|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RAY ISLAM  Dumfries, VA E-mail: [**rayisl78@gmail.com**](mailto:rayisl78@gmail.com) (**US Citizen**) Website**:** [**http://kindidata.com/**](http://kindidata.com/) **Phone: (703) 541-8013** | | | | |
| Years of Exp. | 13+ year – as a Senior Data Scientist and Quantitative Analyst | | | |
| Education | **PhD in Engineering (Cont.)** | | Data Science (ML) | University of Maryland, College Park, MD, USA |
| **MASc in Engineering** | | System Optimization | (UOIT), Ontario, Canada |
| **MSc in International Marketing** | | Global Supply Chain Strategy | Heriot-Watt University, UK |
| **MBA in IT** | | Information Tech. | Luton International College, UK |
| Certification/  Training | * Micro Masters Program on Statistics & Data Science, MIT (Massachusetts Institute of Technology) * Data Science Specialization- Johns Hopkins University (9 courses on Data Science) * Tackling the Challenges of Big Data – MIT (Massachusetts Institute of Technology), MA * Berkeleyx: CS105X Introduction to Apache Spark – University of California, Berkeley * Certified Scrum Master (CSM) * 200-A - [FedRAMP System Security Plan (SSP) Required Documents](https://governmenttraining.blackboard.com/) -Trained | | | |
| Relevant Courses | Advanced Optimization, Advanced Statistics, Engineering Decision Making, Life Cycle Cost Analysis, Advanced Database Management, Project Management, Machine Learning with Python (deep learning) | | | |
| Capabilities | Data Cleaning/ wrangling, Data Analysis, Machine Learning, Predictive Model, Artificial Intelligence, Visualization, Client communication | * Formed and Managed Teams of Data Scientists and Quantitative Analysts * Visualized data analytics and predictive results before Executive committee * Formulated data models and predictive models using ML algorithms for **business forecasting, fraud, waste and abuse** detection, identification of terrorist and **drug traffickers** in financial, retail, insurance and health care (e-prescription) industry and both in government and in private sector * Converted projects from Water fall to **Agile (Scrum and Kanban)** * Drafted customized **Data Quality Assessment Metrics** | | |
| Computer Skills | **Data Analytics/visualization tool:** R, Microsoft Power BI,Microstrategy, Tableau, Pentaho, STATA, NVivo, Informatica, Hadoop, Accumulo (Google big table concept), HBase, Spark, Qlick, Flume, MQ Services, Sqoop, Elastic Search, MapReduce, Amazon S3, Azure, Zeppelin, Yarn, Hive, Python IDE  **Machine Learning:** Artificial Neural Network, Bayesian Network/BBN, Regression, Logistic Regression, Decisions Tree, Elastic-net regularized generalized linear models (built in R), k-NN, SVM, SVDK Clustering, Page Rank and PCA, MCA, MFC, Apriori and other data mining and Deep Learning Algorithms  **Tools:**  Visual Studio, TFS, JIRA, Rally, Version1, HP ALM, Test Trac Pro, Rational team Concert  **Programming: Python,** Linux/Unix, C++/C, Visual Basic, Java, VB Scripting, HTML, R  **Others:** SQL Developer, Oracle R12, 11g, AWS East (Cloud), WinMeger, Splunk, Sort site, Azure, SharePoint (pulse, Alfresco), notepad++, SOAP UI, Putty, Salesforce Cloud, other NoSQL DB | | | |

