

Swarnashree Mysore Sathyendra

Webpage: swarnashree.github.io

Email/Application User ID: msswarnashree@gmail.com

Research Interests

Natural Language Processing, Information Retrieval, Computer Vision, Machine Learning, Multimodal Machine Learning

Education

2014-2018 BE in Information Science and Engineering (ISE) - [PES Institute of Technology, Bangalore, India \(PESIT\)](#)
Department Rank: 1 CGPA - 9.56/10.00

Publications

Real-Time Headgear Detection in Videos using Deep Learning based Feature Extraction with a Supervised Classifier

Swarnashree Mysore Sathyendra*, Rajdeep Pal*, Ranjana Seshadri*, S. Natarajan (* – equal contribution)

24th International Conference on Advanced Computing and Communications (ADCOM 2018), IIIT-Bangalore, 2018

Scholastic Achievements

- Presented a white paper on “**Real-time Text-Search on Encrypted Data**” in association with Goldman Sachs at **Grace Hopper Celebration India 2019**.
- Received the **gold medal** for securing **Department Rank 1**, PESIT 2014-2018.
- Received **C.N.R Rao Scholarship** and **M.R.D Scholarship** with a combined **25%** tuition fee waiver for all 4 years of undergraduate studies (2014-2018).
- Secured **Karnataka State Rank 7** out of approximately 200,000 students in grade 12 examination.
- Awarded for **Academic Excellence** by Carmel School (2012) and Jain University (2014).

Research Experience

Smart FAQ Chatbot for Legal Compliance Policies

Mentor: [Mrunal Shah](#)

Aug 2018 - Feb 2019

Goldman Sachs

- Developed a chatbot to answer questions based on compliance policies and documents.
- Hunspell for efficient autocorrect, state-of-the-art [ConceptNet based semantic parser](#) for concept extraction from documents, **Singular Value Decomposition** to reduce the document-term matrices into lower-dimensional document-concept matrices (reducing the feature vector size by **96%**).
- Devised word-jumbling techniques to reduce the dependency of the algorithm on the ordering of terms.
- The system is currently being used at Goldman Sachs to aid Compliance officers to find answers related to compliance.

Real-time Person Detection in Videos based on Natural Language Description

Advisors: [Prof. S. Natarajan](#), [Prof. Antony L Piriya Kumar Douglas](#)

Jan 2018 - Aug 2018

Bachelor Thesis

Built a complete **multimodal machine learning system** in collaboration with a start up called AllGoVision Technologies Pvt Ltd. The system detects persons in surveillance videos, in real-time, given a natural language description. This project led to a publication in **ADCOM 2018**.

- Dataset acquisition and annotation:** Collected unlabeled data by scraping the web and semi-manually annotated using **K-means** clustering. Derived a dataset for the target coarse-grained classes of head, upper body and lower body by combining masks of appropriate finer-grained classes of the famous **PASCAL-Part VOC 2010** dataset.
- Modelling:** Integrated several modules in the sequence - a simple **bag-of-words** model to extract visual attributes from description, a pretrained **R-CNN** for human figure detection, fine-tuned **YOLO v2** pretrained on PASCAL VOC 2010 dataset for human figure segmentation and a fine-tuned **AlexNet** for clothing type, gender and headgear classification. Gender detection and headgear detection modules had a test set accuracy of **86.47%** and **92%** (with a recall of 0.955 and precision of 0.914) respectively. Manual evaluation of 1845 images used in body segmentation module showed a segmentation accuracy of **96.47%**.

Factoid Based Question Answering System for Biology Coursebooks

Advisor: [Prof. S. Natarajan](#)

Feb 2016 - May 2016

- Designed a custom Q&A system using the QANUS Framework with a grade 11 Biology coursebook as the information source.
- NLP text-processing involved **chunking, parsing, POS Tagging and NER**.
- Devised and implemented **custom chunking and parsing algorithms** to generate candidate answers. The system incorporated a question classification model for categorizing questions into hierarchical classes based on **Li & Roth's two layer taxonomy**
- Used **Lucene** indexing for efficient answer retrieval.

Optical Recognition of Handwritten Isolated Characters

May 2015 - Dec 2015

Advisor: *Prof. Shylaja S.S.*

Studied and implemented classical image processing methods for extracting features from natural images and designed simple heuristic-based algorithms for handwritten character recognition.

