Arnav **Arnav**

Data Scientist · Data Engineer · Software Engineer

530 Alfani Street, Davenport, FL, 33896, USA

(+1) 812-955-9112 | arnav.arnav@alumni.iu.edu | seashiva94.github.io | seashiva94 | seashiva94

Education

Indiana University — *Bloomington, IN*

M.S. in Data Science — *Aug. 2017 - May. 2017*

- Courses: Data Mining, Machine Learning, Computer Vision, Machine Learning for Signal Processing, Learning Theory and Graphical Models, Deep Learning, Big Data Applications

Tezpur University — *Tezpur, Assam, Inia*

B.Tech. in Computer Science and Engineering — *Aug. 2012 - Jul. 2016*

- Relevant Courses: Data Structures, Algorithms, Computational Geometry, OS, DBMS, Compiler Design, NLP, Pattern Recognition

Skills

Programming Languages	Python, R, Scala, C++, SAS, SQL, Javascript, Ruby, Bash
Database and Big Data Technologies	Hadoop, Apache Spark, MongoDB, PostgresQL, Neo4j, Teradata, Hive, Kafka, Dask
Web and Appliation Development	Flask, Django, Ruby on Rails, Swagger Codegen, HTML, CSS, Ember.js
Machine Learning	Scikit-learn, Scipy, Numpy, Pandas, NLTK, SpaCy, CoreNLP, SparkML, Torch, Tensorflow
DevOps	Pyvbox, Openstack API, Boto3, Docker CE, Kubernetes, Ganglia, MQTT, Travis-CI, Git
Data Visualization	Tableau, Matplotlib, Seaborn, Altair-viz, Plotly

Experience

Franchise Data Solutions, LLC. — *Orlando, FL*

Software Engineer — *Apr. 2020 - Jul. 2020*

- Added new functionality to the company's organization, resource and operations management web application platform, using Flask, PostgreSQL, SQLAlchemy, HTML, CSS and Javascript
- Automated data pulls for the platform and integrated custom reporting functionality on the platform based on the data pulled daily from various resources.

The Walt Disney Company — *Orlando, FL*

Decision Science Professional Intern — *Aug. 2019 - Apr. 2020*

- Implemented a state space post processing approach to dynamically adjust product level demand forecasts to match actual behavior based on newly observed data points using Python and SAS that improved 70% of Disney Cruise Line booking forecasts with a mean improvement of 35% in absolute error
- Implemented data preparation, feature engineering, and output monitoring and built models for daily attendance forecasting using Pyspark, Teradata, Tableau and scikit-learn
- Added pyspark and spark-ml support to an internal framework that helps standardize and maintain production ML pipelines across different projects using different machine learning frameworks

Indiana University: Luddy School of Informatics Computing and Engineering — *Boomington, IN*

Associate Instructor — *Aug. 2018 - May. 2019*

- Evaluated student submissions and projects in various modules that cover, Hadoop, Spark, Scala, and Deep learning
- Created course content and setup tutorials for Machine Learning with Spark module and updated content for Hadoop, Scala and Deep Learning Modules

Indiana University: CNS — *Bloomington, IN*

Python Developer — *Oct. 2017 - Dec. 2016*

- Contributed to the development of cloudmesh.pi, a library for prototyping IoT applications on Raspberry pi
- Interfaced different analog and digital sensors and built modules for each of the sensors, that allow plug and use functionality
- Programmed an array of dendrites to move in various patterns controlled by different Raspberry Pis communicating via MQTT
- Link to code: github.com/seashiva94/cloudmesh.pi

Navyug Infosolutions

Noida, UP, India

Software Engineer Intern — *Oct. 2016 - Jul. 2017*

- Developed and deployed an interactive and responsive internal project management (Gofer) web application used by 200 people using Ruby on Rails, Ember.js and JQuery-UI
- Integrated the application with Google APIs to provide Project Project Managers access to auto generated time sheets used for downstream cost analyses and fund allocation
- Added new functionality for a mobile application for an education client to allow instructors to gauge student performance after each class through short quizzes
- Developed Android and iOS application for the web application using Apache Cordova

