JAY SHAH

Santa Clara, CA-95054   (979)213-3271

jayshah5696@gmail.com   jayshah5696   jayshah5696

EDUCATION

Texas A&M University, College Station, Texas   May 2019
Master of Science in Industrial and Systems Engineering, *minor in Applied Statistics* **(GPA 4.00/4.00)**

Gujarat Technological University, Ahmedabad, India   May 2017
Bachelor of Engineering in Mechanical Engineering **(GPA 3.76/4.00)**

CORE COMPETENCIES

Statistical Methods: Regression Analysis, Hypothesis testing and Confidence intervals, ANOVA, Supervised parametric and Non-parametric learning methods, Unsupervised learning methods, Time Series Analysis, Bayesian Analysis, Anomaly Detection, Deep Learning, Dimensionality reduction, and A/B testing.

Programming Tools and Libraries: Python (Pandas, NumPy, Matplotlib, Scikit-Learn, Keras, Tensorflow, PySpark, Dask, Pytorch), R (Tidyverse, Caret, ggplot), SQL (Postgres, ElasticSearch, Django)

PROFESSIONAL EXPERIENCE

Ensemble Energy, Palo Alto, California
Data Scientist   Jun 2019 – Present
- Built a predictive maintenance system for wind assets, to estimate wind turbine's remaining useful life (RUL) and to handle & maintain its logistics using ML model utilizing industrial sensor and fault data with 92% accuracy.
- Developed processes and systems for automated deployment of the models in production with performance tracking using Airflow, Docker, Serverless Functions, Terraform tech on AWS and GCP.
- Architected real-time data processing (ETL) pipeline for 5 customers, and achieved 60% reduction in processing time utilizing multiprocessing, pandas, dask pipeline in Python on AWS serverless cloud server.
- Prototyped and productionized statistical tools to provide insight into power production inefficiency and quantify the energy loss.
- Mentored and guided an offshore team of 2 with providing ml systems principles and designs related strategies.

Data Science Intern   Jun 2018 – Dec 2018
- Implemented a robust anomaly detection system to predict component failure using GBM for 8 components of the wind turbine.
- Estimated bearing type and segmented bearing failures based on 10-min signature profile using K-means clustering to perform RCA.
- Delivered executable insights to customers, by performing physics-based statistical data analysis, and advanced data visualization utilizing ggplot & matplotlib library in Python, that helped to increase 150K $/year in revenue.

Texas A&M University, College Station, Texas
Graduate Research Assistant   Jan 2019 – May 2019
- Researched with Dr. Yu Ding on applying advance machine learning methods to solve and predict wind energy system failure.
- Implemented deep learning methods to predict possible power production and downtimes associated with the failures of wind turbine.

Utilities and Energy Services, College Station, Texas
Student Analyst   Dec 2017 - Jun 2018
- Created weather-controlled building baseline regression models for all digitally metered utilities using enterprise energy module. These models are used to monitor consumption across the campus to prevent sensor issues and energy loss.
- Manipulated Data in SQL to compare baseline modelled consumption with real-time consumption using statistical control limit chart in excel to analyze the average variation related to prediction.

KEY DATA SCIENCE PROJECTS

Power Curve Estimation for Wind Energy Farms
- Implemented regression estimation by applying Boosting model utilizing XgBoost library and achieved cross-validated 0.0310 RMSE with hyperparameter tuning, which is a 30% reduction than the International standard method (IEC Binning).

Portfolio Analysis of New York Cab Data (Data Open – 2018)
- Communicated insights and actions of Uber's strategy that led to increasing its market share. Awarded 2nd Runner-up position at Texas DataOpen – 2018 for recommending portfolio strategies on investment to support decision making.

Real-time detection of Drowsiness-related lane departure using vehicle measures
- Trained CNN with transfer learning in Keras (Python) on time series images using novel feature generation techniques and achieved 7% improved in AUC over state-of-the-art methods with a confidence interval of (0.75-0.86 AUC) using bootstrap significance test.