SENIOR DATA SCIENTIST

Summary

12+ years of machine learning/data science experience with a proven track record of developing and implementing large scale algorithms that have significantly impacted business revenues and user experience. 10+ years of experience applying R, Python, SAS and SQL for algorithm development, data modeling, statistical learnings and data visualization. Ranked a top 1% competitor on Kaggle (worlds largest community of data scientist with over 300k members) Hands-on experience applying several ML/Statistical algorithms to real-world problems: Neural Networks, Gradient Boosted Trees, Random Forest, Clustering, Generalized Linear Models, Simulation models and Gaussian Mixture Models Strong experience building end-to-end machine learning platform using Java and big data technologies (Cassandra, Spark and Hive). Excellent skills at programming in Java/OOP

Highlights

- R Neural Networks
- Java Gradient Boosted Trees
- Python Random Forest
- NoSQL, Cassandra Generalized Linear Models
- Spark, Hive, Pig Optimization
- SQL, CQL Gaussian Mixture Models

Experience

Senior Data Scientist May 2017 to Current

Factset Research Systems Inc. ï¼ Reston , VA Conceptualized, architected and implemented a deep learning model to accurately predict demand for Nvidia's products. Algorithm applies Deep neural networks along with several real time social feeds, econometric indicators, product momentum to forecast demand.

Data Science Lead Jul 2014 to May 2016

Bristol Myers Squibb ï¼ Minneapolis , MN

- Developed and owned the core machine learning feature of the product, a patent-pending behavioral analytics engine for predicting cross-cloud application security threats.
- Architected and implemented the machine learning platform from the ground up in Java, Cassandra, Hive and Spark.
- Analytic platform monitors multiple cloud applications (AWS, Salesforce, Box etc.) and predicts security threats and breaches.
- Implemented the end-to-end platform for performing user behavior analytics using unsupervised machine learning.
- Designed and implemented algorithms for real-time decisioning Implemented process to capture 500+ behavioral profile features Implemented several high IV features for ML algorithm consumption: (Maximum Geo-distance between users login/day, Fast Geolocation Hopping etc.) Owned the feature end-to-end and worked directly with co-founder/CTO to drive product enhancements.

Principal Data Scientist Dec 2012 to May 2014

Lumeris ï¼ Saint Louis , MO

- Lead several big data machine learning initiatives involving the design, development and deployment of advanced machine learning algorithms that impacted PayPal's core products allowing business to grow and scale to over 150 million customers and process over 3 billion online transactions annually.
- The developed solutions have directly helped PayPal lower losses, improve user experience and increase revenue simultaneously.
- Few key PayPal core products worked on are included below.
- PayPal ATO Models: Designed and developed Neural Network models to prevent account takeover activity at the time of user authentication.
- Performed feature selection from over 2000 features, and applied best practices in Neural Network model development.
- Model accurately captured 81% of incoming fraud activity and resulted in 48bps of reduction of losses for PayPal in 2013 freeing up capital to grow and expand margins.
- PayPal Here Mobile Card Reader: Architected the very first anti-fraud ML algorithm that enabled PayPal here product to expand and scale to over million merchants.
- Designed many new geolocation specific features that showed high IV contribution to the implemented model.
- PayPal Wallet: Spearheaded the development of ensemble models for predicting stolen financial usage at the time of transaction completion.
- Developed Gradient Boosted Trees using GBM package in R.
- Implemented new features using IP, IP-GEO, Velocity etc.
- to improve algorithm performance.
- Designed several tools in R to visualize GBT model and developed process to extract model parameters for deployment.
- Partnered with business and product teams to execute and influence business decision around the solution implementation.

Lead Analytic Scientist Jun 2007 to Dec 2012

Cushman & Wakefield Inc ï¼ Stockton , CA

- Managed the design, development and deployment of several predictive modeling solutions that underlie ISO Risk Analyzer portfolio of products which have contributed to significant revenues for the company.
- Few highlights are listed below.
- Lead the development of an expert system to accurately predict a consumer's likelihood of committing a traffic violation.Developed Logistic models for predicting violation probability and Poisson models for predict violation frequency.Incorporated intelligence from 1500 features from 7 different data sources (Claritas, Crime, Traffic Patterns, Business Info, Weather etc.) Tweedie/Compound Poisson Distribution

Model Developed and implemented several Tweedie distribution models to examine environmental factor for predicting personal auto risks for calculating insurance premiums.Developed separate frequency and severity models to improve model accuracy.Captured interaction effects between various environment variables such as traffic generator, traffic pattens, population density, weather etc.Risk Analyzer Credit Scoring Module: Designed and developed first generation credit scoring algorithms to predict a consumer's accident likelihood using credit payment behavior.Performed feature reduction on close to 800 credit attributes and applied Multivariate Adaptive Regression Splines (MARS) to detect 2nd and 3rd order interactions effects.Captured non-linearity using splines and piecewise regression.Worked as the analytic manager for the movie advertisement testing product, a high revenue growth initiative for the company.

Senior Research Analyst Oct 2004 to May 2007
Nbc Universal ï¼ Boston , MA

- Managed 5 statisticians to build predictive models for rating movie trailer/TV spots content on their effectiveness to generate interest.
- Provided analytic insights and advice to maximize the impact and reach of client's resources.
- The product served 4 of the top hollywood movie studios and generated over $13M in annualized revenues for the company.
- Movie Trailer/TV Spots Advertisement Testing Product (Predictive models for rating movie trailers, TV spots, print ads etc.).
- Team worked on over 100 movies for most of the motion picture studios-majors, and assisted in building customized statistical models for several mini-majors.
- Provided strategic insights and advice to Hollywood Studios to maximize the impact and reach of clients marketing resources.
- Box Office Forecasting Models: Developed forecasting models to predict weekend box office revenues to help studios estimate accurately a films potential at the day of the release.
- Model used several audience tracking data to accurately predict weekend box office receipts.
- Fox Movie Sequel Study (Client - 20th Century Fox) ? Developed a series of ordinal logistic regression models to compare characteristics and attributes describing several Fox movies.
- Based on the study results, 20th Century Fox decided to produce a sequel to Die Hard in 2007, which was one of the biggest box office hits of that year.

Researcher Jan 2002 to Jan 2004
University Of Missouri ï¼ City , STATE

- Managed home healthcare initiative for Missouri's Sinclair School that involved developing a AI system based GPS to optimize hospital resources for saving cost.
- The GPS system incorporated new algorithms for solving multiple-traveling salesman problem in a dynamic setting.
- Researched and applied spatial clustering algorithms in the context of constraint optimization problem.

Education
Master of Science , Industrial Engineering/Operations Research 2004 University of Missouri ï¼ City , State , US Industrial Engineering/Operations Research Thesis - "Dynamic Stochastic Vehicle Routing Model in Home Healthcare Scheduling"
Metallurgical Engineering 2002 Indian Institute of Technology Metallurgical Engineering BHU Varanasi, UP, IN Thesis - "Monte Carlo Simulation Modeling of Non-Uniform Stress in Structured Materials" {Secured All India IITJEE rank in the top 0.6% -- ranked among 250,000 applicants}
Skills
ads, AI, big data, clustering, content, Credit, Client, clients, expert system, Fast, features, financial, Forecasting, GPS, Home Healthcare, insurance, IP, ISO, Java, machine learning, machine learning, marketing, MARS, Materials, Office, 2000, Modeling, monitors, Network, Networks, Neural, NoSQL, Optimization, predict, Testing Product, Python, real-time, Routing, Scheduling, Simulation, SQL, strategic, Structured, TV