# SIDDHANTH PILLAY

siddhanthpillay138@gmail.com +1 412-636-6206 linkedin.com/in/sidpillay siddhanthp27.github.io

## EDUCATION

**Carnegie Mellon University - School Of Computer Science** Pittsburgh, PA
Masters in Computational Data Science **(GPA 4.0)** *Dec 2020*
*Graduate Teaching Assistant - Data Science Capstone Seminar, Foundations of Data Science*
*Relevant Courses: Reinforcement Learning, Multimodal Machine Learning, Neural Networks for NLP, Deep Learning, ML for Large Datasets, Machine Learning, Cloud Computing, Computer Systems*

**National Institute of Technology Karnataka** Surathkal, India
Bachelor of Technology in Information Technology **(GPA 9.1)** *May 2019*

## SKILLS

- **Programming Languages**: Python, Java, C, C++, Scala
- **Tools**: Pytorch, Tensorflow, Scikit Learn, NLTK, Numpy, Pandas, Git, MYSQL, HBase
- **BigDataTech/CloudPlatforms**: MapReduce, Spark, Amazon Web Services, Microsoft Azure, Google Cloud Platform, Docker, MongoDB, Kubernetes, Spark

## EXPERIENCE

**Shure Incorporated** Niles, IL
*Software Engineer II, Data Science* *Feb 2021 - Present*
- Developing LipReading models using Tensorflow to improve competitiveness of Shure products in the market
- Developing Neural Network models and robust pipelines to deploy on Edge Audio devices to improve performance

*Data Science Intern (Remote)* *May 2020 - Aug 2020*
- Developed end-to-end Image Segmentation Pipeline in PyTorch and implemented related deployment pipeline
- Deployed Neural Network based signal processing models on x86 processor using Tensorflow Serving and Docker

**Indian Institute of Technology Bombay** Mumbai, India
*Research Intern* *May 2018 - Dec 2018*
- Developed a Patch-based sliding window Neural Network model for biological image segmentation which achieved an accuracy of 87%
- Built pipelines to normalize input images, train, and apply models applications like identifying Cancer Cells in Blood Samples, Malaria Cells in Blood Samples, Tumours in Mammograms, and Blood Cell Classification in Blood Images

## SELECT PROJECTS

**Deep Adaptive Clustering for Cellular Tomography Images** Capstone, *CMU* | *Spring 2020*
- Developed a Semi-Supervised Deep Learning framework in PyTorch to categorize Cell Tomogram images which saves manual effort of annotating data
- Built a Deep CNN Model with a K-Means Regularizer that beat existing unsupervised/semi-supervised models to achieve state-of-the-art accuracy of 95.28%

**Focal Visual Text Attention for Visual Question Answering** Intro to Deep Learning, CMU | *Spring 2020*
- Developed an end-to-end attention-based hierarchical approach based model in PyTorch to process multimodal data and answer question based on input image and metadata

**Attention-based End-to-End Speech-to-Text Deep Neural Network** Intro to Deep Learning, CMU | *Spring 2020*
- Developed a Listener-Speller framework in PyTorch to transcribe input speech into text using Pyramidal Bi-LSTM Networks and Attention Mechanism to achieve Levenshtein mean distance of 12.26

**Popularity Prediction for Reddit Comments** Machine Learning for Large Datasets, *CMU* | *Spring 2020*
- Built an end-to-end pipeline using Azure, Databricks and PySpark/MLLib to explore the usefulness of NLP features from comment body on 55GB Reddit data using Logistic Regression, Random Forest, LinearSVC, Linear Regression

**Cloud Fare Prediction Service** Cloud Computing, *CMU* | *Fall 2019*
- Implemented an end-to-end Fare Prediction Service with custom-built features hosted on Google App Engine using services from Google Cloud Platform

## RESEARCH PAPERS

**Cross Lingual Morphological Inflection for Low Resource languages**
17th ACL SIGMORPHON Workshop on Computational Research in Phonetics, Phonology and Morphology
- Built four attention-based encoder-decoder models in PyTorch using LSTMs and Transformers for ACL SIGMORPHON 2020 shared task
- Analysed the effects of incorporating different techniques such as Hallucination, Language Vector Injection, Sparsemax Loss and Adversarial Language Training