**Krishna**

PH: 502.792.8429 EMAIL: [krishnamlearning@gmail.com](mailto:krishnamlearning@gmail.com)

Professional Summary

Process oriented Data Scientist with 5+ years of experience in, Machine learning algorithms, Python, SQL. Experienced in interpreting and analyzing data to drive growth.

* Proficient in processing large sets of structured, semi-structured and unstructured data and supporting systems/application architecture.
* Experienced in Life Cycle development of data using Python.
* Extensive command over Pandas, NumPy, Matplotlib, Seaborn, SciPy, Sci-kit libraries in Python.
* Proficient in Machine learning algorithms: Supervised, Unsupervised models.
* Expertise in NLP and Deep learning. Extracted different fields from data source using python.
* Experienced in many versions of Python as well as their differences and features.
* Experienced in writing test plans, test cases, test specifications and test coverage.
* Proficient in Tableau data visualization tool to analyze and obtain insights into large datasets, create visually compelling and actionable interactive reports and dashboards.
* Worked on data blending and increased the dashboard performance by extracting data.
* Experienced in analyzing the BI needs, interpreting business user requirements and translating them into reports.
* Excellent interaction skills with clients, understanding Business Applications, Business Data Flow and Data Relations.
* Presented tools and recommendations to directors and executive staff.
* Working Knowledge on Tensor Flow to perform Machine Learning algorithms.
* Expertise in designing survey research studies including questionnaire development, sample selection, checking survey results, conducted Python programming, statistical analyses and data reporting.
* Excellent skills in analyzing complex data, data matching, identify anomalies, and providing usable insight to internal and external data activities and trend.
* Worked in the Agile methodology using Scrum which has its primary focus on the management part of the software development, dividing the whole development period into small iterations (of thirty days) called "sprints".
* Strong commitment to organizational work ethics, value-based decision-making and managerial skills.

Technical Skills

|  |  |
| --- | --- |
| **Data Modelling Tools** | Erwin, ER/Studio, Star-Schema Modelling, Snowflake-Schema Modelling, FACT and dimension Tables, Pivot Tables |
| **Open Source Libraries** | Scikit-Learn, Pandas, NumPy, Matplotlib |
| **Data Mining** | Data reduction, Clustering, Classification, Anomaly detection, Text mining |
| **Machine Learning** | Regression (Linear, Ridge, Lasso and Elastic Net), Classification (Decision Trees, Logistic Regression, Naïve Bayes, k-nearest neighbors), Ensemble methods (Bagging, Adaboost, Functional Gradient Boosting), Clustering (K-means, Mixture Models) |
| **Deep Learning** | Artificial Neural Networks, Recurrent Neural Networks (Gated Recurrent Units, LSTMs), Convolutional Neural Networks |
| **Programming Languages** | Data structures, Algorithms, Python, R., Tensor Flow, SQL, Spark |

Professional Experience

DNAnexus - Mountain View, CA | *June 2018 - Present*

Senior Machine Learning Engineer

* Developed bulk and incremental data pipelines to ingest patents from Big Query tables and US Patent Office resources.
* Using Natural Language Processing, developed high dimensional vector space models for patents and algorithms for document similarity and clustering.
* Train data to manipulate and visualize for solving real world problems in Python.
* Developed APIs to surface Machine Learning models with Graph and Flask.
* Worked on Natural Language Processing with NLTK module of python for application development for automated customer response. Used predictive modeling with tools in SAS, SPSS, R, Python.
* Applied concepts of probability, distribution and statistical inference on given dataset to unearth interesting findings using comparison, T-test, F-test, R-squared, P-value etc.

Accenture - San Francisco, CA | *January 2017 - June 2018*

Machine Learning Engineer

* Conducted research and created a lean algorithm based on reinforcement learning to detect and track the paths of unique people in a live video to run on low powered hardware for a new product line.
* Designed an algorithm for face, emotion detection from live video based on convolutional neural nets for a camera.
* Create Deep Learning Model to predict outcome of Dota 2 match based on hero line up.
* Developed server and camera backend modules with automated tests and deployed a system to perform analytics, generate reports on live events.
* Applied linear regression, multiple regression, ordinary least square method, mean-variance, the theory of large numbers, logistic regression, dummy variable, residuals, Poisson distribution, Bayes, Naive Bayes, fitting function etc. to data with help of Scikit, SciPy, NumPy and Pandas module of Python.
* Optimized the algorithm to reduce the computation by more than 60% with marginal loss in performance.

Sentient Investment Management - San Francisco, CA | *September 2013 - December 2016*

Data Engineer

* Developed applications of Machine Learning, Statistical Analysis, and Data Visualizations with challenging data Processing problems in sustainability and biomedical domain.
* Compiled data from various sources public and private databases to perform complex analysis and data manipulation for actionable results.
* Designed and developed Natural Language Processing models for sentiment analysis.
* Applied clustering algorithms i.e. Hierarchical, K-means with help of Scikit and SciPy.
* Developed visualizations and dashboards using ggplot, Tableau. Worked on development of data warehouse, DataLake and ETL systems using relational and nonrelational tools like SQL, NoSQL.
* Built and analyzed datasets using R, SAS, MATLAB, and Python (in decreasing order of usage).
* Applied linear regression in Python and SAS to understand the relationship between different attributes of dataset and causal relationship between them
* Performs complex pattern recognition of financial time series data and forecast of returns through the ARMA and ARIMA models and exponential smoothening for multivariate time series data
* Pipelined (ingest/clean/munge/transform) data for feature extraction toward downstream classification.
* Used ClouderaHadoopYARN to perform analytics on data in Hive.
* Wrote Hive queries for data analysis to meet the business requirements.
* Expertise in Business Intelligence and data visualization using R and Tableau.
* Expert in Agile and Scrum Process.
* Validated the Macro-Economic data (e.g. BlackRock, Moody's etc.) and predictive analysis of world markets using key indicators in Python and machine learning concepts like regression, Bootstrap Aggregation and Random Forest.
* Worked in large-scale database environments like Hadoop and MapReduce, with working mechanism of Hadoop clusters, nodes and Hadoop Distributed File System (HDFS).
* Interfaced with large-scale database system through an ETL server for data extraction and preparation.
* Identified patterns, data quality issues, and opportunities and leveraged insights by communicating opportunities with business partners.

**Technologies Used**: Machine learning, AWS, MS Azure, Cassandra, Spark, HDFS, Hive, Pig, Linux, Python (Scikit-Learn/SciPy/ NumPy/ Pandas), R, SAS, SPSS, MySQL, Eclipse, PL/SQL, SQL connector, Tableau.

CGI - Lafayette, LA | *August 2011 - August 2012*

Data Analyst

* Increased SQL query performance by 20%, by optimizing the performance of various SQL scripts, stored procedures and triggers by identifying slow running queries using SQL Profiler
* Implemented and developed automated test practices for web application primarily using Visual Studio's Selenium module for Web application which reduced the application testing time from 8 hours to 1 hour