

Sudhir Nallam

Data Scientist

📍 | 444 Saratoga Ave, Santa Clara, CA, 95050
☎ | 312-282-2524
in | <http://www.linkedin.com/in/sudhir-nallam>
✉ | sudhir.nitw@gmail.com

SOFTWARE SKILLS

LANGUAGES	Python, R, Lua, C++, PL/SQL, Octave, Java
PYTHON PACKAGES	pandas, scikit-Learn, numpy, scipy, matplotlib, pymongo
TOOLS & TECHNOLOGIES	Torch, Tensorflow, Hadoop(ecosystem), AWS stack, Airflow, Spark, SPSS, Elasticsearch, AlchemyAPI, Spring, Struts, Hibernate
DATABASES	Mongo DB, DB2, Cognos TM1
WEB TECHNOLOGIES	HTML, CSS, AJAX, Restful services, XML, JSON, Javascript
VISULIZATION	D3, Cognos BI, Tableau
OPERATING SYSTEMS	Linux- RHEL, Ubuntu; Windows

EDUCATION

2015 – 2017	Data Science MASTER OF SCIENCE New York University, New York Courses Taken: Statistics, Machine Learning, Big Data, NLP, PGM, Deep learning
2001 – 2005	Mechanical Engineer BACHELOR OF TECHNOLOGY National Institute of Technology, Warangal, India

WORK EXPERIENCE

	MARCH 2019 – PRESENT
Roku, Los Gatos, CA <i>Data Scientist</i>	
Working in Roku Voice NLU team. Exploring the data corpus using statistical and ML models to create actionable insights. Evaluating the models by defining KPI's and conducting AB tests. Software Stack : Python, Pandas, Hive, ElasticSearch, AWS Stack, Airflow, Tensorflow, scikit-learn, R	
	SEPT 2014 – MAR 2019
Client: IBM, T.J. Watson Research Center, NY (Employer: ProMatrix Corp) <i>Data Scientist</i>	
Developed a predictive analytics tool to help IBM MA team in procuring new acquisitions. Contributed in feature extraction, optimizing hyperparameters and fitting ML models for acquisitions data. Deployed analytic models in production setup. Worked on analyzing the public response on IBM sponsored events by doing sentimental analysis on twitter and blogs data. Software Stack : Java, Python, PL/SQL, SPSS, DB2, Spark, IBM Bluemix Services, Mongo DB, Elasticsearch	
	MARCH 2010 – SEPT 2014
IBM, T.J. Watson Research Center, NY <i>Software Engineer</i>	
Work closely with client teams to identify requirements, estimate projects, and implement prototypes and solutions in an agile, iterative fashion. Analytics (Data Mining) model developed in SPSS is integrated into the system through Python and Clementine Scripting. Designed and developed MAPro (Performance Risk Optimizer), a web application with integrated cognos reports. Developed low latency application through multi-threading and in-memory cache. Software Stack : Java, Python, PL/SQL, SPSS, DB2, Hadoop, Mongo DB, Elasticsearch	
	JULY 2009 - MARCH 2010
Client: Navy Federal Credit Union, Pittsburgh, PA (Employer: Egen Solutions) <i>Software Engineer</i>	

Migrated lotus notes based Reversal application to web based J2EE application in websphere. Connected to mainframe programs (wrapped by Host Bridge) through web service clients using restful webservices.

Software Stack :Struts 2.0, Hibernate 3.0, Restful Webservices, DB2

JANUARY 2008 - JULY 2009

Client: FedEx Ground, Pittsburgh, PA (Employer: Egen Solutions)

Software Engineer

Contributed in increasing the project performance (4 million requests/day). Developed service layer programs and parsers which have optimized time complexity.

Software Stack :Spring 2.5, Hibernate3.2, JSF 1.2, Maven 2.0

GRADUATE PROJECTS

Augmented RNN for Jet Physics

Capstone Project

we have augmented the RNN network for jet classification to handle complex interactions between particles and system uncertainties using adversarial networks.

Software Stack : Pytorch, Python, scikit-learn

Code: <https://github.com/NYU-CDS-Capstone-Project/Voyagers/tree/master/code/jets>

Limitations of Generative Models

Inference & Representation

we have taken two generative models, VAEs and GANs, and understand their power and limitations in approximating various data density estimations. We also studied their data modelling capabilities in various noisy conditions.

Software Stack : Tensorflow, Torch, Lua, python

Code: <https://github.com/sudhirNallam/IRClass.git>

Stacked What-Where Auto-encoders

Deep Learning

Implemented stacked What-Where Auto-encoders to classify MNIST data in unsupervised setting.

Software Stack : Pytorch, Python

Code: <https://github.com/sudhirNallam/SWWAE.git>

Machine Translation with sequence-to sequence model

Deep Learning

Pytorch implementation of Sequence-to-Sequence Learning with Attentional Neural Networks.

Software Stack : Pytorch, Python

Code: <https://github.com/sudhirNallam/seq2seqModel.git>

Predictive Models to Determine Judge Bias in Asylum cases

Machine learning

we developed a predictive model for classifying whether or not a refugee is granted asylum in the United States, and to use that model to determine which features bias judges the most.

Software Stack : Python, Scikit-Learn, R

Code: <https://github.com/sudhirNallam/predictingRefugeeAsylum.git>

Understanding the Complex Interactions in NYC Taxi data and weather data

Big data

We combined NYC taxi data with weather data. From the data we have inferred the correlation that exists between Tip variations with weather conditions and Group riding with weather conditions.

Software Stack : Python, Scikit-Learn, Hadoop Ecosystem, Mongo DB

Code: <https://github.com/sudhirNallam/BigData-Project.git>

AWARDS

2012 **Outstanding Technical Achievement Award**

IBM T.J. Watson Research center