

# Vikrant Deshpande

vikrant.deshpande09876@gmail.com | github.com/vikrantdeshpande09876 | linkedin.com/in/vikrant-deshpande

## Education

### Indiana University Bloomington

Aug 2021 - May 2023

Master of Science, Data Science

- **Relevant Coursework:** Machine Learning, ML for Signal Processing, Statistics, Exploratory Data Analysis, Applied Distributed Systems.

### University of Pune

Aug 2013 - May 2017

Bachelor of Engineering, Information Technology

- **Relevant Coursework:** Database Systems, Algorithms, Machine Learning.

## Experience

### Indiana University

Aug 2021 - Jan 2022

Research Assistant

Bloomington, IN

- Extracted meaningful comparison reports for ongoing genomics-research efforts in mapping cell-type partonomy structures.
- Researched disparate and competing data sources, and incepted a data pipeline to reduce 40% of manual comparison checks.
- Managed to generate dynamic reports after every commit/push, using Github Actions, Unix Bash, and R.

### Deloitte USI Consulting

Jan 2018 - Jul 2021

Consultant - Data Engineer

Mumbai, India

- Specialized in Master Data Delivery for a retail client, and Clinical Data Management for a big Pharma client.
- Spearheaded 3 Proofs-of-Concept, and client presentations for code-walkthroughs and data-storytelling.
- Developed 40+ ETL pipelines for data ingestion, wrangling, and aggregation from a variety of clinical data sources like APIs, AWS S3 Buckets, and SharePoint Lists into a central SQLServer Data Warehouse, with effective SQL query performance tuning and Data Modeling for Type 2 SCD Tables.
- Improved patient-risk monitoring using a LASSO regression model in R, and a Tableau feature to view variations.
- Conceptualized 'Refresh Notification Frameworks' to speed up reporting by 70%, using Tableau metadata.
- Enhanced Master-Data language detection by 40%, by leveraging FastText into the data warehousing project.
- Led a POC to identify a single source of truth by de-duplicating hospital/site entities, and ensure better KPI reporting, using PySpark, R, and SQL.

## Projects

### Distributed Weather Reporter | Python, Java, MongoDB, Docker, Kafka, Circle-CI, Github

Feb 2021 - May 2021

- Leading a team of 3 to create a loosely coupled distributed weather-reporting system using microservices.
- Developed the data-centric Flask service for visualization, and SpringBoot service for audits and authentication.
- Used Kafka for streaming GIF image reports, and Circle-CI for CI/CD pipeline of Docker images ([Link](#)).

### Google Analytics Customer Revenue Prediction | Python

May 2020 - Jul 2020

- Kaggle competition to predict revenue given a huge 15GB GStore unstructured dataset across 3 years.
- Extracted features after Exploratory Data Analysis, and implemented an ensemble predictor using Gradient-Boosting models and cross validation.
- Achieved 0.88 RMSE (top 6%) on the Private LeaderBoard, and won the Bronze Medal ([Link](#)).

### XML/XSD Parser for Pyspark Data-Ingestion | Python, Flask, jQuery, HTML, CSS, Heroku

Jan 2020 - Apr 2020

- Created an app aimed to allow user-friendly ingestion of confidential XML/XSD data into PySpark.
- The app parses XML/XSD data, and translates the chosen tags to a relational schema format.
- Deployed it to Heroku; used by 2 teams in Deloitte-USI to make XML manual parsing 80% quicker ([Link](#)).

### Anonymized credit-card transactions fraud detection | Python

Dec 2021 - Jan 2022

- Identified anomalies like multi swipes or returns in credit-card transactions, and engineered crucial features.
- Implemented baseline models of Logistic Regression, Isolation Forest, and SVM.
- Achieved a minimal Type-II error with 99.9% recall, and 98.5% precision, by finalizing an LGBM Classifier ([Link](#)).

## Skills

- **Languages:** Python, SQL, R, Bash, Java, Javascript.
- **Databases:** SQL Server, MySQL, MongoDB, Postgres.
- **Frameworks & packages:** Flask, PySpark, Tidyverse, SpringBoot, SciKit, Airflow, TensorFlow, Keras.
- **Tools:** Git, Tableau, Informatica, Heroku, Docker, Jira, Bitbucket, Kafka, Postman, AWS- S3, EC2, RDS.