Janish Parikh

609-721-8313 | parikhjanish27@gmail.com | GitHub | LinkedIn | RPubs | janishparikh.com

TECHNICAL SKILLS

- **Development:** Python, SQL, R, PySpark, MS Excel
- **Tools & Framework:** NumPy, Pandas, Scipy, scikit-learn, Selenium, Hadoop, Tableau, PyTorch, Git, skforecast, Prophet
- **Cloud & Engineering:** AWS technologies, Docker, Kafka
- **Theoretical Knowledge:** Time Series Analysis, Ensemble Learning, Big Data, Gradient Boosted Trees, Natural Language Processing, Recommender Systems, Anomaly Detection, Dimensionality Reduction

EXPERIENCE

Data Scientist | *Rutgers University* / *R, SQL, Python*   *Sep 2021 – May 2023*
- Built data pipelines and tools for geospatial analysis and computing metrics for quantifying gerrymandering, providing valuable insights into the socio-political impacts of redistricting.
- Developed projects to provide students with experience in real-world data analysis, machine learning and data storytelling.
- Created anonymized datasets and test environments to simulate and evaluate algorithms for Roommate Matching.

Data Science Intern | *Koch Industries* / *Python, SQL, Tableau, AWS*   *May 2022 – Aug 2022*
- Developed a cascaded ML model to detect failures in an electrical grid network, resulting in a 11% increase in Balanced Accuracy and a 4% increase in the Precision Score.
- Translated business requirements into comprehensive loss functions, and utilized time series models aimed at increasing reliability, and user confidence in the ML predictions.
- Deployed the model into production and enabled live monitoring that allowed predicting vegetation failure beforehand.
- Optimized data preparation pipelines by leveraging Dask, resulting in a 5x reduction in computational costs.
- Led multiple data scientists' collaboration sessions on 'Detecting Drift in Data' and 'Scaling Pandas'.

Data Scientist | *COSGrid Networks* / *Python, AWS, Spark, Kafka*   *Jan 2020 – Aug 2021*
- Designed, and implemented an end-to-end cyber-security tool that was selected amongst the top 6 finalists in 'Cyber Security Grand Challenge!'.
- Researched and deployed ML models for Host/IP-Profiling, IoT Device Classification, and Real-Time Attack Detection.
- Ingested real-time NetFlow data through Spark Structured Streaming, processed more than 10,000 records per minute and alerted the users of any malicious activity under a minute.
- Built dashboards to provide real-time insights on health of the devices and various machine learning model predictions.

EDUCATION

Master of Science in Computer Science   *Sep 2021 – May 2023*
Rutgers University-New Brunswick | CGPA: 4.0   *New Brunswick, NJ*
Awarded Gold Medal for Outstanding Academic Performance

Bachelor of Technology in Computer Science & Engineering   *Aug 2017 – May 2021*
Indian Institute of Information Technology | CPI: 8.78   *Gandhinagar, India*

Relevant Coursework: Machine Learning I & II, Natural Language Processing, Massive Data Mining, Intro to AI, Foundations of Data Science, Data Structures & Algorithms, Software Engineering, Database Management Systems

PROJECTS

Image to Image Translation | *CV, GANs, PyTorch, W&B*   *[Link]*
- Explored I2I Translation using Pix2Pix GAN to translate Street View Images to Aerial View Images and vice-versa.
- Implemented CycleGAN framework for the task of translating Real to Synthetic and Live Pizza Domains.
- Researched the drawbacks of CycleGAN framework and proposed an enhanced CycleGAN by incorporating VGGPerceptual Loss in CycleConsistency that attained a 10% improvement by reducing unrealistic artifacts.

Anomaly Based Intrusion Detection System | *PySpark, Spark MLLib, Random Forest*   *[Link]*
- Trained and persisted a Spark ML Pipeline on UNSW-NB15 dataset using Random Forest Classifier to classify incoming network flows into normal and malicious, attaining an accuracy of 97.28%.

AI Maze Solver | *Deep Learning, PyTorch, Bayesian Inference*   *[Link]*
- Implemented multiple AI agents using Repeated A*, Inference, and Bayesian Networks for optimally searching a hidden target within a maze using Python and NumPy.
- Trained CNN with Dense layers using PyTorch to imitate these agents obtaining an accuracy of 92% in solving mazes.