

PRINCIPAL DATA SCIENTIST

Experienced Data Scientist with a track record of delivering innovative AI-powered solutions. Expertise in leading cross-functional teams, developing advanced ML models, and implementing state-of-the-art algorithms. Transformed organizations, streamlined workflows and established fault tolerant and scalable systems. Accomplished in building robust parsers, enhancing accuracy, and driving multilingual capabilities.

- Predictive Modelling and Analytics
- Natural Language Processing (NLP), Natural Language Generation (NLG)
- Information Extraction and Retrieval
- Reinforcement Learning
- Computer Vision and Audio Transcription
- MLOps and DevOps
- Stakeholder Management
- Conflict Management
- Team Leadership and Incubation

PROFESSIONAL EXPERIENCE

VMOCK INDIA PVT. LTD., Gurgaon, India

7/2018 – present

Career Analytics Platform powered by AI, revolutionizing the candidate-employer-university dynamics. Delivering a comprehensive suite of services including career coaching, instant feedback, intelligent hiring tools, smart editors, and career analytics. Empowering clients with a one-stop platform for unparalleled career success.

Principal Data Scientist, 9/2021 – present

Managed cross-functional Data Science teams (15 Data Scientists & Data Analysts) across business verticals, providing technical and stakeholder management and contributing to research and development of data driven solutions. Mentored teams working on diverse data science problems, including Ranking, Search, Retrieval, Analytics, Entity Recognition (NER), Large Language Models (LLMs), Language Generation (NLG), Grammar Correction (GEC), Computer Vision, Recommendation Systems and Ontologies.

- Restructured Data Science teams and streamlined machine learning development cycle using MLFlow, AWS Sagemaker, EKS (MPS for GPU Partition), resulting in 20-40% improvements in lead & cycle times.
- Implemented containerization of machine learning applications, CI/CD Pipelines, Application Performance Monitoring (APM) & Code Inspection tools, Load Metrics, QA Automations for fault-tolerant and scalable systems thereby reducing infrastructure cost by over 50% for data science teams.
- Mentored team in developing an in-house speech transcription pipeline with VAD (Voice Activity Detection), ASR (Speech Recognition), and Forced Alignment for word level timestamps, saving over few thousands USD/month on 3rd party external APIs and achieving a 29% reduction in word error rates.
- Strategized new initiatives like BestMatch for intelligent hiring with explainable FitScores and SMART Interviews for in person interview practice featuring both verbal and non-verbal feedback.
- Developed a chatbot for VMock utilizing OpenSearch as a vector database and Azure OpenAI APIs (GPT), showcasing conversational process similar to LangChain.

Senior Data Scientist, 12/2019 – 9/2021

Supervised a team of 3 people overseeing all research, development, reviews and deployment for the Resume Parser, powering the flagship Resume Product used across more than 130+ Countries and in leading Educational Institutes such as CMU, Stanford, Chicago Booth, University of Oxford, Columbia, Kellogg, INSEAD etc.

- Finetuned encoder LMs (BERT, RoBERTa) using PyTorch and Hugging Face for superior classification and entity recognition achieving 5-12% improvement over BiLSTM-CRF on various datasets.
- Launched multilingual parsing capability, solving the Reverse Translation and Token Alignment problem using unsupervised IBM Model 1 and HMM models.

- Engineered a highly efficient event-driven task-queue architecture for the Resume Parser, leveraging tools like Celery, Redis, and SNS, significantly reducing infrastructure costs and job fail rates.
- Implemented a high-speed N-gram super-section classifier with outstanding performance, boasting an impressive 81% top-1 accuracy and an exceptional 95% top-3 accuracy.
- Created an in-house Joint Machine Translation and Token Alignment system with a Transformer-based encoder-decoder model, achieving a 44 BLEU Score and a 7 AER.

Data Scientist, 7/2018 – 12/2019

Improved and maintained a robust resume parser that required converting unstructured data into structured data and extracting useful information using techniques of NLP.

- Trained in-house BiLSTM-CRF based NERs with Character Embeddings and Attention Mechanism delivering high recall and precision (>92 % F1 Score) for recognizing various entities as positions, organizations, locations, universities, degrees, section titles etc.
- Executed the migration of a complex codebase of ~50,000 lines from Python2 to Python3, ensuring compatibility and eliminating deprecated features, resulting in enhanced performance and maintainability.

ARTIFACIA INC., Bangalore, India

5/2017 - 7/2017

Research Intern

- Implemented a Machine Learning algorithm using VGG16 ConvNet to extract image features and a LSTM-based recurrent layer for label predictions, achieving a significant 5-6% improvement over the company's best model.
- Designed a text and vector similarity algorithm to filter Instagram images based on user text query and hashtags, incorporating 3rd party image tagging API and user-created hashtags for enhanced accuracy.

TECHNICAL SKILLS

Programming Languages: Python, C, C++, CUDA, OpenMP/MPI, Golang, OCaml

Machine Learning: Tensorflow & Keras, PyTorch, Hugging Face, ONNX, Deepspeed, NVIDIA Triton

DevOp / MLOp Tools: Docker, Kubernetes, MLFlow, DVC, Celery, Redis, Azure Cognitive Services, AWS ECS, AWS Fargate, AWS, Sagemaker, AWS OpenSearch, AWS ElasticCache,

Web Programming: PHP, HTML, CSS, JavaScript, ONNXjs, Laravel, Jekyll

RESEARCH EXPERIENCE

AUTONOMOUS OBJECT TRACKING WITH UAVS [GITHUB], IIT Delhi, 2018 - 2019

- Developed an advanced autonomous agent using computer vision and deep reinforcement learning techniques, including Q-Learning algorithms and experience replay for real-world object tracking.
- Utilized a drone and realistic physics simulations to track a moving car, leveraging a customized version of Microsoft AirSim Simulator for training and testing.

OBJECT DETECTION AND TRACKING FOR DRIVERLESS CARS [GITHUB], IIT Delhi, 2016

- Evaluated Object Detection and Tracking models such as region proposal based classification R-CNN models (Fast, Faster R-CNN) as well as regression based models such as YOLO for detection.
- Utilized detection models and LSTMs to extract features, generate heatmaps, and handle occlusions for robust tracking in sequences/videos.

EDUCATION

Bachelor of Technology, Mathematics and Computing
Indian Institute of Technology, Delhi (IIT-DELHI)