**Akrambaig Saudagar**

Data Analyst\Business Analyst

**PROFESSIONAL SUMMARY**

* Around 7 years of experience in Data Analysis, Data mining with large data sets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Data Visualization.
* Leverage a wide range of data analysis, machine learning and statistical modeling algorithms and methods to solve business problems.
* Experience working in Agile Scrum Software Development.
* Decomposition and documentation of Business Lineage and Technical Lineage.
* Data Lineage and Metadata management Testing Business/Data Analysis
* Knowledge in Cloud services such as Microsoft Azure and Amazon AWS.
* Knowledge about Big Data toolkits like Mahout, Spark ML, H2O.
* Professional working experience in Machine learning algorithms such as Linear Regression, Logistic Regression, Random Forests, Decision Trees, K-Means Clustering and Association Rules.
* Deep expertise with Statistical Analysis, Data mining and Machine Learning Skills using R, Python and SQL.
* Experience in Machine learning using NLP text classification, churn prediction using Python.
* Proficient in managing entire data science project life cycle and actively involved in all the phases of project.
* Extensive experience in implementing metadata, Business Glossary, Data Lineage and data quality.
* Hands on Spark MLlib utilities such as including classification, regression, clustering, collaborative filtering, dimensionality reduction.
* Strong skills in Machine Learning algorithms such as Linear Regression, Logistic Regression, Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, K-Nearest-Neighbors, K-means Clustering, Neural networks, Ensemble Methods.
* Experience in deploying Hadoop cluster on public and private cloud environments like: Amazon AWS, Rackspace and Openstack.
* Working experience in implementing Machine Learning Algorithms using MLLib and Mahout in Hadoop ecosystem and Apache Spark framework such as HDFS, Map Reduce, HiveQL, Spark SQL and PySpark.
* Hands on experience Hadoop, Deep Learning Text Analytics and IBM Data Science work bench tools.
* Hands on experience in Data Governance, Data Mining, Data Analysis, Data Validation, Predictive modeling, Data Lineage and Data Visualization in all the phases of the Data Science Life Cycle.
* Extensively worked for data analysis using R Studio, SQL, Tableau and other BI tools.
* Experience in visualization tools like, Tableau 9.X, 10.X for creating dashboards.
* Used the version control tools like Git 2.X and VM.
* Passionate about gleaning insightful information from massive data assets and developing a culture of sound, data-driven decision making.
* Skilled in Advanced Regression Modelling, Correlation, Multivariate Analysis, Model Building, Business Intelligence tools and application of Statistical Concepts.
* Experienced in Visual Basic for Applications and VB programming languages to work with developing applications.
* Working knowledge of Amazon Web Services (AWS) and Cloud Data Management.
* Experienced in Big Data with Hadoop, MapReduce, Spark 1.6, PySpark, SparkSQL, HDFS, and Hive 1.X.
* Working with large sets of complex datasets that include structured, semi-structured and unstructured data and discover meaningful business insights.
* Knowledge in NoSQL databases such as HBase, Cassandra, and Mongo DB etc.
* Highly skilled in using statistical analysis using R, SPSS, Matlab and Excel.
* Experience working with SAS Language for validating data and generating reports.
* Experience working with Web languages such as Html, CSS, Rshiny etc.
* Strong Data Analysis skills using business intelligence, SQL and MS Office Tools.
* Strong SQL Server programming skills, with experience in working with functions, packages and triggers.
* Hands on experience in implementing LDA, Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis and good knowledge on Recommender Systems.
* Highly skilled in using Hadoop (pig and Hive) for basic analysis and extraction of data in the infrastructure to provide Data Summarization.
* Highly skilled in using visualization tools like Tableau, ggplot2 and d3.js for creating dashboards.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Languages** | Python, SQL, PL/SQL, Unix Shell Scripting (Korn/C/Bourne), Java, J2EE, C, C++ |
| **Documentation** | MS Office Suite (PowerPoint, Word, Excel, Access, Project and Outlook), SharePoint, Rational Requisite Pro, MS Visio |
| **Front end tools** | Microsoft front page, Microsoft Office, Cold Fusion, HTML, XML |
| **Version Tools** | Visual Source safe, SVN, JIRA, Team Foundation Service (TFS), SharePoint |
| **Data Visualization Tool** | Tabular |
| **Reporting Tools** | Business Objects Developer Suite 5.1, Business Objects XI/XIR2/XIR3.1, COGNOS Suite, COGNOS Report Net, Crystal Reports, Oracle Reports 10g/9i/6i |
| **Databases** | Oracle 11g/10g/9i/8i, SQL Server 2008/2005/2000, Sybase, DB2, MS Access |
| **Operating Systems** | UNIX, LINUX, Windows NT/XP/Vista/98/95, Red Hat (RH5 to RHEL 6), MS DOS, IBM AIX, Sun Solaris |
| **Other Tools** | Informatica 9.1.6/8/6/1/8.5.1 (Power Center), Erwin 8.1/7.3, Visio 2007, Oracle Forms and Reports, APEX, Golden Gate, Autosys, Subversion, Revision Control System, BMC Control-M |

