

ANALYTICS ENGINEER

Professional Summary

Accomplished Data Scientist with a passion for generating data and delivering insights for healthcare and financial data through Cloud and Analytical tools. Committed to helping companies advance by helping them to develop strategic plans based on predictive modeling. Bringing forth a proven track record of analyzing complex data sets using big data tools.

Skills

- Scala, SQL, Python, C#, Java, R
- Hadoop, Spark SQL, HBase, Hive, BigQuery
- Databricks, RestAPI, Data Factory, Git, Azure DevOps
- EC2, EMR, RDS, S3
- MS SQL, MySQL, Mongo DB, PostgreSQL
- Dimensional modeling, Business Analysis, Project Management
- Regression, Random Forest, Neural Network, Decision tree, K-Means, kNN
- Tableau, Jupyter Notebook, PyCharm, Eclipse, PowerBI, Teradata, IntelliJ

Work History

Analytics Engineer , 07/2019 to Current

Dnv GL "Essen , DE

- Developed Spark ETL scripts, pipelines on Azure to drive the data transformation for automating manual scanning process, developed dashboards and communicated results with Cloud and Business Intelligence teams
- Developed SQL queries, regression model to forecast insurance sales, ingested data from multiple sources and performed trend analysis on the generated vulnerabilities
- Tested the data pipelines on Azure for claims, analyzed the outliers in Tableau and conveyed seasonality trends to business using clustering and classification
- Developed PySpark programs to generate monthly vulnerability types, executed it using Spark EMR cluster and analyzed data using anomaly detection model to predict vulnerability types
- Performed data cleaning for claims data, performed ETL to compute numbers of monthly and quarterly claims type, that led to 15% improvement in profitability analysis
- Developed and tested new product offerings prior to release to assist development team in bug identification

Data Scientist , 05/2019 to 07/2019

Lockheed Martin Corporation "Burbank , CA

- Developed and optimized existing Teradata queries, developed Scala scripts to process large data, monitored performance of marketing campaigns using random forest, developed new strategies and visualized vaccine results leading to 20% enhancement in reporting
- Ingested data from HBase, MySQL developed PySpark scripts to perform ETL to add fields (yield, expense) performed analysis of vaccines in flu season to predict most used vaccine
- Applied statistical and algebraic techniques to interpret key points from gathered data

Senior Data Analyst , 10/2016 to 04/2019

Rtx "Methuen , MA

- Developed MongoDB queries, fine-tuned existing SQL queries for Risk Analytics Team and generated KPI reporting for personal loans; reduced time it takes to populate data from multiple sources in spreadsheets by 20%
- Converted existing Python scripts to PySpark modules in Spark framework in order to classify fraud accounts into different fraud categories using decision tree
- Contributed to 15% increase in customer loan applications from 2017 to 2018 by conducting exploratory data analysis (booked to approval rate), optimized SQL, SAS queries; analyzed return payment history for customers
- Created data dictionary for MS SQL Server to enhance CBB reporting, analyzed transformed delinquency loss trend data that helped credit risk, marketing team spot emerging trends by comparing data
- Communicated insights on bank's monthly and quarterly loan reporting information to senior management, developed Tableau visualizations and generated KPI metrics using aggregation queries
- Increased annual revenue by 15% by recommending improvements in reporting efficiency
- Researched and resolved issues regarding integrity of data flow into SQL databases
- Mapped current business and operational processes and offered recommendations for improvement

Business Intelligence Analyst , 05/2015 to 12/2015

Alteryx Inc. "Washington , DC

- Analyzed data in Salesforce CRM, performed data mapping, configured changes, and performed data validation that led to 15% increase in efficiency in processing applications
- Designed multidimensional data models, performed ER modeling, analyzed booked data in ETL application using MS SQL server, performed lookups and communicated insights to product team
- Developed intelligence-sharing dashboards, providing company-wide access to collected data

Data Analyst , 01/2012 to 07/2014

Lockheed Martin Corporation "China Lake , CA

- Executed Unix shell scripts on Ab-initio to run ETL workflows, validated loaded data in Oracle DB and sent output data files to development team
- Improved weekly reporting by 15%, by automating tasks for healthcare insurance application using SQL

