

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

Specialization: Machine Learning and Artificial Intelligence

Sep 2021 – May 2023

New Brunswick, NJ

Bachelor of Technology in Computer Science & Engineering

Indian Institute of Information Technology, Vadodara | CPI: 8.78

Aug 2017- May 2021

Gandhinagar, India

Relevant Coursework: Machine Learning I & II, Natural Language Processing, Massive Data Mining, Introduction to Artificial Intelligence, Data Analytics, Data Structures & Algorithms, Software Engineering, Database Management Systems

TECHNICAL SKILLS

- **Development:** Python, SQL, R, Pyspark, MS Excel
- **Tools & Framework:** NumPy, Pandas, Scipy, Seaborn, Matplotlib, scikit-learn, Spark MLlib, ggplot, Geopandas, Shapely, PyTorch, PyOD, Pytest, Dask, Git, Linux, Parquet
- **Cloud & Engineering:** AWS Sagemaker, AWS EMR, AWS Lambda, Docker, Apache Airflow, Kafka, Databricks
- **Databases & Visualization:** MySQL, AWS Redshift, Druid, Amazon RDS, Tableau, Plotly
- **Machine Learning:** EDA, A/B Testing, Ensemble & Multimodal Learning, Big Data, GANs, Time Series Analysis, NLP, Recommender Systems, Anomaly Detection

EXPERIENCE

Graduate Teaching Assistant | Rutgers University / New Brunswick - NJ

Sep 2021 – Present

- I taught students about Data Analytics, Statistical Modelling, and Machine Learning Algorithms using Python and R
- Design projects and curriculum for courses; Data Science Capstone Project and Algorithms in the Wild
- Built anonymized datasets and test environments to simulate matching algorithms for Roommate Matching
- Developed python package and gathered data to perform Redistricting, Geospatial analysis and evaluate Gerrymandering

Data Science Intern | OnPoint - Koch Industries / Wichita - KS

May 2022–Aug 2022

- Developed SOTA Machine Learning Models to predict an incoming equipment failure in electrical network
- Improved the Balance Accuracy Score by 11% and the Precision Score for the critical class by 4% by performing feature engineering, outlier treatment, complex model evaluation and hyperparameter tuning
- Collaborated with multivariate teams and SMEs to deploy trained models and gauge performance improvement
- Streamlined and optimized data analysis/visualization and data preparation pipelines using Dask, Vaex, Plotly and SHAP

Data Science Intern | COSGrid Networks / Chennai - India

Jan 2021–Aug 2021

- Ideated, designed, and implemented an end-to-end product for Cyber-Situational Awareness using big data technologies Spark, Kafka, Druid, Elasticsearch and AWS
- Developed Continual Learning pipelines for 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
- Project was amongst the top 5 finalists in 'Cyber Security Grand Challenge!' organized by Data Security Council of India

PROJECTS

Image to Image Translation

[\[Link\]](#)

- 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 Pizza 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

Maze Solver

[\[Link\]](#)

- Designed multiple AI agents using Repeated A*, Inference, and Bayesian Networks for optimally searching a hidden target within a maze using Python and NumPy
- Optimized these agents to find the targets in $(101) \times (101)$ dimension mazes under 20 milliseconds
- Built a CNN with Dense layers using PyTorch to imitate these agents obtaining accuracy of 92% in solving the mazes.

Conversational Movie Recommendation System

[\[Link\]](#)

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

NYC Citibike Ride Duration Prediction

[\[Link\]](#)

- Examined Citi Bike ridership data 2019, with daily NYC weather data, to study the impact of weather on shared bike usage and generate a predictive model which can estimate the average trip duration of each ride obtaining an RMSE of 5.5 and an adjusted R-Squared score of 0.72