

# MADHURI GUPTA

233 E Wacker Dr, Apt 3009 | Chicago, 60601 | +1 (510) 604-6509

[MadhuriGupta2017@u.northwestern.edu](mailto:MadhuriGupta2017@u.northwestern.edu) | <https://github.com/madhuri333> | <https://www.linkedin.com/in/madhuri-gupta/>

## EDUCATION

**Northwestern University | Evanston, IL** **Sep 2016 - Dec 2017**

**Masters of Science in Analytics | GPA: 3.98/4.00**

*Courses:* Predictive Analytics • Data Mining • Big Data • Data Visualization • Data Warehousing • Java and Python Programming • Text Analytics • Social Network Analysis

**University of Delhi - Netaji Subhas Institute of Technology | Delhi, India** **Aug 2006 - May 2010**

**B.E. in Instrumentation & Control Engineering | GPA: 3.83 | Rank: 2/120 | Awarded Merit Scholarship** for all 4 years

**Data Science Specialization by John Hopkins University (with distinction through Coursera)** **Oct 2015 - Jul 2016**

*Courses:* Getting & Cleaning Data • Exploratory Data Analysis • Statistical Inference • Regression Models • Practical Machine Learning • Developing Data Products • Reproducible Research.

## SKILLS

**Languages/Version Control:** R • SQL • Java • Python • MapReduce • PySpark • GitHub • Bash • AWS • KnitR • RShiny

**Environments:** R Studio • SQL (PostgreSQL, SQL Server Mgmt Studio) • Anaconda • IPython • Tableau • Teradata • Eclipse

**Techniques:** XGBoost • Random Forest • SVM • Neural Net • Naïve Bayes • Regression • K-means • GMM • PCA • CRM • CLV

**Certification:** Certified Scrum Master (CSM)

## WORK EXPERIENCE

**HERE | Analytics Consultant | Chicago, IL** **Sep 2017 - Dec 2017**

- Analyzed temporal and geospatial data and increased Automatic Closure Detection coverage by 68% in Germany.
- Identified distinct highway-like clusters from arterial roads using PCA and k-means in Python.

**Enova | Analytics Intern | Chicago, IL** **Jun 2017 - Sept 2017**

- Developed a bankruptcy prediction model for CashNetUSA products with full project lifecycle experience from building to implementation - defined data paradigm (SQL), performed feature selection and engineering subject to regulatory constraints (R), built multiple models (XGBoost, Random forest, Logistic Regression), corrected for class imbalance, performed scoring/testing while ensuring reproducibility.
- Analyzed long term impact of dormant line of credit accounts to quantify expected return from continued customer engagement.
- Determined predictive power of electronic bank statements in making under-writing decisions for consumer credit.

**Gogo | Analytics Consultant | Chicago, IL** **Sep 2016 - May 2017**

- Built Random forest classification model as a precursor for targeted marketing and optimal pricing. Model utilized web click data (Adobe Analytics) in conjunction with customer and flight information to identify probable customer purchase and significant factors driving it (SQL, R). Other trial models included Neural Networks, Naïve Bayes, SVM.

**Bharat Heavy Electricals Limited (BHEL) | Engineer | Noida, India** **Jul 2010 - Dec 2015**

- Analyzed and identified relevant sensor signals from DCS-SCADA to maximize safe operation and prevent tripping of thermal power plants. Also reduced operational costs (sensor quantity, cable procurement) by INR 900K.
- Recognized as 'Distinguished Employee' for innovative work in standardizing control schemes at firm level.

## KEY PROJECTS AND COMPETITIONS

**Enova's Data SmackDown | 2<sup>nd</sup> Prize (Competition organized by Enova International)** **Nov 9, 2016**

- Employed CRISP-DM and developed regression model to predict likelihood and sales of published books in R.
- Used k-fold cross validation and Lasso regularization to improve robustness in profit estimation.

**Stats LLC | Chicago, IL | (NBA player segmentation)** **Jan 2017 - Mar 2017**

- Performed Exploratory Factor analysis on box score data of 2015 NBA season to identify a player's true role.
- Clustered player performance using K-means and Gaussian Mixture models in R.

**Studio Gang Architects | Chicago, IL** **Sep 2016 - Dec 2016**

- Scraped, cleaned, and analyzed 5000+ user reviews on Studio Gang's architectures from Yelp and Trip Advisor.
- Assigned positive/negative scores to reviews, consequently gaining insights into popular sentiment (R, Python).

**Capstone Project for Data Science Specialization | Chicago, IL** **May 2016 - Jul 2016**

*Next Word Predictor App* ([https://madhuri8933.shinyapps.io/Next\\_Word\\_Predictor/](https://madhuri8933.shinyapps.io/Next_Word_Predictor/))

- Tokenized and cleaned (profanity filtering, stemming) unstructured text data from news, blogs, twitter.
- Generated document feature matrix, n-gram distributions to predict next probable word for a user input phrase.