**SUMMARY:**

* 13 years of professional experience as a Principal/Lead Data Scientist, Data Analyst and Business Development Manager for large scale IT Programs in Aviation, Financial, Insurance, Healthcare and Retail industry; both in government and private sector.
* Worked as a Solution Architect for Big Data and Analytic Projects. Managed teams of data scientists, data analysts, technical developers and architects on Enterprise IT Modernization efforts including **Big Data, Analytics, and Custom Development**.
* Created **data model**, data lakes, data warehouses and worked with **BI** (Business Intelligence) tools to present before the executive team. Worked closely with the clients to identify business requirements and transferring them in to technical requirement.
* Drafted **Data Quality Assessment** Framework for **SQL and NoSQL** data Sets and formulated models using **Data Mining and Machine Learning** Algorithms. Experienced both with industrial and academic practices of machine learning algorithms.
* Worked with multiple teams and coordinated and performed testing activities. Arranged brain storming sessions and identified security control gaps per NIST 800-53 standard, for FedRAMP Ready and CMS (ARS 2.0) certification.
* Managed tasks in a fast-paced environment and coordinated multiple projects with maximum of 15 FTE at the same time.
* Worked as a **practice area lead** for the development of **Big Data**, Analytics and Cloud Computing (Amazon Web Services (AWS) Elastic Cloud Servers, Azure and Salseforce). Drafted and implemented organizational data security controls (Data Masking, Data de-identification, File Encryption, volumetric encryption and other data security controls) and architected data governance rules (Including Rules of Behavior) for FISMA/FedRAMP ready and CMS approval. Provided departmental l**eadership** and direction, developed sponsor contracts and budgets, assigned personnel, implemented and revised standard processes and procedures, and ensured efficient and quality execution of work
* Strong decision-making ability with the aid of analysis, experiences, and expert judgments. A quick learner.
* Proven ability to collaborate successfully with **cross functional**, **cross-regional (virtual) and cross-cultural teams**
* Demonstrated strong presentation skills: one-on-one, groups and C-level executive teams inside and outside of the organization

**TECHNICAL SKILLS:**

* Developed algorithms to search **terrorists and drug traffickers, to identify fraud, abuse, waste, to predict new markets and competitions and to identify KPI** improvement scopes, pricing strategies and others
* In depth technical experiences in all aspects of **Structured and Unstructured Data Management, Data Fusion, Application design and development (Agile and Waterfall); data profiling; Master Data Management (MDM); Big Data, data cleansing, Entity Resolution** and **Entity Analytics, Information / Entity Extraction; data migration** and **COTS Software.**
* Built proof of concept systems for **secure analytics** platform. Established strategic partnerships with technical leadership across functional areas. Made **presentation** to **Executive Team and external audiences** as requested.
* Experienced in **Project Management** principals and **Agile Development** Methodologies (Scrum, Kanban, SAFe, XP).
* Demonstrated the ability to create new and different solutions that align to real work problems and opportunities, by formulating **Secure Analytics Workbench (SAW),** both for AIR’s internal users and for external clients like Centers for Medicate and Medicaid Services (CMS) and US Department of Education (DE)
* Experienced and expert in principles of **machine learning, statistical analysis, predictive modelling, data mining algorithms**, **deep learning, anomaly detection,** and **mathematical segmentation**. Algorithms used includes but not limited to, **Artificial Neural Network (FFNN, BPNN, BNN), Regression Analysis (Logistic and Linear), SVM , SVD, Brute Force, Random Forest,GMM, MCA PCA, FCA, C4.5, KNN, NLP (with decision tree) Clustering** and others. Done pattern recognition and analyzed trends from **Big Data**. Created **conceptual, logical and physical** data model.
* Experienced in applying Machine Learning both in structured and unstructured data for supervised & unsupervised modeling.
* Strong understanding of **complex business challenges**, experienced of designing scientific solutions, manipulating large data sets, using cutting edge machine learning or statistical modeling techniques and synthesizing insights
* Experienced in programming in C++/C, Java, **Scala, R, Python** and MATLAB

**PROFESSIONAL EXPERIENCE:**

### Raytheon (Clients: NASA, DoD), VA, Senior Principal Data Scientist December’17– Present

* Lead a team of 12 analyst and data visualization engineer to enforce data analytic and machine learning practices from scratch
* Developed **predictive models** to predict engineering measures and ethics violation for Directors, VPs and President of Raytheon
* Coded predictive **models R and Python** (Anaconda). Done cost and benefit analysis and predicted ROIs to attract fundings
* Trained team members to become efficient in data wrangling and programming in python language
* Provided **analytics** driven project management supports mainly to NASA and DoD
* Snalyze Engineering measures (KPI). Utilized Jupyter notebook and Microfost power BI for data analysis and visualization.
* Drafted and implemented Data Quality assessment method
* Prepared yearly budget(4 Millions) and architected 2/5 years of roadmap for the Metrics & Analytics team
* Trained team members and implemented agile practice

### GSA (General Services Administration), DC, Senior Data Scientist (Contract) Oct’16 – Nov’18

### Worked closely with the government clients to develop requirements and identify improvement scopes by implementing predictive models for GSA CAP (Common Acquisition Platform) project.

