# AKHIL PUNIA

561 W 143rd St., Apt. 33A NY, 10031 (646)-203-1523 akhil.punia@columbia.edu https://github.com/AkhilPunia

## EDUCATION

COLUMBIA UNIVERSITY, NEW YORK, NY

**Master of Science, Data Science, GPA: 3.25/ 4.00** **Sep 2018 - Dec 2019**

Coursework: Personalisation Theory, Machine Learning, Algorithms, Introduction to Databases, Statistical Inference and Modelling

VELLORE INSTITUTE OF TECHNOLOGY, VELLORE , INDIA

**B. Tech, Electrical and Electronics Engineering, GPA: 9.0/ 10.0** **July 2014 - May 2018**

Relevant Coursework: Neural Networks and Fuzzy Logic, Digital Image Processing, Data Structures, Probability, Operations Research.

## PROFESSIONAL DATA SCIENCE EXPERIENCE

**Columbia University Medical Center, New York**

**Research Assistant** **Nov 2018 – Ongoing**

- Developed a tool for automating classification of 25,000 instructor's comments into 6 predefined Entrustable Professional Activities (EPA) for medical students entering into Residency. Our deep learning based model achieves 91.4% accuracy.
- Designed a pipeline for preprocessing steps including data cleaning and stop word removal followed by generation of sentiment score of these comments using pretrained models. Final results are visualized through a dashboard.
- Work is selected for presentation at NEGEA Conference to be held at University of Pennsylvania in April 2019.

**Inter University Center for Astronomy and Astrophysics, Pune, India**

**Deep Learning Research Assistant** **Dec 2017 – May 2018**

- Worked in the Astronomy Machine Learning research group headed by Prof. Ajit Kembhavi to develop CNN ( AlexNet, VGG-16) based models for Galaxy Morphology classification using data from Sloan Digital Sky Survey.
- Our model achieved 94.5% accuracy in identifying Ring structures in galaxies. We integrated unsupervised learning methods like Autoencoders and RBMs for the purpose of improving the robustness of model over data containing unlabeled images.

## DATA SCIENCE PROJECTS

**Movie Lens Recommendation System** **Oct 2018 – Dec-2018**

- Implemented and compared qualitative and quantitative results of Matrix factorization, Collaborative Filtering and Locality Sensitive Hashing using Spark for handling sparse data in the task of movie recommendation using Movie Lens 20M dataset.
- Addressed the cold start problem by training movie embeddings to offer serendipitous results to new users.

**How America Flies : Exploratory Data Visualization Project** **Oct 2018 – Dec-2018**

- Visualized delays and cancellation patterns of flights across continental united states for the year 2017 using data provided by US Department of Transportation through Tidyverse in R. Choropleth maps were used to convey the spatial information.
- Designed an interactive D3.js based tool for visualizing average delay performance of routes by airlines and airports.

**West Wings Reads Article Recommendation Analysis** **Sep 2018**

- Implemented Topic Modelling to analyse the emergence of prominent issues covered in the 15 months ( June 2017- Sep' 2018) worth of article published by the White House on West Wing Reads as a part of Columbia Data Science Hackathon.
- Scraped data from whitehouse.gov to train a supervised Naïve Bayes classifier with 96.4% accuracy to classify the articles into 14 different news categories. Discovered unique correlations between news source and topics of recommended articles.

## ACHIEVEMENTS AND PUBLICATIONS

**ZS Young Data Scientist Challenge 2017** | Top 35 in India . Implemented an LSTM model for Sequence Prediction Task.

**Future Group Datathon 2017** | Secured 9th position out of 4500+ teams to design a product recommendation system.

Akhil Punia, Ashish Sardana, Monica Subashini. M, **Evaluating Advanced Machine Learning Techniques for Pulsar Classification from HTRU Survey**, IEEE -International Conference on Intelligent Sustainable Systems (8th December, 2017)

## SKILLS

Python ( Scikit-Learn, NLTK, Spacy, Keras, PyTorch, TensorFlow ), R ( Tidyverse, Caret ), Spark, SQL, AWS, Google Cloud, Docker

## VOLUNTEERTING AND HOBBIES

Columbia Data Science Society, Data Science Hackathons, Hiking, Astronomy.