

## Experience

11 years

**Machine Learning Engineer, Sep 2021 – Present, Zillow (Zestimate), Remote**

3.5Y

### **Interactive CMA & Realtime Valuation** (Django, DocumentDB, PyTorch) [Siamese Neural Network]

Architected and led the end-to-end development of an interactive *Comparative Market Analysis* (CMA) platform with Realtime Valuations, Property Embeddings and Comps API, providing agents and buyers with data-driven tools to support home pricing, enhancing client decision-making and driving potential revenue through valuation services.

*Impact: 0 to 1 project to boost engagement and satisfaction, paving the way for **new revenue streams***

### **Zestimate Infrastructure Modernization** (Python, Terraform, AWS, Kubeflow, Metaflow, Docker, Gitlab CI)

Led the modernization of a critical valuation ML infrastructure, transitioning to more cost-effective, containerized technologies, resulting in substantial annual cost savings and enhanced system scalability.

*Impact: Achieved operational improvements and annual cost savings of \$500k.*

### **Technical Innovation & Team Collaboration**

Integrated advanced machine learning tools into team workflows and established coding standards, significantly enhancing collaboration and experiment tracking capabilities. Contributed to open-source projects.

*Impact: Improved overall team efficiency, code quality. Reduced On-Call alerts by **95%***

### **Leadership & Mentorship**

Managed interns and Mentored new hires and junior engineers, fostering technical skill development and guiding them through project contributions.

**Machine Learning Engineer, May 2020 - Sep 2021, OkCupid (Match.com), New York City**

1.5Y

### **Discount Optimization** (Python, Keras, TensorFlow, Weights and Biases) [Wide&Deep]

Lead the efforts to optimize subscription pricing(discounts) to maximize the revenue for OKCupid, Implemented end-to-end ML pipelines, feature engineering, modelling, alerting etc.

*Impact: Increased overall revenue by **6%** through A/B testing against assigned prices*

**Machine Learning Engineer, Apr 2015 - May 2020, FactSet, New York City**

5Y

### **Earning Call Speaker Identification** (Python, TensorFlow, Keras) [Spectrograms, CNN] [Presentation](#)

Developed and deployed an end-to-end speaker identification system to identify speakers in real-time during company quarterly earnings calls using computer vision and deep neural networks.

*Impact: In early testing it was estimated to save around **20%** human-hours*

### **Private-company fact extraction** (Python, Keras, Sagemaker, DataBricks) [ELMo, BiLSTM, Blazingtext]

Lead the efforts to extract 'full company name' with key-people, their titles and biographies etc. from 1.6 million crawled and cached websites of private companies.

### **Fuzzy Duplicate Document Identification Service** (Java, Couchbase) [Shingling, Vector Space Models]

Developed full-stack solution to identify the duplicate documents in real time, given a stream of thousands of documents per day

*Impact: **66%** reduction in compute time for document processing. Also, used by **StreetAccount** to find trending news.*

### **Type-Ahead and Query expansion** (Apache Spark, Java, Python) [Distributed Trie, LogisticRegression]

Lead developer for implementing features like Autocomplete Query(Type Ahead) and suggest similar concepts to expand the formulated query for a 'Financial Document Search Engine'

### **Realtime Formula Ranking** (Apache Spark, Python) [N-gram Language models]

Developed the pipeline to cluster users and rank the formulas in the feature of FactSet terminal

*Impact: Average rank brought down from **5.6(ElasticSearch based)** to **2.3(Language Model based)***

**ML Research Engineer, July 2011 - July 2013, Tata Research Development and Design Centre, India**

2Y

### **Event Detection in Time Series** (Java, Python, RapidMiner) [SVM - RBF]

Wrote an algorithm based on Shape Context for finding frequently occurring patterns and events, with as good results as SAX, DTW etc. with **7%** better results in the particular domain of car sensors.

### **Data Harmonization Framework (DHF)** (Java, Apache Pig)

Implemented an ETL framework that exploits the power of map-reduce and big-databases to fuse incongruous enterprise data from disparate sources in near real time.

# Skills

**Languages** Python • Java • C/C++ • Bash • Javascript • HTML • SQL  
**Frameworks** PySpark • Keras • Metaflow • KubeFlow • TensorFlow • PyTorch • MongoDB • FastAPI • Django

# Publications






 [Google Scholar profile](#)

[Inferring Latent Attributes of an Indian Twitter user using Celebrities and Class Influencers](#)  ACM Hypertext 2015  
[Inferring gender of a Twitter user using celebrities it follows](#) CORR 2014  
[Architecture for Automated Tagging and Clustering of Song Files According to Mood](#) IJCSI, 2010

# Education

**Master of Science** in Computer Science, State University of New York, Buffalo, NY 2014  
**B. Tech.** in Computer Science and Engineering, **IIIT**, India 2010

# Personal Projects / Extra Curricular

 **Organizer @ MUFIn** Committee member, organizer and reviewer to the MUFIn Workshop at top conferences, focusing on innovative approaches to modeling uncertainty in the financial sector (*AAAI2023, PKDD2022*)  
 **Lotion** Unofficial Notion.so Desktop app for Linux (*2K+ GitHub stars / 60K+ Clones & Downloads*)  
 **Romadeva** Tool to convert Roman script to Indic(Devanagari) script (*Used by https://translatorswithoutborders.org*)  
 **jTextBrew** A JAVA library for fuzzy string matching, based on TextBrew algorithm by Chris Brew  
 **Quena** Question and Answering system – Indexed 1.6 Million Wikipedia documents, designed a question parser and a ranking algorithm based on popularity. (*Apache Solr, NER, POS tagger*)

# Professional Focus & Evolution

