

Janish Parikh

New Brunswick, NJ | 609-721-8313 | janishparikh5@gmail.com | [GitHub](#) | [LinkedIn](#) | [RPubs](#) | [Portfolio](#)

EDUCATION

Master of Science in Computer Science

Rutgers University-New Brunswick | CGPA: 4.0

Sep 2021 – May 2023

New Brunswick, NJ

Bachelor of Technology in Computer Science & Engineering

Indian Institute of Information Technology | CPI: 8.78

Aug 2017 – May 2021

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

EXPERIENCE

Graduate Teaching Assistant | *Rutgers University* / *R, Shiny, Python, scikit-learn*

Sep 2021 – May 2023

- Taught over three hundred students about Data Analytics, Statistical Modeling, and Machine Learning using Python/R
- Proposed projects and curricula for courses, Data Science Capstone Project and Algorithms in the Wild
- Created anonymized datasets and test environments to simulate and evaluate algorithms for Roommate Matching
- Designed python packages to perform Geospatial analysis, Redistricting, and evaluate Gerrymandering

Data Science Intern | *Koch Industries* / *Python, SQL, Tableau, AWS*

May 2022 – Aug 2022

- Developed a user focused cascaded ML model to detect failures in an electrical grid network, using an ensemble model and an exponential smoothening model, resulting in a 11% increase in Balanced Accuracy and a 4% increase in the Precision Score for the critical class over the existing
- Translated complex business requirements into comprehensive loss functions, and utilized exponential smoothening and other time series models aimed at increasing reliability, and user confidence in the ML predictions
- Deployed the model into production enabling live monitoring that allowed predicting vegetation failure in NZ beforehand
- Optimized the data preparation pipelines by leveraging Dask's lazy evaluation technique, resulting in a 5x reduction in computational costs and significantly increasing the speed of data processing
- Led two data scientists' collaboration sessions on 'Detecting Drift in Data' and 'Scaling Pandas'

Data Scientist | *COSGrid Networks* / *Python, AWS, Spark, Kafka, SQL, Elasticsearch, Plotly*

Jan 2021 – Aug 2021

- Ideated, designed, and implemented an end-to-end product for cyber-situational awareness using big data technologies that was selected amongst the top 6 finalists in 'Cyber Security Grand Challenge!' organized by the DSCI
- 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 during peak hours and alerted the users of any malicious activity under a minute
- Delivered a customized dashboard using Dash and Plotly, providing stakeholders with real-time insights on network devices and various machine learning model predictions
- Built an ITIL Helpdesk REST API using Django Rest Framework and integrated it with an existing component 'Guider' that resulted in saving 20 man-hours per week

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

Conversational Movie Recommendation System | *Python, NLP, Recommendation Systems*

[\[Link\]](#)

- Created a movie recommendation system leveraging conversational user data, external critic data, and domain adaptation techniques, which is a re-implementation of [paper](#) and obtained a 3% improvement by performing hyperparameter tuning on all three collaborative filtering approaches: KNN, SVD, and SVDpp
- Experimented with neural collaborative filtering approaches employing Neural Matrix Factorization and obtained comparable results of RMSE=1.232 and MAE=0.9569

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%

TECHNICAL SKILLS

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