* Developed an Enterprise Information Management (EIM) Framework detailing all capabilities to manage and implement Data Layer and Advanced Analytics capability and strategy that detailed how the organization needed to manage their data, this covered all aspects of EIM from **Data Governance to Data Integration, to Data Cleansing and Master Data Management**
* Partnered with the leadership team to identify and execute high-impact opportunities to leverage extensive data
* Achieved milestones to accomplish ultimate goal to consolidate GWCM to avoid waste and abuse.
* Developed and implemented customized **Data Quality (DQ) Matrices** for all the data sets related to CAP project
* Led discussions with clients to gather business processes requirements and data requirements to develop varieties of Data Models.
* Designed the Enterprise **Conceptual, Logical, and Physical Data** Model for ‘Bulk Data Storage System’ using Embarcadero ER Studio, the data models were designed in 3NF. Used **Hive** to create on the fly data modeling and data warehousing
* Used **Tableau and Business Object** (SAP)for Business Intelligence tasks (data cleaning, sanitizing, analyzing and creating dashboards) and presented before the White House by the Project In-charge of GSA
* Drafted and reverse engineered multiple Executive dashboard for Government Wide Category Management (GWCM)
* Evaluated integration of tools and data bases ( Hue, SparkR, PySpark, R, Amazon Redshift, Amazon **S3 (Hadoop**), Pentaho) and identified bottlenecks in terms of system throughputs and data transfer error and so on
* Created predictive models using, **PCA, MCA, FAMD, MFA,** **Artificial Neural Network, regression analysis, C4.5 (decision tree), AprioriS** and others. Created test data and training data. Used **Pentaho** to perform **ETL**
* Done coding in **R , Scala and Python** running on **Spark** to run customized predictive models. Co-ordinated data migration from **Oracle (11g and R12) to Amazon RedShift** and assessed anomalies (e.g. missing value, duplicate value )
* Supervised team of BA, Data Scientists and Data Analysts to ensure on time delivery
* Mentored junior data analysts and data scientists. Worked as a Champion Data Scientists for the team and wrote proposals
* Implemented Scrum methodology (Agile Framework) to bring team dynamics
* Architected and implemented Data Quality Assessment Framework (including dimensions, sub-dimensions )for the CAP project
* Collaborated with clinical clients to determine study timelines and milestones

### American Institutes for Research (AIR), DC, Lead Data Scientist July‘12– September ‘16

(Clients: Department of Homeland Security (DHS), [Centers for Medicare & Medicaid Services](https://www.cms.gov/) (**CMS**), US Department of Education and Department of Defense-**DoD**)

