PRIYANSHU S SINHA

D A T A S C I E N C E E N G I N E E R

P 7349575896

E Priyanshu1996@hotmail.com

A #108, Kamdhenu Enclave, Hunsamaranahalli, Bangalore - 562157

W sinha96.github.io/

OBJECTIVE

A Data Scientist having the ability to articulate with both technical professionals and end users to identify and translate business requirements into actionable project plans. Graduated in Data Science and Engineering (DSE) from Great lakes, I have worked on a myriad of datasets where I performed data cleansing, feature engineering, building models, data analysis and data visualizations.

EXPERIENCE

Aug'19 - present
Data Science Engineer, Sumyag Datascience Pvt. Ltd.

- Designed and implemented an end-to-end scalable solution for extracting information from unstructured documents for enterprise clients.
- Achieved cross-domain compatibility by designing a highly configurable data-driven pipeline for domains like commercial insurance, healthcare, automobile, real estate and telecommunication.
- Leveraged AWS auto-scaled infrastructure to build custom statistical machine learning models for generating embeddings to predict entities from a corpus of 60000+ documents, roughly summing up to a terabyte.
- Industrial Projects:
  1. Sygnif.ai - Document Processing Engine:
     a. Implemented a pipeline for extracting raw data from PDF and Image documents. Worked on creating a seamless micro-service for Optical Character Recognition for extracting data from scanned documents.
     b. Designed and collaborated with the team in the development of a multi-phase, massive data

EDUCATION

1. Great Lakes Institute of Management, 2019
Bangalore,
Karnataka,
Post-Graduation Program in Data Science Engineering with 81%
Major:
- Python
- Data Science
- Statistics
- Machine Learning Algorithms

2. Sir M Visvesvaraya Institute of Technology, 2014 - 2018
Bangalore
Karnataka,
Bachelor in Engineering in Electrical and Electronics.

3. Kendriya Vidyalaya, 2001 - 2013
Kanpur
Uttar Pradesh
Completed class 10th and 12th with science stream (Physics, Chemistry, Math)

handling pipeline for enriching features of extracted data points from the document.
   c. Worked on generating 500+ features for data points to create a holistic view of a document by leveraging concepts from Natural Language Processing, Data Wrangling and Text Mining.
   d. Co-developed a large matrix multiplication model using NumPy stack which ensembles 10+ unique sources for classifying various entities in a data point.

2. Symplif.ai - Claims Processing Engine:
   a. Developed an end-to-end pipeline for predicting and extracting valuable insights from a claims document specifically for enterprise clients from the automobile and healthcare industry.
   b. Worked on developing a geometric model for extraction of tabular entities in a document by using asymmetric graphs and mathematical distances among data points.

3. Insurance Learning Pipeline:
   a. Developed and designed a sequential automated pipeline to generate the clean dataset required for supervising custom Bayesian learning models. Generated 10+ distinct models with various interpretations of a data point to generate word embeddings for predicting classification entities
   b. Worked on deploying the learning pipeline on an AWS auto-scaled infrastructure for running a corpus of 25000+ documents across a total of 300 nodes.

4. Resume Genie:
   a. Developed a multi-class classifier for predicting company domain from the candidate's industrial experience extracted from a resume document.
   b. Worked on building machine learning pipelines to predict named entities for resume data extraction

Technologies Stack:
   Python, Machine Learning, Natural Language Processing, EC2, S3, Deep Learning

Jan'19 – Mar'19
Machine Learning Intern, OpenGenus Foundation
My responsibility was to write education content blog on various topic of Machine learning, Data science and deep learning like convolutional neural network, LSTM (long short-term memory) Lidar data processing etc.
Link: Blogs on AI (ML, DS, DL)

PROJECTS

1. Document AI
2. Predicting Salary of Candidates (PGP final project)
3. Quiz AI
4. Face Detection
5. Emotion Detection
6. Chronic Kidney Disease Prediction
7. Power Consumption Prediction
8. Sentiment Analysis
9. Walmart's Sale Prediction
10. Breast Cancer Prediction
11. Handwritten digit Prediction
12. Home Automation using Raspberry Pi (4th year engineering project)
13. Various Kaggle's Dataset analysis

CERTIFICATION

1. Specialization certificate in Deep Learning offered by deeplearning.ai and Coursera (Taught by Andrew NG).
2. Applied Machine Learning in Python offered by University of Michigan and Coursera.
3. Certified Python programmer (level: Advance) offered by HackerRank.

e. PIPELINE DESIGNING

4. CLOUD PLATFORM
   a. AWS S3
   b. AWS SQS
   c. AWS EC2
5. VERSION CONTROLING
   a. BITBUCKET
   b. GITHUB
6. PROJECT MANAGEMENT
   a. JIRA
7. APPLICATION
   a. MICROSOFT EXCEL
   b. POSTMAN
   c. PYCHARM
   d. DATAGRIP
8. DATABASE
   a. MYSQL
   b. POSTGRESQL