***Laxmi***

***Data Science***

*E-mail:* **laxmi10.ds@gmail.com**

*Mobile: (612) 567-8448*

***Professional Summary:***

* *Around* ***8 years*** *of IT experience as a* ***Data Scientist,*** *including profound expertise and experience on statistical data analysis such as transforming business requirements into* ***analytical models, designing algorithms,*** *and* ***strategic solutions*** *that scales across* ***massive volumes of data****.*
* *Experience and* ***domain knowledge*** *in various industries such as* ***healthcare, insurance, retail, banking,*** *media and technology.*
* *Proficient in* ***Statistical Methods*** *like* ***Regression models, hypothesis testing, confidence intervals, principal component analysis and dimensionality reduction.***
* *Experience in* ***Data Analysis, Data Migration, Data Cleansing, Transformation****, I****ntegration, Data Import, and Data Export*** *using multiple ETL tools such as Ab Initio and Informatica Power Center.*
* *Expert in* ***R and Python scripting****. Worked in stats function with* ***Numpy, visualization*** *using* ***Matplotlib/Seaborn*** *and* ***Pandas*** *for organizing data.*
* *Professional working experience in* ***Machine Learning algorithms*** *such as* ***LDA, linear regression, logistic regression, Naive Bayes, Decision Trees, Clustering, and Principle Component Analysis.***
* *Hands on experience with* ***big data*** *tools like* ***Hadoop, Spark, Hive, Pig, Impala, Pyspark, Spark SQL.***
* *Experience working with* ***data modeling*** *tools like* ***Erwin, Power Designer and ER Studio****.*
* *Good knowledge in Database Creation and maintenance of physical data models with* ***Oracle, Teradata, Netezza, DB2, MongoDB, HBase and SQL Server databases.***

***Technical Skills:***

|  |  |
| --- | --- |
| ***Bigdata/Hadoop Technologies*** | *Hadoop, HDFS, YARN, MapReduce, Hive, Pig, Impala, Sqoop, Flume, Spark, Kafka, Storm.* |
| ***Languages*** | *Python, R, HTML5, SAS, Java, Scala, Python* |
| ***Business Intelligence Tools*** | *Tableau server, Tableau Reader, Tableau, Splunk, SAP Business Intelligence and Azure Data Warehouse* |
| ***Development Tools*** | *Microsoft SQL Studio, IntelliJ, Eclipse, NetBeans.* |
| ***Development Methodologies*** | *Agile/Scrum, UML, Design Patterns, Waterfall.* |
| ***Reporting Tools*** | *MS Office (Word/Excel/Power Point/ Visio/Outlook), Crystal reports XI, SSRS, cognos 7.0/6.0.* |
| ***Databases*** | *R Server, Microsoft SQL Server 2012, MySQL 4.x/5.x, Oracle 11g, 12c, DB2, Teradata.* |
| ***Operating Systems*** | *All versions of Windows, UNIX, LINUX, Macintosh HD, Sun Solaris* |

***Professional Experience:***

***Client: Adobe, Califorina Jan 2017 – Till date***

***Role: Data Scientist***

***Description:*** Adobe Systems Incorporated (/əˈdoʊbiː/ ədohbee) is an American multinational computer software company. The company is headquartered in San Jose, California, United States. Adobe has historically focused upon the creation of multimedia and creativity software products, with a more recent foray towards rich Internet application software development.

