Prasham Sheth

Data Scientist

Phone: +1 (516) 707-1668
Email: p.d.sheth@columbia.edu
LinkedIn: http://www.linkedin.com/in/prasham-sheth

SUMMARY

- I currently work as a Data Scientist at the SLB Software Technology Innovation Center (STIC) in Menlo Park, California
- My research interests include Machine Learning and Deep Learning based approaches for solving complex problems in the fields of Computer Vision, Prognostic and Health Management, and Time-Series Analysis. Further, I am focusing on Hybrid modeling techniques involving Physics Informed Machine Learning

EDUCATION

Columbia University New York, NY
Master of Science in Data Science, GPA: 4.08/4.00 Dec 2020
Coursework: Machine Learning, Applied Machine Learning, Applied Deep Learning, Statistical Inference & Modeling, Personalization Theory, Natural Language Processing, Algorithms for Data Science, Computer Systems, Exploratory Data Analysis and Visualization

Nirma University Ahmedabad, India
Bachelor of Technology in Computer Engineering, GPA: 9.50/10, Rank: 2/900 May 2019
Coursework: Machine Learning, Deep Learning, Artificial Intelligence, Linear Algebra, Algorithms

WORK EXPERIENCE

SLB Software Technology Innovation Center (STIC) Menlo Park, CA
Data Scientist Feb 2021 - Present

Physics-Informed Machine Learning and Time Series Analysis

- Developed Physics Informed Machine Learning based Hybrid Framework to create an advisory system that identifies the regions with risky Stick/Slip conditions and outputs an optimal operating window for drilling the future stands
- Devised Hybrid framework using Physics Informed Machine Learning for digitally generating LWD logs (Gamma-Ray logs) in real-time to increase efficiency and robustness of log collection process
- Working towards implementation of Physics Inspired Machine Learning based toolbox for Time Series data

Computer Vision

- Researched and implemented various Computer Vision based use cases to improve Health Safety and Environment
- Evaluated 5+ state of the at solutions for Object Detection, Object Tracking, and Pose Estimation. Fine-tuned different object detection models and integrated them into the pose estimation pipelines for increased performance for field datasets

Media Center of Art and History at Columbia University New York, NY
Graduate Research Assistant Mar 2020 - Dec 2020

- Engineered way to automate process of slide analysis for collection of slides, deploying image processing and ML/DL techniques to detect originality of photographic images in 35mm slides collection
- Formulated pipeline to replicate entire manual process by initially filtering images based on citation used in slides, followed by identifying presence of Halftone patterns (Precision: 95%, Recall:82%)

Unilever Englewood Cliffs, NJ
Data Science Analyst Mar 2020 - Dec 2020

- Developed and deployed an application to streamline the feature extraction and data engineering process for Process Analytics Engine
- Supported development of Process Analytics Engine to get the insights interactively using the data collected from production unit

Swiss Reinsurance America Corporation Armonk, NY
Data Scientist Intern May 2020 - Aug 2020

- Designed and developed Smart Underwriting Framework to generate scores for each submission based on propensity to bind; Prioritizing processing for submissions based on scores improved Cumulative Gains by 20%
- Analyzed Drift in Data using KL Divergence and Logistic Regression-based models, constructed technique to retrain model dynamically in production setting to ensure Domain Adaptability
- Researched various papers, articles, and datasets available for Individual Health Forecasting; Review task resulted in document summarizing 89 research papers involving DL-based approaches for Electronic Health Records

Samsung R&D Institute Noida, India
Research Intern Jan 2019 - May 2019

- Researched various On-Device AI solutions as part of AI core team and contributed to enhancing quality of services provided by Samsung for its mobile devices; Produced ML/DL models by utilizing Scikit-learn, TensorFlow, TF-Lite frameworks
- Devised techniques for Facial Anti-Spoofing System leveraging various ML/DL methods in Python and deployed it as an Android Application
- Analyzed different On-Device AI solutions for health and multimedia services on low-end Samsung smartphones with 1GB of RAM to ensure robustness of solutions

PUBLICATIONS

Journal Publications:

[1] Param Popat, **Prasham Sheth**, and Swati Jain. "Animal/Object Identification Using Deep Learning on Raspberry Pi." In Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 1, pp. 319-327. Springer Singapore, 2019.

