

Data Scientist/Machine Learning Engineer Data Scientist/Machine Learning Engineer Data Scientist/Machine Learning Engineer - EBay San Jose, CA Close to 10 years of expert involvement in IT in which I have 3+ years of knowledge in Data Mining, Machine Learning and Spark Development with big datasets of Structured and Unstructured Data. Data Acquisition, Data Validation, Predictive demonstrating, Data Visualization. Capable in measurable programming languages like R and Python. Proficient in managing entire data science project life cycle and actively involved in all the phases of project life cycle. Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data mining solutions to various business problems and generating data visualizations using R, Python and Tableau. Adept and deep understanding of Statistical modeling, Multivariate Analysis, model testing, problem analysis, model comparison and validation. Skilled in performing data parsing, data manipulation and data preparation with methods including describe data contents, compute descriptive statistics of data, regex, split and combine, Remap, merge, subset, reindex, melt and reshape. Experience in using various packages in R and libraries in Python. Working knowledge in Hadoop, Hive and NOSQL databases like Cassandra and HBase. Hands on experience in implementing LDA, Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis and good knowledge on Recommender Systems. Good industry knowledge, analytical and problem-solving skills and ability to work well within a team as well as an individual. Highly creative, innovative, committed, intellectually curious, business savvy with effective communication and interpersonal skills. I can be able to quickly adapt the new work pace and learning

Work Experience Data Scientist/Machine Learning Engineer EBay - San Jose, CA January 2017 to Present Roles & Responsibilities: Analyzed Trading mechanism for real-time transactions and build collateral management tools. Compiled data from various sources to perform complex analysis for actionable results. Utilized machine learning algorithms such as linear regression, multivariate regression, naive bayes, Random Forests, K-means, & KNN for data analysis. Measured Efficiency of Hadoop/Hive environment ensuring SLA is met. Developed Spark code using Scala and Spark-SQL/Streaming for faster processing of data. Prepared Spark

build from the source code and ran the PIG Scripts using Spark rather using MR jobs for better performance. Analyzing the system for new enhancements/functionalities and perform Impact analysis of the application for implementing ETL changes. Imported data using Sqoop to load data from MySQL to HDFS on regular basis. Developed Scripts and Batch Job to schedule various Hadoop Program. Used TensorFlow to train the model from insightful data and look at thousands of examples. Designing, developing and optimizing SQL code (DDL / DML). Building performant, scalable ETL processes to load, cleanse and validate data. Expertise in Data archival and Data migration, ad-hoc reporting and code utilizing SAS on UNIX and Windows Environments. Tested and debugged SAS programs against the test data. Processed the data in SAS for the given requirement using SAS programming concepts. Imported and Exported data files to and from SAS using Proc Import and Proc Export from Excel and various delimited text-based data files such as .TXT (tab delimited) and .CSV (comma delimited) files into SAS datasets for analysis. Expertise in producing RTF, PDF, HTML files using SAS ODS facility. Providing support for data processes. This will involve monitoring data, profiling database usage, trouble shooting, tuning and ensuring data integrity. Participating in the full software development lifecycle with requirements, solution design, development, QA implementation, and product support using Scrum and other Agile methodologies. Collaborate with team members and stakeholders in design and development of data environment. Learning new tools and skillsets as needs arise. Preparing associated documentation for specifications, requirements and testing. Optimizing the Tensorflow Model for an efficiency. Used Tensorflow for text summarization. Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive. Wrote Hive queries for data analysis to meet the business requirements. Developed Kafka producer and consumers for message handling. Responsible for analyzing multi-platform applications using python. Used storm for an automatic mechanism to analyze large amounts of non-unique data points with low latency and high throughput. Developed MapReduce jobs in Python for data cleaning and data processing.

Environment: Machine learning, AWS, MS Azure, Cassandra, SAS, Spark, HDFS, Hive, Pig, Linux, Anaconda Python , MySQL, Eclipse, PL/SQL, SQL connector, SparkML. Data Scientist/Machine

Learning Engineer Macy's Inc - Duluth, GA August 2015 to December 2017 Roles & Responsibilities: Performed Data Profiling to learn about user behaviour and merged data from multiple data sources. Participated in all phases of data mining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis. Performed K-means clustering, Multivariate analysis and Support Vector Machines in Python and R. Developed Clustering algorithms and Support Vector Machines that improved Customer segmentation and Market Expansion. Professional Tableau user (Desktop, Online, and Server). Data Story teller, Mining Data from different Data Source such as SQL Server, Oracle, Cube Database, Web Analytics, Business Object and hadoop. Providing AD hoc analysis and reports to Executive level management team. Data Manipulation and Aggregation from different source using Nexus, Toad, Business Objects, Power BI and Smart View. In Unix development environment, for Financial application reports used batch processes and models using Perl and Korn shell scripts with partitions and sub-partitions on oracle database. Developed analytics and strategy to integrate B2B analytics in outbound calling operations. Implemented analytics delivery on cloud-based visualization using shiny tool for Business Object and Google analytics platform. SPOC Data Scientist and predictive analyst to create annual and quarterly Business forecast reports. Main source of Business Regression report. Creating various B2B Predictive and descriptive analytics using R and Tableau. Creating and automating ad hoc reports. Responsible for planning & scheduling new product releases and promotional offers. Worked on NOSQL databases like Cassandra. Experienced in Agile methodologies and SCRUM process. Parsing data, producing concise conclusions from raw data in a clean, well-structured and easily maintainable format. Extracted data from HDFS and prepared data for exploratory analysis using data munging. Worked on Text Analytics, Naive Bayes, Sentiment analysis, creating word clouds and retrieving data from Twitter and other social networking platforms. Extensive experience and proficiency in using SAS ODS to create output files in a variety of formats including RTF, HTML and PDF. Worked on different data formats such as JSON, XML and performed machine learning algorithms in R and Python. Environment: R, Python, UNIX Scripting, SAS, Cassandra, Java, Hadoop,

