KASHYAP N RAVAL

Phone: +91 9740032080
kashyap.raval@gmail.com
Bangalore, India
https://github.com/kashyapraval

EDUCATIONAL QUALIFICATION

Integrated Masters in Mathematics and Scientific Computing from IIT Kanpur (2006-2011)

WORK EXPERIENCE

Senior Data Scientist(SMTS), Athenahealth (May'18 -)

Document Classification

- With Classification of documents across classes such as imaging results, medical records, lab results, etc..
- Retraining embeddings and intend to use attention models(recurrent)

Insurance Mapping

- Among the insurance claims that could not be mapped to a particular insurance label(based on a rule engine), the model was developed to map them.
- Significantly accurate results with the use of multi-input-multi-output models neural networks. The model performed better than the Human worker(task was outsourced to BPOs)

Patient Resolution

- With data from different hospitals is hard to identify if two appointments across different hospitals involve the same patient. The patient resolution problem helps identify all clinical encounter details for a patient across specialities and also helps link medication to ailments and provides cross-speciality correlations.

ML Platform

- Deploying models on cloud and monitoring of systems and deployed models across different products.

Data Scientist, Agara Labs (Aug'17 - Mar'18)

Text Classification

- Classification of emails across two large sets of labels (multi-label classification), with classifiers based on n-gram CNN, LSTMs, and also multitask models involving shared layers for the two classification problems. Introduced further branches within a multitask model based on segmentation of data.
- Developed and maintained the models and explored scope for expansion over segments.

Response Generation

- Natural Language Generation based on Language Modeling for Response Generation. The performance was better (than ones trained on wiki-data) due to a smaller and closed vocabulary but not relevant.
- K - nearest neighbour approach when considering the embedding vectors of the emails that were extracted from layers of the classification model.
- This work was presented at **WAIU-2017 at IISc Bangalore**.

Data Scientist, Metripping (Jan'17 - Jul'17)

Destination Rankings

- As a part of travel search and inspiration: based on user inputs, destinations are recommended as an ordered list. This involved data collection across sources and data mining
- Used various machine learning algorithms to cluster destinations (using clustering algorithms like DB- Scan, KMeans along with methods from computational geometry). Similar destinations/experiences were determined based on user generated text data.
- Use of image processing over satellite imagery to develop and quantize features pertaining to a destination. Accounting for imagery differences

R&D Engineer, Synopsys (Feb'14 - Oct'16)

- Optimised spatial configuration of design components to achieve compact designs and reduced design violations. Added support for design rules pertaining to fabrication of wiring on microchips. Product used by Qualcomm, Intel-Arm, nVIDIA for their designs.
- Developed an optimisation solver for non-convex problems using geometric programming after approximate formulation of the problems to one involving posynomial constraints and objective(s).

Quant Analyst, IIFL-Genpact (Jun'11 - Dec'13)

- Design, Analysis and Implementation of strategies for high frequency algorithmic trading.
- Bankruptcy, charge-off forecasting.

PROJECTS (SELF MOTIVATED)

Heart beat Detection with cellphone (2018-)

- Recorded a self video with onePlus5 front camera(any basic camera works) and extracted frames. Inspected minute fluctuations on the pixels corresponding to red frame over the face that weren't present on the background.
- Noise removal based on observing fluctuations in background (FFT)
- Observed the number of times the intensities crossed the standard deviation.
- Testing across friends (needs some more tweaking)

Newsfeed classifier : TOI Vs The Hindu (2016)

- Scraped facebook pages of two leading newspaper publications in India.
- Surveyed various classifiers to compare the accuracies. Random Forest scored highest with 84% accuracy over a test sample. Unsupervised learning: clustering into two clusters yielded an accuracy of over 60%.

TECHNICAL SKILLS

- **Programming**: Python, Keras, Tensorflow, Numpy, Linux
- **Skills**: Deep-Machine Learning, Natural Language Processing, Data Mining, Optimisation, Linear Algebra, Signal processing

RELEVANT COURSES

- Data Structures and Algorithms, Linear Algebra, Probability and Statistics, Machine Learning, Convolutional Neural Networks

WORKSHOPS

- Presented CNN techniques in the "INTRODUCTION TO AI" seminar by **Confederation of Indian Industries** on 12July'2018.
- Held a team-internal workshop at Athenahealth on different ways to train word vectors based (CBOW, skip gram, BERT)
- Held a workshop for application development with VBA, Genpact in 2012.