# Professional Summary:

* Over 8 years of experience as a Professional Qualified **Data Scientist/Data Analyst** in Data Science and Analytics including Machine Learning, Data Mining, and Statistical Analysis.
* Achieved solutions through problem solving and analyzing data.
* Skilled in python, R, SAS, SQL for algorithm development, data modeling, statistical learning and data Visualization.
* Strong exposure to writing simple and complex SQL, **TSQL** queries.
* Strong experience on data analysis, design, planning and implementing of large scaled big data projects in Hadoop Ecosystem. Knowledge of Data Warehousing concepts and Extract Transform and Load (ETL) processes.
* Strong skills in data warehouse, data exploration, data extraction, data validation, reporting and excel.
* Extensively used SQL for accessing and manipulating database systems.
* Hands on experience applying ML/Statistical algorithms to real world problems: **Neural Networks, Random Forest, Clustering, Generalized Linear Models** and excellent skills at **Python, SQL, NumPy, Pandas, R**.
* Proficient at Machine Learning algorithms and Predictive Modeling including **Naive Bayes, Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis, Neural Networks, Ensemble Models, SVM, KNN** and **K-means clustering** and good knowledge on Recommender Systems.
* Experienced with **Docker** container service, created **Dockerized** applications by creating **Docker** images from **Dockerfile.**
* Having good experience in Amazon Webservices (AWS).
* Using tools like **Tableau** and Microsoft Excel for **data analysis** and generating data reports.

# Work History

## **Data Scientist**

Raymond James-St. Petersburg**, FL Sep/2017 to Current**

**Responsibilities:**

* Executed Tasks and objectives remotely with robust efficiency.
* Analyzed raw Data, documents, and scattered information to come to a clear and coherent decision based on AdHoc search algorithms and syntax.
* Worked with parameter tuning and model evaluation techniques Confusion Matrix, Cross validation, AUC-ROC etc. Customer Profiling models using **K-means** and **K-means++** clustering algorithms to enable targeted marketing.
* Worked with **Customer Churn Models** including **Random forest regression, lasso regression** along with pre-processing of the data.
* Worked with various **customer analytics** such as **Customer targeting, campaign sales analysis, KPI analysis, forecasting sales, NLP models.**
* Implemented dimensionality reduction using **Principal Component Analysis and k-fold** cross validation as part of Model Improvement.
* Developed and implemented predictive models using machine learning algorithms such as **linear regression, classification, multivariate regression, Naive Bayes, Random Forests, K-means clustering, KNN.**
* Used **Convolutional Neural Network (CNN)** to perform image classification and object detection.
* Used **Pandas, NumPy, Scikit-learn** in Python for developing various machine learning models.
* Customer segmentation based on their behavior or specific characteristics like age, Weather, region, income, geographical location and applying Clustering algorithms to group the customers based on their similar behavior patterns.
* Consulted with and trained analysts to effectively access data and optimize Tableau dashboards.
* Maintained SQL scripts to create and populate tables in data warehouse for daily reporting across departments.
* Developed ETLs for data sources used in production reporting for marketing and operations teams.
* Created customized reports and processes in SAS and Tableau Desktop
* Assisted the HR director in talent assessment of new Data Scientists.
* Experience in **AWS** (Amazon Webservices) EC2, VPC, IAM, IAM, S3, CloudFront, CloudWatch, CloudFormation, RDS, Route53, SNS, SQS, Elastic Cache.
* Extensively worked on Amazon Sage Maker platform to train, test and deployed deep neural networks/classic models.
* Worked with **Docker** components like **Docker** engine, Hub, Compose and registry for storing **Docker** **images** and files running multiple containers in staging and production environments.
* Worked on **Jenkins, Docker** for continuous integration and end-to-end automation for all build and deployments.
* Used Python 3.X (NumPy, SciPy, pandas, scikit-learn, seaborn) and Spark 2.0 (PySpark, MLlib) to develop variety of models and algorithms for analytic purposes.
* Designed rich data visualizations to model data into human-readable form with **Matplotlib**.

**Environment:**

Tableau, SAS, Matplotlib, Python (NumPy, pandas, SciPy, scikit-learn, seaborn) and Spark 2.0 (PySpark, MLlib), Sage Maker, SQL, Jenkins, Docker, TSQL.

## **Data Scientist**

Rauxa – **San Francisco, CA Dec/2015 to Aug/2017**

**Responsibilities:**

* Developed pipelines to analyze large simulation datasets combining my own Python, Tcl and Shell scripts with established molecular modeling tools.
* Interpreted complex simulation data using statistical methods.
* Used **Pandas, NumPy, Scikit-learn** in Python for developing various machine learning models such Random forest and step-wise regression.
* Worked with ANN (Artificial Neural Networks) and BBN (Bayesian Belief Networks).
* Hands on experience in Dimensionality Reduction, Model selection and Model boosting methods using **Principal Component Analysis** (PCA), **K-Fold Cross Validation** and **Gradient Tree Boosting.**
* Created and maintained reports to display the status and performance of deployed model and algorithm with **Tableau**.
* Implemented cluster services using **Docker** and **Kubernetes** to manage local deployments in **Kubernetes** by building a self-hosted **Kubernetes** cluster using **Terraform** and **Ansible** and deployed application containers.
* Well in-depth experienced in Data Science ML / AI Models (**AWS Sage Maker**, Scala, **Azure** ML Models.
* Developed and builds alignment within Product Operations for frameworks/prototypes that integrate data and machine learning/predictive modeling to make business decisions.
* Identifies and develops solutions that use new areas of data, research and models that can solve business problems.
* Developed project plans for complex and occasionally highly complex development projects
* Managed data and data requests to improve the accuracy of our data and decisions made from data analysis.
* Used and learns a wide variety of tools and languages to achieve results (e.g., R, SAS, Python, SQL)
* Lead work on data and problems across departments to drive improved business efficiencies through designing, building, and partnering to automate models.

