Puneet Ludu

puneet.ludu@gmail.com | New York, NY | +1-(716) 867-4344 | puneet.io

https://github.com/puneetsl
https://www.linkedin.com/in/puneetsl
https://www.kaggle.com/puneetsl

Experience — 10+ years

Machine Learning Engineer, *Sep 2021 – Present*, **Zillow (Zestimate)**, Remote — **4Y**

- **Active Listing Comps Engine** *(Apache Spark, Metaflow, H3 Geospatial)* [Similarity Scoring + Batch/REST APIs]
  Architected similar listings comparison engine powering Zillow Showcase's go-to-market strategy. Built scalable batch pipeline and real-time API using geospatial indexing (H3) and customizable ranking algorithms. System processes millions of listings daily, generating business-critical performance metrics.
  Impact: 30% listing wins, 75% engagement lift for Showcase - metrics used across sales, marketing, and investor communications

- **Interactive CMA & Realtime Valuation Platform** *(Django, DocumentDB, PyTorch)* [Siamese Networks + Explainability]
  Architected 0-to-1 platform integrating real-time valuations, property embeddings, and amenity explanations. Designed APIs for comparative analysis and interactive agent reporting. Led technical decisions on ML integration and service architecture. Platform launching soon.
  Impact: Foundation for new agent revenue products, establishing real-time ML serving patterns

- **Zestimate Infrastructure Modernization** *(Terraform, AWS, Kubeflow, Metaflow, Docker)*
  Drove migration of critical ML pipelines from legacy systems to containerized infrastructure. Established CI/CD patterns, monitoring standards, and deployment practices that reduced operational overhead. Mentored team on modern MLOps practices and infrastructure-as-code principles.
  Impact: $500K annual savings, 80% faster deployments, 95% reduction in on-call alerts through proactive monitoring

- **Next-Gen Explainability: Amenity Adjustment & LLM Tools** *(PyTorch, DeepLift, LangChain)*
  Developing ML explainability service for dollar-value feature attributions and LLM tools for automated Zestimate queries. Building APIs for property data retrieval, comparative analysis, and natural language explanations. Both initiatives in beta, targeting improved user trust and reduced support burden.
  Impact: Foundation for automated customer support and transparent valuations

- **Engineering Excellence & Pipeline Stability**
  Established MR best practices (stacked MRs, pre-submission checklists) and documentation standards. Led weekly bug bashes targeting low-effort, high-impact improvements. Mentored interns and junior engineers.
  Impact: Improved Zestimate pipeline success rate from 75% to 85%, reducing production incidents

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 ● Javascript ● HTML ● SQL

Frameworks   PySpark ● FastAPI ● Metaflow ● KubeFlow ● TensorFlow/PyTorch ● W&B ● DataBricks ● Django

AI Dev & Infra   Cursor ● V0 ● MongoDB ● Pinecone ● Docker ● Kubernetes ● Terraform ● Gitlab CI

Leadership   System Design ● Technical Mentoring ● Cross-functional Collaboration ● Code Reviews ● Architecture Decisions

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, JIIT, 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

Career Responsibility Evolution