# Summary

## Areas of Expertise

* Data Science
* Oil Production and Process
* Natural Language Processing
* Supply Chain Optimization

## Technical Skills

*Data Science and Engineering:*

* Machine/Deep Learning
* Classification and Clustering
* Regression and Time Series
* Data Wrangling and Visualization
* Keras, Tensorflow, CNN, RNN, LSTM, ConvLSTM
* Pandas, Scikit-learn, Scipy, NumPy, PySpark, Matplotlib, Seaborn, Folium
* Big Data, Distributed ML, Databricks, Spark.ML, Hadoop, Horovod, SystemML
* Python, SQL, Java, C++, JSON, Matlab
* Azure, IBM Cloud, AWS
* Google Map, Foursquare, ArcGIS APIs

*Production and Process:*

* OLGA, PIPESIM, SPS (Stoner), PIPEPHASE, PVTSim, Multiflash
* HYSYS, Unisim, ProMax

# Profile

A Professional Engineer with extensive data science experience and profound domain knowledge in oil production, supply chain, natural language processing, and anomaly detection. Developing and implementing large scale data-driven solutions to challenging problems. Strong teamwork and communication skills. Has published and presented 10+ peer-reviewed papers.

# Professional Experience

## Senior Lead Engineer

## Wood Group Plc, 2011 – Present

Energy services company, $10bn revune, 55,000 emplyees.

Lead machine learning and deep learning model development. Use distributed frameworks to solve big data problems. Develop and implement data engineering workflows, practices, and procedures. Interface with other disciplines for the whole data analysis lifecycle. Mentor junior engineers. Estimate man-hours and prepare bid proposals. Deliver great presentations to management, clients, and peers.

*Selected Data Science Projects:*

#### Shale Gas Production Prediction

Developed deep learning models (CNN, RNN) to predict shale gas flow rate. Performed feature engineering analysis and selected the key features. Tuned the model to achieve an R-2 score above .97

#### Morale Analysis Based on Yammer Comments

Built a multiple channel convolutional neural network to classify the comments on Yammer. Improved the accuracy by 3%, compared to a simple CNN or LSTM model.

#### Bottom Hole Pressure Prediction

Built a time series recurrent neural network (RNN) to predict a shale well bottom hole pressure. Reshaped 2-dimensional data for an LSTM layer input. The simulations showed great agreements with test data. Tuned the model for future 2-week pressure prediction.

#### Gathering Flowline Flow Pattern and Pressure Prediction

Cross validated different machine learning classification methods and optimized the XGBoost method to predict the flow pattern within flow lines. Used the predicted flow patterns as feature parameters for machine/deep learning models to improve pressure prediction by 1%.

#### Shale Wellhead Pressure and Liquid Holdup Estimation Using Machine Learning

Studied different regression methods to estimate the wellhead pressure. Performed a hyperparameter tunning for deep learning neural network models to increase the modeling accuracy.

#### Adjacent Well Flow Rate Estimation at Late Life

Tested Deep Neural Network, Gradient Boosting, Random Forest, SVR, SGD Ridge Regression methods to estimate the adjacent well flow rates. Optimized and increased the SVR model score by adding ‘holdup’ as an additional feature

#### Raw Material Demand Forecast

Developed supply chain optimization models to predict raw material and parts demand of process plants. Increased regression model accuracies by considering crude oil price and stock market trends.

#### Oil Well Artificial Lift Method Selection

Built classification machine learning models to determine the optimal artificial lift method. Enriched the features by adding superficial velocity parameters to improve the model accuracy by 4%.

#### Blockage Location Detection Using Machine Learning

Developed Machine learning algorithms to detect the locations of blockages. Optimized collaborative filters and correlation methods to extract blockage features and calculate the locations of blockages.

#### Internal Project to Recommend New Office Location

Used the Foursquare API to get venue information for areas in Houston. Developed a K-means model to cluster the areas. Ranked potential new office area using a recommendation algorithm.

# Previous Experience

## Data Processing Engineer, Circor Energy, 2008-2011

Flow and motion control solution provider, $1.1bn revune, 4,400 emplyees.

Responsible for the provision of data analytics knowledge to managers, project personnel, clients and third parties. Architected and implemented data analytics and visualization for anomaly detection. Developed data acquisition, and data processing workflows. Prototyped and commercialized a pipeline on-line monitoring system which enables detection of blockages and leaks in live subsea pipelines. Conducted JIP projects with universities and industrial clients, and mentored young engineers working towards professional status.

*Selected projects:*

*BP, Bruce Subsea Pipeline Project, North Sea*

*BP, Cleeton Subsea Pipeline Study, North Sea*

*Duke Energy (USA), Cooling Water Pipeline Hydraulic Study, South Carolina*

## Earlier experience, 1999 to 2008

The University of Manchester, 2006 to 2008 – Signal Processing Engineer Control Systems

The University of Nottingham, 2002 to 2006 – Research Assistant

Beijing Oriental Jicheng Company, Ltd., 1999 to 2002 – Process Engineer

# Qualifications

## Education

Postdoctoral Researcher, Electrical and Electronic Engineering, University of Manchester

Doctor of Philosophy (Ph.D.), Chemical and Environmental Engineering, University of Nottingham

Bachelor of Engineering, Processing Engineering, University of Science and Technology, Beijing

## Certifications / Licenses

Licensed Professional Engineer, Texas (#118804)

Chartered Engineer, Engineering Council, UK

IBM Data Science Professional Certifications