

DATA SCIENTIST CONSULTANT

Professional Summary

Data scientist with 1+ years of experience in delivering data-driven insights in large data management enterprise, fast-paced hedge fund and media entertainment conglomerate, passionate and skilled at solving business problems with machine learning models and data analytics.

Skills

- Regression: GLM, Ridge, Lasso, KNN
- Classification: Logistic Regression, Decision Trees, Random Forest, XGB, SVM, Naive-Bayes
- Clustering: K-Means, Hierarchical, DBSCAN
- Statistics:
- A/B Testing, Hypothesis Testing, Bayesian Inference, Probability
- Programming:
- Python, SQL, R, Hive, Spark, Git, Scala, Java
- Ad
- Automate
- Clustering
- Credit
- CRM
- ETL
- Java
- Marketing
- Modeling
- Predict
- Programming
- Promotion
- Python
- Quality
- Sales
- SQL
- Statistics
- Validation

Work History

Data Scientist Consultant , 09/2019 to Current

Deloitte – San Juan , PR

- Implemented and deployed marketing propensity model using XGBoost for customer acquisition on upcoming movies, identified top 70% first-time purchasers with highest product interest scores.
- Constructed ad-hoc SQL queries and performed A/B test for quantifying retention and churn campaign promotion uplifts, collaborated with CRM teams to identify key metrics and evaluate testing results.
- Designed and developed ETL process using SQL and R to automate data validation process for identifying and tracking data quality issues.
- Created user defined function using K-means in Snowflake warehouse to segment customer by user behavior and demographic information.
- Developed and published several interactive and scalable R-shiny visualization dashboards to increase visibility on various KPIs.

Data Scientist Intern , 06/2019 to 08/2019

Ascend Learning – Nashua , NH

- Identified and analyzed problems of cascade effect and data misrepresentation when using credit card transaction to predict companies' sales, provided solutions that reduced out-of-sample MAE error by 75%.
- Partnered with investment team from different sectors to increase size of modeling data using clustering, sampling and rule based method, effectively improved data reliability and reduced geo-bias from various alternative data sources.

Data Scientist Intern , 05/2018 to 09/2018

Orlando , FL

- Designed, developed and deployed automated streamlined procedure in Python for parsing test performance data and building visualization platform using Pyplot and Tableau for multiple drive performance comparison, improving tasks efficiency by 30%.
- Utilized Hive platform to develop an automated pipeline for data query, cleaning and transformation.

Education

Master of Science : Statistical Science , 05/2019

Duke University - City , State

GPA: 3.7

Bachelor of Arts : 08/2016

University Of California, Berkeley - City , State

GPA: 3.92

Work History

Data Scientist Consultant , 09/2019 to Current

Company Name – City , State

- Implemented and deployed marketing propensity model using XGBoost for customer acquisition on upcoming movies, identified top 70% first-time purchasers with highest product interest scores.
- Constructed ad-hoc SQL queries and performed A/B test for quantifying retention and churn campaign promotion uplifts, collaborated with CRM teams to identify key metrics and evaluate testing results.

- Designed and developed ETL process using SQL and R to automate data validation process for identifying and tracking data quality issues.
- Created user defined function using K-means in Snowflake warehouse to segment customer by user behavior and demographic information.
- Developed and published several interactive and scalable R-shiny visualization dashboards to increase visibility on various KPIs.

Data Scientist Intern , 06/2019 to 08/2019

Company Name “ City , State

- Identified and analyzed problems of cascade effect and data misrepresentation when using credit card transaction to predict companies' sales, provided solutions that reduced out-of-sample MAE error by 75%.
- Partnered with investment team from different sectors to increase size of modeling data using clustering, sampling and rule based method, effectively improved data reliability and reduced geo-bias from various alternative data sources.

Data Scientist Intern , 05/2018 to 09/2018

City , State

- Designed, developed and deployed automated streamlined procedure in Python for parsing test performance data and building visualization platform using Pyplot and Tableau for multiple drive performance comparison, improving tasks efficiency by 30%.
- Utilized Hive platform to develop an automated pipeline for data query, cleaning and transformation.

Skills

- Regression: GLM, Ridge, Lasso, KNN
- Classification: Logistic Regression, Decision Trees, Random Forest, XGB, SVM, Naive-Bayes
- Clustering: K-Means, Hierarchical, DBSCAN
- Statistics:
- A/B Testing, Hypothesis Testing, Bayesian Inference, Probability
- Programming:
- Python, SQL, R, Hive, Spark, Git, Scala, Java,
- Ad, automate, Clustering, credit, CRM, ETL, Java, marketing, modeling, predict, Programming, promotion, Python, quality, sales, SQL, Statistics, validation