

Avi Tewari

Irvine, CA | avit@uci.edu | +1(669) 221-0001 | www.linkedin.com/in/avi-tewari/ | <https://github.com/AviTewari>

EDUCATION

University of California, Irvine, CA

Dec. 2024

Master's in Data Science

Relevant Coursework: Database Management System • Statistics And Probability • Linear Algebra • Artificial Intelligence

SRM Institute of Science and Technology, Chennai, TN, India

May 2023

B-Tech in Computer Science Engineering

Relevant Coursework: Data Structures and Algorithms • Data Mining • Advanced Programming Practices

SKILLS AND INTERESTS

Tools & Frameworks: AWS (EC2, S3, EKS, Elastic Map Reduce) • TensorFlow • Keras • PyTorch • Pandas • NumPy • OpenCV • Scikit Learn • NLTK • spaCy • Docker • Kubernetes • MySQL • PostgreSQL

Programming languages: Python • SQL • R • C++ • C • Git (Version Control)

Certifications: AWS Certified Solution Architect - Associate • Tableau Desktop Specialist

WORK EXPERIENCE

UCI School Of Education, *Machine Learning Researcher*, Irvine

Jun. 2024 - Present

- Enhanced **data preprocessing** (tokenization, lemmatization, stop-word removal) and **feature extraction** (TF-IDF, word embeddings), improving classification accuracy of collaborative and cognitive behaviors in classroom interactions by 15%.
- Developed and fine-tuned **Large Language Models (LLMs)** including **BERT** and **GPT** using the **Hugging Face API** for classifying student-teacher interactions, achieving an accuracy of 73% in identifying collaborative and cognitive behaviors.
- Architected **APIs** for real-time **model deployment** of BERT and GPT models into monitoring systems, enabling automated analysis of classroom interactions.

Expand AI, *Data Science Intern*, India

Nov. 2021 - Feb. 2022

- Processed and analyzed a dataset of 32,000 images using **OpenCV** and **Scikit-learn** to enhance image quality and standardize resolution, ensuring high quality inputs for Deep Learning models.
- Constructed a **Convolutional Neural Network (CNN)** using **TensorFlow**, improving data labeling accuracy by 21% by optimizing hyper-parameters such as learning rate, batch size, and dropout rate.
- Established **CI/CD** pipelines using **Jenkins** to streamline the deployment process for the ML model, reducing deployment time by 30%.

LTIMindtree, *Data Science Intern*, India

Aug. 2021 - Sept. 2021

- Directed data collection by web scraping images from diverse online sources to develop an Image Analysis System on **Microsoft Azure** using **Custom Vision**, achieving a 40% reduction in verification time for adherence to organizational dress code policies.
- Implemented **ETL workflow** for data extraction, deduplication, and database loading on a dataset of 5000 customer complaints, and executed exploratory data analysis (EDA) using text mining (Bag of Words and Tokenization) to derive insights.
- Formulated a **Named Entity Recognition (NER)** system using **NLTK** and **spaCy** to streamline complaint classification, facilitating routing to the appropriate department and reducing response time for handling customer feedback by 30%.
- Leveraged **GitHub Actions** to create a **CI/CD** workflow for automating the testing and deployment of the NER system.

PROJECTS

Knowledge Retrieval Assistant

Jun. 2024 - Jul. 2024

- Created a **LangChain** Knowledge Retrieval Assistant app using **LLMs**, leveraging **Pinecone vector database** for efficient embeddings and vector retrieval.
- Implemented a chain mechanism by embedding queries into vectors, retrieving relevant chunks from a **Pinecone vector database**, and using this context to generate accurate answers with minimal API calls, enhancing efficiency and response speed.

Spark-AWS Movie Recommender

Jun. 2024 - Jun. 2024

- Developed a movie recommendation system using **PySpark** and **SparkSQL**, running on a **Spark** cluster deployed on **AWS EMR**.
- Utilized **ETL** processes to extract 1 million ratings data from **AWS S3**, transform the data with advanced analytics to optimize for **collaborative filtering**, loading it into the recommendation system for improved personalization.

Student Attentiveness Monitoring in Online Learning

Mar. 2024 - May 2024

- Fine tuned various architectures like **VGG-16**, **VGG-19**, **DenseNet**, **ResNet** achieving a maximum accuracy of 64% in discerning students' attentiveness in online classroom through facial emotion analysis.
- Built a **CNN** Model in Jupyter Notebook using **Tensorflow**, **Keras** and **Scikit Learn**, attaining 79% accuracy, leading to improved engagement strategies.
- Engineered **REST APIs** for integration of the attentiveness monitoring model with online learning platforms such as Coursera, enabling real-time analysis and feedback on student attentiveness while watching learning online.