MapReduce, HDFS, Pig, Sqoop, Hive, Oracle, Eclipse. Python Developer Myntra - Bengaluru, Karnataka April 2012 to June 2015 Roles & Responsibilities: Worked on the project from gathering requirements to developing the entire application. Worked on Anaconda Python Environment. Created, activated and programmed in Anaconda environment. Wrote programs for performance calculations using NumPy and SQLAlchemy. Wrote python routines to log into the websites and fetch data for selected options. Used python modules of urllib, urllib2, Requests for web crawling. Experience using all these ML techniques: clustering, regression, classification, graphical models. Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data mining solutions to various business problems and generating data visualizations using R, Python and Tableau. Used with other packages such as Beautiful Soup for data parsing. Involved in development of Web Services using SOAP for sending and getting data from the external interface in the XML format. Used with other packages such as Beautiful Soup for data parsing. Worked on development of SQL and stored procedures on MYSQL. Analyzed the code completely and have reduced the code redundancy to the optimal level. Design and build a text classification application using different text classification models. Used Jira for defect tracking and project management. Worked on writing and as well as read data from CSV and excel file formats. Involved in Sprint planning sessions and participated in the daily Agile SCRUM meetings. Conducted every day scrum as part of the SCRUM Master role. Developed the project in Linux environment. Worked on resulting reports of the application. Performed QA testing on the application. Held meetings with client and worked for the entire project with limited help from the client. Environment: Python, Anaconda, Synder (IDE), Windows 7, Teradata, Requests, urllib, urllib2, Beautiful Soup, Tableau, python libraries such as NumPy, SQL Alchemy, MySQLdb. Python developer Credit Sussie - Pune, Maharashtra August 2008 to March 2012 Roles & Responsibilities: Using python libraries for machine learning like pandas, numpy, matplotlib, sklearn, scipy to Load the dataset, summarizing the dataset, visualizing the dataset, evaluating some algorithms and making some predictions. Perform application development, maintenance and ensure adherence to process and information security control. AWS (vpc, EC2, S3, route 53). A key part of API

design and development   Unit test and debugging   Maintain and extend the existing Python/Flask REST   A strong eye on code re-usability and maintainability   Strong background in object-oriented programming, design patterns, algorithms, and data structures.   Used Python scripts to update the content in database and manipulate files.   Generated Python Django forms to maintain the record of online users.   Used Django API's to access the database.   Writing Unit, Functional, and Integration test cases for Cloud Computing applications on AWS.   Writing Python scripts with Cloud Formation templates to automate installation of Auto scaling, EC2, VPC and other services. Designed and managed API system deployment using fast http server and AWS architecture. Developed Restful API's using Python Flask and SQL Alchemy data models as well as ensured code.   Designed and managed API system deployment using fast http server and Amazon AWS architecture.   Designed and developed a horizontally scalable APIs using Python Flask.   Involved in back end development using Python with framework Flask.   Wrote Python modules to view and connect the Apache Cassandra instance.   Created Unit test/ regression test framework for working/new code.   Responsible for designing, developing, testing, deploying and maintaining the web application.   Wrote and executed various MySQL database queries from Python MySQL connector and MySQL Db package.   Involved in debugging and troubleshooting issues and fixed many bugs in two of the main applications.   Which is main source of data for customers and internal customer service team.   Implemented SOAP/RESTful web services in JSON format. Attended many days to day meetings with developers and users and performed QA testing on the application.   Environment: Python, Django, API, HTML, CSS, AJAX, Git, AWS, Apache HTTP, Flask, XML, OOD, Shell Scripting, MYSQL, Cassandra. Education Bachelor ECE in JNTUH University Bachelor's Skills CLUSTERING, DATA VISUALIZATION, ELASTICSEARCH, HADOOP, MACHINE LEARNING,   Expertise: Scikit-learn, NLTK, spaCy, NumPy, SciPy, OpenCv, Deep learning, NLP, RNN, CNN, Tensor flow, Keras, matplotlib, Microsoft Visual Studio, Microsoft Office. (8 years) Additional Information TECHNICAL SKILLS:   Expertise: Scikit-learn, NLTK, spaCy, NumPy, SciPy, OpenCv, Deep learning, NLP, RNN, CNN, Tensor flow, Keras, matplotlib, Microsoft Visual Studio, Microsoft Office.   Machine Learning Algorithms: Linear Regression, Logistic

Regression, Decision Trees, Random Forest, K-Means Clustering, Support Vector Machines, Gradient Boost Machines & XGBoost, Neural Networks. Data Analysis Skills: Data Cleaning, Data Visualization, Feature Selection, Pandas. Operating Systems: Windows, Mac and Linux, Unix. Programming Languages: Python, SQL, R, Matlab, Torch, C, C++, Java ,Octave, Apache Spark, Hadoop, Spark ML. Other Programming Knowledge and Skills: ElasticSearch, Data Scraping, RESTful-API using Django Web Frame work. Tools: Toad, Erwin, AWS, Azure,D3, Mule Soft, Alteryx, Tableau, Shiny, Adobe Analytics, Anaconda

Name: Christopher Johnson

Email: sclark@example.com

Phone: +1-608-997-4594