Natasha Singh

812-650-2032 | Email | LinkedIn | GitHub | Portfolio

Education

Indiana University - Bloomington — IN, USA
MS Computational Linguistics - Artificial Intelligence — *Aug. 2022 – May 2024*

BITS Pilani — Hyderabad, India
Bachelor of Engineering, Minor in Finance — *Aug. 2016 – June 2020*

Experience

Research Scientist - NLP, LLM — Aug. 2023 – May 2024
NLP Lab, Indiana University — *Remote, USA*

- Contributed ⟨elided sentences, reconstructed sentence⟩ pairs in English, Hindi, and Kumaoni languages for The Hoosier Ellipsis Corpus (THEC) project to train LLMs to reconstruct the sentence to address parsers' inability to parse elided sentences.
- Conducted experiments to optimize prompts for LLMs and utilized spaCy to extract triplets by implementing Co-reference resolution, Dependency parsing, and Named Entity Recognition for NLP pipeline.
- Extended goal of the 'Triplet Generation for Knowledge Graphs' project was to train a GPT model that first extracts relevant triplets from the knowledge graph based on the question and then generates an answer using those extracted triplets, thus minimizing hallucinations and providing accurate responses.

Open Source Developer — May 2023 – Aug. 2023
Google Summer of Code — *Remote, USA*

- Developed a rule-based morphological analyzer for the Kumaoni language to capture intricate morphological inflection and case marking, achieving 95% token coverage and 99.78% F1 score. Application includes machine translation, POS tagging, and morphological generation.

Data Analyst II - Machine Learning | *Python, TensorFlow, PySpark, AWS* — Sep. 2020 – Jul. 2022
Epsilon - a Publicis Groupe subsidiary — *Bengaluru, India*

- Developed an attention-based deep learning model for product recommendation at BJ's, leveraging customer demographics data and past purchase history, resulting in 9% increase in sales.
- Implemented an LSTM-based offer recommendation model for Dunkin, achieving a 7.3% increase in sales. Leveraged customer purchase history & offer redemption patterns, considering factors such as offer duration, bonus points, & free items.
- Engineered a BERT-based model to generate personalized email subject lines for Maurices, resulting in a 12% improvement in email campaign effectiveness. Utilized facets like discount, personalization, urgency, etc to maximize open rates.

Projects

Segmentation and Language Identification | *Python, TensorFlow, CNN, LSTM* — Jan 2024 – May 2024

- Designed a CNN model for language identification using spectrogram images of audio, and an LSTM model trained on sequence of MFCC coefficients and PLP features, achieving 98.5% and 92% classification accuracy respectively.
- Enhanced language identification capabilities for language-switching audios, accurately identifying language spans within audio content and achieving a 78% exact match accuracy.

Named Entity Recognition | *Python, PyTorch, Stanza, PyDelphin, Bi-LSTM, GRU* — Jan. 2023 – May 2023

- Investigated the influence of adding syntactic information on BIO tags prediction for Per, Org, Loc, Time and GPE entities by training RNN model on syntactic information from Dependency Grammar & Head-Driven Phrase Structure Grammar, attaining 98% accuracy.

Code Generation: Text to SQL | *Python, PyTorch, BERT* — Aug. 2022 – Dec 2022

- Formulated SQL generation from Natural Language queries as a Question-Answering task, achieving 92% accuracy.
- Extracted text spans from the context using BERT-based Machine Reading Comprehension and Association Rule Mining technique.
- Employed a Slot-filling approach to construct comprehensive SQL queries.

Technical Skills, Awards and Publications

Languages & Frameworks: Python, SQL (Postgres), PySpark, PyTorch, TensorFlow
Libraries: pandas, NumPy, Matplotlib, scikit-learn, spaCy, NLTK, Stanza, SciPy, seaborn
Coursework: Syntactic & Semantic Analysis, Adv. NLP, Applied ML, Deep Learning Systems, Signal & Image Processing, Knowledge Graphs & LLMs, Applications of AI, Operations Research, Calculus, Probability & Statistics
Awards: Student of the Year award by The Times of India (2014)
Publication: IU-NLP-JeDi: Investigating Sexism Detection in English and Spanish. (CLEF 2023 Paper Link)