

AYAN SENGUPTA

@ ayan.sengupta007@gmail.com +91 9706318732 Noida, India <https://victor7246.github.io/>
in <https://www.linkedin.com/in/ayans2017/> <https://github.com/victor7246>

PROFESSIONAL EXPERIENCE

UnitedHealth Group (Optum)

Senior Data Scientist

August 2019 – Ongoing Noida, India

Working with Optum's Risk, Quality & Network Solutions (RQNS) Team

- Working on building NLP solutions to extract structured information from noisy semi-structured scanned clinical charts which along with non-clinical claims information enables Optum in \$1B+ in risk adjustment business
- Handling additional research activities involving Named Entity Recognition (NER) and ML based text correction on noisy texts

Data Scientist

May 2017 – August 2019

- Built and deployed pipeline to detect suspect, screening and hospitalization information for 7M+ member population
- Worked on ML based medical chart scoring and improved false negative rate by 30% leading to reduction of \$70M+ in manual chart coding cost
- Handled various R&D activities on medical fraud detection, voice of customer Analytics and NPS analytics

Corporate accolades received -

- 2019 innovation and 2018 Q4 mastermind awards for innovation
- 3rd place award for NPS analytics in annual data science championship 2018

RELEVANT RESEARCH PROJECTS

Structured Information Extraction from Scanned Medical Charts

Jan 2020 - Ongoing

- Extracting member and provider demographic information from noisy unstructured OCR outputs using Transformers and Sequential models
- Extracting structured medical conditions from scanned medical charts using multi-modal system leveraging image and text data
- Additionally working on text correction using Seq2Seq, Pointer Generator and Denoising Transformers based approaches

Joint Topic-Sentiment Model for Short Texts

Mar 2019 – May 2020

- Developed probabilistic generative model for semi-supervised joint topic-sentiment extraction from texts (paper submitted to *Applied Intelligence*)
- Developed matrix factorization based online topic model for continuous data streams (work submitted to *ICSOC 2020*)
- Developed embedding oriented semantic representation learning from texts (paper submitted to *COLING 2020*)
- Filed US Patent #531179 on Joint Topic-Sentiment Model for Short Texts
- Drafted 2 patents on prioritization of customer complaints and topic detection from streaming feedback data logs

Fraud Detection from Clinical Charts

May 2019 - Aug 2019

- Used clinical embedding based MRF-LDA model for thematic representation learning of OCR-ed charts and statistical methods to calculate semantic alignment of texts
- This project enables detection of false billings, unrelated procedures and testing from medical transcripts. POC was done for PMJAY's Ayushman Bharat

EDUCATION

Post Graduate Diploma in Business Analytics

IIM Calcutta, IIT Kharagpur, ISI Kolkata

Jun 2015 – May 2017

M.Sc. in Mathematics and Computing
IIT Guwahati

Jul 2013 – Jun 2015

B.Sc. in Mathematics and Computer Science

Chennai Mathematical Institute

Jul 2010 – Jun 2013

TECHNICAL SKILLS

- Development, DB

Python R SQL Shell

- ML Frameworks

sklearn PyTorch Tensorflow h2o

- Big Data

PySpark MapReduce Hive Oozie

OTHER RESEARCH ACTIVITIES

- Program Committee member of ECML-PKDD 2020 for Applied DS track
- Paper *A Large-scale Analysis of the Marketplace Characteristics in Fiverr* accepted in HICSS 2016
- Paper *Bayesian Fault Detection Engine for Predictive Maintenance in Data Centre Infrastructure Management* accepted in ICBAl 2016

OTHER DATA SCIENCE ACTIVITIES

- Developer and maintainer of package *jointtsmodel*
- Represented ISI Kolkata in *International Data Science Game 2016* finals in Paris and achieved global rank 14 among 100+ teams
- 2 Bronze medals in Kaggle Multilingual Toxic Comment Classification & Google QA challenge 2020
- Ranked 16 among 50+ in Codalab satellite image classification challenge 2019
- Won \$20K+ across various data science competitions in Topcoder between 2017-2020

OTHER ACCOMPLISHMENTS

- Hold KPMG Six Sigma Green Belt
- Recipient of *National KVPY scholarship* during 2009-2015 for pursuing research (top 200 among 50K+ applicants)
- Cleared *Regional Maths Olympiad* from W.B. region in 2008 (top 50 among 5000+)