Gaurav Sitaula

Data Scientist

Phone: 415-290-7975

Email: gsitaula3001@gmail.com

Professional Summary

Over 5 years of experience in Machine Learning, Deep Learning, Data Mining with large datasets of Structured and Unstructured Data, Data Acquisition, Data Validation, Predictive Modeling,

Experience with AWS cloud computing, Spark (especially AWS EMR), Kibana, Node.js, Tableau, Looker.

Experience in the healthcare domain, or with quality data sets.

Strong technical communication skills; both written and verbal.

Ability to understand and articulate the “big picture” and simplify complex ideas.

Strong problem solving and structuring skills.

Ability to identify and learn applicable new techniques independently as needed.

Ability to create new methods of solution through a combination of foundational research and collaboration with ongoing initiatives.

Experience in stochastic optimization, which has resulted in utilization by commercial applications or open-source algorithms.

Experience formulating and solving discrete and continuous optimization problems.

Experience developing and applying novel methods of stochastic optimization and optimization under uncertainty algorithms for large scale problems, which must include mixed integer type problems.

Expertise with design optimization methods with computational efficiency considerations.

Able to research statistical machine learning methods which must include forecasting, supervised learning, classification, and Bayesian methods.

Conduct complex, advanced research projects in areas of interest to Business Units.

Develop new and advanced cutting-edge techniques and algorithms

Transfer and implement results and technology in hard- and software prototypes and demo systems relevant to the businesses

Survey relevant technologies and stay abreast of latest developments

Draft and submit papers and patents based on research

Contribution to several research projects that combine new data sources and computational tools

Knowledge of remote sensing.

Capable of writing efficient code and working with large datasets

Exceptional mathematical and statistical modeling and computer programming skills

Use of mathematical and statistical modeling and computer programming skills in an innovative manner.

Effectively worked within an interdisciplinary research environment.

Capable of advanced technical sophistication of solutions using machine learning and other advanced technologies.

Technical Skills

Analytic Development: Python, Spark, R, SQL, R, Command Line, Markdown, SAS, SPSS, C/C++

Python Packages: Numpy, pandas, scikit-learn, TensorFlow, SciPy, Matplotlib, Seaborn, Bokeh, Numba

IDE: Jupyter, Spyder, RStudio, Visual Studio, Code::Blocks Version Control: GitHub, Git, SVN

Machine Learning: Natural Language Processing & Understanding, Machine Intelligence, Machine Learning algorithms

Data Query: Azure, Google, Amazon RedShift, Kinesis, EMR; HDFS, RDBMS, SQL 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

Artificial Intelligence: text understanding, classification, pattern recognition, recommendation systems, targeting systems, ranking systems.

Analysis Methods: Advanced Data Modeling, Forecasting, Predictive, Statistical, Sentiment, Exploratory, Stochastic, Bayesian Analysis, Inference, Models, Regression Analysis, Linear models, Multivariate analysis, Sampling methods, Forecasting, Segmentation, Clustering, Sentiment Analysis, Predictive Analytics, Decision Analytics, Big data and Queries Interpretation, Design and Analysis of Experiments, Factorial Design and Response Surface Methodologies, Association Analysis

Analysis Techniques: Classification and Regression Trees (CART), Support Vector Machine, Random Forest, Gradient Boosting Machine (GBM), TensorFlow, PCA, RNN, Regression, Naïve Bayes

Data Modeling: Bayesian Analysis, Statistical Inference, Predictive Modeling, Stochastic Modeling, Linear Modeling, Behavioral Modeling, Probabilistic Modeling, time-series analysis

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

Soft Skills: Excellent communication and presentation skills; ability to work well with stakeholders to discern needs accurately, leadership, mentoring, coaching

Professional Experience

Oct 2016 - Present

Data Scientist

Sensely

San Francisco, CA

Project Summary:

This Virtual Medical Assistant avatar integrates AI to recommend diagnoses based on patient symptoms using a smartphone. The platform uses algorithms trained on large volumes of clinical content, such as medical protocols and chronic disease information, to interpret patient symptoms and to recommend an appropriate diagnosis. Patients can describe their symptoms to the virtual medical assistant named Molly, using speech, text, images and video. The platform can be integrated with electronic health records to provide continuity of care, allowing clinicians to monitor patients outside of the clinical care setting.

The objective of this project is to provide an engaging assistant to a patient having chronic disease with the ability to interact with Sensely's virtual assistant "Molly" for their daily follow-up visits. The platform uses algorithms trained on large volumes of clinical content, such as medical protocols and chronic disease information, to interpret patient symptoms and to recommend an appropriate diagnosis. My role in the project was the development of a Chat-Bot that can recommend the appropriate diagnosis based on the patient's symptoms, using trained algorithms. The Chat-Bot can also update the information of the patient to the existing data storage system.

Details:

Responsible for Data modeling and testing models, and applied predictive techniques to forecast application results.

Use of Predictive Modeling, Data Mining Methods, Factor Analysis, ANOVA, Hypothetical Testing, and Normal Distribution.

Translated 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.

Use of Recurrent Neural Net, 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.

Use of Python programming skills and working with functions.

Development of Logical and Physical Data model and organizing data.

