Ankan Dutta

Machine Learning Engineer

Hoodi, Whitefield, Bengaluru, Karnataka, India, Pin No.- 560066

+(91)9810512279 | ankan.btech14@gmail.com | ankancode | ankandutta

"If you are not willing to risk the usual, you will have to settle for the ordinary"

Education

ITNU(Institute of Technology, Nirma University) — *Ahmedabad, India*

M.Tech in Information and Network Security — *May 2016*

PIET (Priyadarshini Institute of Engineering and Technology), Rashtrasant Tukadoji Maharaj Nagpur University — *Nagpur, India*

B.Tech. in Information Technology — *June 2014*

Technical Skills

Machine Learning Regression Techniques(Linear Regression, multivariate regression), Classification Techniques(Logistic Regression, Naive Bayes Classifier, Decision Trees, Support Vector Machines, Nearest Neighbor), Clustering (K-Means)

Ensemble Methods Bagging, Boosting, Stacking

Natural Language Processing tfidf, Ngram, Stemming, Lemmatisation, Cosine similarity, Bag-of-words, Word2vec, LDA, NMF

Deep Learning Deep Neural Networks(DNNs), Convolutional Neural Networks(CNNs), Recurrent Neural Networks(RNNs) with LSTMs and GRUs units, Generative Adversarial Networks(GANs), Quantization, Pruning

Programming Languages Python, C++, Java

DL Frameworks Pytorch, Tensorflow

Big Data PySpark

Cloud Platform GCP, AWS

Web Frameworks Flask

Software Tools Anaconda, Jupyter Notebook, Spyder, Sublime, Git

Operating Systems Windows, Linux(Ubuntu)

Experience

Argoid — *Bengaluru, India*

Senior Data Scientist — *June 2020 - Present*

- Working on Artificial Intelligence powered Recommender as a Service with an intuitive RESTful API & SDKs
- Working on core Recommendation Engine for a UCG platform.

GoodWorkLabs (Client: Mercedes-Benz R&D) — *Bengaluru, India*

Data Scientist — *Dec. 2018 - June 2020*

- In-House: Worked on a project where we have to detect whether the cat is holding any object in his mouth and Recognition the owner's cat. (Frameworks: Tensorflow and TensorflowLite, Architectures to be used: RetinaNet and FaceNet
- In-House: Working on a project where we are building an Attendance system using Face Recognition. Our solution includes two parts:Face Detection and Face Recognition. Face Detection portion is developed using MTCNN and Face Recognition portion is developed using FaceNet. (Framework used : Pytorch)
- On-Site (Mercedes R&D - Automated Driving): Worked on their Big Data Pipeline for Data gathering from different Sensors of the car. (FrameWork used: PySpark)
- On-Site (Mercedes R&D - Artificial intelligence [Intelligent interiors]): Worked on the Optimization of Data-Preprocessing Pipeline

Quantiphi, Inc.

Mumbai, India

Machine Learning Engineer

Mar. 2018 - Dec. 2018

- Worked on a NLP project for Knowledge graph and Data Dashboard generation over client's data. Applied several Natural Language Processing (NLP) Techniques for Extracting Information from Client's data. (Python Packages Used - NLTK, Stanford NLP and Spacy)
- Worked on an Image Classification Dataset. Performed Data Augmentation on the Images as we had a very limited dataset. Then applied Transfer Learning using ResNet-18 in Pytorch. Got an accuracy of around 92% on the test dataset. (Python Package used: Pytorch)
- Worked on a Neural Machine Translation Project (English to Multiple Languages). Applied Word Embedding Vectors, RNN(With GRU Unit) and teacher forcing Strategy for training the decoder of the RNN. (Python Package used: Pytorch)
- Worked on a neural style transfer using CNNs for abstract art generation (Python Package used: Pytorch) (POC project)
- Worked on Question Answering System using RNNs and Relational Networks (Dataset used SQuAD)
- Worked on Visual Similarity based recommendation system using deep learning for an eCommerce Company. (Architectures Used: Resnet50, Inception V3, Squeeze-net)
- Improved massively In-Category Recommendations of items using 5 layers in-between embedding and then doing continuous sorting using appearance count score and ranking score.

