528 Riverside Drive,
Apt. 3B, New York,
NY 10027

AMEYA KARNAD

(646)492-0972
ak4251@columbia.edu
http://ameyakarnad.github.io

EDUCATION

Columbia University *New York, NY*
Master of Science, Data Science, GPA: 3.71/4 Aug 2018 - Dec 2019 (Expected)
Courses: Algorithms, Exploratory Data Analysis & Visualization, Personalization, Probability and Statistics, Machine Learning, Statistical Inference & Modeling, Applied ML, Data Science & Public Policy, Analysis of Social Networks and Crowds, Applied Deep learning (Audit), Capstone Project (Bloomberg - Editorial Classifier)

Visvesvaraya Technological University *Hubli, IN*
B. V. Bhoomaraddi College of Engineering and Technology Jul 2012 - Jun 2016
Bachelor of Engineering, Information Science, GPA: 9.42/10 (Silver Medalist)

SKILLS

Languages and technologies: Python (pandas, numpy, scikit-learn, nltk, tensorflow, networkx, genism, spacy), R (ggplot2, dplyr, data.table, R-shiny), JavaScript (SVG, D3), SQL, MongoDB, Docker, Kubernetes, AWS

PROFESSIONAL EXPERIENCE

Data Science Research Engineer - *EdLab, Columbia University* (Data Science Scholar's Program) Jun 2019 - Present
Technologies: *Apache PredictionIO, MongoDB, EZProxy, Docker*
- Development and research on Search and Recommendation systems, Network analysis, and Topic modeling.
- Currently working on Auto-tagging of documents using deep learning and building a python package for metadata retrieval.

Teaching Assistant – Applied Data Science, *Columbia University* *New York, US*
Awards: *"Excellence in Course Assistantship" for the course. Link* Jan 2019 - May 2019
- Guided students in the use of Data preprocessing, Feature engineering, Machine learning, and Communicating results in various fields like Education, Healthcare, Marketing, Advertising and Social Media using R.

Software Engineer - Micro Focus (spun-off from Hewlett Packard Enterprise) *Bangalore, IN*
Awards: *2nd place in a site-wide Hackathon* **Technologies**: *Kubernetes, Docker, Python, SQL* Sept 2016 - Jun 2018
- Worked with agile teams to design, test and automate REST APIs for providing analytics insights into software security compliance and risk data for a data-center automation product.

SELECT DATA SCIENCE PROJECTS (GITHUB LINK)

Editorial Classifier, Bloomberg - Columbia University Ongoing
Languages: *Python(Pytorch, spacy), R (ggplot)* **Applications**: *NLP, Classification, Deep Learning*
- Worked with unstructured news sources to build classifiers to identify editorial content for Bloomberg Terminal.
- Developed traditional classification algorithms with text based features like n-grams, Parts of speech, Named entities etc.
- Trained Bidirectional LSTM, BERT, and XLNET Models for classification using Pytorch Link.

Educational Recommendation System, Edlab
Languages: *Python (nltk, predictionIO), SQL, MongoDB, AWS* **Applications**: *Recommendation systems*
- Developed a multi-platform database pipeline for physical and digital user behavior, and research content using AWS.
- Built and evaluated a hybrid recommendation system on multiple digital learning platforms.

Does Climate Change and Natural Resource Storage Cause Conflict? *Columbia University*
Languages: *Python (Scikit-learn), R (Shiny, data.table)* **Applications**: *Regression, SVM, Decision Trees, Dashboard*
- Developed Machine Learning models to find correlation between Natural, and Economic factors and Conflicts in Senegal Link.
- Built a Dashboard to be used by policymakers to take effective decisions to avoid natural resources shortages and conflicts.

How America Flies? - Creating Insightful Visualizations on Airline Performance Data, *Columbia University*
Languages: *R (ggplot2, dplyr), JavaScript (D3, SVG)* **Applications**: *Data visualizations, Exploratory data analysis*
- Conducted Exploratory data analysis and found insights on Airline On-time performance data Project Link.
- Designed a Visualization tool in JavaScript and D3 to find flight delays between top airports Tool Link.

Beer Personalization and Recommendation System, *Columbia University*
Languages: *Python (surprise, nltk, scikit-learn)* **Models**: *Matrix factorization, NLP, Nearest Neighbours*
- Developed a comparative analysis of the models mentioned above with regards to accuracy and time.
- Tackled recommendation concerns such as cold- start, variety and serendipity Link.