Xenoss
High-load and big data engineering

Alex D.

Senior Machine Learning Engineer

Summary

- 7 years of experience in a machine learning engineer position
- Experiencing in managing AI engineering teams and coordinating with product teams
- Extensive experience in creating, testing, and fine-tuning models for MarTech and AdTech
- Deep understanding of the AdTech/MarTech business domain
- Experience in multiple machine learning domains: Deep Learning, Computer Vision, Recommendation systems, Demand forecasting, etc.
- Guiding projects throughout the entire life cycle, from concept design to development, testing, and scaling the solution

Skills

Fundamentals: Concurrency, Networking (TCP, HTTP, UDP), High-load, Big data...

Programming languages: Python, R, Matlab, C++, CSS3, HTML5, JavaScript

Platforms: ML/MLOps libraries and platforms: TensorFlow, PyTorch, scikit-learn, Tensorflow, Caffe, Theanos

Frameworks: ReactJS, Spring Boot

Areas of expertise: Reinforcement Learning (RL), Large Language Models (LLMs), Robotics, Computer Vision and Graphics (including Generative Algorithms)

Distributed storage and database systems: SQL or NoSQL, MySQL, or Cassandra

Big data: Spark, Hadoop, Parquet

Clouds: AWS (EC2, Route 53), GCP (BigQuery, BigTable)

Machine learning hardware: Tensor GPUs or AI chips

Devops: Docker, Kubernetes, Terraform

Job experience

Data Scientist at Xenoss

Sift Media, Jan 2021 – Now

Product

Ad-Lib.io is a cutting-edge platform that leverages advanced machine learning technology to revolutionize digital advertising. It focuses on personalizing and optimizing ad content in real-time, using AI algorithms to analyze user data and context. The platform's capabilities enable continuous improvement of ad performance, learning from user interactions to refine targeting strategies and creative elements to maximize the impact and efficiency of digital advertising campaigns.

Role description

My role as an AI Engineer involved developing and refining the machine learning algorithms crucial for the dynamic personalization of advertisements. I focused on creating models capable of analyzing large datasets of user interactions and demographics, enabling the system to understand and predict consumer behavior and preferences accurately. This involved processing and interpreting complex patterns in real-time data streams, including browsing habits and engagement metrics, to tailor ad content to individual users. Additionally, I worked on optimizing the algorithms for scalability and efficiency, ensuring they could adapt and respond swiftly in a high-traffic, real-time bidding environment.

Responsibilities and achievements

- Using Machine Learning and LLMs to improve the quality of contextual targeting.
- Working towards multimodal intent understanding, video understanding, entity mining.
- Use algorithmic optimization, big data analysis, simulation pipelines, and embedded devices.

Tech stack

- Cloud Infrastructure: Amazon Web Services (AWS), Google Cloud Platform (GCP)
- Data Storage and Databases: MongoDB, PostgreSQL, Apache Kafka
- Machine Learning and Data Processing: TensorFlow and PyTorch,
- Data analysis: Pandas and NumPy

Data Scientist at GlobalLogic

Building a content personalization system, Feb 2020 – Jan 2021

Product

I was involved in designing an AI-driven content personalization platform that leverages machine learning algorithms to curate and recommend media content tailored to individual user preferences, such as movies, TV shows, and music. The platform analyzes user behavior, viewing history, and engagement patterns to deliver highly relevant and engaging content recommendations.

Tech solution

My work involved crafting and refining Large Language Models (LLMs) for both visual understanding and natural language understanding, ensuring a seamless and intuitive user experience. Throughout the project, I was responsible for authoring detailed specifications for new features and improvements, ensuring our solutions not only met but exceeded industry standards. This role required a blend of technical expertise and creative problem-solving as we pushed the boundaries of what's possible in AI-driven content discovery and distribution.