* Lead team of data scientists, data architect, data analysts and testes to develop an Accumulo based Data Framework for 14 DHS (including IC) agencies and 21 data sets
* Architected Data Quality Matrix (identified and mapped dimensions, sub-dimensions and criteria) and implemented Quality process for DHS Data framework to provide **sanitized** data to our I&A team at **Booz Allen Hamilton**.
* Developed a flexible **scoring model** based on multivariate regression analysis for Data Quality Matrix
* Identified data defect categories and developed Data Quality Matrix in association with the team to identify accuracy, completeness, breadth and depth of data. Developed training data sets to train mathematical models.
* Developed search algorithms **(Brute-force search, Fibonacci Search technique, binary search methodology, Regression analysis, Decision dress (SVM) and others) to identify terrorists** and **drug traffickers.**
* Expanded AIR’s footprint with clients like Department of Education **(DE)** and Center for Medicare and Medicaid Services (**CMS),** specifically in **Machine Learning, Predictive Modelling, Advanced Analytics** and Systems integration.
* Worked on Electronic Benefits Transfer **(EBT)** Retailer Transactions **(ALERT)** System of US Department of Defense **(DoD)** for **fraud detection** and decision support system to monitor and track electronically conducted transactions completed by Supplemental Nutrition Assistance Program (SNAP) recipients in authorized meal program and food retailer locations.
* Managed the Enterprise wide Data Correlation working group; in this capacity I lead the requirements gathering, design and testing activities for data correlation matching algorithm. Additional responsibilities included leading working sessions across the user community, and briefing the development contractor on the requirements and design.
* Used Business Intelligence and Data Visualization tools: **Tableau, Microstrategy**, **Pentaho, Qlik View**
* Compiled Technology growth strategy Business Development plans related to target growth markets, staffing plans, analytics tools requirement, and account Management, and presented to Firm Leaders – this is now being executed.
* Participated in writing of Business Plan to create and expand a range of **Big Data, Data Fusion, and Information Management & Data Analytics** Capabilities across the firm.
* This business plan detailed new technology capabilities such as Information Extraction, Data Ingestion, Entity Resolution, **Geospatial Data Management**, RDF Triple Stores, specialized Data Warehouse appliances’ such as Cloudera’s **Hadoop** and EMC’s **Greenplum** Data Computing Appliance DCA, Semantic Web / Ontology’s and Complex Event Processing, needed by the Firms’ clients and subsequently added as service offerings.
* Led the development of methodologies, best practices, and lessons learned based on the implementation of the aforementioned technologies that resulted in reusable Intellectual Capital for the Firm.
* Architected “Secure Analytics Workbench”, an end to end Architecture / System proof of concept that was demonstrated to clients and firm leaders for sales to their Client Accounts. This also resulted introduction of machine learning and incorporation of static and dynamic predictive tools e.g. integration of **Artificial Neural Networks** **(ANN), BBN, KNN,**
* Extensive experience on advanced statistical methods, **Bayesian learning techniques, pattern recognition and outlier detection algorithms, and predictive modeling** methods including **clustering, decision trees, induction, naïve bayes, hiden decision tree, brute-force, GMM, regression analysis,** **Fuzzy C-means,** K-Nearest Neighbors **and random** **forest** approaches using advanced statistical tools like **R, SAS, SPSS, STATA, NVivo, Informatica and micro Strategy**
* Worked with SQL and NoSQL database infrastructure including **Oracle, SQL Server, Amazon S3 and Hadoop Clustering**
* Used **Pig** for data modeling and warehousing: **Yarn (at DoD),** **Sqoop(DHS), Flume** for streaming data in to **Accumulo** at DHS.
* Worked as a Technical Volume Lead, Pricing Volume lead / solution architect on a $2M response for **Secure** Data Fusion / Analytics Platform for AIR. Developed the technical architecture and co-authored the technical sections of the response with topics covering, Information Extraction, Entity Resolution, Data Integration, Cloud Computing, and Oracle’s Analytics Products Endeca
* Design an Earned Value Management (EVM) measurement process to complete the Level of Effort (LOE) and Work Breakdown Structure (WBS) for project planning purposes and to track the Earned Value against the LOE
* Coordinated alliances between COTS vendors for tools such as Amazon Web Services (AWS), Palantir, Informatica, InTTensity, Symantec, Microstrategy, STATA, SAS, and Teradata, SageNet and Vormetric to develop Secure Workbench Analytics
* Worked with the Chief Information Security Officer (CISO) by drafting initial Rules of Behavior (ROB), Data governance rules, PHI/PII and by filling up appropriate section of SSP (System Security Documents) and others, following NIST standards, for FedRAMP/FISMA and CMS security certification to generate IDIQ contracts with different Government entity
* Worked with large and complex databases containing millions to **billions** of records (megabyte to petabyte)
* Ensured quality standards are met for patient data in clinical database through data cleaning processes and data audits
* Architected **Master Data Management** framework. Researched on healthcare environment for trends and opportunities.
* Drafted and edited artifacts like Data Tagging Workbook, Schema, ERD, Data Dictionary and others
* Engaged and collaborated among US Department of Education, CMS, AIR Advanced Analytics team and Marketing Team for synergies and cross pollination, development and resource efficiencies and to identify independent variables for predictive analysis
* Evaluated and used Microsoft Azure with Zeppelin Notebook

