Anisha Yidala

(858)-261-8590 anishayidala123@gmail.com /in/anisha-yidala

Education

MS in Computer Science | George Mason University, USA January 2021 - December 2022

B.Tech. in Computer Science & Engineering | GITAM University , India June 2015 – April 2019

Theory and Application of Data Mining, Machine Learning, Deep Learning, Advanced Natural Language Processing, Artificial Intelligence, Mining Massive Datasets with MapReduce

Work Experience

Machine Learning Software Engineer | Tech Talent Connect , California , USA. May 2023 – Present

- Spearheading the research and development for a suite of Machine Learning systems including intent and entity recognition, and sentiment analysis to empower a **Text-Based Virtual Assistant** based on state-of-the-art **Large Language Models (GPT)**
- Developed a hybrid flow that leverages **GPT** to understand the context and uses **MLP** for Intent Recognition that is performing with **83% accuracy** on an internal dataset of 103k texts.
- Designed a **data analytics wash boarding tool** to monitor networked devices in a Hybrid Infrastructure environment that directly resulted in **40% reduction** in system downtime by helping network admins monitor resources more accurately.

Software Engineer | Tata Consultancy Services, Maharashtra , India. January 2020 – January 2021

- Designed diverse set of Validation Controls for robust form validation based on **jQuery** and **JavaScript** while developing dynamic webpages in **HTML5, CSS3, AJAX, JSON, and XML**.
- Drove development and deployment of numerous Cloud based Micro Services on **Azure** in a **DevOps** environment with special emphasis on automated **CICD pipelines** for seamless deployment and maintenance.
- Developed efficient QA systems for unit and integration testing for DataPower services with automated DIT testing, code discovery, and unit test coverage reporting.
- Orchestrated comprehensive system studies, requirements consolidation while establishing effective team communication and client satisfaction through **Agile methodologies**.

Projects

Identifying Italian Language Dialects with Transformers November 2022

Python , PyTorch ,NumPy ,NLTK, BERT

- Developed a **BERT based transformer** architecture to identify 11 unique Italian dialects from a corpus of **260k documents** and achieved a cross category **accuracy of 81%.**
- Leveraged NLP techniques such as **tokenization** and **custom learned embeddings** to pre-process a Wikipedia corpus to train the model at scale on cloud GPUs.
- Developed baseline models by **fine-tuning pre-trained transformers** on **HuggingFace** to understand and compare varying levels of Italian dialect text identification capability of our model.

News Authenticity Prediction August 2022

Python , Pandas , Sklearn ,Spacy

- Designed a **2 stage classification pipeline** to separate embedding and actual classification task for predicting news' authenticity.
- Developed a separate **embedding pipeline** starting from off the shelf **GloVe embeddings** and eventually fine-tuning a large pre-trained **transformer model** on our corpus to extract embeddings from tokenized news text.
- Achieved **78.7% accuracy** in the final News Authenticity prediction task by training an **LSTM based classifier** that accepts a fixed length embedding and outputs a 2 class prediction label.

Poetry Generation March 2022

Python , HMM , NLTK

- Crowd sourced poems from the internet of 2 poets from different eras and developed a corpus of over 1000 poems per poet.
- Developed a **tokenized and cleaned dataset** from the raw 2231 poems and employed exploratory data analysis to understand implicit trends such as excessive use of tones, words, or arrangement of words.
- Developed a learned classifier using **Hidden Markov Models** to classify given a stanza from the poem, who the most likely poet is by learning the probability distribution of words used by each poem.
- Extended the Hidden Markov Models to generate poems given a cue input for each poet and a target topic. The poems generated followed proper grammar and aligned closely with the writing style of each of the poets, providing us with meaningful novel poetry.

Credit Risk Mitigation System September 2021

Python , Pandas , Matplotlib , Sklearn

- Developed an **ensemble system** for identifying credit fraud risk per credit card application using a combination of **SVM, Decision Trees, Random Forests, and XGBoost Classifiers** and achieved a **confidence of 87.5%** in identifying potential credit fraud risks.
- Designed a **dashboard** to understand which factors are affecting an individual's credit rating by performing **correlation analysis** on input data to provide interpretability to the system such that it becomes easier for the user the understand why they were flagged as potential risk.

Skills

Languages and Frameworks: Python, PyTorch , TensorFlow, Java, JavaScript, SQL, HTML, CSS.

Software and Tools: AWS, Azure, Hugging Face , COSMOS, PyTest, Git, GitHub, VS Code, PyCharm, Linux.

Libraries: Pandas, NumPy, Matplotlib, Scikit-learn, NLTK, SPacy, CNTK, fastText , Flair, Seaborn, Flask, Keras.