Education

Master of Science : Management Information Systems , 07/2016

Northeastern University - City , State

- Coursework in Big Data and Data Science

Bachelor of Engineering : Electronics and Telecommunication , 2008

Mumbai University - City

Skills

- Scala, SQL, Python, C#, Java, R
- Hadoop, Spark SQL, HBase, Hive, BigQuery
- Databricks, RestAPI, Data Factory, Git, Azure DevOps
- EC2, EMR, RDS, S3
- MS SQL, MySQL, Mongo DB, PostgreSQL
- Dimensional modeling, Business Analysis, Project Management
- Regression, Random Forest, Neural Network, Decision tree, K-Means, kNN
- Tableau, Jupyter Notebook, PyCharm, Eclipse, PowerBI, Teradata, IntelliJ

Work History

Analytics Engineer , 07/2019 to Current

Company Name " City , State

- Developed Spark ETL scripts, pipelines on Azure to drive the data transformation for automating manual scanning process, developed dashboards and communicated results with Cloud and Business Intelligence teams
- Developed SQL queries, regression model to forecast insurance sales, ingested data from multiple sources and performed trend analysis on the generated vulnerabilities
- Tested the data pipelines on Azure for claims, analyzed the outliers in Tableau and conveyed seasonality trends to business using clustering and classification
- Developed PySpark programs to generate monthly vulnerability types, executed it using Spark EMR cluster and analyzed data using anomaly detection model to predict vulnerability types
- Performed data cleaning for claims data, performed ETL to compute numbers of monthly and quarterly claims type, that led to 15% improvement in profitability analysis
- Developed and tested new product offerings prior to release to assist development team in bug identification

Data Scientist , 05/2019 to 07/2019

Company Name " City , State

- Developed and optimized existing Teradata queries, developed Scala scripts to process large data, monitored performance of marketing campaigns using random forest, developed new strategies and visualized vaccine results leading to 20% enhancement in reporting
- Ingested data from HBase, MySQL developed PySpark scripts to perform ETL to add fields (yield, expense) performed analysis of vaccines in flu season to predict most used vaccine
- Applied statistical and algebraic techniques to interpret key points from gathered data

Senior Data Analyst , 10/2016 to 04/2019

Company Name " City , State

- Developed MongoDB queries, fine-tuned existing SQL queries for Risk Analytics Team and generated KPI reporting for personal loans; reduced time it takes to populate data from multiple sources in spreadsheets by 20%
- Converted existing Python scripts to PySpark modules in Spark framework in order to classify fraud accounts into different fraud categories using decision tree
- Contributed to 15% increase in customer loan applications from 2017 to 2018 by conducting exploratory data analysis (booked to approval rate), optimized SQL, SAS queries; analyzed return payment history for customers
- Created data dictionary for MS SQL Server to enhance CBB reporting, analyzed transformed delinquency loss trend data that helped credit risk, marketing team spot emerging trends by comparing data
- Communicated insights on bank's monthly and quarterly loan reporting information to senior management, developed Tableau visualizations and generated KPI metrics using aggregation queries
- Increased annual revenue by 15% by recommending improvements in reporting efficiency
- Researched and resolved issues regarding integrity of data flow into SQL databases
- Mapped current business and operational processes and offered recommendations for improvement

Business Intelligence Analyst , 05/2015 to 12/2015

Company Name – City , State

- Analyzed data in Salesforce CRM, performed data mapping, configured changes, and performed data validation that led to 15% increase in efficiency in processing applications
- Designed multidimensional data models, performed ER modeling, analyzed booked data in ETL application using MS SQL server, performed lookups and communicated insights to product team
- Developed intelligence-sharing dashboards, providing company-wide access to collected data

Data Analyst , 01/2012 to 07/2014

Company Name – City , State

- Executed Unix shell scripts on Ab-initio to run ETL workflows, validated loaded data in Oracle DB and sent output data files to development team
- Improved weekly reporting by 15%, by automating tasks for healthcare insurance application using SQL