

DATA SCIENTIST / PROJECT MANAGER/ CEO

Summary

I am a data scientist/engineer/analyst with a passion for statistics/machine learning/deep learning and focused on providing insightful data-driven solutions. I have 6+ years experience in data analytics and data engineering in various fields of manufacturer, finance and infrastructure. My current research and development interests include data engineering, machine learning, deep learning, NLP, and experimentation.

Skills

Data Scientist, Data Analyst, Business System Analyst, Oracle PLSQL developer. Data Visualization: D3.js, Tableau, R visualization packages, Microsoft Excel. Data Analytics Tools/Programming: Python (numpy, scipy, pandas, scikit-learn, gensim, keras), R (caret, weka, and ggplot), SQL, Microsoft SQL Server, MYSQL, Oracle PLSQL, Hadoop, Hive, and MapReduce. Machine Learning Algorithms: classification, regression, clustering, feature engineering. Big Data Tools: Hadoop, MapReduce, SQOOP, Pig, Hive, NOSQL, Cassandra, Spark. Others: Deep Learning, NLP, Topic Modeling, SkLearn, Graph Mining, Text Mining, C, C++, PHP, JavaScript, ASP, Shell Scripting.

Experience

07/2017 to Current

Data Scientist / Project Manager/ CEO Apex Systems i¼ Albuquerque , NM

- Develop Tableau map report to show new born puppy/kitten data distribution in US for better visualization for customers on pet online purchase website.
- Optimized Recommendation System of our E-Commerce website by deeply integrating with user's background information, user's subconscious and personality information through his/her social network, and user's pet information, to better predict and recommend pet/pet product/pet service to users based on Collaborative Filtering using Python.
- Optimized pet supplies website product search engine for users searching by a customized sentence instead of key word, by using NLP Tokenization, Normalization, POS Tagging, Parsing, Named Entity Recognition and Terminology extraction, topic modeling, automatic summarization.
- Analyze the user's tagging, comments, retweet, and like history of Pet Social Forum on our website for a whole week, to predict hot pet topic/trend for coming week.
- Abstract the pet topic related dataset from twitter's dataset, do text mining on Twitter's Pet dataset using topic modeling and sentimental classification, by integrating users' background information using deep learning algorithms like CNN/RNN/LSTM/GRU, compared with Random Forest/SVM to predict hot topic/trend among pet owners, to further adjust our website sales strategy to recommend popular pet/pet product/service to customers.
- Assisted in development of www.pettigpet.com website of PC version and responsive mobile version using PHP, MYSQL, HTML, CSS, and JavaScript, based on framework of Thinkphp.
- Assisted in development of Mobile PettingPet service app on IOS using Swift and Android using Angular.
- Recruiting, assigning and training personnel to fill roles as needed on project team, handling project execution from pre-planning progress through implementation and evaluation.
- Design, plan and manage the PettingPet project by establishing comprehensive work plan by preparing the business /development/ technical requirement documents to specify the project scope, objectives and criteria for evaluation.
- Identifying sequencing, assigning, monitoring and tracking individual activities, tasks, and project performance as required.
- Reviewing work progress, including quality and quantity with team members at regular intervals, ensuring project deliverables are completed in-time and to-standard.
- Overseeing and managing project progress in a manner that complies with company, policies and procedures.

09/2016 to 09/2017

Data Engineer Waukesha , WI

- Develop Facebook DataSwarm pipelines using Python and Hiveql, for hardware data processing/load between MySQL database, Hive Data warehouse, Scuba system and ODS system, liking refactoring cables/reboots pipeline to have less tasks, setting up automatic email alert for SAS dashboard, grab ODS uptime to Hive, developing pipelines to parse NVME metrics, etc.
- Daily Hardware Health Dashboard reporting using Line chart, bar chart, histogram, Bullet Graph, Heat map, Map, creating data sets, groups, bins, annotations and mark labels, filters, trend lines, writing Tableau function/formulas, etc.
- using Tableau, for hardware health data maintenance like Server reboot dashboard, hardware component failure dashboard, SAS&SATA& NVME HDD/ Flash Drive/IPMI /OOB/Cables health dashboard for data centers in Department of Infrastructure in Facebook.
- Create Thermostat dashboard for rack level sensors & failures data using Line chart, bar chart, histogram by creating sets, groups, filters and writing Tableau formulas, connect server and publish the dashboard with scheduling set up.
- Machine learning application in hardware product validation cycle.
- Data Analysis/Prediction on Infrastructure Hardware data using Statistical Method and Machine Learning algorithms and deep learning algorithms using Python and R.
- Identify/break/fix any issues or bugs involved in data pipelines of data from FB infrastructure hardware health regarding setup, performance, functionality, and workflow stuck, etc.

05/2010 to 09/2014

Business System Analyst Canon USA i¼ City , STATE

- Operational support for Canon Americas Mercury system, includes data adjustment/research, batch data loading, system migration, technical and functional specification documentation, reports, business process alignment, workflow stuck, and reconciliation, etc.
- Break/Fix any issues or bugs collected from client and development regarding setup, performance, functionality, and workflow stuck, etc.

- System Enhancement regarding functionality and performance, etc.
- Review and analyze ASP code for UMC Mercury application web development for data research and system feature fix and enhancement using Visual Interdev 6.0.
- Query real-time data regarding Canon Americas new item request, item status inquiry, item data disclosure and import, model configuration and maintenance, warranty maintenance, and model configuration inquiry, etc.
- Using complex SQL queries on Oracle 9i Canon mercury database.
- Data loading using Imp/exp, data pump, and external tables from Americas Mercury system to S21-CUSA merchandise master, to Ross-CUSA retails system, and Global Mercury system.

Education and Training

2016

Master of Science : Computer Science University of California i¼ City , State , USA Computer Science GPA: 3.78 Courses:-Statistics Programming -Databases and Knowledge Bases -Graphs and Network Flows -Language and Thought -Current Topics in Computer Theory -Computer Science Classics -Basic Data Science -Data Mining and Big Data Analytics -System Security

2010

Master of Science : Electrical Engineering University of Bridgeport i¼ City , State , USA Electrical Engineering Electrical Engineering 3.62 Computer Networks -Database Management Systems -Data and Computer Communications -Data Structures

2007

Bachelor of Science : Telecommunications Engineering Jilin University i¼ City , State , China Telecommunications Engineering GPA: 3.50 Skills

Big Data, business process, C++, Cables, clustering, Data Analyst, Data Analysis, data processing, Data Mining, Data Visualization, Data warehouse, Database Management, E-Commerce, Electrical Engineering 3.62, PHP, JavaScript, Machine Learning, Microsoft SQL, MYSQL, NOSQL, Oracle 9, Oracle, PLSQL, Python, sales, SAS, scheduling, Shell Scripting, SQL, System Analyst, Tableau, validation, web development