Kartik Dhawan

+1 (812)-369-5282 | kartikdhawan09@gmail.com | GitHub | Medium | LinkedIn

Summary

Full-Stack data scientist experienced in developing and deploying machine learning models into cloud environments with a focus on revenue generation and marketing. Currently in Detroit, open to relocation and remote.

Professional Experience

Rocket Companies, Detroit, MI | Data Scientist March 2020- Present

- Collaborated across three revenue generation data science teams that accounted for a total of $50 billion in retail loan volume over 20 months. This is 10% of the total Quicken Loans lead flow.
- Utilized Python and AWS Sagemaker to train an XG-Boost model that predicts client's probability to close, thus prioritizing a salesman's pipeline and increasing conversion by 15%.
- Developed A/B testing strategy to validate model performance in production, increasing volume scored by 35%.
- Generated $1 million in additional monthly revenue by building propensity model to cross-sell personal loans to denied mortgage leads. **Won department-wide award for 'Biggest Impact'.**
- Saved 1 hour of analyst time per week by architecting an anomaly detector system on a 200 million record data pipeline by utilizing chi-square distributions and SparkML.

Tech Stack: Python, SQL, Spark, XG-Boost, TensorFlow, Pandas, NumPy, Scikit-Learn, AWS (S3, Athena, Sagemaker)

Rocket Companies, Detroit, MI | Sr. Big Data Engineer Jan 2018- March 2020

- Architected and Deployed Big Data orchestrator ETL on AWS using PySpark, Glue and Lambdas to predict solution eligibility for 40 million clients used to create leads, generate marketing campaigns and pre-fill data.
- Created data pipeline that imputes data utilizing third party sources and statistical techniques, resulting in 15 million additional fresh sales leads. **Nominated for an all-company award as a 'Special Achievement'**.
- Reduced Spark data pipeline run time by 75% through implementing delta loads and calculating projected benefit in-memory through statistical techniques, avoiding expensive API calls on millions of records.
- Leveraged NiFi and Attunity to automate batch ingestion of files into Cloudera Data Lake, created Hive tables on data stored in S3 to allow seamless querying of data.
- Analyzed 100 gigabytes of semi-structed data in Spark to troubleshoot solution discrepancies.
- Upgraded legacy Python code to config driven PySpark ETL application with CI/CD enabled enabling new feature deployments within minutes.

Tech Stack: Python, Hadoop, NiFi, Hive, Oozie, AWS (S3, Lambdas, Glue, Step Functions), Circle-CI

Rocket Companies, Detroit, MI | SQL Data Engineer Sep 2016- Jan 2018

- Integrated source data into dimensional models by implementing ETLs in SSIS, scheduled automated jobs to save analysts 2 hours per day on manual data pulls.
- Authored highly- performant T-SQL code to power Tableau and Power BI dashboards.
- Optimized SQL queries by utilizing query execution plans. Reducing run time by 40% on average.
- Collaborated with business stakeholders to perform data pulls and automate analyst queries.

Tech Stack: MS SQL Server, Tableau, Power BI, SSIS, SSRS

Education

Georgia Institute of Technology *Master of Science in Analytics* Aug 2019- July 2021

Indiana University Bloomington *Bachelor of Science in Informatics* Aug 2012- May 2016

Projects/Publications

- **Towards Data Science: Exploring the world of Ramen through Text Analytics!**
  - Leveraged bag of words and n-grams through NLTK to discover new varieties of instant ramen.
- **Towards Data Science: My Quarantine, Wrapped.**
  - Gained insights into my music listening patterns inspired by Spotify's 'Wrapped' feature.
- **Placement Optimization of Public EV Charging Stations in Cleveland using Car Sensor Data**
  - Proposed potential hotspots for stations by utilizing DBScan to cluster terabytes of car sensor data