### Surescripts LLC, VA, Principal Data Scientist Jan ‘10 – June ‘12

* Was contracted as a Technical Project Manager and then got converted as a Principal Data Scientist within 4 months
* Developed models to predict healthcare **fraud, waste, and abuse (FWA)** using structured and semi structured data (from Oracle DB, Mongo DB(JSON and Hadoop). Processed image, pdf, text, videos and xml.
* Led data intelligence practice with cross-cutting business and functional focus on Data Analytics and Business Intelligence.
* Peer reviewed design and development of artifacts and provided expertise on implementation of the Cognos TM 1 software.
* Participated on authoring an Enterprise Wide Information Extraction (IE) Strategy and COTS IE Evaluation Plan. The IE Strategy focused on the Extraction of 4 types of unstructured data Entities [e.g. Patient identity, Places, Prescription no., service provider], Temporal / Time; Geographic; and the Relationships between those data
* As a Chief Data Scientist, I was responsible for completing a data quality assessment of several Surescripts databases and drafted data modeling, data cleansing and a data management Strategy
* Performed as a Task Lead to implement data mining algorithms predictive model and to assess data security (PII/PHI), business intelligence, and document management; e.g. Classification methods: **Logistic Regression, Decision Trees, SVM, Random Forest, Neural Network (ANN),** Regression methods utilized - Linear, Nonlinear, Boosted Regression Trees, ensemble methods; Clustering methods: **K-means, Hierarchical Clustering, Mixture Modelling**
* Additional tasks performed included weekly status reports for clients and executive team, resolving data quality issues, maintaining project hours utilized, and interfacing with client as requested.
* Drove the collection of data and refinement from multiple high-volume data sources. Researched methods to improve statistical inferences of variables across models and developed statistical, mathematical and predictive models.
* Prepared documents and made presentations to make informed, data-driven decisions to meet business objectives.
* Technologies used, included but not limited to**: SQL, R, Python, NLP, Spark, Hive, Oxdata H20, MapReduce, Hadoop,**
* Experienced working w/ IT to facilitate Surescripts builds to enable real time predictive analytics.
* Conducted instillation of regular **business intelligence** process and visualized key prioritization, velocity, quantity and frequency
* Conducted cluster analysis and Euclidean methods to build objective global segmentation models and analytics to provide Surescripts user segmentations with focused, authenticated and encrypted health care related information
* Time-series Modelling/Forecasting - AR, ARMA, GARCH, Exponential Smoothing
* Provide information and analytical support for critical decision-making with clinical, cost and care delivery data, including demand and census forecasting. Drafted predictive models to identify waste and abuse reduction waste.

**AAMC, DC, Technical Manager (Data Science) Nov ‘08 – Jan ‘10**

* Coordinated technical architecture and authored the technical sections of the response with topics covering, Information Extraction, Entity Resolution, Data Integration, RDF Triple Store and specialized Data Warehouse appliances, e.g. Netezza
* As a Technical Project Manager, managed Data Collection, Storage & Dissemination (DCSD) and was responsible for managing a Software Development team of 12 FTEs, and two task orders (ERAS re-engineering for PDWS and DWS)
* ERAS was a program that was several months behind schedule, lacked requirements, design and software development delivery approach. Upon taking over I led a 90 day project recovery effort that assessed project health and to created revised delivery plan
* Lead teams of data analysts, quality analysts, DBAs and Developers
* Prepared presentation for Manager, Data Science, to present analytical findings to non-analytical audiences
* Led the successful delivery of on time and on budget of the project, which contained four (sub-systems) and an 11 years of Data Conversion effort from an ERAS mainframe, through Requirements, Design, Development, Testing and deployment
* Done Management of all contract modifications, funding increments, deliverables, and Bi-weekly status reporting to the Executive Committee including Program Directors, Chief Information Officer, Contracting Office (CO) and COTR.
* Coordinated weekly meetings with the Chief Information officer and Associate Directors of Operations to report status, discuss action items, risks, issue, program reviews, contractual items and delivery status.
* Managed the staffing, hiring, and project evaluations of all staff, including interviewing to ramp up and hire developer staff
* During this task order due to staffing shortages, served as the lead for the overall System Architecture and Design, Requirements, Testing (SQL based Back end test support), Configuration Management, Address Validation and Geocoding, Entity Resolution, and ETL design activities prior to hiring a lead architect and development manager.
* Performed project closeout activities, transitioned the IT system to the Operations & Maintenance (O&M Team) and transitioned the project to another project manager for Release 2.
* Authored Data Quality Plan and Legacy System Retirement Plan and co-authored the Entity Resolution Plan, these were presented to AAMC senior management with details on how the project team would address these areas.
* Coordinated Data driven and **ETL (Informatica)** driven development
* Developed data collection tools to clean and integrate complex datasets for users including Program directors, student applicants for admission and residence and others
* Used analytics tools to analyze and interpret large datasets on applicant’s activities (E.g. R, STATA, NVivo, Informatica)
* Performed and co-ordinated analysis of large amounts of data (i.e. transactional, click stream, social media) to build **predictive models** (supervised and unsupervised ) with Structured and Semi-structured data.