**Environment:**

Python (NumPy, pandas, SciPy), SQL, TSQL, Tableau, SAS, Docker,Scikit-learn, R, Sage Maker, Scala, Azure.

## **Machine Learning Engineer**

Well care– **Tampa, FL Sep/2014 to Nov/2015**

**Responsibilities:**

* Validate the Data by using the PySpark programs.
* Processed large amounts of raw data into a meaningful context for healthcare clients.
* Developed a machine learning model to match the claims with the supporting documents to decrease the manual intervention.
* Worked with several proofs of concept models using Deep Learning, Neural networks.
* Performed feature engineering including feature intersection generating, feature normalize and label encoding with Scikit-learn preprocessing.
* Developed machine learning & python scripts for day to day business activities.
* Worked with various regression algorithms like **Random Forest Regression, Decision Tree regression, Polynomial Regression, Binomial Regression and Support Vector Models** to forecast machinery failures in auto-motive industries.
* Worked with various **customer analytics** such as **Customer targeting, campaign sales analysis, KPI analysis, forecasting sales, NLP models.**
* Worked on **Personalized marketing** models to implement simplicity and targeted marketing for specific customers.
* Worked with **Clustering algorithms** to target specific group of customers to generate profitable revenue.
* Used **market basket analysis**, **association rules** analysis to identified patterns, data quality issues and leveraged insights.
* Created various types of data visualizations using **Matplotlib**, **Seaborn** and **Tableau**.
* Used **Tableau** to convey the results by using dashboards to communicate with team members and with other data science teams, marketing and engineering teams.
* Created visualization maps for a diabetic insurance claims using a **heatmap** to find correlations with help of **seaborn** visualization.

**Environment:**

Python (NumPy, pandas, SciPy, scikit-learn, seaborn) and Spark 2.0 (PySpark, MLlib), SQL, TSQL, Tableau

## **Data Analyst**

Huawei Telecommunication India Pvt Ltd – **Navi Mumbai, India Mar/2012 to Aug/2014**

**Responsibilities:**

* Designed and maintained databases using **Python** and developed Python based **API**
* Analyzed the source data coming from different sources (**SQL Server, Oracle and also from flat files like Access and Excel**) and working with business users and developers to develop the Model.
* Have Used **Informatica** Data Quality as **ETL** tool to transform the data from various sources and bring them into one common format and load them into target database for the analysis purpose from Data Warehouse.
* Executed SQL queries to validate actual test results and match expected results as per financial rules.
* Responsible for maintaining the integrity of the SQL database and reporting any issues to the database architect.x v
* Design and model the reporting data warehouse considering current and future reporting requirement
* Involved in the daily maintenance of the database that involved monitoring the daily run of the scripts as well as troubleshooting in the event of any errors in the entire process.
* Creating database schema for **MySQL** Database and helped to draw ER-Diagrams using Microsoft Visio.
* Managed code versioning with **GitHub** and deployment to staging and production servers.

**Environment:**

MySQL, SQL, TSQL, Oracle, ER-Diagrams, Microsoft Visio, Informatica, ETL Tools, Excel, GitHub

## **Business Analyst**

Huawei Telecommunication India Pvt Ltd – **Navi Mumbai, India Feb/2010 to Mar/2012**

**Responsibilities:**

* Gathered business requirements and documentation using use case specifications, Visio diagrams and class diagrams. Worked with an IT team and analyzed the business and functional requirements. Documented and delivered Functional Specification Document **(FSD)** to the project team.
* Practiced various **SDLC** methodologies. Used effective **Agile Methodology** techniques to conduct weekly team meetings to ensure the development activities are streamlined. Worked with Software Developers to help translate Project requirements into functional specifications (SDLC) and User Stories **(Agile/SCRUM).**
* Assisted and coordinated integration and User Acceptance Testing **(UAT)** prior to the release of new data feeds into production.
* Provided analysis, understanding and business perspectives on many financial and operational issues. Gathered Functional and Data Requirements, analyzed workflows and created **Use Cases**, Requirement Specifications, Report Specifications, Data Requirements, Data Mappings and **Data Flow Diagrams**.
* Develop business process models using **Agile Methodology** by document user stories existing and future business processes. Analyze **"AS IS"** and **"TO BE"** scenarios, designed and documented new process flows, business process and various business scenarios.
* Gather business requirements and documentation using use case specifications, Visio diagrams and class diagrams. Develop Business Requirement Document **(BRD)** that supported overall strategy, goals and objectives.
* Create System Requirement Document **(SRD)**, document, User Requirement Specification **(URS)** and Change Request **(CR)** document for system application development. Extract email addresses and sent emails using SharePoint and document details from within a workflow.
* Good knowledge on **Requirement Gathering Techniques.** Participate in b**rainstorming sessions** to prioritize and distribute tasks to the UAT team for effective execution.
* Good understanding of Java, J2EE Design Patterns and Knowledge in Microservices architecture.

**Environment:**

Microsoft Office, MS-Project, SharePoint 2013, MS-Visio 2010, HTML, Java, version one.

# Education

|  |  |
| --- | --- |
|  | P.H.D: Information Technology **2017-current**  Master of Science: Computer Science |