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, ggplot2 and d3.js software and publishing and presenting dashboards, Storyline on web and desktop platforms.

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.

Use of Spark 2.1, Spark SQL and PySpark, Apache Kafka for stream processing.

Normalization & De-Normalization for optimum performance.

Responsible for Data Analytics, Data Reporting, Ad-hoc Reporting, Graphs, Scales, PivotTables and OLAP reporting.

Interacted with data from Hadoop for basic analysis and extraction of data in the infrastructure to provide data summarization.

May 2015 – Oct 2016

Data Scientist

Synchrony

Draper, UT

Project Summary:

Synchrony provides credit card services to major vendors. Many transactions to initiate credit, and credit applications are done on the web or on intranet systems. The project involved working on a system and software for identifying fraudulent behavior in applications and transactions. Analysis involved looking at differences and changes in user behavior, actions on a website, and anomalies in responses and transactions. This comprised plurality of fields or input parameters that identify the actions performed on a website including fields related to previous actions by that user or other users of the website. The fields or input parameters are represented in a vector format where vectors represent different sessions of activity on the website, pages of the website, users of the website, or other attributes of the use of a website. Analysis is performed to determine if new sessions are similar or dissimilar to previously known sessions and if a session is converging or diverging from known sessions based on the velocity and direction of the velocity of the vectors in the vector space.

Project Points:

Applied advanced analytics skills, with proficiency at integrating and preparing large, varied datasets, working with big data engineers, and communicating findings to stakeholders such as C-level executives.

Developed analytical approaches to strategic business decisions.

Performed analysis using predictive modeling, data/text mining, and statistical tools.

Collaborated cross-functionally to arrive at actionable insights.

Synthesized analytic results with business input to drive measurable change.

Assisted in continual improvement of AWS data lake to provide refine and structure data sets.

Identifying, gathering, and analyzing complex, multi-dimensional datasets utilizing a variety of tools.

Performed data visualization and developed presentation material utilizing Tableau.

Developed data visualization tools using ggplot2, d3.js, R Shiny, bokeh.

Responsible for defining the key business problems to be solved while developing, maintaining relationships with stakeholders, SMEs, and cross-functional teams.

Used Agile approaches, including Extreme Programming, Test-Driven Development, and Agile Scrum.

Provided knowledge and understanding of current 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.

Use of applied machine learning algorithms and performance optimization.

Worked with applied statistics and applied mathematics in the development of analysis tools.

Facilitated the 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.

November 2013 – May 2015

Data Scientist

Cerebri AI

Austin, TX

Project Summary:

Uses AI to predict customer behavior and help to figure out the customer commitment to a brand or product. Cerebri AI developed ­software — using IBM Corp.’s Watson supercomputing platform — designed for Fortune 500 companies. We are tasked to build the predictive models to meet the client requirements based on the historical data. The predictive model dynamically predicts "Next Best Actions" of the customer behavior to improve the customer engagement which eventually improves the customer success and ROI of the organization. This also helps users to make decisions during the sales process. As such, it’s expected to reduce risk and increase sales revenue.

Project Points:

Identifying and executing process improvements, hands-on in various technologies such as Oracle, Informatica, and Business Objects.

Identified, gathered, and analyzed datasets for customer analysis to create specific machine learning algorithms.

Performed data visualization and developed presentation material utilizing Tableau and “R programming”.

Worked with stakeholders to define business case desired results and determine steps to implement.

Developed large data sets from structured and unstructured data and developed data models.

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, skikitlearn, pandas)

Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes.

Validated the machine learning classifiers using ROC Curves and Lift Charts.

Extracted data from HDFS and prepared data for exploratory analysis using data munging.

Updated Python scripts to match training data with our database stored in AWS Cloud Search, so that we would be able to assign each document a response label for further classification.

August 2012 – November 2013

Data Analyst

Citibank

Irving, TX

Project Summary:

Worked on Data Analytics team to manage customer data, analyze customer profiles and develop programs to recommend products that would result in a high degree of sales. Data was derived from Citi big data system and my analytics team supported Sales and Distribution, matching the right product to the right segment, (client relationship management overview), Decision Management, supporting the decision making process with the right and accurate data, whether it is to grow or wind down specific portfolios, or automating reporting to regulators, eliminating manual work and errors that can lead to heavier impairments.

Project Points:

Simplified and digitized Citi's global consumer banking system as part of Customer Management Systems (CMS) team.

Generated SQL scripts to retrieve data from multiple tables and to load data into UAT and production environment.

Handling large databases in Dev, UAT and Production environments. Maintained high security while sharing the data within internal, external team.

Performed data analysis to measure individual performance and analyzed priorities of tickets to draw insights.

Developed reports, dashboards for daily/weekly/monthly performance metrics by using SQL, MS Excel, MS PowerPoint and share point.

Recognizes the connection between business operations and analytics to influence business strategies and solutions.

Gathers and prepares data from multiple sources to support information analytics.

Monitor and track data reporting details for all data sources.

Document requests for data enhancements and proactively review data for areas requiring improvement, change, or infill.

Education

MS in Applied Statistics with a Specialization in Business Analytics, Bowling Green State University, Bowling Green, OH

BS in Electronics and Communication Engineering, Tribhuvan University, Kathmandu, Nepal