**Royal Bank of Canada (RBC), Ontario, Quantitative Analyst June ’07 – Oct ‘08**

* Lead a team of 5 FTE including Quantitative Analysts and Business Analysts
* Interpreted and translated Business and System Requirements and interacted with users, product owners and developers.
* Administered business intelligence systems and done business data analysis, visualization and reporting.
* Performed End-to-end financial data flow testing and performed test data collection, integration and quality control.
* Created predictive model based on **Regression, Logit Regression and Decision Tree**
* Validated of formulated model through **residual analysis, R-square value analysis, normal distribution** and so on
* Collected structured data from various sources, done data cleaning and created synthetic data to address missing values
* Closely communicated with the client (business team/ Product Owner) to understand business requirements and to transfer them as test cases and test steps to implement and facilitate Test Driven Development (TDD)
* Analyzed business flow of the application and used MS Project and JIRA as tracking tools
* Worked with the team from the beginning to the end of the SDLC in an agile (scrum) environment
* Wrote **SQL (2008)** queries for RDBMS (Oracle 10g and SQL) to update tables and pull data and performed analysis.
* Done data driven testing and Created meta data for testing purpose
* Used **Informatica** to perform **ETL** processing and Interfaced with data systems to perform focused reviews.
* Communicated defects, encountered during regression test and followed-up with developers until all issues were resolved.
* Done API testing with XML scripts, used SOAP UI and REST Client. Performed ad-hoc queries with **PL SQL**
* Collaborated with Project Manager to determine data driven reporting needs and tracking measures
* Done content management by checking out/updating contents, comparing versions, merging files, editing change requests, generating reports & charts, creating & organizing requirements, etc. using Pulse and knowledge link
* Used source code control systems such as GitHub and SVN to track and verify defect fixes

**Manulife Insurance, Ontario, Senior Analyst (Business Intelligence/Quantitative Analysis) Aug ‘05 – June ‘07**

* Communicated Clients on Business requirements, performed quantitative analysis and worked closely with Data Analysts to prepare presentations based on fact and figure for non-technical stakeholders
* Worked as a member of Manulife’s Business Intelligence Practice with cross-cutting Business and Functional focus on Technology Strategy & Architecture, Systems Integration, Business Analytics and Information / Data Management technologies and methodologies providing solutions. Provided expertise in both client delivery of Technology solutions and new business development and captured leadership on future opportunities.
* Analyzed business requirements to transform in to technical requirements. Created **Traceability** Matrix for Business requirements and test cases
* Used **Cognos 8BI** and **Micro Strategy** to analyze and present data analytics and predictive models for Business Units and relevant Stakeholders
* Used Fusion Charts to and Spreadsheet for data visualizing before the business entitiy.
* Used **SQL and PL/SQL** to pull data to prepare reports. Drafted Metadata for data driven testing
* Worked with the development team and test team during development and testing phase to define acceptance criteria, definition of done and to perform functional and data driven testing

**Academic Experience:**

**University of Maryland, College Park, MD Research Assistant/Teaching Assistant [Part-time] Jan ‘06 – Present**

**George Washington University, DC/University of Toronto, ON**

* Doing research on formulation of prognostic and dynamic Software Defect Prediction model with integration of Machine learning (Artificial Neural network) at University of Maryland, College Park [PhD program, Part-time]
* Done Design of Experiments (DOE). Analyzed Data generated from SEM, AFM, Custom PVT device using MathLab
* Done FMEA (Failure Mode Effects Analysis) to identify possible relationship and failures in Nano Composites (PNC)
* Ran Data Simulations (Genetic Algorithm) to optimize material vs gas content

**Clearance:** Active Public trust, Top Secret Clearance (in process, sponsored by DHS)

**References:**  Can be provided if requested.