Data scientist and machine learning professional, experienced in advanced statistical analysis, manipulating and mining meaningful data from large datasets. Working knowledge of Mathematical, Data Science and Machine Learning concepts. Able to formulate solution strategies to data science problems, apply exploratory analysis to identify abnormalities in data and utilize an appropriate set of algorithms.

**PROFESSIONAL SUMMARY:**

* 6+ years of experience in Data Analysis, Data Analyst, Python and Project Management.
* Over 4 years of experience in all phases of diverse technology projects specializing in Data Science and Machine Learning.
* Involved in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling and data visualization with large data sets of structured and unstructured data, created ER diagrams and schemas.
* Proficient in process research which requires analytic models, data inputs and output, analytic metrics and user interface needs.
* Developed predictive models using Decision Tree, Random Forest, Naive Bayes, Logistic Regression, Cluster Analysis, and Artificial Neural Networks.
* Experienced with machine learning algorithms such as Logistic Regression, Random Forest, KNN, Support Vector Machines, Neural Networks, Linear/Non-Linear Regression and K-means
* Expertise in employing techniques for Supervised and Unsupervised learning(Clustering, Classification, PCA, Decision trees, KNN, SVM), Predictive Analytics
* Excellent understanding of Agile and Scrum development methodology
* Experienced in Python to manipulate data for data loading and extraction and worked with python libraries like Matplotlib, Numpy and Pandas for data analysis.
* Hands on experience of Data Science libraries in Python such as Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Keras.
* Equipped with experience in utilizing statistical techniques which include Correlation, Hypothesis modeling, Inferential Statistics as well as data mining and modeling techniques using Linear and Logistic Regression, Clustering, Decision Trees, and K-mean clustering.
* Mitigated risk factors through careful analysis of financial and statistical data. Transformed and processed raw data for further analysis, visualization, and modeling.
* Experience in Big Data technologies like Spark 1.6, Spark SQL, pySpark, Hadoop 2.X, Hive 1.X
* Experience in visualization tools like, Tableau 9.X, 10.X for creating dashboards
* Automated recurring reports using SQL and Python and visualized them on BI platform like Tableau.
* Worked on Artificial Neural Networks and Deep Learning models using Theano and Keras packages using Python.
* Worked with No SQL Database including Cassandra and Mongo DB.
* Experienced in writing SQL queries, working knowledge of RDBMS like SQL Server 2008, No SQL databases like Mongo DB 3.2
* Experience working on Microsoft SQL Server, Oracle, Hadoop/Hive.
* Analytical, performance-focused, and detail-oriented professional, offering in-depth knowledge of data analysis and statistics
* Utilized complex SQL queries for data manipulation.
* Excellent Team player and self-starter possess good communication skills.
* Experience in various phases of Software Development life cycle (Analysis, Requirements gathering, Designing) with expertise in writing/documenting Technical Design Document (TDD), Functional Specification Document (FSD), Test Plans, GAP Analysis and Source to Target mapping documents.
* Used the version control tools like Git 2.X and build tools like Apache Maven/Ant
* C, C++, Java, Python, SQL programming skills, with experience in working with functions, packages and triggers.
* Skilled in Advanced Regression Modeling, Correlation, Multivariate Analysis, Model Building and application of Statistical Concepts.

**EDUCATIONAL QUALIFICATION:**

**Bachelors in Computer Science, JNTU**

**TOOLS AND TECHNOLOGIES:**

|  |  |
| --- | --- |
| **Programming Languages** | Python (2 &3), SQL, Java, C/C++ |
| **Machine Learning** | Regression analysis, Decision Tree, Random Forests,  Naïve Bayes, SVM, Neural Network, K-Means Clustering and KNN. |
| **Frameworks and Toolkits** | NumPy and Pandas. |
| **Statistical Methods** | Descriptive statistics, Hypothesis Testing, ANOVA, Chi-square testing,  F-tests, Confidence Intervals, Bayes Law, Dimensionality Reduction,  ROC curve. |
| **Databases** | Oracle 10g/9i/8i, MS SQL Server 2008, MS Access, MySQL 5.x, |
| **Operating Systems** | Windows NT/2000/XP/7,8&10, Windows Server 2008. |

**PROFESSIONAL EXPERIENCE:**

**PPD – Raleigh, North Carolina Feb 2018-Present**

**Role: Data Scientist**

**Responsibilities:**

* Responsible for predictive analysis of CRO trial data to predict patient yield and recruitment timeline for clinical trials across APAC & EMEA regions.
* Data was extracted extensively by using SQL queries and used python packages for the data mining tasks.
* Performed Exploratory Data Analysis, Data Wrangling and development of algorithms in Python for data mining and analysis.
* Extensively used Python's multiple data science packages like Pandas, NumPy, Matplotlib, Scipy, Scikit-learn, Keras and Theano.
* Worked on exploratory data analysis, data cleaning, visualization, Statistical Modeling using Python 3.5 and Tableau
* Performed Data Cleaning, handled missing data, outliers, feature scaling, features engineering.
* Used Python-based data manipulation and visualization tools such as Pandas, Matplotlib and Seaborn to clean corrupted data before generating business requested reports.
* Developed extension models relying on but not limited to the Random forest, logistic, linear regression, Stepwise, Spline model, Support Vector Machine, Naive Bayes classifier, K-Centroid clusters.
* Used Machine Learning to build various algorithms (Random Forest, Decision trees, Naive Bayes) classification models.
* Used Python programming language to graphically analyze the data and perform data mining.
* Performed extensive data mining to find out relevant features in an anonymized dataset using Python.
* Explored supervised Machine Learning algorithms (Regression, Random Forest, SVM, Decision tree, Artificial Neural Networks) and used parameters such as Precision/Adjusted R-Squared /residual splits to select the winning model.
* Developed a Tableau based dashboard from oracle database to present to the business team for data visualization purpose.

