

# Rahul Sangole, Data Science Manager.

rsangole@u.northwestern.edu | rsangole.netlify.com | github: rsangole | 812.390.5166

## Profile

Data science manager with Six Sigma Black Belt certification and 11 years of work experience delivering value in a Fortune 500 company. 5 years of experience leading teams solve business problems in areas of engineering, service, quality and product management.

## Education

2016 – 2019 Master of Science in Predictive Analytics  
Northwestern University

2006 – 2007 Master of Science in Mechanical Engineering  
University of Michigan, Ann Arbor

## Experience

Cummins, *Data Science Manager*

Aug 2016 - Present

- Lead a team of data scientists to deliver algorithms that alert engineers upon anomalous engines behavior reducing unscheduled operational downtime for mining customers resulting in financial savings of \$1m+
- Develop unsupervised anomaly detection schemes using statistical tests, robust regression filters, process control methods, cusum control charts, ranked permutation transformation on multivariate datasets
- Deployed algorithms in production-ready R packages using functional, defensive programming using *trycatch*, *assertive*, *testthat*, *log4r*, *RevoScaleR* on an Azure HDInsights cluster with Hive and Azure Blob Storage
- Generated new insights on engine usage patterns across customers by application of t-SNE to high dimensional datasets that impacted the solution strategy of a major prognostics initiative
- Developed a failure prediction model on high class imbalance problem using lasso regression and xgboost models
- Developed analytics functional excellence practices: project chartering processes, CRISP-DM adoption, coding guidelines, R repos, common utility packages, version control practices (*gitflow*), company-wide technical sessions, monthly trainings etc
- Co-leading an initiative to investigate research publications for engine specific prognostics and anomaly detection methods
- Developed data science position profiles and competency definitions required to setup a new data science function within Cummins

Cummins, *Six Sigma Black Belt*

Nov 2014 – Aug 2016

- Led high complexity and big impact projects using 6 sigma for quality, product management, supply chain and engineering with financial impact from \$150,000 to \$10 million. Received numerous awards.
- Solved business problems using inferential statistics, null-hypotheses testing, regression, measurement systems analysis, control charts etc.

Cummins, *Product Validation*

Jan 2008 – Nov 2014

- Drove \$2 million cost reduction via improved engine component designs on numerous new product launches.
- Developed calibrated finite element analyses (2-10% error) to predict on-engine failures.

# Resume of Rahul Sangole, Continued.

## Academic Projects

- Sales price prediction for Ames housing dataset using multilinear regression modeling using *lm*
- Wins per game predictive modeling for 186 years of baseball data using regression and decision trees using *glm*, *broom*
- Wine case purchase volume predictive modeling using poisson, negative binomial, and hurdle models using *glm*, *caret*, *broom*
- Model to predict donation amounts for a not-for-profit marketing campaign using a variety of machine learning modeling approaches including boosting, bagging, random forest, PCR and elastinet using *caret*
- Miles per gallon prediction on the *ISLR::auto* using *flexmix* modeling
- Time series forecasting for item level forecasts for Russian software firm competition (1C). Top-down approach using TSLM, ARIMA, Prophet, STL models using *xts*, *forecast*, *TSclust*, *mice*
- Time series forecasting for the DengAI, disease spread competition, utilizing transfer entropy, method of analogues, and single layer LSTM models using *xts*, *forecast*, *TransferEntropy*, *keras*
- Text analysis of aviation safety data using tSNE, TF-IDF and structural topic modeling in R using *tm*, *topicmodels*, *stm*, *tidytext*
- Developed fully connected neural network model developed using *numpy* and *pandas*, to classify MNIST dataset
- Customer segmentation modeling using tSNE, hierarchical agglomerative clustering and k-means followed by market segmentation profiling
- Discrete choice experiment modeling using Hierarchical Bayes Multinomial Logit to select product design
- Model to predict target market for campaign using random forests and naïve bayes models

## Coursework

Regression & Multivariate Analyses, Generalized Linear Models, Applied Machine Learning, Advanced Modeling Techniques, Deep Learning, Marketing Analytics, Experimental Design & Process Control, Text Analytics, Statistical Quality Control, DB Systems & Data Preparation

## Skills

Languages & Cloud – R, Python, SQL, Azure  
Visualization – ggplot2, lattice, matplotlib, seaborn  
Packages – tidyverse, caret, xts, pandas, scikit-learn  
Other – RStudio, Anaconda, git, SAS JMP, Minitab, PostgreSQL

## Achievements

- 2015 Chairman's Quality Award
- 2014 Chairman's Quality Award nomination
- Best Practice Awards for four Six Sigma Green Belt projects
- 2009 Most Valuable Player Award
- KVPY Scholar, Aero Department, IIT Bombay, 2003 - 2006

## Extra Curricular

- University of Michigan Recruiting Team, 2009 - Present
- Contributor and maintainer of *ProjectTemplate* R package
- Mechatronic projects using Arduino
- Yearly community involvement activities
- SAE Indiana Board Member, 2011 - 2016