Omkar Acharya

1713 Crest Road, Raleigh NC 27606 | 443-917-8052 | oachary@ncsu.edu | linkedin.com/in/acharyaomkar01 | github.com/omkaracharya

EDUCATION

North Carolina State University | *Raleigh, NC* *Aug 2016 – Dec 2017 (Expected)*

Master of Computer Science (Specialization in Data Science) *GPA: 4.00/4.00*

- *Coursework: Foundations of Data Science, Algorithms for Data Guided Business Intelligence, Automated Learning & Data Analysis, Graph Theory, Artificial Intelligence, Databases, Design & Analysis of Algorithms, Computer Networks*

Pune Institute of Computer Technology (University of Pune) | *Pune, India* *Aug 2012 – Jun 2016*

Bachelor of Engineering (Computer Engineering) *GPA: 3.78/4.00*

- *Coursework: Data Mining, Natural Language Processing, Discrete Structures, Data Structures, Software Engineering*

EXPERIENCE

Quantworks Inc. | *Graduate Intern - Backend* **|** *Raleigh, NC* *May 2017 – Present*

- Building a smart assistant abstractor for the extraction of structured data from medical records in REDCap database
- Shadowing the human abstractors to understand the workflow and get the feedback on the smart assistant's work
- Writing the Adobe Acrobat VC++ plugins for Optical Character Recognition (OCR), PDF highlighting and bookmarking
- Automating the abstraction process to bring down the PDF processing time from 3-4 hours to 15-20 minutes
- Performance improvement, bug fixing of a live, production Operation Research system for a Fortune 500 company

Iknowlation Research Lab | *Machine Learning Intern* **|** *Pune, India* *Aug 2015 – April 2016*

- Developed a system to describe the video content in English using Convolutional and Recurrent Neural Networks
- Finetuned two of the Nvidia's CNN caffemodels (Google-Net and VGG-Net) on ImageNet dataset of 1.2M images
- Trained Long Short-Term Memory (LSTM) models on MSCOCO dataset of 120K images with Chainer framework
- Evaluated the performance of the model and finetuned the hyperparameters to improve its accuracy from 63% to 77% based on the existing conventional systems
- Won the 3rd prize for paper presentation (out of 35) at Birla Institute of Technology, Goa

PROJECTS

Business Intelligence Project - Yelp Restaurant Photo Classification *(Python, Scikit-Learn, Caffe, H5Py, Pandas, Numpy)*

- Implemented a three-stage transfer learning pipeline with Convolutional Neural Nets and Support Vector Machine
- Extracted features from with 234,842 training images and 1,190,225 testing images using BAIR Reference Caffemodel
- Trained a Support Vector Machine (SVM) One vs All classifier using these features for multilabel learning
- Ranked 134th out of 355 in the Kaggle competition with an F1 score of 0.764 on testing data (0.797 on validation)

Machine Learning for Images - GMM + RBC for Settlement Mapping *(Python, R, Weka, QGIS, Scikit-Learn, Numpy)*

- Developed an unsupervised learning system of exploring the settlement activities for planning and disaster recovery
- Implemented Gaussian Mixture Models (using Expectation Maximization algorithm) and 1-Holt Rule Based classifier
- Achieved ~85% accuracy on the validation set containing the different patches from a satellite image of 176MB
- Evaluated its performance with Weka's RIPPER (Rule Based Classifier) and Scikit-Learn's in built GMM model
- Designed and presented a poster along with a paper of the project workflow and its implementation details

Machine Learning for Sentiment Analysis *(Python, Apache Spark, Kafka, Scikit-Learn, Matplotlib, Doc2Vec, Pycharm)*

- Built a sentiment analysis mini-project using the database containing the real-time tweets and IMDB reviews
- Approach 1: Logistic Regression with important words as features (Accuracy: Twitter 60.3%, IMDB 84.8%)
- Approach 2: Doc2Vec with Artificial Neural Nets and document vectors (Accuracy: Twitter 63.56%, IMDB 83.49%)
- Used Twitter's Streaming API to fetch and analyze the tweets at runtime and stored them using Apache Kafka buffer

Database Application - Personal Health Manager *(Java, Oracle, SQL, JDBC, IntelliJ)*

- Developed a system to assist people to manage the health information, recommendations and emergency alerts
- Used Java Database Connectivity (JDBC) to interfacing with the Oracle SQL 11g database containing the tables for patients, health history, their health supporters, and diseases
- Implemented multiple triggers in Oracle to notify the patient and his/her health supporters about the unusual event