Prashant Gupta

Senior Data Scientist

📱 +91-9406987509 | ✉ prashantgpt91gmail.com | 🐙 github.com/prashantgpt91 | in linkedin.com/in/prashantgpt91

Summary

As a Data Science professional with 8 years of Industry experience, I work cross-functionally to deliver data science solutions. I ensure success by fostering alignment across Product, Design & Engineering and advocating for customer value. I have lead teams through the full research and development lifecycle, while supporting individuals' professional growth. My research interest lies in Model Explainability & Causal Learning.

Skills

Machine Learning	Supervised and unsupervised classification and regression. Object detection, Entity extraction, reinforcement learning
Data Engineering	Architecting high-volume ETL and machine learning pipelines on AWS using FastAI, Celery, Redis, uWSGI, glue, lambda
Leadership	Leading 8 member team. Scoping complex tasks and acting as a data advocate across multiple teams.

Education

Masters in IT (Indian Institute of Information Technology ~)   *M.P, India*
B.Tech + M.Tech, Dual Degree   *Jul. 2010 - May 2015*
- Data Mining, Multiobjective Optimization methods, Artificial Intelligence, Machine Learning, Statistical Learning

Experience

TOTAL EXPERIENCE 8 YEARS, RELEVANT EXP 7.5 YEARS

TCS - Envestnet D&A   *Bengaluru, India*
Senior Scientist   *Nov 2022 - Present*
- Heading Data Science & Engineering team of 8 members including big data developers, data analyst & data scientists. Driving innovation in Ingestion platform having per cobrand ingestion of 10 millions users & billions of transactions via serverless architecture on AWS.
- Part of User level enrichment team built on Knowledge graph + Graph Database empowering 15Mn user nodes, 4Mn Merchant nodes and 35k geo nodes to facilitate business queries.
- Developed forecasting model for oktospend app team with Mean Absolute Percent Error < 10% which identifies recurring income and forecast financial obligations while accounting for fluctuations leveraged from historic data.
- Leveraged PySpark to perform advanced data processing and analysis, enabling optimized data transformations and data cleansing operations.
- Developed and maintained scalable data pipelines using Snowflake, ensuring efficient data ingestion, transformation, and loading processes for large-scale datasets.
- Designed and implemented complex workflows using Airflow, automating data pipeline orchestration, scheduling, and monitoring for seamless data integration and processing.
- Utilized AWS Lambda functions to enable serverless computing and efficiently process real-time data streams, ensuring timely and accurate data insights.
- Created and managed data processing pipeline and metadata repositories using AWS Glue, facilitating data discovery, cataloging for enhanced data validation.
- Implemented robust serverless workflows using AWS Step Functions, orchestrating and coordinating various AWS services to automate data processing tasks, improving overall workflow efficiency and reliability.
- **Technical Skills:** AWS Glue, Lambdas, Step function, Airflow, EMR, Parquet, Pyspark, pytest

Envestnet Data & Analytics   *Bengaluru, India*
Principal Engineer / Team Lead   *July 2018 - Nov 2022*
- Heading Data Science team of 5 members including data analyst & data scientists. Driving innovation in Information extraction of a platform with 33 mn Active users per month.
- Lead Data quality initiative across 4 containers spanning across 400+ entities & achieving weighted average accuracy of 96% for top high volume 20 entities.
- Handled 29% Service request volume through autohealing pipeline. Population percentage of 73 & Accuracy percentage of 60 was achieved in Site conversion, which increased the velocity of site conversion pipeline by at least 2X
- Analytics for Credit Risk Modeling, where lending decisions were made by incorporating masked data, outlining the income & spending behaviors of prospective borrowers throughout their everyday lives.
- Pushing Data quality accuracy metrics to 97.7 for top 20 entities. Increasing SR % volume from 29 to 35.
- Transaction type identification & Transaction to holding mapping. Achieved an accuracy of +94% in both the modules.
- **Technical Skills:** Python with PyTorch, NumPy, Matplotlib, Pandas, Scikit-learn, C++, Ubuntu Linux, Git.

StegoSOC formerly Cloudadic Inc.

Bengaluru, India

Machine Learning Engineer

Mar 2016 - Jul 2018

- Worked with a team of four people building OCR engine. We worked on building Receiptbot, which is an OCR processing framework, right from prototyping to production. My role is to mentor each module plus developing label detection & training deep neural nets for Text Recognition & Text Detection.
  Tools | Tech : Python, OpenCV, Keras, RNN, Transfer Learning, Computer Vision, NLP for NER, pytorch, airflow, JIRA
- Built Topic Modelling Model for our product to enable business domain classifier. I was responsible for end to end delivery of this Model.
  Tools | Tech : LDA, Python
- Build Address detector and recogniser from any arbitrary webpage using a hybrid approach of regex and decision tree classifier for Ziwid. I was responsible for end to end delivery of this Module apart from integration to main product.
  Tools | Tech : Python, C++, Decision trees, docker, gitlab, agile
- Successfully made a streaming pipeline that fetches the list of top twitter handle's that produces huge content on a particular topic & the consumers that consumes a large part of it. Also identifying how diverse & monotonous is the content of those top twitter handle's.
  Tools | Tech : Python, MongoDB, Flask, uWSGI, NGINX
- Built an AI powered SOC capable of analyzing logs in real time and alert the SOC analyst instantly for StegoSOC with a team of 3 people.
  Tools | Tech : Clustering, Supervised Learning, Cyber attack graph, CyberSecurity | Brief Article

Abyeti Technologies Inc.

Bengaluru, India

Technical Analyst

July'15 - March'16

- Worked on CodeGrounds - Hiring platform. Written Logging & other modules independently.
  Tools | Tech : InfluxDB, Grafana, MySQL, version control
- Worked on Precise & SQL workload Analyzer. These are monitoring tools for all types of Databses, Dotnet Applications & J2EE. Worked on feature addition of both the products.
  Tools | Tech : Core Java
- Uptime - Another server Monitoring tool from Idera is a lightweight application as compared to Precise. Worked on feature addition & bug resolving in the team.
  Tools | Tech : Salesforce Ticket Resolving | Core Java

—

Patent & Publications

Operations Research	Worked on multi-objective evolutionary Algorithms [NSGA-II & TDEA] under Prof. M.K Tiwary, IIT Kharagpur and Implemented three variants of network manufacturing problems		Github
Btech Thesis	A Weighted Mutual k-Nearest Neighbour for Classification Mining. Submitted this work in IEEE Conference (International Conference on Computational Systems in Engineering and Technology under Advancing Technology for Humanity)	Github	arxiv
Master Thesis	Detecting Original & Duplicate YouTube videos. Incorporating the features of YouTube.com and using attribute relational graph of an image, a robust technique was developed to detect copyright infringement of videos	Github	Thesis report

References available upon request.