# Professional Summary

Professional in the field of Data Science with 6+ years of experience in statistical analysis, data analytics, data modeling, and creation of custom algorithms. Application to the disciplines of machine learning and neural networks using a variety of systems and methods in training algorithms with different could platform. Industry experience includes predictive analytics in finance, marketing, advertising, geospatial and Internet of Things (IoT). Use of NLP and Computer Vision technologies.

* Experience with a variety of NLP methods for information extraction, topic modeling, parsing, and relationship extraction
* Familiarity with developing, deploying, and maintaining production NLP models with scalability in mind
* Worked on Natural Language Processing with NLTK, SpaCy and other module for application development for automated customer response
* Wrote automation processes using Python and the AWS Lambda service
* Utilized Docker to handle deployment on heterogeneous platforms such as Linux, Windows, OSX, and AWS
* Reviewed the use of MongoDB, node.js, and Hadoop to automate the data ingestion and initial analysis processes
* Reviewed and deployed the infrastructure on AWS to minimize cost while providing the required functionality
* Scale analytics solutions to Big Data with Hadoop, Spark/PySpark, and other Big Data tools
* Experience with Public Cloud (Google Cloud, Amazon AWS and/or Microsoft Azure)
* Experience working with big data infrastructure with tools such as Hive, Spark and h2o, sparkling water
* Implementing solutions with common NLP frameworks and libraries in Python (NLTK, spaCy, gensim) or Java (Stanford CoreNLP, NLP4J)
* Experience with knowledge databases and language ontologies
* Quantitative training in probability, statistics and machine learning
* Experience in the application of Neural Network, Support Vector Machines (SVM), and Random Forest.
* Creative thinking and propose innovative ways to look at problems by using data mining approaches on the set of information available
* Identifies/creates the appropriate algorithm to discover patterns, validate their findings using an experimental and iterative approach
* Applies advanced statistical and predictive modeling techniques to build, maintain, and improve on multiple, real-time decision systems. Closely works with product managers, service development managers, and product development team in productizing the algorithms developed
* Experience in designing star schema, Snow flake schema for Data Warehouse, ODS architecture
* Experience in designing stunning visualizations using Tableau software and publishing and presenting dashboards, Storyline on web and desktop platforms
* Experience in working with relational databases (Teradata, Oracle) with advanced SQL programming skills
* In-depth knowledge of statistical procedures that are applied in Supervised / Unsupervised problems
* Basic-Intermediate level proficiency in SAS (Base SAS, Enterprise Guide, Enterprise Miner) & in UNIX
* Track record of applying machine learning techniques to marketing and merchandizing ideas
* Experience in Big Data platforms like Hadoop platforms (Map-R, Hortonworks & others), Aster and Graph Databases
* Experience in operations research / optimization.
* Experienced in working with advanced analytical teams to design, build, validate and refresh data models that enable the next generation of sophisticated solutions for global clients
* Excellent communication skills (verbal and written) to communicate with clients and team, prepare + deliver effective presentations
* Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design

# Technical Skills

**Data Science Specialties:** Natural Language Processing, Machine Learning, Internet of Things (IoT) analytics, Social Analytics, Predictive Maintenance

**Analytic Skills:** Bayesian Analysis, Inference, Models, Regression Analysis, Linear models, Multivariate analysis, Stochastic Gradient Descent, Sampling methods, Forecasting, Segmentation, Clustering, Naïve Bayes Classification, Sentiment Analysis, Predictive Analytics, Stochastic Analytics

**Analytic Tools:** Classification and Regression Trees (CART), H2O, Docker, Support Vector Machine, Random Forest, Gradient Boosting Machine (GBM), TensorFlow, PCA, RNN, Linear and non-Linear Regression

**Analytic Languages and Scripts:** R, Python, HiveQL, Spark, Spark SQL, Hadoop, Scala, Impala, MapReduce, Millib

**Languages:** Java, Python, R, Command Line, C++/C, JavaScript, SQL, SAS

**Python Packages:** Numpy, Pandas, Scikit-learn, Tensorflow, SciPy, Matplotlib, Seaborn, Plotly, NLTK, Scrapy, Gensim

**Version Control:** GitHub, Git, SVN

**IDE:** Jupyter Notebook, VS Code, Intellij IDEA, Spyder, Eclipse

**Data Query:** Azure, Google Bigquery, Amazon RedShift, Kinesis, EMR; HDFS, RDBMS, SQL, MangoDB, HBase, Cassandra and NoSQL, data warehouse, data lake and various SQL and NoSQL databases and data warehouses.

**Deep Learning:** Machine Perception, Data Mining, Machine Learning Algorithms, Neural Networks, TensorFlow, Keras

**Soft Skills:** Experienced with delivering presentations and technical reports; collaboration with stakeholders and cross-functional teams, and advisement on how to leverage analytical insights. Developed analytical reports which directly address strategic goals.