Personal achievements

- Created and fine-tuned deep learning models for image and video analysis to power the visual search component.
- Implemented and optimized Large Language Models (LLMs) to improve the system's ability to understand and process user queries in natural language.
- Integrated AI models into existing infrastructure to enhance the search and discovery experience without disrupting current operations.
- Prepared detailed documentation on model development processes and reporting on performance metrics to stakeholders.

Tech stack

- Programming languages: Python, Java, SQL
- Frameworks and libraries: TensorFlow, PyTorch, Keras, OpenCV, Pandas, NumPy, Scikit-learn, NLTK, SpaCy, Hugging Face's Transformers
- Data management and processing: Apache Kafka, Apache Spark, Hadoop
- Cloud and DevOps: AWS, Docker, Kubernetes, Git, Jenkins
- Database technologies: PostgreSQL, MongoDB, ElasticSearch
- Monitoring and logic: Prometheus, Graphana, ELK Stack

Senior software engineer at TCS

Personalized targeting and content delivery, Jan 2019 – Jan 2020

Product

This product harnesses machine learning and data analytics to offer hyper-personalized ad targeting and content delivery. Its core features include a real-time bidding system that optimizes ad placements for maximum viewer engagement and ROI, proprietary machine learning models that analyze, consumer behavior, preferences, and sentiment, as well as data visualization tools.

Tech solution

The system harnessed real-time data processing, using Apache Kafka streams to capture and process user interaction data continuously. This data, along with historical user data stored in a scalable MongoDB database, was pivotal in training our models. For feature engineering and data manipulation, tools like Pandas and SQL were employed, handling datasets with millions of rows and numerous attributes.

Our model's predictive capabilities were enhanced by incorporating NLP algorithms using SpaCy and GPT-4, enabling the system to understand and interpret user queries and content with remarkable accuracy. This was crucial for the content recommendation engine, which dynamically adjusted the content displayed to each user based on their unique profile and interaction patterns.

Personal achievements

- Built ML models for content targeting that increased content personalization accuracy by 40%,
- Engineered a real-time data processing pipeline using Apache Kafka, which reduced data latency, enabling quicker and more efficient user data analysis for personalized content delivery
- Pioneered the integration of NLP for content generation

Tech stack

- Programming Languages: Python, JavaScript, SQL, R
- Databases: MongoDB, PostgreSQL, Redis
- Machine Learning Frameworks: TensorFlow, PyTorch, Scikit-Learn: For classical machine learning algorithms.
- Natural Language Processing: SpaCy, NLTK, GPT-3
- Data Processing and Analysis: Pandas, NumPy, Apache Spark
- Real-time Data Processing: Apache Kafka, Apache Flink
- Containerization and Orchestration: Docker, Kubernetes

Machine Learning Intern at DataRobot

Oct 2018 – Jan 2019

Product

During a 3-months internship, I worked on multiple projects in multiple machine learning domains. Here are the key milestones of the program:

- Assisted in the design and implementation of a machine learning-based recommendation system,
- Conducted data preprocessing, including cleaning, normalizing, and transforming large datasets for model training using Python and its libraries.
- Collaborated in the development of neural network models using TensorFlow, significantly enhancing the system's predictive accuracy.
- Gained practical experience with AWS for deploying and testing models, understanding AI systems' scalability and real-world application.
- Developed a feature engineering strategy that effectively enhanced model performance.
- Participated in cross-functional team meetings, gaining insights into the end-to-end product development process.
- Created a dashboard for model evaluation and analytics, using Python visualization tools like Matplotlib and Seaborn.
- Contributed to the documentation of model development processes, ensuring clarity and future usability.

This internship was a helpful learning experience, providing me with a comprehensive understanding of AI engineering in a commercial setting, and the opportunity to contribute to building machine and tesing learning models.

Few older projects have been skipped not to clutter up the CV

Education

Doctor of Philosophy - PhD, Data Science

Bachelor of Science | Computer Science

Certifications

Google Professional Machine Learning Engineer

NVIDIA Deep Learning Institute Certifications

TensorFlow Developer Certificate by TensorFlow