Suraj Deshmukh

Personal Data

Name:	Suraj Dilip Deshmukh
Place and Date of Birth:	Kotul, Maharashtra | 20 Dec 1992
Address:	9, Shivmudra-2, Near RTO, Nashik
Phone:	+91 9503269604
email:	surajdeshmukh96@gmail.com
Linkedin:	https://in.linkedin.com/in/suraj-deshmukh-86245069
github:	https://github.com/suraj-deshmukh

Summary

Currently working as Assoiciate Technology at Synechron Technology pvt ltd, have 2+ years of experience in field of Machine Learning, Deep learning and Natural Language Processing. Have done 2 years full time Master of Technology in Modeling and Simulation that covered modules majorly on machine learning and deep learning as well as programming languages like C,C++,R and python. Have experience in developing python api using swagger api framework, Django based web applications, distributed computing in AWS as well as developing shiny(R)/bokeh(Python) dashboards.

Work Experience

Nov 17 - till date	Synechron Technology Pvt Ltd, Pune Associate Technology
Jul 16 - Oct 17	Tech Mahindra, Pune Associate Software Engineer
Jan 16 - Jul 16	S.S.Userworks, Koregaon Park, Pune Data Science Intern

Education

June 2016	Master of Technology in Modeling and Simulation, **Savitribai Phule Pune University**, Pune Institute: Centre For Modeling and Simulation(CMS), Pune Major: Machine Learning Class: A grade(GPA = 4.88/6)
July 2014	Bachelor of Engineering in Electronics and Telecommunication **Savitribai Phule Pune University**, Pune Institute: G.E.S.R.H.Sapat College Of Engineering, Management Studies And Research, Nashik Class: First Class(60%)
February 2010	Higher Secondary Certificate at **Pune Board** Institute: B.G.P. Sahyadri Junior College, Sangamner Class: Distinction(80.67%)
March 2008	Secondary School Certificate at **Pune Board** Institute: D.G.S. Vidyalaya, Sangamner Class: Distinction(83.84%)

PROJECTS

ORGANIZATION	PROJECT
SYNECHRON	**London Stock Exchange Group(LSEG) Crawler** Currently working on developing an application which would extract data files such as pdf/msdoc files from given seed urls up to certain depth and classify the extracted documents into relevant and non-relevant category. I am currently working on Scrapy framework to crawl data files from seed urls and also on creating cluster of aws ec2 instances to crawl the sites in distributed manner and next part would be to create a machine learning model for classification of documents. **client:** London Stock Exchange Group(LSEG)
SYNECHRON	**London Business Unit(LBU) Dashboard** LBU Dashboard for managing resource related tasks such as project/account allocation or deallocation of particular employee etc. I worked as Django developer and developed back-end components such as Django database, Django forms and tables and also supported the front-end developers by integrating their work with my bank-end work. **client:** Synechron London Business Unit
SYNECHRON	**Client Onboarding KYC Renewal** POC for Client Onboarding KYC Renewal for HSBC bank. I was assigned a task of extracting information from scanned passport and utility bills in pdf format. I worked on two machine learning models for these two tasks. First machine learning model was based on OCR(Optical Character Recognition) to extract passport details given a scanned copy of passport and second machine learning model was based on NLP(Natural Language Processing) to extract and parse free form address text into structured way. Also worked on swagger api framework to expose these two models as rest api **client:** HSBC London
SYNECHRON	**Intelligent Resume Search** Worked on Intelligent Resume Parser and Recommendation based on job description using deep learning and nlp techniques. The main objective is to find out relevant candidate resume using given job description. Also I worked on user interface part of this project using python's bokeh framework. **client:** Synechron(HSBC Inhouse account)
SYNECHRON	**Fraudulent Email detection** POC on Fraudulent Email detection for CITY bank. For this task I have used different fraud email datasets and combined them to form binary classification problem (ham and spam class) **client:** CITY
TECH MAHINDRA	**TACTiX (TechM Actionable Intelligence Extended) Platform API (Machine Learning API):** Worked on developing **machine learning** api having classification, regression and clustering services. API's are developed in python using swagger api framework
TECH MAHINDRA	**Active Learning** It is a semi supervised learning machine learning. The motivation behind active learning is to avoid the manual labeling process for creating dataset for classification purpose. This can also be used to reduce the size of huge dataset by keeping only those examples having maximum information. There are ways in active learing algorithm to find out the informativeness of the examples in dataset.

Projects

S.S Userworks   Image Captioning
The task is to generate natural language descriptions of images
and their regions. I have used deep learning techniques(LSTM/RNN) to generate
sentence descriptions from images. The name of the dataset is flickr8k containing 8k
images and their corresponding captions.

S.S Userworks   Visual Question Answering:
Given an image and a natural language question about the image,
the task is to provide an accurate natural language answer.

S.S Userworks and
CMS   Multi Label Image Classification
Multi Label Image classification using Neural Network. Convolutional Neural
Network(CNN) for feature extraction from images and classification using
Neural Network
github repo:
https://suraj-deshmukh.github.io/Keras-Multi-Label-Image-Classification/

CMS   Epidemic Model
An epidemic model is a simplified means of describing the transmission of
communicable disease through individuals. The modeling of infectious diseases
is a tool which has been used to study the mechanisms by which diseases spread,
to predict the future course of an outbreak and to evaluate strategies to control
an epidemic.
Shiny Web App : https://mywebapps.shinyapps.io/SIRM/

Other Work   BBC News Classification
Classification of 2225 documents from the BBC news website corresponding to stories
in five topical areas (business, entertainment, politics, sport, tech).
github repo:
https://github.com/suraj-deshmukh/BBC-Dataset-News-Classification

Kaggle Machine Learning Competitions Achievements

- Porto Seguro's Safe Driver Prediction
  Predict if a driver will file an insurance claim next year.
  Learderboard: 231/5169 (Top 5%)
  Medal Secured: Silver

- Home Depot product search relevance
  Predict the relevance of search results on homedepot.com
  Learderboard: 292/2125 (Top 14 %)

Research Paper

- Diabetic Retinopathy Detection using soft computing Techniques:
  Diabetes, referred to as diabetes mellitus, describes a group of metabolic diseases in
  which the person has high blood glucose. Possible complications that can be caused by
  badly controlled diabetes: Eye complications, Foot complication, Skin complications,
  Heart problems, Hypertension, etc. Diabetic retinopathy (DR), a common complication
  of diabetes, affects the blood vessels in the retina. It is due to the retina not receiving

enough oxygen. In this work, publically available MESSIDOR database with 1200 fundus images are considered. Due to the textural changes in the retinal fundus images texture analysis methods like statistical moments and GLCM are used and extracted 40 features. The unbalanced dataset decreases the accuracy so oversampling is done before classification. Classifiers such as Support Vector Machine (SVM), Random Forests, Gradient boost, AdaBoost, Gaussian Naive Bayes used to detect the DR.

- **Date of Conference:** 9-10 Sept. 2016
- **Date Added to IEEE Xplore:** 16 March 2017
- **Publisher:** IEEE
- **Conference Location:** Pune, India
- **Link:** http://ieeexplore.ieee.org/document/7877573/

Programming Skills

- C, C++
- R(Shiny Web Application), Python (Object Oriented Programming and Functional Programming, API)

Languages

- Marathi (Native Proficiency)
- Hindi (Professional Working Proficiency)
- English (Professional Working Proficiency)