Xiaojun Zhao
Data Scientist

Tel: (+1)860-208-6944
Email: zhaoxiaojun106@gmail.com
Linkedin: https://www.linkedin.com/in/xiaojunzhao/
Decision Support & Machine Intelligence Group,
United Technologies Research Center, East Hartford,CT

Profile

- 4+ years of extensive quantitative research/data science experience in various areas, such as Predictive Analysis, Prognostics Health & Monitoring, Cybersecurity, NLP, etc, with a proven track record of developing and implementing statistical/machine learning algorithms that have significantly impacted business revenues, and solve a variety of real world industry problems.
- 4+ years of hands on experience applying ML/Statistical algorithms to real-world problems: Random Forest, Regression models, Neural Networks (LSTM, CNN), XG boost, Clustering, Gaussian Mixture Models.
- Strong experience building production-grade end-to-end data pipelines for internal and external client.

Experience

Data Scientist 2016-Present
United Technologies Corporation Research Center, East Hartford, CT

Interact with **United Technologies Research Center** (UTRC) Program Offices, **United Technologies Corporation** (UTX) business units (**Pratt&Whitney ,Climate Control & Security ,OTIS Elevator, UTC Aerospace Systems**), external research organizations (**DARPA, DHS**, etc) and universities to enhance existing programs and provide technical expertise for decision support method development.

After-Market Pricing Analytics of Engine Maintenance Costs (Pratt & Whitney* 2016-2017*)

- Built and maintained the data warehouse (Parquet) of 5+ years of shop visiting records containing 150,000+ engine parts using data ETL (extract, transform, load) process.
- Developed and validated a novel predictive model for billion dollar V2500 portfolio managed by P&W Engine Services organization, and provided significance performance uplift in reducing the prediction error of the engine maintenance cost by 90%.
- Spearheaded the development of the end-to-end data pipeline and operationalized by the P&W after 2017Q1.

Prognostics and Health Monitoring (United Technologies Corporation Aerospace Systems* 2017*)

- Identify drivers of best/worst in class airline performance with respect to removals of a particular 787 line-replaceable unit (LRU) by applying feature selection and LDA analysis.
- Developed a machine learning model for removal predictions of an Embraer 170/190 component by incorporating air quality, weather information and flight history with operationalizable accuracy.

Mobile Application Security Program Project *(Department of Homeland Security* Current*)*

- *Funded by DHS ($1.45 million) focuses on detecting Android Malware detect malicious and vulnerable apps of varying risk severity levels.*
- *Trained adversarial deep learning model over 3+ million mobile applications to accurately classify samples built with majority of benign code and less malicious code.*
- *Produce a detailed risk-assessment report that includes an explanation of why an app is considered malicious using the LIME Analysis (Local interpretable Model-Agnostic Explanations).*

Data Scientist Fellow 2015.9-2015.11

The Data Incubator Fellowship Program, New York City, NY

- Selected (< 2% acceptance rate) among thousands of scientists and engineers with advanced degrees after intensive three round technical interview (more than 80% of the candidates hold PhDs).
- Rigorous eight week full-time data science post-graduate training fellowship.
- Completed various data science and data analytics projects, using techniques including web scraping, data wrangling, SQL database building and querying, predictive modeling, machine learning, interactive visualization, NLP, Hadoop MapReduce, Spark, Tensor-Flow, etc.

Graduate Researcher 2014 -2015

University of Connecticut, Storrs, CT

- Researched and explored the classification task for 3D point cloud models by incorporating supervised machine learning approaches with powerful shape descriptors (Light-Field descriptors, Angular Radial Transform, and Zernike Descriptors) that have traditionally been used for classifying polygonal models.

Education

Master of Science, Mechanical Engineering, GPA: **4.0 (A)** 2013-2015

University of Connecticut - Storrs, CT, US

Research Thesis-"Machine Learning Approaches to 3D Model Classification"

Skills

Programming: Python (4+ yrs), SQL/NoSQL(3+ yrs), Scala(2+yrs), R(2 years)

Data Science/Engineering/Visualization: Apache Spark, Django, AWS, Hive, Bokeh, Seaborn, D3.js, Tableau

Machine/Deep Learning: Scikit-Learn, Tensorflow, Keras, Gensim, SpaCy

Publications

1. **X. Zhao,** H. Ilies, "**Learned 3D Shape Descriptors for Classifying 3D Point Cloud Models"**, Journal of CAD and Application, 2017
2. D. Mehta, **X.Zhao,** E. Bernal, D. Wales, "**The Loss Surface of XOR Artificial Neural Networks**" Journal of Physical Review E, 2018
3. G. Ekladious, **X. Zhao**, H. Mostafa, R. Georgescu "**Dynamic Modeling of Maintenance Costs**" 2017 Annual PHM Conference