Raman Shinde

https://github.com/Raman-Raje

Email : raman.shinde15@gmail.com

Mobile : +91 9595161238

Career Summary

- Currently pursuing my career as a Data Scientist with core expertise in entity extraction, pdf parsing, data analysis and building Machine Learning-based solutions.
- My work has revolved around architecting ML-driven solutions in products/platforms and getting them live into production.
- Expertise in various ML algorithms(KNN,K-means,Naive Bayes,Logistic Regression,SVM,DT,Random Forest,GBDT etc.) and scikit-learn.
- Experience and expertise in building Deep Learning models using Neural Networks such as LSTM/RNN, CNN, Transformer, Bert, Autoencoders, Memory Networks, Seq2Seq, etc.
- Experience in implementing research papers and model building using TensorFlow, Keras, Pytorch and Flask API's for model serving.
- Well versed with libraries such as Numpy, Pandas, Matplotlib, OpenCV, SciPy. Experience in using GCP/AWS when there's a need for high computing power.

Experience

- **Xpanxion** Pune, India
  Data Scientist *Jan. 2020 - Present*
  - **Anthem - Symphony(Python/DL/ML/NLP)**
    - Extracting information from the various medical documents with help of AI/ML.
    - Processing the text information to extract benefit, rates, drug details and plans details.
    - Document Classification to classify documents into different categories.
    - Entity and candidate extraction from classified documents with the help of libraries like Fonduer, Camelot, Tesseract, AWS textract, OpenCV.
  - **Digital Access Hub(Python/DL/ML/NLP)**
    - Building various reusable AIML components as a part of Innovation team
    - Extracting information from the reciepts/tables with the help of Optical character recognition(OCR)
    - Recommendation system using the content-based/collaborative algorithm
    - Worked on Computer Vision use cases like Object localization and detection, Image segmentation and Gesture recognition.
    - CCA (Canonical Correlation Analysis), Factor analysis to find out the relation between the features and then use the ML algorithm to find out the insights.
- **Siemens R&D** Pune, India
  Product Development Engineer *Dec. 2018 - Dec 2019*
  - **Automation Designer(Python/ML/C++)**
    - Developing an application for designing of manufacturing sequence.
    - Debugging and fixing the issue in the existing code.
    - Implemented model for Question answering from knowledge base.
- **TCS** Pune, India
  Software Developer *Dec 2015 - Nov 2018*
  - **Application Developer(Python)**
    - Developed an application for client NCRA for Monitoring and Controlling.
    - Monitoring and debugging the issue and modifying the GUI as per requirements.
    - Build a model for the false triggering of alarm and anomaly detection.
  - **Production Management**
    - Worked in Production Management for client Morgan Stanley.
    - In QAPM, I was part of the EDS team which deals with different financial applications like CRD, SRD and TS.

Projects

- **Deep Learning**
  - **Neural Machine Translation using Attention mechanism (NLP)** Task is to implement Machine Translator.Attention was used to deal with longer sequences. Data cleaning and output labels were padded with start and end tokens before feeding to n/w.
  - **Nueral Question Answering(NLP + Attention + Machine Reading Coprehension)** Objective is to find correct answer for given question and context pair. Implemented Standford Attentive Reader.SQUAD v1 dataset was used for this project. Various binary and NLP features were used to get the best results. Compared the final results with fine tuned BERT model.
- **Machine Learning**
  - **Netflix Movie Recommendation System (Collaborative based recommendation)** Objective was for the given movie and user predict the rating given by him/her to the movie.The dataset was obtained from kaggle. Matrix factorization was used to get similarity matrices. Tried and tested various ML models to get minimum Root Mean Square.
  - **Stack Overflow Tag Prediction** Objective is to predict as many as tags possible with high Precision and Recall.The dataset was obtained from kaggle. The given problem is **multi-label classification problem**. The dataset contains features such as Id, Title, Body and Tags. Data preprocessing and cleaning was done to remove html tags and hyperlinks. Micro-Averaged F1-Score was used as performance metric as mentioned on Kaggle.

Technical Skills

- **Languages:** Python, C++, R(basics)
- **Database:** MySql,MongoDB
- **Data Analysis:** Pandas, Numpy, Matplotlib, Seaborn, openCV
- **ML/DL Toolkit:** Keras, scikit-learn , tensorflow, pytorch
- **Others:** GitHub, svn, Jira, Jenkins, Service Now etc.,

Certifications/Internship

- Applied Machine Learning course at Applied AI. (Jan 2018 to May 2019)
- Completed Standford Statistical Learning (Self-Paced) course.
- Completed Deep Learning Specialization course from Coursera
- Internship at IARE, Aurangabad on Industrial automation. (May 2014 - Jun 2014)

Education

- **B.Tech** in Electronics and Telecommunication from SGGSIE&T, Nanded with **CGPA 7.7** (2011 - 2015)
- Class Xll (HSC), form Maharashtra State Board of Education with **83.33%** (2009 - 2011)
- Class X (SSC), form Maharashtra State Board of Education with **90.92%** (2008 - 2009)