Tanmayi Balla

TensorFlow Certified Developer

Data Scientist & Machine Learning

Piscataway, NJ 08854 • (812) 778-4651
tballa@iu.edu • https://tanmayiballa.github.io/
https://www.linkedin.com/in/tanmayiballa/

Available to work immediately • Open to relocation within the US.

- Passionate Data Scientist with a knack for solving cutting-edge computer vision problems in 2D/3D object detection, classification, and segmentation. Proven expertise in deploying image recognition in real-time systems **(4 years of research)**.
- Exploring the frontiers of Large Language Models (LLMs), addressing knowledge representation and hallucination issues for downstream conversational chatbots. Proficient in prompt engineering, fine-tuning, and building knowledge graphs through Natural Language Processing (NLP), and demonstrated experience in working with Graph DBs (KGs) - Neo4j and SparkQL.
- Extensive experience building end-to-end machine learning systems for production, with expertise in supervised, and unsupervised models. Gained a deep understanding of the math intricacies of ML algorithms, enabling innovation & research.
- Proven track record of performing in-depth data cleaning, manipulation, and advanced analytics to shed key insights on business strategies and organizational decision-making. One such innovative model is currently **LIVE on the eBay website**.

TECHNICAL SKILLS

Programming Languages:	Python | C++ (Problem-solving, coding) | Java | C | MATLAB | R (Statistics) | JavaScript
Machine Learning	Regression Models (Linear, Logistic, Decision Trees, Random Forest) | Classification Models (K-means Clustering, Logistic Regression, KNN) | XGBoost | Gradient Boosting Methods | Ensemble Models | Support Vector Machine | Time Series Analysis (ARIMA, LSTM, Prophet)
Deep Learning:	BERT Language Models | NLP | Large Language Models (GPT, LLAMA, Falcon, Claude AI) | Neural Networks (Convolutional Neural Networks (CNN), Deep NN, Graph NN, Recurrent NN) | Word/Image Embedding Techniques | LSTMs, GRU | Gen AI (Generative Adversarial Networks, Transformers, Self-attention models, Vision Transformers).
Computer Vision	Object detection, image classification, semantic segmentation (R-CNN Family, Mask R-CNN, YOLO, SSD) | EfficientNet, ResNet, VGG | Vision Transformers | Image Registration.
ML and DL Tools:	PyTorch | TensorFlow | Keras | HuggingFace | OpenCV | NLTK | Spacy | Numpy | Pandas | Matplotlib | Seaborn | scikit-learn/sklearn | Scipy | CUDA | XGBoost | Prophet | Plotly | Theano
Data Engineering & cloud:	SQL | PostgreSQL | Apache Spark | PySpark | Hive | Hadoop | MapReduce | Salesforce (Marketing Cloud, Automation Studio, Analytics QL, CRM) | Azure (Databricks, Data Factory, Data Lake) | AWS (S3, EC2, Lambda, Sagemaker) | Google Certified ACE - GCP.
Data Visualization & other tools:	Tableau | Power BI | Statistical Analysis using R (ggplot) | Hypothesis testing | Microsoft Office (Advanced MS Excel) | JIRA | CUDA | Git | GitHub | GitLab | Putty | WinSCP | Docker

WORK AND RESEARCH EXPERIENCE

Data Analytics - Marketing Data Analyst CRM **|** ***Part-Time***, Indiana University CRM, Bloomington, IN **Dec 2022 – May 2024**

Optimized data management processes for **350+ business units** in Salesforce Cloud, utilizing SQL to streamline dataflow. Developed **100+ automated data pipelines**, reducing manual efforts by **45%** and generating efficiencies across university-wide email marketing.

- Boosted **CTR by 34%,** uncovered key insights of campaign efficiency in prompting payments using data mining and EDA.
- Enhanced email personalization with target user segmentation by optimizing SQL queries for a **15% faster data retrieval**.
- Spearheaded Salesforce business unit migration and achieved **20%** reduced operations overhead and enhanced collaboration.
- Performed large-scale data analytics on relational databases using SQL through email marketing metrics, and conversion rates to synthesize insights from analysis and strategically influence campaigns to achieve a **24% boost in open rates**.
- Handled the creation and deployment of **150+ reports** in CRM and MC. Constantly monitored analytics dashboards and performed thorough root-cause analysis to identify, resolve, and prevent discrepancies for over **180+ requests** in **JIRA.**
- ***Key Skills:*** Python, SQL scripting, APIs, JavaScript, Salesforce (Cloud, Analytics Query Language, CRM), data engineering.

Data Science - Machine Learning Intern, eBay Inc., Bellevue, WA **May 2023 – Aug 2023**

Collaborated with cross-functional teams to perform rigorous research and development and implemented scalable machine learning

predictive models with **98% accuracy**, offering solutions for personalized recommendation applications to global buyers at eBay based on customer purchase behavior and demand forecast. ***Tech Stack:*** Python, SQL, XGBoost, BERT, ARIMA, LSTM, K-means.

