

## Contact

xxx-xxx-xxxx

[www.linkedin.com/in/md-faisal-akbar](https://www.linkedin.com/in/md-faisal-akbar) (LinkedIn)  
[dsfaisal.com](https://dsfaisal.com) (Personal)  
[public.tableau.com/profile/faisalakbar](https://public.tableau.com/profile/faisalakbar) (Other)

## Top Skills

R

Python

Tableau

## Languages

English

Bengali

## Certifications

Data Science Specialization

Tableau Author

HarvardX Data Science - Probability

Python Data Science Toolbox (Part 2)

Data Science: Visualization

# Faisal Akbar

Data Analytics Consultant at Healthfirst  
Jamaica, New York

## Summary

Passionate, curious and with strong enthusiasm and problem solving skills, always looking to deliver data driven solutions and provide high quality analysis.

## Experience

Healthfirst

Data Analytics Consultant

December 2019 - Present (7 months)

New York, New York

e-Zest Solutions

Data Scientist/ Senior Tableau Developer

June 2018 - December 2019 (1 year 7 months)

Irving, TX

- Working with health insurance PBM claims data-sets of varying degrees of size and complexity to identify and analyze the match mismatch scenarios for paid and reject claims and finding the root cause using Hadoop, Spark, HIVE, Machine learning and statistical modelling, and report analysis output using Tableau.
- Designed, developed, maintained, optimized, and deployed workflow tableau dashboard in tableau server, integrate the dashboard with web application via tableau JavaScript API and test in development and production environment and ensure tableau row level security.
- Developed smart triage tableau dashboard for claims comparison and validation based on all the scenarios, Functionalities and reject codes.
- Developed tableau KPI dashboard and cost benefit analysis dashboard for claims execution, mismatch claims, reject and paid claims, mismatch by top features, functionality, top NDC, top drug type, copay comparison by state, city and pharmacy level.
- Automated and optimized all the daily, weekly, monthly reports including claim execution, mismatch, scenarios, reject codes, ESI claims versus CVS claims, mismatch by field, user activity and executive reports using tableau and saved more than 120 hours of work per week.

- Developed claims surveillance dashboard to predict and flag potential fraudulent claims and anomalies detection, outliers, and trend changes.
- Developed completely responsive mobile application tableau KPI dashboard for executive.
- Analyze and select several potential appropriate modeling approaches for a given analytic problem (machine learning methods such as Ensemble and classification models, decision trees; logistic regression, clustering, Principal Component Analysis (PCA), operations research; statistical modeling such as multivariate techniques).
- Developed Machine learning and statistical analysis dashboard using R Shiny for defect prediction, predictive analysis and claims to plan prediction.

## Capital One

### Data Scientist

November 2016 - June 2018 (1 year 8 months)

Greater New York City Area

- Conducted analysis on billions of customer transactions to identify sales and promotions scopes using Hadoop, Spark, HIVE and AWS suite
- Continuously built machine learning models from development through testing and validation to their 30+ million customers in production. ML used includes but not limited to Regression Analysis, Clustering, Boosting, XG Boost, SVM, and Random Forest, Bayesian HMM, Classification, Principal Component Analysis (PCA), MCA, Feed Forward Neural Network (FFNN)
- Created Dashboard using Tableau to communicate complex ideas to executive committee
- Used R (RStudio) and Python (PySpark) programming for data cleaning, data visualization, risk analysis and predictive analytics.
- Created Complex Tableau calculations, table calculations, Level-of-Detail (LOD), context filters, actions, time period functions, parameters based on requirements.
- Experienced with table design and data management using HDFS, Hive, Sqoop, MySQL, and Kafka.
- Mined geospatial data (GIS). Used R to create visualizations of different measures. Libraries used: ggmap, rgdal, gdal, rgeos, etc
- Done NLP and text mining experience to create statistical predictive models with unstructured text data.
- Worked on preprocessing the text data using pandas and nltk. Like, removing stop words, stemming, lemmatization etc.
- Used Convolution Neural Networks and toolkits including CAFFE, cuda-convnet, and others.

- Created predictive models with the text data using nltk and comparing these classification models (naive Bayes, svm, max entropy etc). Application involved: Sentimental analytics, sentence meaning extraction and topic modeling
- Performed scoring and financial forecasting for collection priorities and predict fraud.
- Developed a model for predicting repayment of debt owed to small and medium enterprise (SME) businesses.
- Created SQL scripts and analyzed the data in MS Access/Excel and Worked on SQL and SAS script mapping

## JPMorgan Chase & Co.

### Data Scientist (Consultant)

December 2015 - October 2016 (11 months)