**WORK EXPERIENCE**

**Santander Bank, Boston, MA December 2018 – Present**

**Data Analyst\Business Analyst**

**Responsibilities:**

* Responsibilities involved the analyzing end user requirements and communicating and modeling them to the development team.
* Retrieved data from Hadoop Cluster by developing a pipeline using Hive (HQL), SQL to retrieve data from Oracle database and used ETL for data transformation.
* Performed data wrangling to clean, transform and reshape the data utilizing pandas library. Analyzed data using SQL, R, Java, Scala, Python, Apache Spark and presented analytical reports to management and technical teams.
* Established source-to-target mapping based on the above analysis and maintain the data lineage in MSSQL database.
* Worked with different datasets with complexity including both structured and unstructured data and Participated in all phases of Data mining, Data cleaning, Data collection, variable selection, feature engineering, developing models, Validation and Visualization.
* Developed predictive models on large scale datasets to address various business problems through leveraging advanced statistical modeling, machine learning and deep learning.
* Developed SSIS package to extract data from the lineage database and transformed it to the required format to be loaded onto IBM IIS.
* Responsible in maintaining the Enterprise Metadata Library with any changes or updates
* Analyzed Historical data by using various machine learning algorithms such as clustering, multiple linear regression, logistic regression, SVM, Naive Bayes, Random Forests, K-means, & KNN for data analysis.
* Conducted exploratory data analysis using Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, SciPy, NLTK in Python for developing various machine learning algorithms.
* Implemented Data Quality validation techniques to validate data and identified many anomalies. Extensively worked on statistical analysis tools and adept at writing code in Advanced Excel, R and Python.
* Created data dictionary, Data mapping for ETL and application support, DFD, ERD, mapping documents, metadata, DDL and DML as required.
* Enforced model Validation using test and Validation sets via K- fold cross validation, statistical significance testing.
* Developed automated data pipelines from various external data sources (web pages, API etc) to internal data warehouse (SQL sever, AWS), then export to reporting tools like Datorama by Python.
* Worked with various kinds of data (open-source as well as internal). I have developed models for labeled and unlabeled datasets, and have worked with big data technologies, such as Hadoop and Spark, and cloud resources, like Azure and Google Cloud.
* Enforced F-Score, AUC/ROC, Confusion Matrix, Precision, and Recall evaluating different model’s performance.
* Multi-layers Neural Networks built in Python Scikit-learn, Theano, TensorFlow and keras packages to implement machine learning models.
* Created Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data.
* Created complex charts and graphs with drill downs that will allow various divisions to quickly locate outliers and
* correct any anomalies.
* Developing Stored Procedures, Functions, Views and Triggers, complex SQL queries using SQL Server, TSQL and Oracle PL/SQL.
* Worked with various data sources with multiple relational databases like Oracle11g /Oracle10g/9i, MS SQL Server;
* Relational and Flat Files into the staging area, ODS, Data Warehouse and Data Mart.
* Designed and developed standalone data migration applications to retrieve and populate data from Azure Table / BLOB storage to Python, and Power BI.
* R programming language for graphically critiquing the data and performed data mining. Interpreting Business requirements, data mapping specifications and responsible for extracting data as per the business requirements.
* Participated in features engineering such as feature generating, PCA, Feature normalization and label encoding with Scikit-learn preprocessing. Data Imputation using variant methods in Scikit-learn package in Python.
* Utilized Informatica toolset (Informatica Data Explorer and Data Quality) to inspect legacy data for data profiling.
* Used Teradata utilities such as Fast Export, MLOAD for handling various tasks data migration/ETL from OLTP Source Systems to OLAP Target Systems.
* Created reports, dashboards and data Visualizations by using Tableau, to explain and communicated data insights, significant features, model's score and performance to perfectly elucidate for both technical and business teams.

