

Job Seeker

Sr. Data scientist - Verisign

Parsippany, NJ - Email me on Indeed: [indeed.com/r/3dc212e66fc27472](https://www.indeed.com/r/3dc212e66fc27472)

- > Over 8 years of Experienced working as a Data Scientist/Data Architect/Data Analyst/Data Modeling with emphasis on Data Mapping, Data Validation in Data Warehousing Environment.
- > Extensive experienced on business intelligence (and BI technologies) tools such as OLAP, Data warehousing, reporting and querying tools, Data mining and Spreadsheets
- > Worked on different type of Python modules such as requests, boto, flake8, flask, mock and nose
- > Extensive experienced on business intelligence (and BI technologies) tools such as OLAP, Data warehousing, reporting and querying tools, Data mining and Spreadsheets
- > 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 applications
- > Worked on different type of Python modules on modules such as requests, boto, flake8, flask, mock and nose
- > Strong understanding of when to use an ODS or data mart or data warehousing.
- > Experienced in employing R Programming, MATLAB, SAS, Tableau and SQL for data cleaning, data visualization, risk analysis and predictive analytics
- > Adept at using SAS Enterprise suite, R, Python, and Big Data related technologies including Hadoop, Hive, Pig, Sqoop, Cassandra, Oozie, Flume, Map-Reduce and Cloudera Manager for design of business intelligence applications
- > Ability to provide wing-to-wing analytic support including pulling data, preparing analysis, interpreting data, making strategic recommendations and presenting to client/product teams.
- > Hands on experienced with Machine Learning, Regression Analysis, Clustering, Boosting, Classification, Principal Component Analysis and Data Visualization Tools
- > Strong programming skills in a variety of languages such as Python and SQL
- > Familiarity with Crystal Reports, and SSRS - Query, Reporting, Analysis and Enterprise Information Management
- > Excellent knowledge on creating reports on Pentaho Business Intelligence.
- > Experienced in Database using Oracle, XML, DB2, Teradata15/14, Netezza, server, Big Data and NoSQL.
- > Worked with engineering teams to integrate algorithms and data into Return Path solutions
- > Worked closely with other data scientists to create data driven products
- > Strong experienced in Statistical Modeling/Machine Learning and Visualization Tools
- > Experienced in working with large-scale data set
- > Expert at Full SDLC processes involving Requirements Gathering, Source Data Analysis, Creating Data Models, and Source to target data mapping, DDL generation, performance tuning for data models.
- > 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
- > Experienced working with large, real world data - big, messy, incomplete, full of errors
- > Hands-on experienced with NLP, mining of structured, semi-structured, and unstructured data
- > Experienced and in-depth knowledge of the SAS Enterprise Miner, Python programming language

WORK EXPERIENCE

Sr. Data scientist

Verisign - Reston, VA -

January 2016 to Present

Responsibilities:

- > Built data pipelines for reporting, alerting, and data mining. Experienced with table design and data management using HDFS, Hive, Impala, Sqoop, MySQL, Mem SQL, Grafana/Influx DB, and Kafka.
- > Worked with statistical models for data analysis, predictive modelling, machine learning approaches, recommendation and optimization algorithms.
- > Working in Business and Data Analysis, Data Profiling, Data Migration, Data Integration and Metadata Management Services.
- > Worked extensively on Databases preferably Oracle 11g/12c and writing PL/SQL scripts for multiple purposes.
- > Built models using Statistical techniques like Bayesian HMM and Machine Learning classification models like XG Boost, SVM, and Random Forest using R and Python packages.
- > Worked with data compliance teams, data governance team to maintain data models, Metadata, data Dictionaries, define source fields and its definitions.
- > Worked with Big Data Technologies such Hadoop, Hive, MapReduce
- > Developed MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming.
- > Setup storage and data analysis tools in Amazon Web Services cloud computing infrastructure.
- > A highly immersive Data Science program involving Data Manipulation & Visualization, Web Scraping, Machine Learning, Python programming, SQL, GIT, Unix Commands, NoSQL, MongoDB, Hadoop.
- > Performed scoring and financial forecasting for collection priorities using Python, R and SAS machine learning algorithms.
- > Handled importing data from various data sources, performed transformations using Hive, MapReduce, and loaded data into HDFS
- > Managed existing team members, lead the recruiting and on boarding of a larger Data Science team that addresses analytical knowledge requirements.
- > Worked directly with upper executives 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.
- > Developed a generic model for predicting repayment of debt owed in the healthcare, large commercial, and government sectors.
- > Created SQL scripts and analyzed the data in MS Access/Excel and Worked on SQL and SAS script mapping.
- > Developed a legal model for predicting which debtors respond to litigation only.
- > Created multiple dynamic scoring strategies for adjusting the score upon consumer behavior such as payment or right-party phone call.
- > Rapid model creation in Python using pandas, numpy, sklearn, and plot.ly for data visualization. These models are then implemented in SAS where they are interfaced with MSSQL databases and scheduled to update on a timely basis.
- > Data analysis using regressions, data cleaning, excel v-look up, histograms and TOAD client and data representation of the analysis and suggested solutions for investors
- > Attained good knowledge in Hadoop Data Lake Implementation and HADOOP Architecture for client business data management.
- > Identifying relevant key performing factors; testing their statistical significance
- > Above scoring models resulted in millions of dollars of added revenue to the company and a change in priorities of the entire company.

Environment: R, SQL, Python 2.7.x, SQL Server 2014, regression, logistic regression, random forest, neural networks, Topic Modeling, NLTK, SVM (Support Vector Machine), JSON, XML, HIVE, HADOOP, PIG, Sklearn, SciPy, GraphLab, No SQL, SAS, SPSS, Spark, Hadoop, Kafka, HBase, MLib.

EDUCATION

Master of science in Data analytics

Clarkson university

SKILLS

APACHE HADOOP HDFS (2 years), APACHE HADOOP MAPREDUCE (2 years), APACHE KAFKA (2 years), AWS (2 years), DATABASES (2 years)

ADDITIONAL INFORMATION

Technical Skills:

Data Modelling Tools Erwin r9.6, 9.5, 9.1, 8.x, Rational Rose, ER/Studio, MS Visio, SAP Power designer.
Programming Languages Oracle PL/SQL, Julia, UNIX shell scripting, Java. Python (NumPy, SciPy, Pandas, Gensim, Keras), R (Caret, Weka, ggplot)

Big Data Technologies Hadoop (Hive, HDFS, MapReduce, Pig, Kafka, Spark), Cassandra, MongoDB, AWS (RDS, Dynamodb, Redshift)

Reporting Tools Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x,
AWS Cloud EC2, S3, RDS, DynamoDB, Kinesis, Redshift

ETL Informatica Power Centre, SSIS.

Project Execution Methodologies

Ralph Kimball and Bill Inmon, data warehousing methodology, Rational Unified Process (RUP), Rapid Application Development (RAD), Joint Application Development (JAD)

Operating Systems Windows, UNIX, LINUX, Mac OS

Databases Oracle [...] SQL Server, MySQL, PostgreSQL