Synopsys Inc

Noida, India

Post Graduate Engineer Trainee (PGET)

Dec. 2016 - Feb. 2018

- Worked on development and of AutoChar (Memory IP Characterization Software) and WebChar (Web Wrapper of AutoChar).
- Worked on the interpolation problem of Autochar. Used Time Series Regression Model to solve it.
- Development using C++, Python, Django, Tcl, Sklearn, Tensorflow

Projects

Face Recognition based Attendance System

Bengaluru, India

April 2020 - April 2020

Used Pre-Trained MTCNN and Facenet to build this Attendance System.
– `https://github.com/ankancode/Face_Recognition_based_Attendance_System`

I Know Everything

Bengaluru, India

Dec. 2019 - Jan. 2020

Used Pre-Trained Google BERT to build this Question-Answering system.
Version of BERT used : bert-large-uncased-whole-word-masking
– `https://github.com/ankancode/Question_Answering_System_Using_BERT`

Multi-Label Text classification for Household grocery Items

Bengaluru, India

July. 2018 - July. 2018

I attempted to solve the multi-label classification problem using 2 layer Bi-LSTM . (For the full network configuration see model.py)
For the Loss Function I have used BCEWithLogitsLoss
Optimizer used : Adam
– `https://github.com/ankancode/Multi-Label_Text_classification_for_Household_grocery_Items`

Auto Timeline

Mumbai, India

Nov. 2018 - Nov. 2018

Created a event timeline over all the articles extracted from wikepedia for the information we want over a particular topic.
Displayed only the best n (NOT necessarily first n articles) sentences which are in chronological order.
Also The created a flask application wrapper over it.
– `https://github.com/ankancode/Wikipedia_Chronological_Search`

Predicting Loan Application Scores Based on Historical Data

Mumbai, India

Oct. 2018 - Oct. 2018

Checking the skewness in Dataset, Dropping the columns which have Null value count above 11000, Normalizing the data
Performing Imputation on columns which have Null Values below 11000 using Decision Tree Regressor, Applying Synthetic Minority Over sampling Technique (SMOTE) to balance the data
Used the Random Forest classifier for feature selection and Used the important feature collected from Random Forest in training another XGBoost classifier with 200 estimators and max_depth 5.
– `https://github.com/ankancode/Predicting-Loan-Application-Scores-Based-on-Historical-Records`

Detecting near duplicate images

Mumbai, India

Sept. 2018 - Sept. 2018

Used image embedding from a pre-trained Resnet50 trained over imagenet dataset for detecting the near duplicate images. Also used histogram based strategy like Hellinger Distance and different hashing techniques like phash and dhash for the same purpose but getting image embedding from a pre-trained Resnet50 performed better.
– `https://github.com/ankancode/Detecting_near_duplicate_images`

Clothes Classification Based Sleeves

Mumbai, India

Sept. 2018 - Sept. 2018

I applied transfer learning using a pre-trained Resnet50. Used the weights of imagenet. Fine-Tuned the ResNet50 on my cloths dataset. After fine-tuning for 30 epochs got these results : Train Accuracy: 82.03 % Test Accuracy: 72.37 %
– `https://github.com/ankancode/Clothes_Classification_Based_Sleeves`

Email Spam Classification using (RNN-LSTM)

Mumbai, India

Jul. 2018 - Jul. 2018

Implemented a RNN-LSTM spam classifier which predicts whether given an email content is Spam or not. Trained the model for only 2 epochs, and got a accuracy of 95.9% on the final test data.
– `https://github.com/ankancode/RNN-LSTM`

Topic Modelling and Summarization

Mumbai, India

Jul. 2018 - Jul. 2018

Build a Summarization system that intends to summarize around given theme/topic from the document.
Following tasks where performed,
Found the possible top level topics - Clustering all the documents related to each of the topics - Summarized the content around each identified topic.
Algorithms used LDA and NMF.
– `https://github.com/ankancode/Topic_Modelling_and_Summarization`

Certifications

- Recommendation Systems with TensorFlow on GCP

`https://www.coursera.org/account/accomplishments/certificate/DPRNLAEUA84G`

- AI for Medical Diagnosis

`https://www.coursera.org/account/accomplishments/certificate/QBHJTGQQYXWN`

- Mathematics for Machine Learning: Linear Algebra

`https://www.coursera.org/account/accomplishments/certificate/H6P7B26ULFZ3`

- Taming Big Data with Apache Spark and Python - Hands On!

`https://www.udemy.com/certificate/UC-1QA0YM8A/`

Publication

Enhancing the Performance of Audio Visual Speech Recognition Using Deep Learning Techniques, International Journal of Computer Science & Communication (ISSN: 0973-7391),
`http://www.csjournals.com/?p=1143`
DOI:10.090592/IJCSC.2016.121

Achievements

- Successfully cleared GATE in 2014 with a score of 392 (All India Rank - 9480)