


## Experience


10+ years

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


4Y

 **Listing IQ: Interactive CMA Platform** (*Django, DocumentDB, PyTorch*) [Real-time Valuations + Embeddings]  
Led proof-of-concept and core architecture for platform now live as Listing IQ CMA – AI-powered tool enabling agents to customize property comparisons with map-based filtering, editable valuations, and amenity-level explanations. Designed APIs integrating real-time Zestimate valuations, property embeddings, and comparative analysis. Mentored junior engineer through production launch.

**Impact:** *POC → shipped product powering agent pricing decisions; foundation for new agent revenue products*


 **Active Listing Comps Engine** (*Apache Spark, Metaflow, H3 Geospatial*) [Similarity Scoring + Batch/REST APIs]  
Architected and built similar listings comparison engine from ground up – batch pipeline and real-time API using H3 geospatial indexing and customizable ranking algorithms. Processes 3M+ listings, powers [Zillow Showcase](#) dashboards (8M monthly views) helping agents demonstrate listing performance vs non-Showcase homes. Later led infrastructure migration and algorithm improvements.

**Impact:** *95% complaint reduction, coverage 90%→98.5%; metrics cited in sales, marketing, and investor communications*

 **Infrastructure & Engineering Leadership** (*Terraform, AWS, Metaflow, Docker*)

Drove CI/CD V4 migration and ML pipeline modernization with zero production incidents. Led alert reduction initiatives and bug triage coordination. Conduct technical interviews, mentor junior engineers, and lead code reviews across team.

**Impact:** *\$500K annual savings, 61% alert reduction (641→249 alerts YoY)*


 **AI/ML Initiatives (In Progress)** (*PyTorch, CLIP, LangChain, Databricks, MLflow*)

Leading Image Embeddings project – designing DualLossAutoEncoder to predict property condition from CLIP embeddings (3.3M listings) for Zestimate integration (DRI, cross-team coordination with computer vision and valuation teams). Co-developed customer care tool alpha – React UI and LLM integration enabling agents to generate personalized, AI-powered Zestimate explanations.

**Impact:** *Advancing visual ML and LLM capabilities for valuation products; first LLM tool deployed to customer care*

**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](#)]

Led efforts to optimize subscription pricing (discounts) to maximize revenue for OkCupid. Implemented end-to-end ML pipelines, feature engineering, modeling, 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

 **ML-Powered Financial Data Extraction** (*Python, TensorFlow, Keras, Sagemaker*) [CNN, ELMo, BiLSTM]

Led multiple ML initiatives: (1) Speaker identification system for earnings calls using spectrograms and CNNs, (2) Private company fact extraction from 1.6M websites using ELMo/BiLSTM. Built end-to-end pipelines from data collection to deployment.

**Impact:** *20% reduction in human-hours for earnings call processing, automated extraction from millions of documents* 



 **Financial Document Search & Ranking Systems** (*Apache Spark, Java, Python*) [Distributed Trie, N-gram LM, Vector Space]

Architected search infrastructure: (1) Type-ahead and query expansion for document search, (2) Real-time formula ranking using language models, (3) Duplicate document detection service. Scaled to handle thousands of queries/documents per day.


**Impact:** *Improved formula ranking from 5.6 to 2.3, 66% faster document processing, powered StreetAccount trending news*

**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 • SQL

**Frameworks** PyTorch/TensorFlow • PySpark • FastAPI • Metaflow • Databricks • MLflow • Django

**AI/ML** CLIP Embeddings • LangChain • RAG • Feature Engineering • A/B Testing • Model Monitoring


**Infra** Docker • Kubernetes • Terraform • GitLab CI • AWS • W&B • Pinecone

**Leadership** System Design • Technical Interviews • Mentoring • Code Reviews • Cross-team Coordination

## 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

 **Workshop Organizer @ MUFin**

Program committee member, paper reviewer at top ML conferences – Workshop on Modeling Uncertainty in the Financial Sector (*AAAI 2023, ECML-PKDD 2022*)

 **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

