AMLA SRIVASTAVA

155 Claremont Avenue, 766, New York, NY 10027 | as5196@columbia.edu I (913) 278-7823 |
https://www.linkedin.com/in/amla-srivastava/ | www.github.com/amla26

EDUCATION

Columbia University, New York, NY **Expected Dec 2017**
M.S. in Data Science GPA: 4.0

- *Relevant courses*: Machine Learning, Applied Machine Learning, Natural Language Processing, Data Visualization, Probability and Statistics, Algorithms for Data Science, Exploratory Data Analysis, Big Data Analytics
- *Teaching Assistant*: Introduction to Computing for Engineering/Applied Sciences (Python) under Prof. Daniel Bauer

PEC University of Technology, Chandigarh **May 2015**
B.E. in Computer Science & Engineering GPA: 9.4

- *Relevant courses:* Algorithms, Databases, Data Structures, Operating Systems, Neural Networks, Artificial Intelligence
- SPDC & DASA scholarships by Govt. of India

SKILLS

- **Programming:** Python (scikit-learn, NumPy, Pandas, keras, nltk), R (dplyr, tidyverse), SQL, C, C++, Java, PHP, HTML
- **Data Visualization:** Python (matplotlib, seaborn), R (ggplot2, shiny), D3.js, CartoDB, Tableau, Plotly
- **Big Data/Other:** Spark, H20, Hive, Trifacta Wrangler, RapidMiner, Git, MS Office

WORK EXPERIENCE

NBC Universal, Data Science Intern, New York, NY **Sep 2017 - Present**

- Developing Spark scripts to aggregate Nielsen ratings data, building a R Shiny app to perform inferential analysis and clustering to explore demographic variations across shows, networks, part of day, etc.

The Hartford, Data Science Intern, Hartford, CT **May – Aug 2017**

- Developed an unconstrained loss model to enhance current Class Plan using H2O and Python; improved Gini score by 4%
- Built a web application for recommending Increased Limits Factors (ILFs) to determine premium using R and Shiny

McKinsey & Company, Jr. Research Analyst, Gurgaon **Jun 2015 – Jun 2016, Jan –Jul 2014**

- Provided research and analytics services for a global Consumer Electronics client; analyzed consumer behavior and competitive market trends
- Conducted study on 'Text Mining', built models for resume classification, sentiment analysis in R, RapidMiner and Tropes
- Developed a request classification tool in Excel, VBA and RapidMiner to track team performance with an accuracy of 94%

DATA SCIENCE PROJECT EXPERIENCE

Classifying traffic complaints in Boston : *Text Classification & Topic Modeling* **May 2017**

- Built a logistic regression model to classify traffic complaints in Boston using scikit-learn; achieved a final accuracy of 92%
- Improved model using Latent Dirichlet Allocation(LDA), Non-Negative Matrix Factorization(NMF) and k-means clustering

Predicting Consumer Conversion for Banking : *Ensemble Methods* **Apr 2017**

- Constructed an ensemble model using logistic regression, SVMs, gradient boosted trees and PCA in scikit-learn
- Predicted subscription rates for the bank's direct marketing campaign; achieved an AUC score of 0.83

Visualizing Vehicular Collisions in NYC : *Exploratory Data Analysis* **Mar - Apr 2017**

- Created interactive heatmaps, bar charts, etc. using R, Tableau and CartoDB to explore NYPD accident data
- Visualized trends in accident types, seasonality, causes, and risks between 2013-16

Music Recommendation System : *Big Data Analytics* **Mar 2017**

- Developed a recommendation engine using collaborative filtering on the AudioScrobbler dataset with PySpark and MLlib, predicted top artists for over 20,000 users

Predicting most popular products: *CDSS Microsoft Data Science Hackathon (2nd Place, $3000 prize)* **Oct 2016**

- Built a softmax regression model in Python and Azure ML Studio to predict most popular product using Walmart store features, enabled stores to identify substitutes in case of product unavailability

OTHER

- **Organizations:** Columbia Engineering Graduate Student Council, PEC Entrepreneurship Cell
- **Status:** U.S. Citizen