[2] **Prasham Sheth**, Priyank Thakkar, and Praxal Patel. "Optimal Location Prediction for Emergency Stations Using Machine Learning." International Journal of Operational Research. 2022.

[3] **Prasham Sheth**, Sai Shravani Sistla, Indranil Roychoudhury, Mengdi Gao, Crispin Chatar, Jose Celaya, and Priya Mishra. "Real-Time Gamma Ray Log Generation from Drilling Parameters of Offset Wells Using Physics-Informed Machine Learning." SPE Journal (2023): 1-11.

Conference Publications:

[1] **Prasham Sheth**, Indranil Roychoudhury, Crispin Chatar, and José Celaya. "A Hybrid Physics-Based and Machine-Learning Approach for Stick/Slip Prediction." In IADC/SPE International Drilling Conference and Exhibition. OnePetro, 2022.

[2] **Prasham Sheth**, Sai Shravani Sistla, Indranil Roychoudhury, Mengdi Gao, Crispin Chatar, Jose Celaya, and Priya Mishra. "Real-Time Digital Log Generation from Drilling Parameters of Offset Wells Using Physics Informed Machine Learning." In SPE/IADC International Drilling Conference and Exhibition. OnePetro, 2023.

ACADEMIC PROJECTS

Energy Efficient AI on Edge Devices (Master Thesis Project) Sep 2020 - Dec 2020

- Developed techniques for compressing Deep Learning Models for faster inference on edge devices and reduced carbon footprint in association with GE Research
- Researched Post-training quantization, Quantization Aware Training (QAT), and model Pruning techniques for model compressing
- Designed low latency compressed models with minimal accuracy drop for formulated models for multiple datasets

Hybrid approach combining Content and Model-based Techniques for Recommendation Sep 2019 - Dec 2019

- Created Restaurant Recommendation System based on Yelp dataset (2019) (6,685,900 reviews, 192,609 businesses, 200,000 pictures, ten metropolitan areas, over 1.2 million business attributes: hours, parking, availability, and ambiance)
- Developed Recommendation System utilizing various techniques: Alternating Least Square based Matrix Factorization, Factorization Machines, Content-based Recommendation capturing information from Images/Text-based reviews

- Accelerated performance of Recommendation Engine by engineering way to combine results from various approaches to cultivate strengths of each model and get more generalized recommendations

Optimal Location Prediction for Emergency Stations Jul 2018 - Jan 2019

- Identified and engineered various influencing parameters namely location attributes, population density, traffic, navigation, frequency of ambulance calls, and weather conditions
- Formulated various ML/DL Models to identify best suitable for Travel Time Estimation between different locations of city; Travel Time estimated from XGBoost is used to drive K-Medoids for predicting optimal locations
- Conceptualized approach demonstrated promising test results - Decreased turnaround time along with reduced utility of resources; for Staten Island: average time reduced by 6 seconds utilizing 14 Fire Stations in comparison to 19 actual ones

End-to-End Sentence Level Lipreading Jan 2018 - Apr 2018

- Designed model for performing end-to-end sentence level lip-reading, rather than approaches of detecting individual words by simultaneously recognizing spatiotemporal visual features and sequence model; Improved performance by approximately 10% over the LSTM Baseline
- Structured method to process frames (captured at 25 fps from GRID Corpus) leveraging combination of CNN and Bidirectional Gated Recurrent Units (Bi-GRUs) to enhance performance of end-to-end sentence formation

SKILLS

- **Programming Languages:** Python, SQL, R, Java, C++, C
- **Tools and Technologies:** Scikit-Learn, NumPy, Pandas, Statsmodels, PyTorch, OpenCV, Scipy, Google BigQuery, Oracle DS, MongoDB, Google Cloud Platform, GitHub, LaTeX