**Environment:** Python 3.6.4, R Studio, MLLib, Regression, NoSQL, SQL Server, Hive, Hadoop Cluster, ETL, Spyder 3.6, Agile, Tableau, Java, NumPy, Pandas, Matplotlib, Power BI, Scikit-Learn, Seaborn, e1071, ggplot2, Shiny, TensorFlow, AWS, Azure, HTML, XML, Informatica Power Center, Teradata.

**SVC Links, Pittsburgh, PA December 2016 - December 2018**

**Data Analyst/Engineer**

**Responsibilities:**

* Performed data wrangling to clean, transform and reshape the data utilizing pandas library. Analyzed data using SQL, R, Java, Scala, Python, Apache Spark and presented analytical reports to management and technical teams.
* Worked with different datasets which includes both structured and unstructured data and Participated in all phases of Data mining, Data cleaning, Data collection, variable selection, feature engineering, developing models, Validation and Visualization.
* Developed SQL scripts and execute it against various systems listed in the lineage for the defective data element and documented to check/validate whether it is getting its value from the source data elements according to the derivation logic.
* Integrate Data stage Metadata to Informatica Metadata and created ETL mappings and workflows.
* Developed predictive models on large scale datasets to address various business problems through leveraging advanced statistical modeling, machine learning and deep learning.
* Implemented public segmentation using unsupervised machine learning algorithms by implementing K-means algorithm by using PySpark using data munging.
* Designed, configured and deployed Amazon Web Services (AWS) for a multitude of applications
* Experience in Machine learning using NLP text classification, churn prediction using Python.
* Worked on different Machine Learning models like Logistic Regression, Multi-layer perceptron classifier and K-means clustering.
* Lead discussions with users to gather business processes requirements and data requirements to develop a variety of conceptual, logical and Physical Data models.
* Expertise in Business intelligence and Data Visualization tools like Tableau.
* Handled importing data from various data sources, performed transformations using Hive, MapReduce and loaded data into HDFS.
* Good knowledge in Azure cloud services, Azure Storage to manage and configure the data.
* Used R and Python for Exploratory Data Analysis to compare and identify the effectiveness of the data.
* Created clusters to classify control and test groups.
* Analyzed and calculated the life cost of everyone in a welfare system using 20 years of historical data.
* Developed triggers, stored procedures, functions and packagers using cursors associated with the project using PL/SQL.
* Used Python, R, SQL to create statistical algorithms involving Multivariate Regression, Linear Regression, Logistic
* Regression, PCA, Random forest models, Decision trees, SVM for estimating and identifying the risks of welfare dependency.
* Designed and implemented a recommendation system which leveraged Google Analytics data and the machine learning models and utilized Collaborative filtering techniques to recommend policies for different customers.
* Performed analysis such as Regression analysis, Logistic Regression, Discriminant Analysis, Cluster analysis using SAS programming.
* Worked on No SQL databases including Cassandra, Mongo DB, Mark Logic and HBase to access the advantages and disadvantages of them for a goal of a project.

**Environment:** Hadoop, HDFS, Python 3.x (Scikit -Learn/ Keras/ SciPy/ NumPy/ Pandas/ Matplotlib/ NLTK/ Seaborn), R (ggplot2/ caret/ trees/ arules), Tableau (9.x/10.x), Machine Learning (Logistic regression/ Random Forests/ KNN/ K-Means Clustering / Hierarchical Clustering/ Ensemble methods/ Collaborative filtering), GitHub, Agile/ SCRUM