# Professional Experience

**Nov**

**Wells Fargo | San Francisco, California**

**Project Summary:**

Wells Fargo is committed to satisfy customers financial needs and help them succeed financially throughout the world. Data scientist was responsible for recognizing opportunities to improve the prediction of misconduct through the use of Machine Learning (ML) and Natural Language Processing (NLP) techniques. They are expected to build models that evaluate and improve new compliance regulations implemented by the Feds, Consumer Financial Protection Bureau (CFPB), and the Options Clearing Corporation (OCC) by combining both unstructured data, using text analytics and Natural Language Processing (NLP), and structured data.

My role on the project was responsible for performing text analysis, building advanced predictive models and analytics supporting LOBs across the enterprise, and render business intelligence reports from varied data. The insights gleaned from this data were used to determine target markets and their issues, plan supply and demand of financial products and ways to save on costs with predictive analytics as well as reduce risk and examine strategic partnerships.

This project includes using machine learning techniques and programming languages to derive relevant analysis and metrics, including building proof of concepts to determine value of implementing in future projects. As part of a team focused on understanding the customer experience across the organization, I worked with business partners to understand the business objectives, explore data sources and build NLP/ML solutions. I was able to think outside the box to uncover new ways to analyze unstructured data leveraging our Open Source Data Science Platform (OSDS) and other analytical tools to solve complex business objectives.

**Project Points:**

* Applied business analytics skills, integrated and prepared large, varied datasets and communicated results.
* Worked with specialized database architecture and computing environments.
* Developed analytic approaches to strategic business decisions.
* Performed analysis using predictive modeling, data/text mining, and statistical tools.
* Built predictive modeling using Machine Learning algorithms such as Random Forests, Na ve Bayes, Neural Networks, MaxEnt, SVM, Topic Modeling/LDA, Ensemble Modeling, GB, etc.
* Used common NLP techniques, such as pre-processing (tokenization, part-of-speech tagging, parsing, stemming)
* Performed semantic analysis (named entity recognition, sentiment analysis), modeling and word representations (RNN / ConvNets, TF-IDF, LDA, word2vec, doc2vec)
* Worked with big data infrastructure and tools such as Hive and Spark
* Collaborated cross-functionally with team to develop actionable insights.
* Synthesized analytic results with business input to drive measurable change.
* Performed data visualization and developed presentation material using Tableau.
* Responsible for defining the key business problems to be solved while developing, maintaining relationships with stakeholders, SMEs, and cross-functional teams.
* Provided knowledge and understanding of current best practices and emerging trends within the analytics industry.
* Participated in product redesigns and enhancements to know how the changes will be tracked and to suggest product direction based on data patterns.
* Applied statistics and organizing large datasets of both structured and unstructured data.
* Worked with applied statistics and applied mathematics tools for performance optimization.
* Facilitated data collection to analyze document data processes, scenarios, and information flow.
* Determined data structures and their relations in supporting business objectives and provided useful data in reports.
* Assisted in continual improvement of AWS data lake environment.
* Used Agile approaches, including Extreme Programming, Test-Driven Development, and Agile Scrum.
* Promoted enterprise-wide business intelligence by enabling report access in SAS BI Portal and Tableau Server.

**June 2015 – Oct 2016**

**Data Scientist**

**Shell Oil | Houston, Texas**

**Project Summary:**

Predictive Analytics can enhance oil field production and cuts costs by finding optimal well settings and forecasting equipment failures and potential problems. The data spanned several years tracking oil wells in every major North American basin. The data included information on drilling and operational data from thousands of wells and hundreds of miles of low-pressure pipelines. Analysis of the data revealed critical issues with field deployed equipment.

This project built “digital twins” — computer models replicating above and below ground well behavior for artificial lift equipment. Input from sensor readings was applied to specific field issues: 1) improving plunger timing to realize well potential; 2) predicting preventive equipment maintenance to prevent failure in rods and submersible pumps; 3) reducing overuse of chemicals in wells.

**Project Points:**

* Worked in Git development environment
* Experienced in Data Integration Validation and Data Quality controls for ETL process and Data Warehousing using MS Visual Studio, SSIS, SSAS, SSRS
* Adept at using SAS Enterprise suite, Python, and Big Data related technologies including knowledge in Hadoop, Hive, Sqoop, Oozie, Flume, Map-Reduce
* Proficient in Predictive Modeling, Data Mining Methods, Factor Analysis, ANOVA, Hypothetical Testing, and Normal Distribution
* Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data
* Professional competency in Statistical NLP / Machine Learning, especially Supervised Learning- Document classification, information extraction, and named entity recognition in-context
* Worked with Proof of Concepts (POC's) and gap analysis and gathered necessary data for analysis from different sources, prepared data for data exploration using data wrangling
* Designed Physical Data Architecture of New system engines
* Hands on experience in implementing neural network skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis and good knowledge on Recommender Systems
* Strong SQL Server and Python programming skills with experience in working with functions
* Efficient in developing Logical and Physical Data model and organizing data as per the business requirements using Sybase Power Designer, ER Studio in both Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) applications
* Experience in designing star schema, Snow flake schema for Data Warehouse, Operational Data Store (ODS) architecture.
* Experience and technical proficiency in Designing, Data Modeling Online Applications, Solution Lead for Architecting Data Warehouse/Business Intelligence Applications
* Worked with languages like Python and Scala and software packages such as Stata, SAS and SPSS to develop neural network and cluster analysis
* Designed visualizations using Tableau software and publishing and presenting dashboards, Storyline on web and desktop platforms
* Developed Logical Data Architecture with adherence to Enterprise Architecture
* Used dplyr in R and pandas in Python for performing Exploratory data analysis
* Experience working with data modeling tools like Power Designer and ER Studio
* Adept in statistical programming languages like R and Python including Big Data technologies like Spark, Spark SQL, PySpark Hadoop, HIVE, HDFS, and MapReduce
* Well experienced in Normalization & De-Normalization techniques for optimum performance in relational and dimensional database environments
* Skilled in System Analysis, E-R/Dimensional Data Modeling, Database Design and implementing RDBMS specific features
* Responsible for Data Analytics, Data Reporting, Ad-hoc Reporting, Graphs, Scales, PivotTables and Online Analytical Processing (OLAP) reporting
* Interacted with data from Hadoop for basic analysis and extraction of data in the infrastructure to provide data summarization
* Created visualization tools and dashboards with Tableau, ggplot2 and d3.js
* Worked with and extracted data from various database sources like Oracle, SQL Server, and DB2

