Sr. Data Scientist/Machine Learning Engineer Sr. Data Scientist/Machine Learning Engineer Sr. Data Scientist/Machine Learning Engineer - SEI Investments Developed Map Reduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming.   Perform Exploratory analysis, hypothesis testing, cluster analysis, correlation, ANOVA, ROC Curve and build models in Supervised and Unsupervised Machine Learning algorithms, Text Analytics & Time Series forecasting   Prepared the model data and built machine learning algorithms using Python Pandas, scikit learn, numPy, keras etc. libraries using Anaconda Jupyter & Programming Linear, Logistic Regressions, KNN, K-Means Clustering, Sentiment/Text Analytics, NLP, Na ve Bayes, Time Series forecasting using lm, glm, Arima, Apriori, Forecast.   Extracting data from Big Data Hadoop Data Lake, Excel, Analyzing, Cleaning, Sorting, Merging Reporting and creating dashboards using Base SAS, SAS Macros, SQL, Hive, SAS VA, SAS, and Excel.   Developing MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming.   Used pandas, numpy, seaborn, scipy, matplotlib, scikit-learn, NLTK in Python for developing various machine learning algorithms and utilized machine learning algorithms such as linear regression, multivariate regression, naive bayes, Random Forests, K-means, & KNN for data analysis.   Currently working on building clustering and predictive models using Mllib to predict fault code occurrences using Spark and Mllib.

  Conducting studies, rapid plots and using advance data mining and statistical modelling techniques to build solution that optimize the quality and performance of data.   Demonstrated experience in design and implementation of Statistical models, Predictive models, enterprise data model, metadata solution and data life cycle management in both RDBMS, Big Data environments. Stored and retrieved data from data-warehouses using Amazon Redshift and designed and implemented system architecture for Amazon EC2 based cloud-hosted solution for client. Developed Simple to complex Map Reduce Jobs using Hive and Pig and developed multiple Map Reduce jobs in java for data cleaning and preprocessing.   Analyzing large data sets apply machine learning techniques and develop predictive models, statistical models and developing and

enhancing statistical models by leveraging best-in-class modeling techniques.   Developed Map Reduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Worked with various Teradata15 tools and utilities like Teradata Viewpoint, Multi Load, ARC, Teradata Administrator, BTEQ and other Teradata Utilities.   Utilized Spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, Caffe, TensorFlow, MLLib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc.   Developed LINUX Shell scripts by using NZSQL/NZLOAD utilities to load data from flat files to Netezza database. Working on Information extraction from different kinds of text documents using NLP, text mining and regular expressions.   Worked extensively on Tableau Desktop, apply filters, drill downs, and generate Data visualizations, interactive Dash Boards, that can interact with views of data and worked on several options like query, display, analyze, sort, group, drill down, organize, summarize and generate charts, monitor and measure goals, identify patterns Work Experience Sr. Data Scientist/Machine Learning Engineer SEI Investments - Malvern, PA March 2017 to Present Responsibilities: Developed Map Reduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming.   Perform Exploratory analysis, hypothesis testing, cluster analysis, correlation, ANOVA, ROC Curve and build models in Supervised and Unsupervised Machine Learning algorithms, Text Analytics & Time Series forecasting   Prepared the model data and built machine learning algorithms using Python Pandas, scikit learn, numPy, keras etc. libraries using Anaconda Jupyter & Programming Linear, Logistic Regressions, KNN, K-Means Clustering, Sentiment/Text Analytics, NLP, Na ve Bayes, Time Series forecasting using lm, glm, Arima, Apriori, Forecast. Extracting data from Big Data Hadoop Data Lake, Excel, Analyzing, Cleaning, Sorting, Merging Reporting and creating dashboards using Base SAS, SAS Macros, SQL, Hive, SAS VA, SAS, and Excel.   Developing MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming. Used pandas, numpy, seaborn, scipy, matplotlib, scikit-learn, NLTK in Python for developing various machine learning algorithms and utilized machine learning algorithms such as linear regression,