NTPC NETRA Advanced Computing Lab

Greater Noida, UP, India

Software Engineer Intern — *May. 2015 - Jul. 2015*

- Studied the various components of the system and the specific configuration used at the site to enable computationally intensive tasks. Used Ganglia to monitor the system and detect anomalies using data logs
- Implemented various machine learning algorithms like anomaly detection, regression and classification using previously collected power plant data in Python

Projects

Deep Gaussian Processes for Representation Learning

Indiana University, Bloomington

Learning Theory and Graphical Models — *Jan. 2019 - May 2019*

- Implemented a hierarchical Gaussian Process Latent Variable model for representation learning for supervised and unsupervised tasks using GPFlow and Python
- Tested the performance of the learned representations on image reconstruction and classification tasks on oil flow, MNIST handwritten characters and Frey faces data sets. Link to code: github.com/seashiva94/DeepGPLVM

Speaker Identification and Verification from Audio

Indiana University, Bloomington

Machine Learning for Signal Processing — *Aug. 2018 - Dec. 2018*

- Trained a Siamese neural network based on VGGVox model on the VoxCeleb dataset using Python, PyTorch on AWS and achieved 0.78 precision and 0.84 recall on the data
- Developed a terminal application using python and shell for speaker identification and verification
- Link to code: github.com/seashiva94/speaker-identification

Open Domain Information Extraction

Indiana University, Bloomington

NLP Lab — *Jul. 2018 - Dec. 2018*

- Extract object-predicate relationships from text using various NLP modules and store them in Neo4J to enable semantic search, through a Flask based web page
- Linked the extracted entities to various existing knowledge graphs like DBpedia, and enriched the knowledge graph by adding information from various reliable sources such as Concept Net ans MS Concept Graph
- Created audio recordings and annotated words, parts of speech and prosody elements in English using Praat for augmenting existing databases to be used for Speech Pragmatics research

Swagger Flask service for Openstack Cluster Management

Indiana University, Bloomington

Advanced Cloud Computing — *Jan. 2018 - May. 2018*

- Wrote swagger YML specification for openstack (Devstack) virtual machine instances instances
- Used Swagger-codegen and Flask to create REST API endpoints for openstack services (start, stop, delete, create, and list VMs) and potentially provide an additional layer for user privilege and access management
- Used shell scripts and Makefile to build and test the APIs and deploy the application with docker on chameleon cloud.Link to code: github.com/seashiva94/hid-sp18-503

Raspberry-Pi Remote Monitoring Application

Indiana University, Bloomington

Big Data Applications and Analytics — *Oct. 2017 - Dec. 2017*

- Used lightweight MQTT protocol to remotely control a Raspberry-Pi robot car over WiFi to aid monitoring and surveying applications and stream live video frames from the Raspberry Pi on-board camera to the controlling machine, for remote navigation using Python
- Used shell scripts along with make for setup of the MQTT server, the controller and the raspberry-pi remote application. Link to code: github.com/seashiva94/hid201

American Sign Language Recogntion from Video

Tezpur University, Tezpur, Assam, India

Final Year Project — *May. 2016 - Aug. 2016*

- Built a desktop based application to recognize American Sign Language gestures from video and convert them to text sentences using Python and WxWidgets
- Used OpenCV extract features from videos of of signers in the American Sign Language Lexicon Video Dataset(ASLLVD) maintained by Boston University, and trained a deep neural network classifier for recognizing different gestures
- Achieved 85% accuracy in individual word detection and 60% accuracy in translating sentences

Inventory Management Application for Local Business

Tezpur University, Tezpur, Assam, India

Database Management Systems

Aug. 2015 - May. 2016

- Built an Inventory management application for a local business using Django and PostgreSQL that allowed inventory tracking and employee schedules
- Deployed the application on an AWS instance and trained the sales employees on using the application, leading to increased throughput, and reduction in monthly inventory processing times

Awards

2017 **Data Science Fellowship**, Indiana University Bloomington *MS, Data Science*