**Jan 2014 – May 2015**

**Data Scientist**

**Swift Transportation | Norfolk, Virginia**

**Project Summary:**

The transportation industry uses Big Data analytics to transform supply chain logistics -- to optimize routing, to streamline factory functions, and to give transparency to the entire supply chain. The data comes from a variety of sources, including enterprise systems, traffic analysis, weather forecasts, location information, mobile internet and GPS enabled smartphones. Logistics is a dynamic and complex process which is prone to bottlenecks, particularly in the last mile of a supply chain.

The analytics group analyzed and addressed logistics optimization. The goal was to save money and time, avoid late delivery and manage utilization of resources. This included planning routes, managing employment and fleets. Data was utilized on destination locations, shipping area, parking and time of delivery to prevent backlogs at the unloading point, and ensure deliveries were made at a time when they could be received.

**Project Points:**

* Identified and executed process improvements, hands-on in various technologies such as Oracle, Informatica, and Business Objects
* Designed both 3NF data models for operational data store (ODS), online transaction processing (OLTP) systems and dimensional data models using Star and Snow Flake Schemas
* Developed large data sets from structured and unstructured data. Perform data mining.
* Partnered with modeling experts to develop data frame requirements for projects
* Performed Ad-hoc reporting/customer profiling, segmentation using R/Python
* Created statistical models for the collected data, exploratory, pre-processing, to provide conclusions with decision guides
* Programmed a utility in Python that used multiple packages (scipy, numpy, pandas)
* Implemented Classification using supervised algorithms: Logistic Regression, Decision trees, KNN, Naive Bayes
* Validated machine learning classifiers using ROC Curves and Lift Charts
* Extracted data from HDFS and prepared data for exploratory analysis using data wrangling
* Updated Python scripts to match training data with database stored in AWS Cloud Search and assign each document a response label for further classification

**May 2012 – Dec 2013**

**Data Scientist**

**Axiom Tech Group | Chicago, Illinois**

**Project Summary:**

Axiom provides enterprise advisory solutions and risk management. The firm relies on big data analytics to provide strategic insights and reporting to outside clients. I worked on analytics projects for various clients; cleaning data and performing analysis and reporting.

**Project Points:**

* Business Data Analytics - Involved in ETL/BI requirement gathering and conversion into useful functional requirements
* Source to target data Mapping document preparation
* Developed report wireframes along with SQL schema data element definitions
* Worked with Data Warehouse architecture and wrote SQL queries
* Applied dimension modelling to identify dimension & fact tables and associated data elements
* Familiar with wealth and asset management concepts as well as trading life cycles
* Conducted in-depth data analysis on the reports/dashboards to identify gaps
* Involved in data governance to find authoritative sources for the critical data elements used in the governance reports
* Data profiling to validate data quality issues for the critical data elements
* Participated in user acceptance testing to ensure software satisfied all requirements before it was deployed to production
* Knowledge in BFSI domain and financial markets
* Well versed with Agile process, scrum and sprint concepts
* Knowledge on creation of user stories in JIRA
* Worked as an ETL/Reporting Tester
* Familiar with Test plan & strategy document preparation and Test case preparation based on the requirements
* End to end testing in DWH projects
* Validation of ETL jobs against requirements by running through Control-M scheduler
* Validating target tables structure, constraints against ETL requirements
* Validating target data against source data based on ETL requirements
* Involved in test data preparation
* Report & Dashboard testing against target tables using SQL queries
* Worked with module testing including defect capturing in ALM
* Experienced with complete software development life cycle (SDLC) and software testing life cycle (STLC) life cycles
* Worked extensively with on-time delivery, process improvement, regular interaction with client and mentoring the team

# Education

**Degree**

University of California, Los Angeles (UCLA)

MA. Applied Economics

**Degree**

University of Colorado, Boulder

BA. Economics and Mathematics (double major)