***Responsibilities:***

* *Extracted data from* ***HDFS*** *and prepared data for exploratory analysis using* ***data munging***
* *Built models using Statistical techniques like Bayesian HMM and Machine Learning classification models like* ***XGBoost****,* ***SVM****, and* ***Random Forest.***
* *Participated in all phases of* ***data mining, data cleaning, data collection, developing models, validation, visualization*** *and* ***performed Gap analysis.***
* *A highly immersive Data Science program involving* ***Data Manipulation*** *&* ***Visualization****,* ***Web Scraping****,*
* ***Machine Learning, Python programming, SQL****,* ***GIT****,* ***MongoDB, Hadoop****.*
* *Data Manipulation and Aggregation from different source using* ***Nexus, Business Objects, Toad, Power BI*** *and* ***Smart View.***
* *Skilled in using* ***dplyr*** *and* ***pandas*** *in* ***R*** *and* ***Python*** *for performing exploratory data analysis.*
* *Developed scalable model using Spark (RDD, MLlib, Ml, Data frames) in Scala*
* *Implemented Classification using supervised algorithms like* ***Logistic Regression, Decision trees, Naive Bayes, KNN.***
* *As Architect delivered various complex* ***OLAPdatabases/cubes, scorecards, dashboards*** *and* ***reports****.*
* *Updated* ***Python*** *scripts to match training data with our database stored in* ***AWS Cloud Search****, so that we would be able to assign each document a response label for further classification.*
* *Used Teradata utilities such as* ***Fast Export, MLOAD*** *for handling various tasks data* ***migration/ETL*** *from* ***OLTP*** *Source Systems to* ***OLAP Target Systems***

*Environment: Python, BI, ER Studio 9.7, JSON XMA, XGBoost.*

***Client: BBH, Newyork Nov 2015 – Dec 2016***

***Role: Data Scientist***

***Description:*** *We make ambitious ideas for ambitious clients. Using the power of creativity for outsized impact on culture and business growth, we work with some of the most innovative marketers in the world and are consistently honored with the highest strategic and creative accolades as a result.*

***Responsibilities:***

* *Responsible for analyzing large data sets to develop multiple custom models and algorithms to drive innovative business solutions.*
* *Perform* ***preliminary data analysis*** *and handle anomalies such as missing, duplicates, outliers, and imputed irrelevant data.*
* *Performed* ***Market Basket Analysis*** *to identify the groups of assets moving together and recommended the client their risks*
* *Experience in determine trends and significant data relationships using advanced Statistical Methods.*
* *Implemented techniques like* ***forward selection, backward elimination and step wise*** *approach for selection of most significant independent variables.*
* *Performed* ***Feature selection and Feature extraction*** *dimensionality reduction methods to figure out significant variables.*
* *Used* ***RMSE score, Confusion matrix, ROC, Cross validation and A/B testing*** *to evaluate model performance in both simulated environment and real world.*
* *Developed interactive* ***dashboards****, created various Ad Hoc reports for users in Tableau by connecting various data sources.*

**Environment:** Python, SQL server, Hadoop, HDFS, HBase, MapReduce, Hive, Impala, Pig, Sqoop, Mahout, Spark, Tableau, ETL, Linux.

***Client:* Safeway Pleasanton, CA *Jan 2014- Oct 2015***

***Role:* Data Scientist**

***Description:*** *Safeway is independent organization owned by Albertsons companies. Albertson’s Marketing team need visibility into Sales, Promotional and marketing data to me merged with rest of the Organization. EDW data from United markets was extracted, mapped and integrated with Albertson’s data.*

***Responsibilities:***

* *Responsible for performing Machine-learning techniques regression/classification to predict the outcomes.*
* *Responsible for design and development of* ***advanced R/Python*** *programs to prepare transform and harmonize data sets in preparation for modeling.*
* *Created data visualization with* ***ggplot2*** *in R to understand annual sales pattern.*
* *Applied clustering algorithms i.e. Hierarchical, K-means with help of Scikit and Scipy anddeveloped* ***visualizations and dashboards using ggplot2, Tableau.***
* *Executed SQL queries from* ***R/Python*** *on complex table configurations.*
* *Prepared data frames by using* ***Gsub () function in R*** *for identifying missing data that used for production data analysis.*

***Environment:*** *MS Excel, R, Python, SAS, SQL, MS Word, MS Excel and Tableau*

***Client:* *Medtronic Inc. Santa Ana CA Oct 2012- Dec 2013***

***Role: Data Analyst***

***Description:*** *Medtronic Public Limited Company is a medical device company. Its headquarters are in Dublin, Ireland. Its operational headquarters are in Fridley, Minnesota. Medtronic is among the world's largest medical equipment development companies.*