- Achieved a high clustering accuracy of **95%**, trained an unsupervised ML (K-means) to segment US regions and forecast top-ranked items (eBay Motors) per segment, thereby identifying buyer needs and triggering seasonal recommendations.
- Employed XGBoost and ARIMA to forecast item category spikes (peak months), enabling targeted inventory strategies for eBay sellers. This regional forecasting anticipated a **30% improved seller experience** (high sales, low item burn rate).
- Scaled the ML model using Spark, enabling data profiling of eBay's large-scale high dimensional datasets **(50M+ rows)**, thereby transforming raw data into a structured form and boosting the efficiency of multiple ML models of internal teams.
- Boosted the model performance to **98.5%** and extracted insights into key performance indicators (KPIs) by utilizing quantitative analytics, statistical modeling, time series forecasting, and deep learning techniques on the transformed data sets.
- Delivered a **22% increase in user preference matching** and partnered cross-functionally to collect, clean, and transform large-scale raw unstructured data, and strategically designed quantitative experiments to extract meaningful features.
- Demonstrated the model's business potential to the leadership using data visualization, thereby expanding its scope at eBay.
- ***Key Skills:*** Supervised, unsupervised learning, Spark, PySpark, SQL, NLTK, tf-idf, spacy, pandas, NLP, parametric (ARIMA, ES), and non-parametric distributions (LSTM, SVR, NN), statistical analysis using hypothesis testing (ADF, KPSS).

Data Science - Imagineer, Fractal Analytics, Bengaluru, India **May 2021 – Aug 2022**

Accomplished high client satisfaction by delivering an end-to-end deep learning solution with an average **efficiency of 97%**, flexible for **7** different product categories. Extensively utilized state-of-the-art object detection, object segmentation, and object classification.

- Led end-to-end training, development, and execution of computer vision classifiers using product cover images, for product tracking in retail store shelving. Defined a custom loss function to reduce the **FPR by 12%** and improve **precision by 9.5%.**
- Crafted PySpark codes in Databricks to analyze tables stored in Azure Data Lake. Worked collaboratively to develop data pipelines to perform ETL and testing strategy to generate **Key Performance Indicators (KPIs)** for the PowerBI dashboard.
- Delivered a UI that accelerates image data analytics by **45%**, simplifying product recognition and statistics of item sales.
- Established a workflow in Azure Form Recognizer, using NERs to automate license extraction from multilingual documents.
- Performed large-scale AI training and inference to create **12+ robust object detectors**, competing with SOTA algorithms.
- Utilized Image registration, generation, and reconstruction to build a multi-modal classification model to validate rooftop damages for an insurance client. Trained using geospatial images from different modalities (DSM, DTM, IR).
- ***Tech Stack:*** PyTorch, TensorFlow, Pandas, GPUs - CUDA, Git, EasyOCR, Keras, Azure, Docker, JSON XML data parsing.

Computer Vision - Jr. Research Assistant, Humanitarian Technology (HuT) Labs, Kollam, India **Sept 2018 – May 2021**

- Formulated and modeled innovative algorithms for selective pixel extraction, leveraging color prediction using the Mask R-CNN's semantic segmentation mask, thereby outperforming SOTA by **20% in time and 8% in accuracy**. Journal.
- Enhanced the **Human-Robot interaction accuracy to 89%** by implementing a leg-tracking algorithm to identify humans.
- Trained **5000+ images** and deployed object detection and classification using YOLOv3 and CUDA, for a real-time robot.
- **AI-based** autonomous robot for the elderly and bedridden: Implemented real-time 2D/3D vision capabilities, speech-to-text (Google STT), navigation (object tracking using SLAM), and developed an end-to-end pipeline using Python, C++, and ROS.
- **Handled the development of real-time Robotics code in Python and C++** to program the robot to perform vegetable sorting, chopping, item picking, opening a door, welcoming guests, handshake, etc. using speech and GUI instructions.
- ***Tech Stack:*** Tensorflow, PyTorch, Python, C++, ROS,Git, Linux, Docker, Natural Language Processing, Speech Recognition.

EDUCATION

Master of Science in Data Science (Computational Sciences), Indiana University Bloomington, Indiana **Aug 2022 - May 2024**

Coursework: Machine Learning (math-based), Deep Learning, Computer Vision, Statistics (R), Knowledge Graphs and LLMs, Competitive Programming, Algorithms, Advanced Database Systems, Engineering Cloud Computing. ***GPA***: 3.97/4.0

Natural Language Processing Lab **| Research Assistant | Volunteer | Computer Science Dept.** **Aug 2023 - Present**

- Architected and implemented robust testing pipelines for LLMs in Natural Language Understanding, achieving a **25% performance boost** and an **85% reduction in hallucination rate** through relation extraction, prompting, and fine-tuning.
- Deployed LLMs: Llamav2, GPT, using Nvidia Tesla A100 GPUs and leveraged its performance using fine-tuning, prompt engineering, and RAG. Designed an end-to-end chatbot using Neo4j KG and OpenAI APIs to optimize information retrieval.
- Web scraped **1000+** SEC 10-K filings to build a knowledge base, assisting investors in identifying potential business risks.
- ***Skills:*** OpenAI, GPT, LLAMAv2, Falcon 180B, Claude AI, Python, langchain, Neo4j, Cypher, NetworkDB, LoRA, RAG.

Bachelor of Technology in Electronics and Communication Engineering, Amrita University, India **July 2017 - May 2021**

- ***Robotics Lab*:** Integrated deep learning stack to real-time robots and got shortlisted for the RoboCup German Open 2020.

PUBLICATIONS & CERTIFICATIONS

- **TensorFlow Developer Certificate, issued by TensorFlow, valid until May 2027.**
- **Journal: Optimized and Efficient Color Prediction Algorithms Using Mask R-CNN, Electronics, MDPI**
- **ROS Based GUI Controlled Robot for Indoor Mapping and Navigation, Springer, (ICCNCT-2020)**
- **Speech and Kinova Arm-Based Interactive System with Person Tracking, Springer, (ICCNCT-2020)**