**V Sharath**

**678-214-3443**

Data scientist, Machine leaning engineer and AI developer with experience on manipulating and mining meaningful data from large sets of structured, semi-structured and unstructured data. Great knowledge of mathematical, data science and machine learning concepts. Able to formulate a solution strategy to data science problems, apply exploratory analysis to identify abnormalities in data and utilize appropriate set of algorithms. (regression, SVM, decision tree, clustering and deep learning)

**PROFESSIONAL SUMMARY:**

* Over **8 years** of experience in Data Analysis/Business analysis, ETL Development, and Project Management.
* Over **5 years** of experience in all phases of diverse technology projects specializing in Data Science and Machine Learning.
* Proven expertise in employing techniques for Supervised and Unsupervised (Clustering, Classification, PCA, Decision trees, KNN, SVM) learning, Predictive Analytics, Optimization methods and Natural Language Processing(NLP), Time Series Analysis.
* Experienced in Machine Learning Regression Algorithms like Simple, Multiple, Polynomial, SVR(Support Vector Regression), Decision Tree Regression, Random Forest Regression.
* Experienced in advanced statistical analysis and predictive modeling in structured and unstructured data environment.
* Strong expertise in Business and Data Analysis, Data Profiling, Data Migration, Data Conversion, Data Quality, Data Governance, Data Lineage, Data Integration, Master Data Management(MDM), Metadata Management Services, Reference Data Management (RDM).
* Hands on experience of Data Science libraries in Python such as Pandas, NumPy, SciPy, scikit-learn, Matplotlib, Seaborn, BeautifulSoup, Orange, Rpy2, LibSVM, neurolab, NLTK.
* Solid understanding of AWS(Amazon Web Services) S3, EC2, RDS and IAM, Azure ML, Apache Spark, Scala process and concepts.
* Good Understanding of working on Artificial Neural Networks and Deep Learning models using Theano and Tensorflow packages using in Python.
* Experienced in Machine Learning Classification Algorithms like Logistic Regression, K-NN, SVM, Kernel SVM, Naive Bayes, Decision Tree & Random Forest classification.
* Hands on experience on R packages and libraries like ggplot2, Shiny, h2o, dplyr, reshape2, plotly, RMarkdown, ElmStatLearn, caTools etc.
* Efficiently accessed data via multiple vectors (e.g. NFS, FTP, SSH, SQL, Sqoop, Flume, Spark).
* 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.
* Excellent understanding of Hadoop architecture and Map Reduce concepts and HDFS Framework.
* Strong understanding of project life cycle and SDLC methodologies including RUP, RAD, Waterfall and Agile.
* Very good knowledge and understanding of Microsoft SQL server, Oracle, Teradata, Hadoop/Hive.
* Strong expertise in ETL, Data warehousing, Operational Data Store (ODS), Data Marts, OLAP and OLTP technologies.
* Experience working on BI visualization tools (Tableau, Shiny & QlikView).
* Analytical, performance-focused, and detail-oriented professional, offering in-depth knowledge of data analysis and statistics; utilized complex SQL queries for data manipulation.
* Equipped with experience in utilizing statistical techniques which include Correlation, Hypotheses modeling, Inferential Statistics as well as data mining and modeling techniques using Linear and Logistic regression, clustering, decision trees, and k-mean clustering.
* Expertise in building Supervised and Unsupervised Machine Learning experiments using Microsoft Azure utilizing multiple algorithms to perform detailed predictive analytics and building Web Services models for all types of data: continuous, nominal, and ordinal.
* Expertise in using Linear & Logistic Regression and Classification Modeling, Decision-trees, Principal Component Analysis (PCA), Cluster and Segmentation analyses, and have authored and coauthored several scholarly articles applying these techniques.
* Mitigated risk factors through careful analysis of financial and statistical data. Transformed and processed raw data for further analysis, visualization, and modeling.
* Proficient in research of current process and emerging technologies which need analytic models, data inputs and output, analytic metrics and user interface needs.
* Assist in determining the full domain of the MVP, create and implement its relevant data model for the App and work with App developers integrating the MVP into the App and any backend domains.
* Insure REST-based API including all CRUD operations integrate with the App and other service domains.
* Installing and configuring additional services on appropriate AWS EC2, RDS, S3 and/or other AWS service instances.
* Integrating these services with each other and insuring that user access to data, data storage and communication between various services.
* Excellent Team player and self-starter, possess good communication skills.

