

# Ujjwal Peshin

📞 (917) 847 3872 • ✉ up2138@columbia.edu • 🌐 ujjwal95.github.io • 👤 ujjwal95  
in ujjwal-peshin

## Education

**Columbia University**, New York, NY **Aug 2018 – Dec 2019**  
*Master of Science in Data Science* GPA: 3.73

Relevant Coursework: Machine Learning, Neural Networks and Deep Learning, Data Analytics Pipeline, Personalization (Recommendation Systems), Exploratory Data Analysis and Visualization, Algorithms, Statistics

**NSIT, University of Delhi**, Delhi, India **Aug 2013 – Jun 2017**  
*Bachelor of Engineering in Computer Engineering* CGPA: 7.63

Relevant Coursework: Artificial Intelligence, SQL, Algorithms and Data Structures

## Technical Strengths

**Skills** Python (Pandas, Numpy, Scikit-learn, TensorFlow, Keras, SpaCy, NLTK), R, SQL, NoSQL, AWS, Spark, Scala

## Data Science Experience

**Capital One**, New York, NY **Feb 2020 – Current**  
*Senior Data Scientist*

- Building the next generation of Small Business **customer risk ML models**, providing **15% improvement** over existing models.
- Built a modeling framework to allow easy building of models, local and on EMR clusters, and link up with SQL databases.
- Built a custom metric to measure effectiveness of a risk model, which was used in grid search to select the best performing model.
- Shipped best performing real-time analytical model to production and worked with them to validate inputs and model scores.

**Capital One**, New York, NY **Jun 2019 – Aug 2019**  
*Data Science Intern*

- Built a database of USA businesses, used to reach out to potential customers, and assist in fraud detection and KYC verification with the help of algorithms, Big data, and ML models, for a **10% higher coverage**.
- Applied graph algorithms to evaluate use cases of graphs in finding new potential customers and businesses.

**ZS Associates**, Gurugram, India **Jul 2017 – Jul 2018**  
*Data Scientist*

- Formed a state space search model to optimize Inclusion/Exclusion patient pool for clinical trials.
- Prepared a webapp using **Django** to display physician records using various data sources, which was presented at a conference, to provide an enhanced view to the client regarding each physician record.
- Produced multiple ML models to find prevalence of AFib, providing an upscale of **\$30 million** to the client.

## Academic Projects

**Bloomberg L.P.** **Sep 2019 – Dec 2019**

- Predict a news article as an Editorial or a regular article, which would replace the rule-based filter on the Bloomberg terminal.
- Built classifiers using **LightGBM and XLNet and TF-IDF, Bag-of-words, and VADER (sentiment analysis)** featurizers.

**Earth Institute, Columbia University** **Sep 2018 – May 2019**

- Analyzed physician-patient interactions of medical students to determine students who need guidance.
- Used NLP techniques like **TF-IDF** and **word vectors** to build an ML model for various criteria of interest.

## Leadership & Teaching

**Columbia University**, New York, NY **Sep 2018 – May 2019**  
*Graduate Course Assistant*

- Teaching assistant for COMS W4995 Applied Machine Learning under **Prof. Andreas Mueller** and Elements of Data Science under **Prof. Bryan Gibson**.
- Provided guidance to graduate students and created and evaluated assignments and projects in Python.
- The courses provide a foundation of analysis of data and discuss the application of machine learning models on real world data, with hands-on experiments with Python.

## Publications

- Ujjwal P., Tanay J. and Shivam K.**, *Touchless Recognition of Hand Gesture Digits and English Characters Using Convolutional Neural Networks* in Lecture Notes in Computer Science, Vol. 11407. Springer, 2019.