Dheeraj Singh

dheeraj2444.github.io/ | dhsingh@iu.edu | +1 (812) 361-7212 | Dheeraj2444 | dheeraj2444

EDUCATION

Indiana University, Bloomington, IN — August 2017 - May 2019
Masters of Science, Data Science — GPA: 3.8/4.0
Key skills: Machine learning, Deep learning, Algorithms, Statistics, Data mining, Analytics, Data cleaning/Data processing

Indian Institute of Technology (IIT), Kharagpur, India — August 2009 - May 2013
Bachelor of Technology (Honors), Engineering — GPA: 7.1/10.0

deeplearning.ai — April 2018 - Present
Deep Learning Specialization

TECHNICAL SKILLS

- **Programming/Scripting**: Python, R, SQL, Cython, MATLAB, Bash, Linux
- **Software**: 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
- **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

Machine Learning Engineer Intern | Altair Engineering Inc., Troy, Michigan — May 2018 - August 2018

- Developed a 3D shape recognition system using voxelization and 3D Convolutional Neural Nets (CNN). Trained the network on Princeton ModelNet10 (CAD models) dataset and achieved 87% accuracy.
- Designed a model to predict the failure of an Air Pressure System (APS) for Scania Trucks in order to minimize maintenance cost (Industrial Challenge for IDA, 2016). Proposed solution outperformed 2nd & 3rd ranked teams.
- Trained deep neural nets on simulation dataset to predict a reduction in mass for a given geometry & load condition in order to achieve an optimized structure and achieved a significant reduction in required manual labor.

Senior Project Associate | Department of Computer Science, IIT Kanpur, India — April 2016 - May 2017

- Developed a software system to identify vehicle license plates using template matching. Formulated rules to build an OCR for character identification. Improvised prediction accuracy by 7% on 5000 real-world images.
- 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.

Senior Business Analyst | Tinyowl Technologies, Mumbai, India — May 2015 - February 2016

- Developed a model to predict the probability of users from different clusters returning back to the platform for targeted & channelized marketing and reduced marketing expenditure by 11%.
- Segmented users by k-means clustering using attributes like spending behavior & order pattern for targeted marketing.
- Enhanced customer service & engagement by building a system to collate and quantify user sentiments on Twitter.
- Developed an internal dashboard to track and report pre-defined business metrics and trends using *shiny* package in R.

Analyst | Ipsos Research, Bangalore, India — June 2013 - December 2014

- Performed marketing mix analysis to measure the impact of all business drivers for fortune 500 clients to improve their business & marketing performance through integrated analytics and actionable insights.
- Analyzed marketing ROI to highlight over-allocated & inefficient investments for efficient marketing budget allocation.
- Collaborated and led cross-functional teams across geographies to deliver client deliverables.

DATA SCIENCE GRADUATE PROJECTS

Text Independent Speaker Verification System *[Signal Processing, STFT, Siamese Network, PyTorch]*

- Trained VGGVox model based Siamese network on VoxCeleb dataset (100K utterances for 1,251 celebrities) on AWS. Achieved 0.78 precision & 0.84 recall. Developed a terminal application for identification and verification in real time.

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. AlexNet-78%, VGGNet-85%, ResNet-83% on 10K test images.

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

- 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 and achieved 81% accuracy.

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.