**EDUCATIONAL QUALIFICATION:**

**Bachelors in Computer Science, JNTU 2009**

**Masters in Computer and Science, UARK 2013**

**PROFESSIONAL EXPERIENCE:**

**Coca-Cola Company – Tampa, FL. July2017– Till Date**

**Role: Data Scientist/Machine Learning Engineer**

**Responsibilities:**

* Studied the data to create/customize models to analyze/visualize important KPIs for the project
* Developed and execute processes for accurate data capture across all clients to obtain key insights and relationships to overall business objectives using Statistical Hypotheses Modeling
* Exploratory data analysis using R to deep dive into internal and external data to diagnose areas of improvement to increase efficiency
* Utilized Boosted Decision Tree, Linear and Bayesian Linear Regression Machine Learning models in Microsoft Azure to develop and implement interactive Webservice predictive models
* Led the company's machine learning and statistical modeling effort including building predictive models and generate data products to support customer segmentation, product recommendation and allocation planning; prototyping and experimenting ML/DL algorithms and integrating into production system for different business needs.
* Used Hadoop for performing various Analysis Services. Connected Analysis Services to Hadoop/Hive via the relational SQL Server engine, enabling Klout to reduce data latencies, eliminate maintenance overhead and costs, move aggregation processing to Hadoop, and shorten development cycles dramatically.
* Acquired and cleaned using Talend and structure data from multiple source including external and internal databases.
* Performed data extraction, manipulation, cleaning, analysis, modeling and data mining using R programming in R Studio.
* Use Classification, Clustering Algorithms models using Python for Text Analytics for Text Data, grouping Data.
* Use Python Pandas, seaborn, matplotlib, Tensorflow, Keras, Text Data Clustering, RNN.
* Designed 10+ dashboards in Tableau for sales managers with instant access to personalized analytics portal, so they can access key business metrics such as time to close opportunity, delay-to-contract, resulting in increased customer satisfaction and improving client’s standing in the Sales Performance management industry
* Studied the data to create/customize models to analyze/visualize important KPIs for the project
* Developed and execute processes for accurate data capture across all clients to obtain key insights and relationships to overall business objectives using Statistical Hypotheses Modeling
* Exploratory data analysis using R to deep dive into internal and external data to diagnose areas of improvement to increase efficiency
* Utilized Boosted Decision Tree, Linear and Bayesian Linear Regression Machine Learning models in Microsoft Azure to develop and implement interactive Webservice predictive models
* Led the company's machine learning and statistical modeling effort including building predictive models and generate data products to support customer segmentation, product recommendation and allocation planning; prototyping and experimenting ML/DL algorithms and integrating into production system for different business needs.
* Used common data science toolkits, such as R, Python, NumPy, Keras, Theano, Tensor flow, etc.
* Designed, built and deployed a set of R modeling APIs for customer analytics, which integrate multiple machine learning techniques for various user behavior prediction (CLTV, marketing funnel propensity models etc.) and support multiple marketing segmentation programs.
* Built models using decision trees, segmentation, regression and clustering intelligent decision models to analyze customer response behaviors, interaction patterns and propensity

**Disney – Seattle, WA. Nov2016-June 2017**

**Role: Data Scientist/Machine Learning Engineer**

**Responsibilities:**

* Work independently and collaboratively throughout the complete analytics project lifecycle including data extraction/preparation, design and implementation of scalable machine learning analysis and solutions, and documentation of results.
* Performed statistical analysis to determine peak and off-peak time periods for ratemaking purposes
* Conducted analysis of customer data for the purposes of designing rates.
* Identified root causes of problems, and facilitated the implementation of cost effective solutions with all levels of management.
* Application of various machine learning algorithms and statistical modeling like decision trees, regression models, clustering, SVM to identify Volume using scikit-learn package in R.
* Worked on different data formats such as JSON, XML and performed machine learning algorithms in Python.
* Involved in transforming data from legacy tables to HDFS, and HBASE tables using Sqoop.
* Research on Reinforcement Learning and control (Tensorflow, Torch), and machine learning model (Scikit-learn).
* Hands on experience in implementing Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, Principle Component Analysis.
* Performed K-means clustering, Regression and Decision Trees in R.
* Work independently or collaboratively throughout the complete analytics project lifecycle including data extraction/preparation, design and implementation of scalable machine learning analysis and solutions, and documentation of results.
* Partner with technical and non-technical resources across the business to leverage their support and integrate our efforts.
* Partner with infrastructure and platform teams to configure, tune tools, automate tasks and guide the evolution of internal big data ecosystem; serve as a bridge between data scientists and infrastructure/platform teams.
* Worked on Text Analytics and Naive Bayes creating word clouds and retrieving data from social networking platforms.
* Pro-actively analyze data to uncover insights that increase business value and impact.
* Support various business partners on a wide range of analytics projects from ad-hoc requests to large-scale cross-functional engagements
* Prepared Data Visualization reports for the management using R
* Approach analytical problems with an appropriate blend of statistical/mathematical rigor with practical business intuition.
* Hold a point-of-view on the strengths and limitations of statistical models and analyses in various business contexts and can evaluate and effectively communicate the uncertainty in the results.
* Application of various machine learning algorithms and statistical modeling like decision trees, regression models, SVM, clustering to identify Volume using scikit-learn package in python, MATLAB.
* Worked on different data formats such as JSON, XML and performed machine learning algorithms in Python.
* Approach analysis in multiple ways to evaluate approaches and compare results.

