

<b>Md Faisal Akbar</b>			
<b>Queens, NY</b>	<b>E-mail: mfakboo@gmail.com (US Citizen)</b>		<b>Phone: (929) 278-1431</b>
<b>Years of Exp</b>	7+ year – as a Data Scientist and Data Analyst		
<b>Education</b>	Master's in Information Systems, CGPA- 3.77/4.00, Graduation: 2018	MIS	The City College of New York, NY
	Bachelors Engineering, CGPA- 3.51/4.00, Graduation: 2008	Chemical Eng.	Shahjalal University of Science and Technology, BD
<b>Certification/ Training</b>	<ul style="list-style-type: none"> <li>➤ Data Science Specialization- Johns Hopkins University (9 courses on Data Science)</li> <li>➤ Certified Scrum Master (CSM)</li> </ul>		
<b>Capabilities</b>	Data Mining, Data Quality Assessment, Data Analysis, Machine Learning, Predictive Model, Data Visualization, Client Communications	<ul style="list-style-type: none"> <li>• Visualized data analytics and predictive results before clients</li> <li>• Worked in Agile (Scrum/Kanban) environment</li> <li>• Done data integrity testing and backend testing</li> </ul>	
<b>Computer Skills</b>	<p><b>Machine Learning:</b> Artificial Neural Network, Bayesian Network/BBN, Regression, Logistic Regression, Decisions Tree, K-NN, SVM, SVDK Clustering, and PCA, MCA, MFC, and other data mining and ML algorithms</p> <p><b>Operating Systems:</b> Windows, UNIX, LINUX, Mac OS</p> <p><b>Technical:</b> R   Shiny   Python   Oracle SQL Developer   SQL Server   MySQL   Tableau   Power BI   QlikView   Gretl   Github   Markdown   C++   HTML, CSS   JavaScript   Bootstrap   WordPress   Sqoop   Elastic Search   MapReduce   Amazon S3   Azure   Jupyter   Yarn   Seahorse</p> <p><b>Tools:</b> Excel- VLOOKUP, Pivot Table, IF Statement, Data Analysis, Excelvba   PowerPoint   Word   Visio   Google Analytics   SEO   CDN   FireFTP   Firebug   Putty</p> <p><b>Networking Skills:</b> Server Administrator   Configuration and Maintenance (Active Directory, DNS, DHCP, Mail, NAT)   Router Configuration   Troubleshoot IP Addressing</p>		

#### SUMMARY:

- Over 7 years of experiences as a Data Scientist/Data Analyst/Data Modeler
- Experienced with business intelligence (and BI technologies) tools such as Tableau, Qlik View, Microsoft Power BI, Micro Strategy and spreadsheets
- Worked on different types of **Python** modules such as requests, boto, flake8, flask, mock and nose
- Efficient in developing logical and physical **data model** and organizing data as per the business requirements using **Sybase Power Designer, Erwin, ER Studio in both OLTP and OLAP application**
- Experienced in employing **R & Python** programming, MATLAB, SAS, Tableau, Power BI and SQL for data cleaning, data visualization, risk analysis and predictive analytics
- Experienced on Big Data related technologies including but not limited to **Hadoop, Hive, Pig, Sqoop, Cassandra, Oozie, Flume, Map-Reduce and Cloudera** Manager for Data manipulation and data analytics.
- Hands on experienced with **Machine Learning** and deep learning algorithms
- Familiarity with Crystal Reports, and SSRS - Query, Reporting, Analysis and Enterprise Information Management
- Used Pentaho to create Business Intelligence report
- Experienced with other data repository including, but not limited to Oracle, XML, DB2, Teradata15/14, Netezza, SQL Developer
- Worked with team of engineers and analysts to integrate algorithms and data into Return Path solutions
- Experienced in working with large-scale data set (Billions of entities)
- Extensively used Agile methodology as the Organization Standard to implement the data Models
- Experienced with machine learning tools and libraries such as Scikit-learn, R, Spark and Weka
- Hands-on experienced with **NLP**, mining of structured, semi-structured, and unstructured data
- Experienced using **SAS Enterprise Miner**
- Developed R shiny projects (dsfaisal.com/shiny)

#### PROFESSIONAL EXPERIENCES:

##### Capital One, NY, Data Scientist

November 2016 - Present

- Conducted analysis on billions of customer transactions to identify sales and promotions scopes using Hadoop, Spark, HIVE and AWS suite
- Wrote SQL and HQL to clean and investigate large, messy, data sets of numerical and textual data
- 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

- Investigated new technologies on the future of digital banking
- Used R (RStudio) and Python (PySpark) programming for data cleaning, data visualization, risk analysis and predictive analytics.
- Worked with financial analysts to integrate algorithms and data into Return Path solutions
- Extensively used Agile methodology (Scrum/Kanban)
- Built data pipelines for reporting, alerting, and data mining.
- 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.
- Extracted data from web sources using web scrapping and web crawling in python and created text corpora.
- 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
- Done image processing: in this method we analyzed images taken from different cameras in both 2D and 3D using computer vision techniques. Mainly used open CV and tensor-flow to analyze the image to create recognition models (face or object). This was a pilot project
- Performed scoring and financial forecasting for collection priorities and predict fraud
- Assisted in coding of MapReduce to configure our Hadoop system.
- Assisted in setting up storage and data analysis tools in Amazon Web Services (AWS) cloud computing infrastructure.
- Worked directly with executive stakeholders to define requirements of scoring models.
- Developed a model for predicting a debtor setting up a repayment rehabilitation program for student loan debt.
- 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
- Used JIRA as tracking tool

#### **JP Morgan & Chase, Columbus, OH, Data Scientist (Consultant)**

December 2015- October 2016

- 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
- Worked on different type of modules such as requests, boto, flake8, flask, mock and nose
- 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
- At the same time extracted and manipulated data from SQL and Oracle based transactional data base system
- Wear the hat of a data engineer and 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
- Drafted Business Intelligence level analytics on Pentaho and presented before clients
- Explored other databases including DB2, Teradata15/14 and Netezza,
- Worked with machine learning tools and libraries such as Scikit-learn, R, Spark and Weka
- Explored green plum and SAS Enterprise Miner
- 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
- Identified meaningful insights from chargeback data. Communicated analysis to engineers, product and stakeholders.
- Analyzed high-volume data to investigate, identify and report trends linked to fraudulent transactions
- Utilized Sqoop to ingest real-time data. 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
- Used RTC as tracking tool

## **Medicare and Medicaid (CMS), VA, Data Scientist (Consultant)**

October 2013- November 2015

- 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.
- Developed Logical Data Architecture with adherence to Enterprise Architecture.
- 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, C, C++, Java, 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, MD, Senior Data Analyst**

February 2011– September 2013

- 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

**Clearance:** Public trust

**References:** Available upon request.