**Fiserv, Dallas, TX October 2015- November 2016**

**Data Analyst/ Engineer**

**Responsibilities:**

* Responsible for gathering requirements from Business Analyst and Operational Analyst and identifying the data sources required for the request.
* Enhanced Data collection procedures to include information that is relevant for building analytic systems and created a value from data by performing advanced analytics and statistical techniques to determine to deepen insights, optimal solution architecture, efficiency, maintainability, and scalability which make predictions and generate recommendations.
* Creating automated anomaly detection systems and constant tracking of its performance  
  Strong command of data architecture and data modeling techniques.
* Maintained and developed complex SQL queries, stored procedures, views, functions, and reports that qualify customer requirements using Microsoft SQL Server 2008 R2.
* Support Sales and Engagement's management planning and decision making on sales incentives and production by, developing and maintaining financial models, reporting and sensitivity analysis by customer segment.
* Worked with the ETL team to document the transformation rules for Data migration from OLTP to Warehouse environment for reporting purposes.
* Used Pandas, NumPy, Seaborn, SciPy, Matplotlib, Sci-kit-learn.
* Worked on data modeling and produced data mapping and data definition documentation.
* Having experience with visualization technologies such as Tableau.
* Data mining using state-of-the-art methods.
* Extending company's data with third party sources of information when needed.
* Enhancing data collection procedures to include information that is relevant for building analytic systems.
* Advanced and developed test plans to ensure successful delivery of a project. Employed performance analytics predicated on high-quality data to develop reports and dashboards with actionable insights.
* Précised Development and implementation of several types of sub-reports, drill down reports, summary reports, parameterized reports, and ad-hoc reports using SSRS through mailing server subscriptions & SharePoint server.
* Generated comprehensive analytical reports by running SQL queries against current databases to conduct data analysis.
* Generated ad-hoc reports using Crystal Reports 9 and SQL Server Reporting Services (SSRS).
* Generated the reports and visualizations based on the insights mainly using Tableau and developed dashboards for the company insight teams.
* Worked closely with data architect to review all the conceptual, logical and physical data base design models with respect to functions, definition, maintenance review and support data analysis, Data quality and ETL design that feeds the logical data models.
* Created financial package that supports 3-Year financial plan for all AWS cloud services infrastructure expenses.

**Environment**: SQL Server 2008R2/2005 Enterprise, SSRS, SSIS, Crystal Reports, Windows Enterprise Server 2000, DTS, SQL Profiler, Tableau, Qlik View, Django, ad-hoc, SharePoint and Query Analyzer.

**Info Logic Systems, Hyderabad, India Feb 2013– May 2015**

**Data Analyst**

**Responsibilities:**

* Coached the stakeholders regarding the agile practices.
* Mentored the team and helped in forging a strong relationship.
* Conducted requirement analysis by carrying out GAP Analysis.
* Evangelized the benefits of Scrum to ensure its smooth adoption and tracked velocity and sprint progress
* Managed stakeholder expectations, status, risks, issues, and created product release plan.
* Handled issues/ risks proactively and worked with the Senior Management Teams to deliver the project
* Created Project initiation document (PID) and established Requirements traceability Matrix (RTM) using Rational Requisite Pro to trace the completeness of requirements
* Created UML diagrams to convey the requirements to the team members.
* Effectively conveyed weekly project status, burn down charts to the stakeholders.
* Conducted Sprint planning and retrospective activities through Adobe Connect.
* Conducted coaching sessions regarding Agile development methods
* Participated and hosted meetings, technical discussion, reviews and release planning
* Resolved impediments by coordinating and collaborating with the cross-functional teams.
* Created functional and non- functional requirement documents and technical specification documents for the Hadoop environment
* Understood all the Hadoop architecture and drove all the meetings
* Conducted safety check to make sure that my team is feeling safe for the retrospectives
* Aided in data profiling by examining the source data
* Performed data mappings to map the source data to the destination data
* Developed Use Case Diagrams to identify the users involved. Created Activity diagrams and Sequence diagrams to depict the process flows
* Experienced with NOSQL database
* Loaded and transformed large sets of structured, semi structured and unstructured data
* Coordinated cluster services through Zookeeper
* Involved with data analysts during the ETL operations in identifying the source files/databases from OLTP systems. Implemented the transformation logic in order to achieve data uniformity
* Performed SQL Stored Procedures for complex queries
* Provided Knowledge Transfer to the team. Coached my team members (one-to-one), if they were stuck anywhere
* Displayed information using Information radiators
* Guide the stakeholders in performing UAT (User Acceptance Testing)

**Environment:** Scrum, Version One, Oracle, HTML5, Tableau, MS Excel Ideaboardz, Server Services, Informatica PowerCenter v9.1, SQL, Microsoft Test Manager, Adobe Connect, MS Office Suite, LDAP, Kerberos, Knox, Ranger, Atlas, Hive, Spark, Pig, Oozie, Zookeeper, Zeppelin, Sqoop, Kaka, Nifi.

**Education:**

* Masters in computer information systems/California University of management and sciences/2017
* Bachelor’s in Bio technology/Achrya Nagarjuna University/2007