Columbus, Ohio Area

- Created Dashboards using Microsoft Power BI and Tableau.
- Developed Logical and Physical Data model and organized data as per the business requirements using Sybase Power Designer, Erwin, ER Studio in both OLTP and OLAP applications
- Done data cleaning and development of predictive model using RStudio and Python IDE (Anaconda)
- Utilized Hadoop based storage system, Cassandra, Hive, Pig and Sqoop to design business intelligence applications
- Used Flume for auto streaming of data
- Provided end to end support by extracting data, preparing analysis, interpreting data, making strategic recommendations and presenting before internal client team
- Wrote SQL scrips to extract & clean data and to join from different data sources
- Familiar with Crystal Reports, and SSRS - Query, Reporting, Analysis and Enterprise Information Management
- Explored other databases including DB2, Teradata15/14 and Netezza,
- Worked with machine learning tools and libraries such as Scikit-learn, R, Spark and Weka
- Reviewed suspicious activity and complex fraud cases to help identify and resolve fraud risk trends and issues.
- Clearly and thoroughly documented investigation findings and conclusions.
- Analyzed of customer data to tune rules, exposes patterns, research anomalies, reduce false positives, and build executive and project-level reports

- Analyzed high-volume data to investigate, identify and report trends linked to fraudulent transactions
- Done Exploratory Data Analysis, trying to find trends and form clusters.
- Built models using techniques like Regression, Tree based ensemble methods, Time Series forecasting, KNN, Clustering and Isolation Forest methods, C4.5, SVM and others
- Worked both with structured and semi structured data
- Heavily utilized pandas, keras and other libraries and maintained RDD's using SparkSQL.
- Communicated and coordinated with other departments to collection business requirement

## Medicare and Medicaid (CMS)

### Data Scientist (Consultant)

October 2013 - November 2015 (2 years 2 months)

Virginia

- Built sentiment analysis models that scored all customers interactions with the organization and a topic model that automated classification of text conversations.
- Worked on customer journey analytics, enhancing knowledge on various channels, hence, improved the overall customer satisfaction and reduce operational cost.
- Implemented data mining and machine learning solutions to various business problems, hence, resolving issues.
- Applied LSTM networks to predict cost of healthcare services for individual customer
- Used QlikView to design dashboards and to do high level BI.
- Analyzed and extracted relevant information from large amounts of data to help automate for self-monitoring, self-diagnosing, self-correcting solutions and optimize key processes.
- Done data architecture design, development, maintenance for Windows and Android device applications.
- Done advanced SAS programming such as PROC SQL (JOIN/ UNION), PROC APPEND, PROC DATASETS, and PROC TRANSPOSE.
- Built predictive/machine learning models using Regression, Boosting, GBM, NNs, HMMs, CRFs, MRFs, other deep learning, Decision Tree, Random Forest, Naïve Bayes, Correlation, Multivariate Analysis, Logistic Regression and Cluster Analysis.
- Done coding in R, Python, SQL and UNIX,
- Strong Data Warehousing ETL experience of using Informatica

- Done data mining in various data sources including but not limited to Oracle, MS SQL Server, DB2, ODS, and Hadoop based data storage systems
- Done data mining, Factor Analysis, ANOVA, Hypothetical testing, normal distribution and other advanced statistical techniques
- Participated in full software development lifecycle and Scrum methodologies.
- Excellent knowledge on Normalization (1NF, 2NF, 3NF and BCNF) and De-normalization techniques for improved database performance in OLTP, OLAP and Data Warehouse/Data Mart environments.
- Excellent track records of working with complex data sets and translating data into insights to drive key business decisions

## Kaiser Permanente

### Senior Data Analyst

February 2011 - September 2013 (2 years 8 months)

Maryland

- Built defection risk models to be put into production internally. These models were being used by those who interact with our members.
- Done cluster analysis to quickly understand the varied needs of our insurance members. These population groupings gave quick understanding on customer experience with kp.org
- Done data conversion requirements based on existing business requirements
- Wrote SQL queries for data extraction, manipulation and formation of Table.
- Wrote HQL (on Spark) to extract data from Hadoop based storage system and to form table
- Maintained a Python/Django web application
- Done coding in R and Python to form, test and tun predictive model
- Created test data and train data.
- Wrote a python-based SQL generator that helped speed up a weekly reporting from several days
- Data bases used are mainly oracle based and Hadoop based.
- Done data modeling for big data and transactional data bases (oracle based) system
- Analyzed the business requirements and transferred high level requirements to technical requirements
- Assisted in tuning of the database, which includes indexes, and optimizing SQL statements, monitoring the server.
- Collaborated to form the data mapping document from source to target and the data quality assessments for the source data.

- Co-ordinated with various business users, stakeholders and SME to get Functional expertise, design and business test scenarios review, UAT participation and validation of financial data.
- Used vision to design and develop Use Case, Activity Diagrams, Sequence Diagrams, OOD (Object oriented Design) using UML and Visio.
- Used JIRA as tracking tool

---

## Education

### The City College of New York

Master's degree, Masters in Information Systems · (2016 - 2018)

### Shahjalal University of Science and Technology

Bachelor of Engineering - BE, Chemical Engineering · (2005 - 2008)