RISHABH SOLANKI

+1(508) 717-5407
rishabh01solanki@gmail.com ◇ LinkedIn ◇ GitHub

SUMMARY

AWS Certified Machine Learning Specialist and Data Scientist specializing in the development and deployment of machine learning models, with a strong proficiency in AWS services, seeking to leverage expertise to deliver business value.

WORK EXPERIENCE

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

- As part of a squad resolving overfitting issues in machine learning predictions, I effectively used Apache Spark and Hadoop in tandem with AWS Glue ETL on EMR clusters for rigorous data cleaning. This approach led to the development of robust models, culminating in a 2% enhancement in their performance and a 10x growth in lead generation.
- Devised an automated data extraction workflow using AWS Glue crawlers, big data tools like Hive and Impala, reducing the data preparation time by 40%. Leveraging this efficient workflow, I developed models using TensorFlow, Keras, and XGBoost on the SageMaker platform, and conducted evaluations using visualization libraries like Matplotlib and Seaborn.
- Addressed scalability issues in our machine learning models by strategically employing Amazon SageMaker for distributed training and optimization of models. I effectively used S3 for data storage, EC2 for compute, and Lambda for automation. Integrated Apache Spark and Snowflake on EMR clusters, leading to a significant 20% reduction in deployment time.
- Developed pipeline workflows using big data tools like Kafka for real-time data processing, thus bolstering stakeholder trust in machine learning models. Applied statistical models, predictive models, and time series analysis, and utilized NLP and Deep Learning techniques on unstructured data for better model performance. Employed Agile methodologies using Jira and Git for efficient collaboration and version control.

Data Science Intern June 2017
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 data modeling and predictive analysis solutions. Leveraged SQL and APIs for data manipulation and utilized Hive for efficient management of structured and unstructured data, significantly enhancing research productivity.

PROJECTS

Sentiment Analysis on Movie Reviews using NLP (TF-IDF) (available here) July 2023

- Utilized various machine learning algorithms, including Logistic Regression, Random Forest, Naive Bayes, and SVM, for sentiment analysis on a corpus of movie reviews, classifying sentiments as 'positive' or 'negative'.
- Implemented extensive data preprocessing techniques such as tokenization, stopword removal, stemming, and vectorization (TF-IDF), enhancing the performance of the machine learning models.
- Achieved superior results with the SVM model, demonstrating successful automation of sentiment analysis. The project has potential applications in areas such as customer feedback interpretation and product recommendation. Future enhancements include exploration of different machine learning techniques and hyperparameter tuning.

Stock Price Forecasting using Support Vector Regression (SVR) and Bootstrap June 2023

Implemented Support Vector Regression (available here) with an RBF kernel to predict stock prices. Incorporated bootstrapping with 1000 samples for model training and calculated prediction intervals. Visualized actual and predicted stock prices along with their associated uncertainty, showcasing the model's accurate predictions and valuable insights.

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

- Developed and implemented a machine learning algorithm for analyzing white dwarf merger processes, securing a $550k NASA grant.
- Used Spark and Hadoop for optimizing large-scale simulations, leading to a 20% reduction in computation time and handling datasets 35% larger than before.

- Adopted Agile methodologies, Test-Driven Development (TDD), and Git for effective version control, reducing deployment time by 25%.

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	Natural Language Processing, Decision Trees, Classification, Regression, SVM
Data Visualization Libraries	Matplotlib, Seaborn, Plotly
SDLC	Jira, Git, Agile methodologies
Big Data Tools	Apache Spark, Hadoop, Kafka, Hive, Impala and Snowflake
Cloud Services	AWS (Glue, Lambda, SageMaker, Kinesis), Docker, Amazon Kubernetes Service

CERTIFICATIONS

- AWS Certified Machine Learning – Specialty July 2023
- AWS Certified Cloud Practitioner June 2023

EXTRA-CURRICULAR ACTIVITIES

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