**Environment:** Python 3, SQL, Pandas, Seaborn, Decision Trees, Random Forest, Support Vector Machines, Artificial Neural Networks, Tableau.

**Vertiv – Columbus, OH. Apr 2017 – Jan 2018**

**Role: Data Scientist**

**Responsibilities:**

* Implemented end-to-end systems for DataAnalytics, Data Automation and customized visualization tools using Python, Hadoop and MongoDB.
* Used Pandas, NumPy, Seaborn, Matplotlib, scikit-learn in Python for developing various machine learning algorithms.
* Worked on csv, Json, excel different types of files for the  Data cleaning and Data analysis
* Used Python for statistical operations on the  data and Matplotlib for the visualizing the data
* Ensured that the model has a low False Positive Rate.
* Managed large datasets using Pandas data frames and MySQL.
* Built various graphs for business decision-making using Python Matplotlib library.
* Identified root causes of problems, and facilitated the implementation of cost-effective solutions with all levels of management.
* Performed Data Cleaning, handled missing data, outliers, feature scaling, features engineering.
* Application of various ML algorithms and statistical modeling like decision trees, regression models, random forest, SVM, clustering to identify Volume using different packages in python.
* Performed data visualization with Tableau and generated dashboards to present the findings
* Created and designed reports that will use gathered metrics to infer and draw logical conclusions from past and future behavior.
* Worked independently and collaboratively throughout the project lifecycle including data extraction/preparation, design and implementation of scalable machine learning analysis and solutions, and documentation of results.
* Performed Classification using supervised algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes.
* Performed data profiling to merge the data from multiple data sources.
* Knowledge of other relational database platforms such as Oracle, NoSQL

**Environment:** Python 3, MySQL, Matplotlib, Seaborn, Linear Regression, Logistic Regression, Random Forest, Support Vector Machines, KNN, Tableau.

**Zen3 Info Solutions, Hyderabad, India Aug 2014 - Dec 2016**

**Role: Data Analyst**

**Responsibilities:**

* Communicated with business users, data architects and developers to identify needs, define project scope and detailed functional and non-functional requirements.
* Gathered and documented requirements throughout the Software Development Life Cycle (SDLC)
* Prepared Data Mapping documents, Use Cases, Requirements Traceability Matrix and worked with the User experience team (Wireframe preparing team)
* Created requirement documentation, reviews with user representatives, recommended priorities, gave presentations and walkthroughs and obtained user sign-off.
* Driven daily stand up meetings offshore, iteration planning meetings and requirement gathering meetings.
* Extracting data from different public data repositories and from different databases and creating datasets for statistical analysis.
* Identified Multiple Dimensions and Fact tables. Used advance data modeling concepts of degenerated dimension, sub-dimension, Factless fact table, Aggregate fact tables in the Multidimensional model.
* Extensively used Tab Admin and Tab Cmd commands in creating backups and restoring backups of Tableau repository.
* Administered user, user groups and scheduled instances for reports in Tableau.
* Hands-on development in creating and modifying worksheets and data visualization dashboards.
* Designed & developed various analytical reports from multiple data sources by blending data on a single worksheet in Tableau Desktop.

**Environment:** Windows 7 & 8, MS SQL Data Base, MS Excel, MS Word, MS PowerPoint, MS Outlook, SharePoint, MS Visio, MS Project, Oracle 11g, Tableau Server and Tableau Desktop v8/v9, MS Excel, SQL, ERWIN, SSIS, SSRS.

**HTC Global Services India Private Ltd, India Sep 2013 - July 2014**

**Role: Data Analyst**

**Responsibilities:**

* Involved in all the phases of the project, communicated with the client and prepared Technical Specification Document.
* Planned monthly and quarterly business monitoring reports by creating Excel Pivot summary reports to include System Calendars.
* Broadly designed easy to follow visualizations using Tableau software and published dashboards on web and desktop platforms.
* Maintain a holistic view of all business processes and users in the system to understand cross-functional impacts regarding configuration, process and workflow.
* Executed quantitative analysis that translate data into actionable insights, therefore provides analytical ,data-driven decision-making support for key projects.
* Compiled and analyzed sales and marketing reports using SQL and Statistical models.
* Established procedures for manipulating the data reading from various data sources and consolidate them as single CSV File and update the content in the database tables.
* Lavishly created numerous processes and flow charts using MS Visio to meet the business needs and interacted with business users to understand their data needs.
* Involved in analysis, design and documenting business requirement specifications to build data warehousing extraction programs, end-user reports and queries.
* Worked closely with Associates to find the problems and getting solutions on the tool.

**Environment:** Tableau, MS Access, MS Excel, MS Visio, UML diagrams, SQL Server.

**Augusta Hi-tech Soft Sol, Hyderabad, India. May 2013 - Aug 2013**

**Role: Java Developer**

**Responsibilities:**

* Developed web application using Struts, JSP, Servlets, Java beans that uses MVC design pattern
* Created user-friendly GUI interface and Web pages using HTML, CSS and JSP.
* Used Eclipse as IDE tool for creating Servlets, JSP, and XML.
* Implemented the Struts framework based on MVC design pattern.
* Wrote SQL for JDBC prepared statements to retrieve the data from database.
* Worked with Database Administrators to solve the problems generated while creating tables for application
* Developed and implemented customized maintenance plans to meet the needs and keep up to date with the glitches and fixes for performances issues.
* Developed various UML diagrams like use cases, class diagrams, interaction diagrams (sequence and collaboration) and activity diagrams
* Involved in build and deploying the application using ANT

**Environment:** Java, SQL, XML, HTML, CSS, JSP, JDBC, Eclipse IDE.