

Junior NLP Data Scientist

Liam Patel

Professional summary

Innovative Junior Data Scientist specializing in Natural Language Processing (NLP). Skilled in designing and training machine learning models to analyze unstructured text data. Experienced in Python, TensorFlow, and NLP libraries such as SpaCy and NLTK.

Experience

NLP Data Scientist Intern

June 2024 - December 2024

LingoAI / San Francisco, CA

- Built a named entity recognition (NER) model using SpaCy, achieving 92% accuracy on annotated datasets.
- Developed a sentiment analysis pipeline for customer reviews, improving customer satisfaction metrics by 10%.
- Deployed machine learning models on AWS, streamlining API integration for clients.
- Conducted fine-tuning of transformer models (BERT) for text classification, increasing accuracy by 15%.

Graduate Assistant (NLP Research)

September 2022 - May 2023

Stanford University / CA

- Conducted research on transformer-based models (BERT, GPT) for text summarization tasks.
- Designed and implemented an unsupervised topic modeling algorithm, reducing processing time by 30%.
- Created custom preprocessing scripts to clean noisy text data from social media platforms.
- Published research findings in Computational Linguistics Quarterly.

Projects

- Sentiment Analysis Dashboard.** Built a real-time sentiment tracker for social media, leveraging MongoDB and Flask.
- Topic Modeling for Legal Documents.** Created a Latent Dirichlet Allocation (LDA) pipeline to categorize legal texts.
- Chatbot Development.** Designed a chatbot using Rasa and TensorFlow to assist customers with FAQ support.

(415) 555-1234

liampatel@gmail.com

San Francisco, CA

Links

GitHub

LinkedIn


Education


Master of Science in Computer Science (Specialization: Machine Learning)


Stanford University - Stanford, CA


Graduated: Dec 2024

Skills

Programming: Python (NLTK, SpaCy, TensorFlow, PyTorch) 

NLP: Sentiment Analysis, Named Entity Recognition (NER), Topic Modeling 

Data Preprocessing: Text Cleaning, Tokenization, Lemmatization 

Databases: MongoDB, SQL 

Tools: Git, Jupyter Notebook, AWS 