Professional Experience

Goldman Sachs *Technical Associate (SDE II)* (Fast-track promotion from SDE I in two years)

May 2018 - Present

- 1) Currently building an ingestion pipeline and one-stop store for division-wide data, reducing consumption time for downstream applications from 48 hours to **less than 1 hour**. 2) Performed detailed analysis on the older DB utilization, table usage management and provided **optimized query time** by leveraging **Presto** as the query engine and **HDFS** partitioning for data milestone for a bi-temporal data store. 3) Leveraged **Spark** optimizations and **Yarn** queues for resource management. 4) Allowed for event driven data ingestion using **kafka messaging system**.
- Designed and implemented a process chain management platform with features like **intelligent logging, process chain re-runability and version controlling of intermediate data**. Modelled process chains as **directed graphs (DAG)** to prune out unnecessary dependent processes, **reducing run time** of process chains significantly.
- Worked on building a search engine plugin for real-time text search on encrypted data using **AES Encryption, Key Management Servers, n-grams**.
- Created a personalized dashboard application with **real-time** updates to view, action on and close various automated tasks generated for each Compliance officer. Developed an ingestion pipeline to aggregate, store and maintain **vertically fragmented distributed data consistency**.

Goldman Sachs *Technical Intern*

Jan 2018 - April 2018

Automated and streamlined the existing ticketing system for requesting creation of new conferences and sessions as part of a project for the Human Resources Division.

Mast Global (L Brands) *Technical Intern*

May 2017 - July 2017

Built a **Sandbox Sync Tool** for automatically syncing code sandboxes across different production teams. The tool was incorporated in Mast Global's technical stack and is used by teams across L Brands.

Snaptrude *Technical Intern*

Aug 2016 - Nov 2016

Snaptrude provides automated and real-time visualization for real estate. Developed a framework for **automatic detection of doors and windows** on the scanned floor/site plan using image preprocessing steps that include noise removal, image binarization, skeletonization and median blurring using OpenCV.

Inking Thoughts *Full Stack Developer for the company's official website*

June 2016 - July 2016

Developed the **official website** using MEAN Stack for Inking Thoughts - a startup to promote budding artists, performers and poets. Implemented features such as forms, upload feature for uploading documents, a dedicated gallery page to view all pictures.

Technical Talks

- "Real-time Text-Search on Encrypted Data", Grace Hopper Celebration India, October 2019
- "Real-Time Headgear Detection in Videos Using Deep Learning Based Feature Extraction with A Supervised Classifier", ADCOM 2018 at IIIT Bangalore, September 2018
- *Machine Learning Classification and Clustering Algorithms and Basics of NLP*, Goldman Sachs, September 2018
- *Gradient Descent and Stochastic Gradient Descent: A Deep Dive*, PES Institute of Technology, October 2017

Miscellaneous Projects

- Implemented a Feed Forward Neural Network for predicting movie success from features such as actor and director rating, genre and run length. Studied the effects of changing hyperparameters such as learning rate, optimizers and activation functions.
- IoT based Smart Pill Bottle using Raspberry Pi with a desktop page and android application for doctors and patients to use.
- Built a prototype for IoT edge analytics using LIOTA by VMWare at the Cloud Computing and Big Data Lab in PESIT.
- Built a mechanical ball shooter that worked on the principle of projectile motion using Arduino, servo and DC motors.

Technical Skills

- **Programming Languages:** Python, Java, C, R, Matlab
- **Tools & Frameworks:** TensorFlow, PyTorch, Keras, NLTK, Numpy, scikit-learn, OpenCV, Hadoop, Presto, Spark
- **Relevant Coursework:** Machine Learning, Introduction to Natural Language Processing, Data Analytics, Computer Vision, Algorithms for Intelligent Web, Linear Algebra, Finite Automata and Formal Languages

Positions of Responsibility

- **Technical Head** for the annual cultural fest of PES Institute of Technology – Aatmatrishra 2017.
- **Center Head** for Igniting Young Minds Project of Leaders For Tomorrow (NGO) - Led and supervised 50 volunteers to impart soft skills to underprivileged children.