

Dheeraj Singh

🏠 dheeraj2444.github.io/ ✉ dhsingh@iu.edu ☎ +1 (812) 361-7212 🔊 [Dheeraj2444](#) 📺 [dheeraj2444](#)

EDUCATION

Indiana University, Bloomington, IN

Master of Science, Data Science

August 2017 - May 2019

GPA: 3.9/4.0

Key skills: Machine learning, Deep learning, Algorithms, Statistics, Data mining, Analytics, Data cleaning/Data processing

Indian Institute of Technology (IIT), Kharagpur, India

Bachelor of Technology (Honors), Engineering

August 2009 - May 2013

GPA: 7.1/10.0

TECHNICAL SKILLS

- **Programming:** Python, R, SQL, Cython, MATLAB, Bash, Linux
- **Libraries:** NumPy, Pandas, Scikit-Learn, SciPy, PyTorch, Keras, TensorFlow, NLTK, spaCy, dplyr, Spark, Git, Vim
- **Data Visualization:** Matplotlib, Seaborn, ggplot, Jupyter Notebook
- **Databases:** PostgreSQL, MySQL, MongoDB, Neo4j
- **Machine Learning:** Regression, Classification, Clustering, Dimensionality Reduction, Ensemble Methods, PCA, NLP, HMM, Recommender System, Computer Vision, Statistical Learning, Predictive Modeling, Math, Linear Algebra

INDUSTRY EXPERIENCE

Altair Engineering Inc., Troy, Michigan

May 2018 - August 2018

Machine Learning Engineer Intern

- Developed a 3D shape recognition system using voxelization and 3D Convolutional Neural Nets (CNN) resulting in 87% accuracy on the Princeton ModelNet10 (CAD models) dataset.
- Built a multi-layered neural network to predict failure of an Air Pressure System (APS) for Scania Trucks in order to minimize maintenance cost (Industrial Challenge for IDA, 2016); outperforming 2nd & 3rd ranked teams.
- Built a system by training deep neural nets on simulation datasets to predict reduction in mass for a given geometry and load condition in order to achieve an optimized structure.

Department of Computer Science, IIT Kanpur, India

April 2016 - May 2017

Senior Project Associate

- Developed a software system to identify vehicle license plates using template matching. Formulated rules to build an OCR for character identification based on the pixel arrangement, resulting in 83% accuracy.
- Employed Bootstrap framework, HTML, CSS, and PHP to design a web application for real-time management & visualization of data stored in MySQL. Defined the database schema, configured & deployed in phpMyAdmin.

Tinyowl Technologies, Mumbai, India

May 2015 - February 2016

Senior Business Analyst

- Implemented logistic regression to predict probability of users from different clusters returning back to the platform.
- Performed k-means clustering to segment users based on attributes such as spending behavior, ordering pattern.
- Built a system to collate and quantify user sentiments on Twitter, thereby improving customer service & engagement.
- Developed an internal dashboard to track and report pre-defined business metrics and trends using *shiny* package in R.

Ipsos Research, Bangalore, India

June 2013 - December 2014

Analyst

- Performed Market Mix Modeling & Pricing Analysis for client specific marketing strategies in a variety of domains.
- Performed multivariate regression analysis to quantify ROI from marketing expenditures and provide analytical insights.
- Coordinated on a daily basis with a global front desk team managing clients at New York and gained experience of the entire project life-cycle from conceptualization to deck delivery on client-side by working on entirety of a project.

DATA SCIENCE GRADUATE PROJECTS

Deep Learning Specialization [Artificial Intelligence (AI), Neural Networks, RNN, CNN, LSTM]

- Implemented forward & backward propagation in NumPy. Tweaked hyper parameters, optimization techniques, initializations, regularization techniques. Experimented with different deep learning architectures.

CIFAR-10 Classification [Deep Learning, PyTorch, Transfer Learning, Artificial Intelligence (AI)]

- Trained different architecture (AlexNet, VGGNet, ResNet) for image classification to benchmark on CIFAR-10 using transfer learning and fine tuning in PyTorch. AlexNet-78%, VGGNet-85%, ResNet-83% on 10K test images.

Spark Streaming [DStreams, RDD, MapReduce, Twitter API, PySpark, Tweepy]

- Developed a live dashboard to visualize trending hashtags for a given topic using Twitter API, PySpark, and Tweepy.

Sequence Classification [Natural Language Processing, Word Embeddings, Multi-label Classification]

- Built a stacked bidirectional LSTM Network and trained on a TREC question dataset (5952 sentences) to classify a given question into one of the six possible question types.