**Objective** – Seeking roles in the field of Data Science specializing in Machine Learning and Big Data Analytics

**Education**

* **Master of Science - Language Technologies Institute (School of Computer Science)** **Aug’16 - May’18**

Carnegie Mellon University, PA

* **Bachelors in Computer Engineering** **Aug’10 - May‘14**

University of Pune, IN. Division: First Class with Distinction

**Relevant Coursework**

* Intro To Machine Learning
* Intro to Deep Learning
* ML for Signal Processing
* Machine Learning for Large Datasets
* Language & Statistics
* Big Data in Practice

**Technical Skills**

|  |  |  |
| --- | --- | --- |
|  | **Proficient** | **Familiar** |
| **Core Languages:** | C, Python, Core Java, Visual Basic, | C++, R, Javascript, SAS, MATLAB, Scala |
| **Databases:** | Oracle, MySQL, Hive, Cassandra | MS Access, MongoDB |
| **Development/Productivity Tools:** | Turbo C, Informatica, Excel, Anaconda | VS’10 & ‘14, Jupyter Notebook, Hadoop, Spark |

**Academic Projects**

# Machine Learning Engineer. Noble.AI June’19 – Current

# Built offline ML Pipeline to extract & structure information in semi-structured docx as a part of UIE using Transfer Learning.

# Worked on preprocessing the documents and creating client visualization for the unstructured documents presented to clients

# Created Data Visualization for R&D experiment dataset showing various issues like variance in the dependent variable.

# Built first working MVP for Intelligent Recommendation Engine.

# Tools Used: Python, sklearn, matplotlib, luminoth, Tensorboard, Django

# Real Time Audio Event Detection on Edge (RA - Prof Yuvraj Agarwal, Synergy Labs, <mites.io>) Jan’18 – May’18

# Built from scratch the entire ML and Data Pipeline, stages include – Feature Extraction, Feature Engg, Hyper Parameter Tuning etc.

# Ran Multiple Experiment using classical ML algorithms like Logistic Regression and SVM’s automatically detect Audio Events like Vacuum Cleaner, Drill Machine, Faucet Running etc.

# Built a parallel pipeline running multiple experiments for each label tuning hyperparameter.

# Performed Data Analysis to debug ML algorithm performance using dimensional reduction algo like PCA.

# Tools Used – python, librosa, sklearn, jupyter.

# Speech Recognition using Wall Street Journal Data (Professor Bhiksha Raj) Jan’18 – May’18

# Used the WSJ labelled dataset at frame and phoneme level to recognize unlabeled speech signal.

# Built a 3 layer Neural Network on frame level data to train & make predictions resulting in accuracy of 56% for 136 labels.

# Built a 4 layer CNN Model on phoneme level data to train and make predictions resulting in 80% accuracy for 46 labels.

# Preprocessed data to deal with issues like variable length phoneme representation for CNN inputs.

# Built an end-to-end ASR using Listen-Attend-Spell Architecture with the CMUSphinx language model.

# Tools Used – Tensorflow, Pytorch, Python

# Audio Forensic for Maritime Recognition (Carnegie Mellon University – Prof. Rita Singh and Prof. Bhiksha Raj) Aug’17 – Dec’17

* + Built a system to automatically identify maritime audio signatures like Boat and Helicopter sound which can be used in Hoax Call Identification, solve criminal cases etc.
  + Collected audio recordings from Youtube 8M dataset using automatic scripts and parsing video description.
  + Used feature representations like Constant-Q. Correlograms , Modulation Spectrograms. Also used a pretrained CNN model to extract proxy features using the fully connected layer of CNN architecture.
  + Achieved accuracy of 73% using decision trees and 77% using Adaboost. Also proposed a full end to end architecture which could help in a more detailed analysis of sounds like make/type of helicopter and boat engine.

# Tools Used – Python, Sklearn, Spark, MATLAB.

**Data Science Intern Walmart Labs June’17 – Aug’17**

* Working on the Walmart Performance Ads team to optimize the current model used by Walmart to display relevant ads.
* Predicting Click through Rate(CTR) of ads using contextual information resulting in increase in the revenue based.
* Feature Engineering, identifying new features & performing experiments to tune parameters of current model.
* Deployed Models into production to run A/B test & validated model performance for comparing the online and offline evaluation results like NLL, P/R & ROC AUC and Click Through Rate.
* **Tools Used – Python, Spark(MLlib), Scala, Hive, Cassandra, Weka**

**Fake/Real News Classification (Carnegie Mellon University – Prof. Roni Rosenfeld) Oct’17 – Dec’17**

* + Built a system to classify Fake news from the real Broadcast News Articles(1992-1996) using different statistical techniques.
  + Extracted various Statistical, Vectorized, Contextual, Semantic Features.
  + Used the KenLM Language model to extract the Tri-gram and 5-gram perplexity resulting in 89% accuracy on the development set and 90% accuracy with all features combined.

