

Prateesh Reddy

Data Scientist / ML Engineer

prateesh.reddy34@gmail.com | +1 (650)-260-8525

SUMMARY

- Data Scientist with a Master's degree and 3 years of extensive experience in Data Extraction, Data Modelling, Data Wrangling, Statistical Modeling, Machine Learning, and Data Visualization.
- Well-versed and experienced in all stages of the ML lifecycle from defining, Proof-of-concept (PoC), processing, model building, evaluation, deployment, and monitoring to re-train using data-driven insights.
- Proficiency in AWS, and Azure Databricks in packaging, collaborating, tracking, and deploying ML models.
- Experienced in tackling optimization problems using Linear Programming, with hands-on expertise in GurobiPy and CPLEX.
- Expertise in advanced measurement, utilizing tools like Propensity Score Matching and Difference-in-Differences for causal inference
- Gained e-commerce and online marketplace experience during my tenure at Alexa AI, Amazon, understanding intricate user behaviors, preferences, and market dynamics, which underscore the world's leading online shopping platform.
- Proficient in using git for version control, drive success as a CODE OWNER for a million dollar generating repo by managing pull requests, conflict resolutions, interactive rebasing.
- Proficiency in machine learning algorithms such as Regression, Decision Trees, Supervised Learning, Unsupervised Learning, Classification, SVMs, Decision Trees, Random Forests, Naive Bayes, KNN, K Means, and CNN.
- Proficient in using various Python packages for Data Science such as NumPy, SciPy, Pandas, Matplotlib, MLOps, AutoML, PySpark, and Scikit-learn.
- Proficient in designing visualizations using Tableau and Power BI software and Storyline on web and desktop platforms, publishing and presenting dashboards.
- Well-versed in design, management, and visualization of databases using MySQL and SQL Server.

EDUCATION

Masters of Science in Data Science | Indiana University Bloomington, Bloomington, USA
Bachelors in Computer Science | GITAM University Hyderabad

SKILLS

Methodologies: SDLC, Agile, Waterfall

Language: Python, R, Pyspark, Java, JavaScript, SQL, SAS

IDEs: Visual Studio Code, PyCharm, AWS Sagemaker, AWS EMR, Databricks

ML Algorithm: Linear Regression, Logistic Regression, Decision Trees, Supervised Learning, Unsupervised Learning, Classification, SVM, Random Forests, Naive Bayes, KNN, K Means, CNN

Packages: NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Spark, Pytorch

Cloud Technology: Amazon Web Services (AWS), Azure

Visualization Tools: Tableau, Power BI, AWS QuickSight

Database: SQL Server, MySQL, PostgreSQL, Oracle DB, MongoDB

Other Tools: MS Excel, Git, Apache Spark, Apache Airflow, Jenkins

Operating System: Windows, Linux

EXPERIENCE

Toyota, NC | Feb 2023 – Present | Sr. Data Scientist

- Implementing and following Agile development methodology within the cross-functional team and acting as a liaison between the business user group and the technical team.
- Implemented data quality checks and data validation processes within Databricks, ensuring the accuracy, consistency, and integrity of data for downstream analytics and reporting purposes.
- Led the project from the design stage to the production stage by working on the Machine learning pipeline and scaled it up based on requirements.
- Guided new hires on the workflow followed at Toyota and helped by defining tasks.
- Worked on Jenkins CI/CD setup and SonarQube setup to our code reviews in github.
- Worked on the creation of DAGs in Airflow based on our architecture.
- Spearheaded a \$150 million every year optimization initiative at Toyota, further worked on proof-of-concept (POC) predictive model independently. The POC provided valuable insights, prompting further exploration and expansion of the optimization project.

Alexa AI Amazon, Santa Barbara, CA | May 2022 - Aug 2022 | Data Scientist Intern

- Low-frequency Alexa questions had higher confidence intervals/margin of error. Performed Stratified sampling using various distributions and presented their differences. The result led to a change in Alexa tables' data pipeline.
- Improved Alexa categorizer model by extracting relevant data based on the features using AWS Glue to load them with the existing S3 buckets.
- Created a dashboard for historical comparison of Alexa's successful answering metric with the target and baseline which is now being used in Alexa-wide weekly webinars to discuss the overall performance.

Cyient Ltd, India | Aug 2019 - Dec 2021| Data Scientist

- Extracted road sign boards through object detection from terrestrial imagery to minimize manual efforts of data annotation for North American-based clients.
- Achieved a hit rate of 92%, resulting in a saving of 12 FTEs.
- Performed data profiling and data quality assessment to identify and resolve data issues during the ETL process.
- Design and develop ETL processes to extract data from various sources, transform it into a suitable format, and load it into a data warehouse or data lake for analysis.
- Designed, developed, and implemented Power BI Dashboards, Tableau reports, Scorecards, and KPI Reports.
- Developed the necessary Stored Procedures and created Complex Views using Joins for robust and fast retrieval of data in SQL Server.
- Performed all necessary day-to-day Git and GitHub support for version controlling different projects and designs.

PROJECTS

Cognitive Search Engine by NLP:

- Provided Cognitive search capability for **Eli Lilly and company** to search against databases like FDA and EMA via natural language questions and return relevant results to help with accelerating regulatory submissions for Eli Lilly.

- Performed abstraction-based Natural language generation methods using BERT and BART Transformer.

Pixel Gram Application:

- Designed a scalable distributed photo-sharing application for researchers to manage and share photos.
- Architected and developed 5 microservices for application with Rest APIs and RabbitMQ as intra-service communication.
- Implemented API gateway as middleware to validate/authenticate requests through JWT tokens and to act as a single endpoint.
- Deployed dockerized micro-services into OpenShift cluster scripting resource files through Jenkin pipelines on IU Jetstream cloud.
- Achieved up to 1800 concurrent requests per instance, tested through load test in JMeter.