multivariate regression, naive bayes, Random Forests, K-means, & KNN for data analysis. Currently working on building clustering and predictive models using Mllib to predict fault code occurrences using Spark and Mllib. Conducting studies, rapid plots and using advance data mining and statistical modelling techniques to build solution that optimize the quality and performance of data. Demonstrated experience in design and implementation of Statistical models, Predictive models, enterprise data model, metadata solution and data life cycle management in both RDBMS, Big Data environments. Stored and retrieved data from data-warehouses using Amazon Redshift and designed and implemented system architecture for Amazon EC2 based cloud-hosted solution for client. Developed Simple to complex Map Reduce Jobs using Hive and Pig and developed multiple MapReduce jobs in java for data cleaning and preprocessing. Analyzing large data sets apply machine learning techniques and develop predictive models, statistical models and developing and enhancing statistical models by leveraging best-in-class modeling techniques. Developed MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Worked with various Teradata15 tools and utilities like Teradata Viewpoint, Multi Load, ARC, Teradata Administrator, BTEQ and other Teradata Utilities. Utilized Spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, Caffe, TensorFlow, MLLib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. Developed LINUX Shell scripts by using NZSQL/NZLOAD utilities to load data from flat files to Netezza database. Working on Information extraction from different kinds of text documents using NLP, text mining and regular expressions. Worked extensively on Tableau Desktop, apply filters, drill downs, and generate Data visualizations, interactive Dash Boards, that can interact with views of data and worked on several options like query, display, analyze, sort, group, drill down, organize, summarize and generate charts, monitor and measure goals, identify patterns Environment: Python, SQL, Oracle 12c, Netezza, SQL Server, Informatica, Java, SSRS, PL/SQL, T-SQL, Tableau, MLLib, regression, Scala NLP, Spark, Kafka, MongoDB, Workday, logistic regression, Hadoop, Hive, TensorFlow, Teradata, IDE, random forest, OLAP, Azure, MariaDB, SAP CRM, HDFS, ODS, NLTK, SVM, JSON, Tableau, XML, AWS Redshift,

Pandas, Cassandra, MapReduce, AWS, Tableau, Caffe. Sr. Data Scientist/ Machine Learning Engineer UHG - Brooklyn, NY January 2015 to February 2017 Responsibilities: Implemented end-to-end systems for Data Analytics, Data Automation and integrated with custom visualization tools using R, Hadoop and MongoDB, Cassandra. Building predictive models using tools such as SAS, R with very granular data stored in big data platform. Worked on different data formats such as JSON, XML and performed machine learning algorithms in R and used Spark for test data analytics using MLLib and Analyzed the performance to identify bottlenecks. Involved working with Machine Learning Algorithms such as Decision Trees, Random Forest, Gradient Boosting, Support Vector Machines, K Mean Clustering, Na ve Bayes, Bayesian Belief Networks and Artificial Neural Networks. Developed Predictive models, Machine learning (Supervised and non-Supervised) using R for Machine Motor. Creating various B2B Predictive and descriptive analytics using R and Tableau and performed data cleaning and data preparation tasks to convert data into a meaningful data set using R. Used R to verify the results of Mahout on small data sets. Developed missing but important features of ML algorithms to the Mahout. Utilized Spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, MLLib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc Created partitioned and bucketed tables in Hive. Involved in creating Hive internal and external tables, loading with data and writing hive queries which involves multiple join scenarios. Performed K-means clustering, Multivariate analysis and Support Vector Machines in R. Create analytical models using analytics algorithms like regression, decision trees, clustering, text mining etc. and leveraging tools like R, Tableau etc. to deliver actionable insights and recommendations. Developed multiple Spark jobs using Scala for data cleaning and preprocessing Designed the schema, configured and deployed AWS Redshift for optimal storage and fast retrieval of data. Used External Loaders like Multi Load, T Pump and Fast Load to load data into Teradata14.1Database. Involved in Troubleshooting and quality control of data transformations and loading during migration from Oracle systems into Netezza EDW. Used S3 Bucket to store the jar's, input datasets and used Dynamo DB to store the processed output from the input data set. Worked on classification/scripting of multiple attribute