# Tools Used – Python, Sklearn, KenLM.

**Movie Recommendation System using MovieLens Dataset (Carnegie Mellon University) May’17 - June’17**

* Used the Matric Factorization Technique to recommend movies to users following the Netflix Prize Winner’s Strategy on the Movie Lens Dataset consisting of 1 million ratings as training set.
* Implemented the Alternating Least Squares Optimizing Techniques to solve the “RMSE” Objective Function.
* Performed Experimental Analysis to tune hyperparamaters like K, lambdas etc.
* **Tool Used: Spyder, Python (NumPy, matplotlib, SciPy)**

**Home Depot Product Search Relevance(Carnegie Mellon University) May’16 - June’16**

* Used various NLP techniques to perform feature engg on the unstructured dataset - Product Description & Attributes.
* Used ML Algos like RandomForest Regressor and Linear Regression to score each search query with the result.
* **Tool Used: Python (NumPy, matplotlib), Big Data/Distributed Sytems** -**Spark – Pyspark, MongoDB**

**Super Fridge: Automated Grocery List using Object Detection in Refrigerator Mar’17-Apr’17**

* Built an app on Raspberry Pi to detect objects in a Refrigerator and creating a Grocery List for missing items.
* Built modules for Clarifai API and Pi Camera used for object detection & upload the grocery list to google drive for users.
* **Tools Used: Python, Raspberry Pi & Camera, Calrifai API (Object Detection), Google Drive API**

**Musicon: Music playing based on User Activity Recognition: SteelHacks’17**  **24hr – Hackathon (Feb’17)**

* Built an Android app which used Google’s Accelerator(Motion Sensor) to determine User Activity(Brisk Walk, Jogging, etc.).
* Integrated the User activity recognition module with Spotify API, which played song based on user activity.
* **Tools Used: Android JDK, Java, Google Accelerator (motion sensor) API, Spotify API**

**Image classification to classify proteins into subcellular localization patterns (CMU) Aug’16 - Dec’16**

* Built an Active Learning Framework containing Pool Based Data Access Model, Uncertainty based Querying Strategy and different base learners like SVM, Gaussian NB, KNN and Logistic Regression
* Accuracy score of 0.97 was achieved on dev data using SVM as base learner.
* **Tool Used: Spyder, Python (sklearn, NumPy, matplotlib, SciPy)**

**Using Probabilistic Graphical Model to forecast Stock Prices for Time Series data (CMU) Aug’16 - Nov’16**

* Transformed into stationary TS by using log space to remove unequal variances & difference to handle trend component.
* Checked stationarity using Dickey-Fuller Test (Features daily Stock prices of - Apple, MS, Hecla, NEM Mining, GM, Ford)
* Created precision matrix using transformed features and marginalized Precision Matrix for missing data.
* Conclusively was able to predict with minimal error rate the stock prices for Apple using only 3 days of data for companies.

**Professional Experience**

**Business Operation Associate ZS Associates Inc. Sept’14 – June’16**

* **Master Data Management - Role (Data Steward)**
* Automated processes like loading client data and QCing client deliverable and performed Ad-hoc analysis.
* Automation of Processes to reduce response time for file processing by over 80%
* **Technologies Used – Python, MS Excel, VBA, Informatica Siperian, PL/SQL**
* **Smart Data Quality Management – (Quest ’15) 24hr – Hackathon (Oct’15)**
* Participated & won in Quest’15 organized by ZS Associates which had 44 participating teams.
* Designed Prod Arch detailing flow & control of Data pipeline. Implemented “Thomson Tau Method” to detect outliers
* **Technologies Used – R, MS Excel, VBA and MS Access**

**Summer Intern Softkoash Solutions Pvt. Ltd May ’12 – July ’12**

* Implemented MS NerdDinner project as a POC. Fixed bugs to proprietary ERP Solution used by customers in production.
* **Technologies Used – C#, Microsoft’s .NET Framework, HTML, CSS and JavaScript**

**Co-curricular Activities**

* Won Quest ‘15 (Hackathon at ZS Associates India)
* 2nd Best Project - PICT’s “Impetus & Conceptus’14
* 2nd Prize MITCOE TechFest - ‘Network Raptors’
* Best Project in Ops Excellence (ZS Global Office)