***Responsibilities:***

* *Experience in working on Spark* ***SQL queries, Data frames,*** *import data from Data sources, perform transformations, perform read/write operations, save the results to output directory into HDFS.*
* *Worked on installation of Kafka on* ***Hadoop cluster*** *and to use it for streaming & cleansing of raw data and have extracted useful information using Hive and stored the results in Hbase.*
* *R and Python languages used to identify chemical performance via Classification, tree map and regression models along with visualizing data for interactive understanding and decision-making.*
* *Identified outliers, anomalies and trends in any given data sets by using R.*
* *Provided daily change management process support, ensuring that all changes to program baselines are properly documented and approved, maintained, managed and issue change schedules.*
* *Developed, installed, maintained and monitored company databases in high performance/high availability environment with supported configuration, performance tuning to ensure optimal resource usage.*
* *Performed data imputation using* ***Scikit-learn package in Python.***

***Environment:*** *Machine learning, AWS, HDFS, Hive, Pig, Linux, Python, R, SAS, SPSS, MySQL, Eclipse, PL/SQL, SQL connector, Tableau 14.*

***Client:* ezDI LLC, Ahmedabad, India *Jan 2011- Sep2012***

***Role:*****Data Analyst**

***Description:*** *ezDI helps hospitals reduce and maintain their DNFB at or below 4 days and eliminates their dependence on costly and outdated coding software by automating and optimizing the entire coding workflow – allowing CDI staff, coders and auditors to work together in perfect unison on a single platform.*

***Responsibilities:***

* *Worked with several* ***R packages including dplyr, Spark, Causal Infer, spacetime****.*
* *Implemented end-to-end systems for Data Analytics, Data Automation and integrated with custom visualization tools using* ***R and Hadoop****.*
* *Performed Exploratory Data Analysis and Data Visualizations* ***using R, and Tableau****.*
* *Worked with Data governance, Data quality, data lineage, Data architect to design various models and processes.*
* *Independently coded new programs and designed Tables to load and test the program effectively for the given POC's using with Big Data/Hadoop.*
* *Designed data models and data flow diagrams using Erwin and MS Visio.*
* *Strong programming skills using R, Elastic Search & Machine Learning Algorithms.*
* *Designed and implemented machine learning algorithms to enhance existing data mining capabilities.*

***Environment:*** *Unix, Python 3.5, MLLib, SAS, regression, logistic regression, Hadoop, Teradata, OLTP, OLAP, HDFS, JSON, XML.*

***Client: Echidna Software Pvt. ltd - Bengaluru, INDIA Jun 2009 - Dec 2010***

***Role: Python Developer***

***Description:*** *Echidna began when a small group of commerce leaders knew there had to be a better way to do commerce. So, they branched off and created a new kind of agency -- one that combines amazing UX, enterprise-level technology implementation, and value-added marketing and analytics services.*

***Responsibilities:***

* *Involved in the design, development and testing phases of application using* ***AGILE methodology****.*
* *Designed and maintained databases using Python and developed* ***Python*** *based* ***API*** *(RESTful Web Service) using* ***Flask, SQL Alchemy and PostgreSQL****.*
* *Designed and developed the UI of the website using* ***HTML, XHTML, AJAX, CSS*** *and* ***JavaScript****.*
* *Participated in requirement gathering and worked closely with the architect in designing and modeling.*
* *Created and managed all hosted or local repositories through Source Tree's simple interface of* ***GIT*** *client, collaborated with* ***GIT*** *command lines and Stash.*
* *Developed consumer-based features and applications using* ***Django****,* ***HTML****, behaviour Driven Development (BDD) and pair-based programming.*
* *Designed and developed the UI for the website with* ***HTML, XHTML, CSS, Java Script and AJAX***
* *Used* ***AJAX****&****JSON*** *communication for accessing* ***RESTfulweb services*** *data payload.*
* *Designed dynamic client-side JavaScript codes to build web forms and performed simulations for web application page.*
* *Created and implemented* ***SQL Queries, Stored procedures, Functions, Packages*** *and* ***Triggers*** *in SQL Server.*

***Environment:*** *Python 2.5, Java/J2EE, Django1.0, HTML, CSS Linux, Shell Scripting, Java Script, Ajax, JQuery, JSON, XML, PostgreSQL, Jenkins, ANT, Maven, Subversion, Python*