**Global Atlantic Financial Group – Brighton, MA Feb2014- Aug2016**

**Role: Data Scientist**

**Responsibilities:**

* Responsible for predictive analysis of credit scoring to predict whether or not credit extended to a new or an existing applicant will likely result in profit or losses.
* Data was extracted extensively by using SQL queries and used R packages for the data mining tasks.
* Performed Exploratory Data Analysis, Data Wrangling and development of algorithms in R and Python for data mining and analysis.
* Extensively used Python's multiple data science packages like Pandas, NumPy, matplotlib, SciPy, Scikit-learn, Tensorflow and NLTK.
* Worked on exploratory data analysis, data cleaning, visualization, Statistical Modeling using Python 3.5, R Studio, R Shiny 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, Seaborn to clean corrupted data before generating business requested reports.
* Developed extension models relying on but not limited to Random forest, logistic, Linear regression, Stepwise, Spline model, Support Vector machine, Naive Bayes classifier, ARIMA/ETS model, K-Centroid clusters.
* Used Machine Learning to build various algorithms (Random Forest, Decision trees, Naive Bayes) classification models.
* Extensively used R packages like (GGPLOT2, GGVIS, CARET, DPLYR) on huge data sets.
* Used R programming language to graphically analyses the data and perform data mining.
* Did extensive data mining to find out relevant features in an anonymized dataset using R and Python. Used an ensemble of Xgboost (Tuned using Random Search) model to make predictions.
* Explored 5 supervised Machine Learning algorithms (Regression, Random Forest, SVM, Decision tree, Neural Network) and used parameters such as Precision/Adjusted R-Squared /residual splits to select the winning model of the 5 different models.
* Developing Data Mapping, Data Governance, Transformation and Cleansing rules for the Master Data Management (MDM) Architecture involving OLTP, ODS and OLAP.
* Developed Tableau based dashboard from oracle database to present to business team for data visualization purpose.

**Pearson – Iowa City, IA Mar2013- Dec2013**

**Role: Data Analyst**

**Responsibilities:**

* Thoroughly analyzed Financial accounts and statements – Income Statement, Cash Flows and Balance Sheets of several companies for recommending the retail investors
* Improvised various statistical models to help in better decision-making on sector growth, company forecasts, etc.
* Performed secondary research on companies and interpreted their financial reports and other supporting documents
* Regularly generated Data Models using R transform Financial Data into useful information for the Stocks team
* Performed data analysis by gathering, analyzing and deploying spatial data from its pristine form to derive financial projections
* Used Google Fusion Tables and Tableau to publish visualizations.
* Rated stocks based on Fundamental Analysis, quarterly performance and future outlook
* Prepared Equity Research and Quarterly Earnings reports for companies under assigned sectors
* Developed Industry Spotlight reports for retail investors on a regular basis
* Developed category segmentation using R which provides customizable view of market share and led to decreased labor cost by 50%.
* Actively pursued data quality compliance, assessed risk factors and generated models and scenarios for forecasting operational risk.
* Analyzed Learning Management System (LMS) for client and implemented improvements for process variation thereby reducing lead time by 30%.
* Developed a dashboard using Tableau to provide the management with an overall understanding of resource optimization, attaining incremental revenue worth $1.2 million.

**C-DAC – India. May2009- June2012**

**Role: Business Data Analyst**

**Responsibilities:**

* Developed R programs for manipulating the data reading from various Oracle data sources and consolidate them as single CSV File, and update the content in the database tables.
* Created monthly and quarterly business monitoring reports by writing complex SQL queries to include System Calendars, Inner Joins and Outer Joins to retrieve data from multiple tables.
* Designed easy to follow visualizations using Tableau software and published dashboards on web and desktop platforms.
* Critically examined RFPs (Request for Proposals) and EoIs (Expression of Interest) and assisted in preparing various research proposals.
* Created numerous processes and flow charts using Visio and iGrafx to meet the business needs and interacted with business users to understand their data needs.
* Monitored dashboards & application performance.
* Made recommendations to management for the coordination of the daily workflow of the mapping stage and establish standard performance benchmarks for the timely processing of core mapping stages.
* Involved in analysis, design and documenting business requirement specifications so as 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.