LeiYang

Engineering | Data Science

contact

4304 Ellsworth Blvd
Malta, NY 12020

781.534.9264

lei.yng@gmail.com

technology

• Hadoop • Spark
• Storm • Hive
• PostgreSQL • D3
• MongoDB • Tableau

programming

• Python • R
• C# • Java
• JavaScript • SQL

language

chinese mother tongue
english fluency

link

Github:// leiyang-mids
LinkedIn:// leiyng

affiliation

IEEE, ASME

summary

- Engineering background (PhD Mech. Eng.) with specialty in industrial data analysis, Master of Information and Data Science, seeking data scientist position.
- Experience in managing end-to-end data analysis pipeline from data exploration, feature engineering, model building, performance evaluation, and visualization.
- Expertise in application of data mining techniques, including machine learning and statistical modeling.
- Experience in data analysis at large scale, with specialty in parallel computing with MapReduce framework.
- Proficiency in programming with Python, R, Javascript, C#, Java, SQL, Matlab.
- Familiarity with various analytic and visualization technologies, including Hadoop, Spark, Storm, PostgreSQL, Hive, scikit-learn, D3, Tableau, ggplot, and matplotlib.

education

2015–2016 **Master** of Information and Data Science — UC Berkeley School of Information

2005–2011 **Ph.D.** Mechanical Engineering — University of Cincinnati
Specialization in anomaly detection and fault prediction

2004–2005 **M.S.E** Mechanical Engineering — University of Michigan, Ann Arbor

1999–2004 **B.S** Mechanical Engineering — Shanghai Jiao Tong University
Graduate with honors

experience

2010–Now **GLOBALFOUNDRIES** — Malta, New York
Member of Technical Staff - R&D fault prediction
• Conduct research for equipment fingerprinting with time-series data; develop a comparing system using **k-Nearest Neighbor**. • Conduct research on equipment fault prediction; develop an analytic pipeline with **Partial Least Square** for failure predictability evaluation, component remaining useful life estimation, and degradation root-cause identification. • Implement engineering solutions into factory infrastructure and maintain application throughout the life-cycle.

2008–2009 **Advanced Micro Devices** — Austin, Texas
Senior Factory Automation Engineer - R&D anomaly detection
• Develop fault detection and classification solution with **Principal Component Analysis** algorithm. • Implement enhancement for real-time model adaptation with recursive updating strategy. • Research on parameter sharing between models from different context groups, using statistical analysis.

2007 **Advanced Micro Devices** — Austin, Texas
Co-Op Engineer - R&D predictive maintenance
• Analyze maintenance log, evaluate impact and severity of different failure modes, and develop a procedure to prioritize maintenance strategy.

projects

2016 **Semantic Health:** A Search Engine for Choosing Obamacare Plan Github
• Using ElasticSearch, built a search engine for CMS.gov public use file of ACA plans. • Implemented a pair-wise **Learn to Rank (LETOR)** algorithm to improve search ranking based on user click-through rate data. • Using Flask and SearchKit, built semantic search front-end for health plan browsing.

2015 **Facial Keypoints Recognition:** A Machine Learning Competition Kaggle
• Built a **Convolutional Neural Network** to detect 15 different keypoints on the face. • Applied various feature engineering techniques to enhance training data, including histogram stretching, Gaussian blurring, and image cropping and flipping. • Ranked 13th on the Kaggle leaderboard at submission time.

2015 **Path Finder:** A Cycling Segment Analytics Tool Github
• Built a web app based on **Lambda Architecture** to demonstrate data storage and retrieval techniques. • Batch retrieved history data on HDFS (Hive) for cycling segments, real-time streaming analysis using Spark. • Provided web user interface for summary and search functions such as popularity ranking, filtering based on category, distance, elevation, zip etc. • Visualize results with D3.

2015 **Commodity Price:** A Visualization Approach Github
• Analyzed functional gaps of current commodity trade process. • Implemented a **D3 dashboard** of 3 graphs to provide overview of trading price, drill-down with open interest and volume of a specific contract, and detailed view of one day trading. • Conducted usability testing of the solution, addressing visualization challenges in workflow streamlining, loading speed, and web app design etc.

2015 **Penny Wise and Pound Foolish:** A Social Experiment Report
• Conducted **randomized social experiment** to evaluate round number impact on gift card Valuation. • Designed survey question and collect inputs from Amazon Mechanical Turk. • Analyzed and estimated treatment effect with linear regression for various covariate.

2005–2010 **Dissertation Research** PDF
• *Qualitative algorithm comparison:* developed an Analytical Hierarchy Process based method for prognostic algorithm selection with user survey results. • *Semiconductor equipment diagnosis:* investigated and demonstrated the application of **Bayesian Belief Networks** on semiconductor fault classification and root-cause identification. • *Prognostic methodology design:* developed a unified approach for data analysis pipeline, including functions for feature extraction, failure prediction, and root-cause diagnosis.

publications

- L. Yang and J. Lee. A toolbox approach for prognostic development and deployment, *Proceedings of 62th meeting of the MFPT society: Failure Prevention for System Availability. May 2008.*
- L.Yang and J.Lee. Bayesian Belief Network-based Approach for Diagnostics and Prognostics of Semiconductor Manufacturing Systems. *Journal of Robotics and Computer-Integrated Manufacturing, 2011.*
- R.Good, R.Madani and L.Yang. A Fab-Wide Chamber Matching Solution for High-Mix Semiconductor Manufacturing Based on Equipment Fingerprinting, *Proceedings of SEMICON Europa, September 2012.*