques	Deep Learning, Natural Language Processing (NLP), Random Forest, Naive Bayes
Data Visualization Libraries	Matplotlib, Seaborn, Plotly
SDLC	Jira, Git, Agile methodologies
Big Data Tools	Apache Spark
Cloud Services	AWS (EC2, S3, Lambda, SageMaker, Bedrock)

CERTIFICATIONS

- [AWS Certified Cloud Practitioner](#) Issued June 2023

EXTRA-CURRICULAR ACTIVITIES

- Member of Society of Physics Students, Hiker, Badminton player