CHAITRA VISHWANATHA HEGDE

cvh255@nyu.edu | https://www.linkedin.com/in/hegdechaitra/ | https://github.com/HegdeChaitra
551-689-6433 | 212 Beacon Avenue, Jersey City, New Jersey - 07306

EDUCATION

MASTER OF SCIENCE, DATA SCIENCE - GPA: 3.819/4.0 **May 2019 (Expected)**

New York University, Center for Data Science **New York, USA**

- Relevant Course Work: **Deep Learning** for Medicine, **Machine Learning**, Big Data, Probability and Statistics for Data Science, **Natural Language Processing, Natural Language Understanding, Contextual Cognitive Modeling**
- Received **Moore-Sloan research grant** to conduct research on image segmentation task using Deep Learning
- Former Marketing and Student relation representative at NYU CDS Leadership Circle

BACHELOR OF ENGINEERING, COMPUTER SCIENCE - GPA: 3.8/4.0 **2013 - 2017**

Visvesvaraya Technological University **Karnataka, India**

- Relevant Course Work: Algorithms and Data Structure, Object-Oriented Programming in C++, Graph Theory, Operating Systems

EXPERIENCE

Data Science Intern **Sep '18 - Present**

Bombora Inc, NYC, USA

Keywords : NLP, Deep Learning, Classification, Convolutions, GRUs, Hierarchical Attentions

- Built a convolution-based model to classify **English text** documents as one of the **5800 classes**. The model achieved an accuracy of 73.5% on balanced validation set while the model in use was giving 62.2% accuracy. Hence, 11% accuracy gain making a significant difference.
- Built a **attention network** for the same English text classification task. Currently, analyzed the attentions of words for given document and utilized it for other subtasks.
- Implemented the Facebook's Fasttext model (bag-of-n-gram with TF-IDF) in Pytorch for an **NLP** task of classifying **French and English documents**.

Research Intern [link to repo] [Abstract] **June '18 - Aug '18**

NYU Langone Health, NYC, USA

Keywords : Computer Vision, Deep Learning, Health Care, Segmentation, Encoder-Decoder, Convolutions

- Presented at **ISMRM** workshop held at Washington DC, a well-known workshop among MRI community
- Worked as an AI research intern on segmenting 200 brain tissues in the **MRI** images to replace time-consuming tools like the Freesurfer
- Used **Deep Learning** models like - coordinated U-net with dice loss + cross entropy loss with a various weighing mechanism to make the model achieve the best performance of (Dice Score = 0.94) on the very imbalanced dataset.

Junior Data Scientist [link to repo] **Jan '18 - Aug '18**

NYU Center for Data Science, NYC, USA

Keywords : Regression, Machine Learning, Random Forrest

- Predicted geo-temporal variables by studying massive amount of highly sparse and unorganized datasets from NYC Open Data
- Enhanced model's performance by adding novel taxi-demand relevant features from multiple open source datasets, such as- Foursquare venue dataset, Foursquare user check-in dataset and NYC Taxi dataset. Identified demand drivers using feature importance maps.
- Random Forest algorithm was used for modeling and various evaluation metrics like R^2 score, Mean Absolute Error and Mean Square Error were used for comparing multiple models. GeoPandas and Matplotlib were used for visualizing the results.

Developer Intern **Jan '16 - Mar '16**

Grinions.com, Bangalore, India

- Worked on a user interactive Web Application for collecting contributions for social events from attendees
- Wrote Well organized codes using Python's Flask framework and MySQL to handle huge traffic on the website
- Built an attractive user interface for the website using Google's Material Design Lite, HTML-5, CSS and JavaScript

TECHNICAL SKILLS

Programming Languages :	Python, C/C++, SQL, HTML, CSS
Machine Learning Libraries :	PyTorch, PySpark, Sci-kit learn, Pandas, Matplotlib, Numpy, Seaborn, GeoPandas, TorchText
Cloud and Facilities :	AWS, HPC, CUDA, UNIX, GitHub, Google Cloud

RELEVANT DATA SCIENCE PROJECTS

Neural Machine Translation Engine [link to repo] **Sep - Dec' 18**

Keywords : NLP, Machine Translation, Sequence-Sequence Models, GRUs, Attentions, Encode-Decoder

- Built two language translation engines. One translates **Vietnamese to English** and other translates **Chinese to English**, using Auto-Encoder-Decoder type model with **Attention**.

- Experimented and reported performance on models with different types of encoder/decoder, such as : GRUs, Convolution and Transformer style encoder and decoders

---

Knee Cartilage Segmentation Using Deep Learning [link to repo] **Feb - May '18**

Keywords : Computer Vision, Deep Learning, Health Care, Segmentation, Encoder-Decoder, Convolutions

- Performed highly imbalanced multi-class segmentation of knee cartilages in MRIs using novel Deep Learning models like U-Net, V-Net and ensemble of similar such models. Final model achieved state-of-the-art results and surpassed a human radiologist level
- Performed Perturbation analysis to understand important features of MRI. Built confidence maps that showed the model's confidence of the voxel-level predictions.

NLP Mini Projects **Sep - Nov'18**

- **Sentiment Analysis on IMDB review** [link to repo] : Built a bag-of-n-gram model to classify the given IMDB review as of either a positive or a negative sentiment
- **SNLI inference Problem** [link to repo] : Built a Convolution-based model and a GRU-based model to find-out whether the given pair of sentences entail, contradict or neutral to each other

---

Dataset Search Engine [link to repo] **Feb - May '18**

Keywords : BigData, SQL, PySpark

- Built a sophisticated search engine to search for datasets in NYC Open Data portal that contains approximately 1500 datasets (~650 GB)
- Developed 11 advanced search functionalities, currently are not supported by NYC Open Data website, to help the user find relevant datasets
- Used PySpark to parallelize, hence speed up the execution of code. Created data summaries to quickly speed up the search process

---

Study on Factors influencing NYC Housing Prices [link to repo] **Apr - Apr '18**

Keywords : Regression, Machine Learning, Random Forrest, Neural Networks

- Runner-up of "Enigma Datathon", conducted at NYU Center for Data Science
- Extracted and Processed 5 NYC open datasets to characterize a location in NYC in term of approachability, safety, standard of living, etc.
- Combined these location relevant features with the housing price dataset to enhance house price prediction model's performance and understand factors driving the house prices in NYC

Bank Marketing Campaign Analysis [link to repo] **Sep - Dec '17**

Keywords : Regression, Machine Learning, Tree-based Models

- Analyzed the prior marketing campaigns of a Portuguese Bank using various ML techniques like Logistic Regression, Random Forests, Decision Trees, Gradient Boosting and AdaBoost and predicted if the user will buy the Bank's term deposit or not
- **Recommended**, the marketing team, ways to better target customers using feature importance maps and business intuition

---

Yelp Recommendation Engine [link to repo] **Sep - Dec '17**

Keywords : Machine Learning, Matrix Factorization, Collaborative Filtering

- Winner of "Best First Year Project" award at NYU CDS Academy Awards
- Built a recommendation engine for recommending restaurants to Yelp users using traditional models like the Cosine similarity-based model, SVD and Alternating Least Square model; Rating Matrix was very sparse with the sparsity of 99.4%