models by applying text-mining, NLP, SVM and Regular Expressions given product features like title, description etc. & predicting product attribute values using Python/R   Worked on different data formats such as JSON, XML and performed machine learning algorithms in R.   Used Spark for test data analytics using MLLib and Analyzed the performance to identify bottlenecks and used Supervised learning techniques such as classifiers and neural networks to identify patters in these data sets   Developed Tableau visualizations and dashboards using Tableau Desktop. Tableau workbooks from multiple data sources using Data Blending.   Developing new data warehousing system based on spark 2.x and spark streaming, utilizing Scala and Java 8   Strong Knowledge on concepts of DataModeling Star Schema/Snowflake modeling, FACT& Dimensions tables and Logical&Physical data modeling. Environment: R3.x, Erwin 9.5.2, MDM, QlikView, MLLib, PL/SQL, Tableau, Teradata 14.1, JSON, HADOOP (HDFS), MapReduce, SQL Server, MLLib, Scala NLP, SSMS, ERP, CRM, Netezza, Pandas, SAS, SPSS, Java, IDE, Cassandra, SQL, PL/SQL, AWS, SSRS, Informatica, PIG, Spark, Azure, R Studio, MongoDB, MAHOUT, JAVA, HIVE, AWS Redshift. SQL Developer against ODS Western Digital - Burlington, NJ April 2012 to December 2014 Burlington, NJ April 2012 to December 2014 Responsibilities: Design database, data models, ETL processes, data warehouse applications and business intelligence (BI) reports through the use of best practices and tools, including Erwin, SQL, SSIS, SSRS and OLAP, OLTP.   Transformed Logical Data Model to Physical Data Model ensuring the Primary Key and Foreign Key relationships in PDM, Consistency of definitions of Data Attributes and Primary Index Considerations.   Validated the data of reports by writing SQL queries in PL/SQL Developer against ODS.   Developed MapReduce programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in the EDW.   Involved with Data Analysis primarily Identifying Data Sets, Source Data, Source Meta Data, Data Definitions and Data Formats Data Analyst Macy's Inc - Duluth, GA January 2010 to March 2012 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. Python Developer HCL India April 2008 to December 2009 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, Sypder (IDE), Windows 7, Teradata, Requests, urllib, urllib2, Beautiful Soup, Tableau, python libraries such as NumPy, SQL Alchemy, MySQLdb. Skills SQL (10+ years), APACHE HADOOP MAPREDUCE (8 years), MapReduce (8 years), MAPREDUCE (8 years), PL/SQL (8 years) Additional Information Skills APACHE HADOOP MAPREDUCE (7 years), MapReduce (7 years), OLAP (7 years), ONLINE ANALYTICAL PROCESSING (7 years), PL/SQL (7 years) Technical Skills: Data Analytics Tools/Programming:

Python (numpy, scipy, pandas, Gensim, Keras), R (Caret, Weka, ggplot), MATLAB, Microsoft SQL Server, Oracle PLSQL, Python, SQL, PL/SQL, T-SQL, UNIX shell scripting, Java, SAS.   Big Data Techs: Hadoop, Hive, HDFS, MapReduce, Pig, Kafka, HBase, Cassandra, MongoDB.   Analysis and Modeling Tools: Erwin, Sybase Power Designer, Oracle Designer, BPwin, Rational Rose, ER/Studio, TOAD, MS Visio   ETL Tools: Informatica Power Center, Data Stage 7.5, Ab Initio, Talend   OLAP Tools: MS SQL Analysis Manager, DB2 OLAP, CognosPowerplay   Languages: SQL, PL/SQL, T-SQL, XML, HTML, UNIX Shell Scripting, C, C++, AWK   Databases: Oracle12c/11g/10g/9i/8i/8.0/7.x, Teradata14.0, DB2 UDB 8.1, MS SQLServer 2012/2008/2005, Netezaa and Sybase ASE 12.5.3/15, Informix 9, HBase, MongoDB, Cassandra, Amazon Redshift.

Operating Systems: Windows 2007/8, UNIX (Sun-Solaris, HP-UX), Windows NT/XP/Vista, MSDOS   Project Execution Methodologies: Ralph Kimball and Bill Inmon data warehousing methodology, Rational Unified Process (RUP), Rapid Application Development (RAD), Joint Application Development (JAD)   Reporting Tools: Business ObjectsXIR2/6.5/5.0/5.1, Cognos Impromptu 7.0/6.0/5.0, Informatica Analytics Delivery Platform, MicroStrategy, Tableau.   Tools: MS-Office suite (Word, Excel, MS Project and Outlook), VSS   Others: Spark MLLib, Scala NLP, MariaDB, Azure, SAS, IDE, Microsoft Azure, AWS, Data Scientist/Machine Learning Engineer - Macy's Inc Duluth, GA   Close to Eight 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

Name: Julie Salinas

Email: amber15@example.org

Phone: 820.909.4636x3606