Pratik Dani

+918233773653 | danipratik91@gmail.com | Ram Dev Road, Pali, Rajasthan

Summary

Highly skilled Data Science professional with a proven track record of leveraging advanced NLP techniques and machine learning models to optimize various business processes. Experienced in fine-tuning models such as DistilBERT, Albert, and XLNet for tasks including text classification, named entity recognition, and document classification. Proficient in deploying models via Flask APIs and building automation pipelines using tools like Bitbucket, Docker, Jenkins, and OpenShift. Demonstrated success in increasing efficiency and productivity, with a strong focus on delivering tangible business outcomes. Passionate about driving innovation and continuous improvement in data-driven decision-making processes.

Work History

Accenture | Gurgaon, Haryana
Data Science Consultant | 08/2021 - Present

- Developed a natural language to SQL query generation tool for strategy team using OpenAI GPT model.
- Developed a tool for extracting and creating Job responsibilities, summaries, skills, other attributes. Helping strategy team to create JD's more efficiently.
- Used Langchain, vector database to create a streamlit application for getting results based on user query.
- Employed fine-tuning techniques on a **DistilBERT** model, a **text classification model**, to categorize individual sentences obtained from the data provider into job summary, responsibilities, and other sections of the JD's.
- Utilized a fine-tuned **Albert** Named Entity Recognition (**NER**) model to extract named entities from the input data, followed by classification using an **XGBoost** model to categorize the extracted entities into behavioral, functional, and technical skills.
- Leveraged the same fine-tuned **Albert NER** model to extract information pertaining to certifications, qualifications, and experience from the input data.
- The JPA Orgstudio Accelerator revolutionized HR Professionals and Talent Acquisition Teams' efficiency, **increasing JD creation speed by 200%** and **reducing time-to-hire by 50%**.
- Employed embeddings from **vector databases** to capture semantic representations of skills and job titles, optimizing the extraction process for enhanced accuracy and efficiency in generating comprehensive job descriptions.
- Extracted data and tables from **images** and **PDF** using **Detectron2, Tesseract** and **OpenCV**.

DataMagnet | Gurgaon, Haryana
CEO and Founder | 10/2023 - Present

- Developed scalable RESTful APIs using Python, deployed on Google Cloud Platform (GCP).
- Created a robust frontend using Next.js, facilitating user-friendly sign-up and API key management.
- Integrated AWS Cognito for secure user authentication.
- Architected real-time API log monitoring and implemented Stripe for payment processing.
- Executed targeted cold email campaigns, leveraging automation tools, resulting in a 300% increase in user sign-ups and a 200% boost in website traffic.
- Negotiated and closed high-value contracts with data vendors, securing annual agreements exceeding $10,000.

Capgemini | Gurgaon, Haryana
Data Scientist | 11/2019 - 08/2021

- Employed finetuning techniques on an **XLNet** model using the **MNLI** dataset augmented with labeled data from Subject Matter Experts (SMEs) to classify documents, discerning compliance status accurately. We were able to achieve **97+ F1 score**.
- Achieved robust performance in document classification, ensuring precise identification of **compliant** and **non-compliant** documents.
- Deployed **XLNet model** via the **Flask API** for efficient and scalable compliance assessment.
- Wrote a **bulk scheduling pipeline** to process **1000+** documents on weekend.
- Transformed the document classification process from manual review to **automated assessment**, resulting in a significant **180% increase in efficiency.**
- Deployed the Flask API using a seamless pipeline comprising **Bitbucket**, **Docker**, **Jenkins**, and **OpenShift**, ensuring efficient and automated deployment processes.

Xceedance | Gurgaon, Haryana
Data Science Analyst | 06/2018 - 11/2019

- Trained an **LSTM NER** model to extract 27 named entities from email bodies, achieving an **F1 score of over 94**.
- Preprocessed email data to enhance model performance by removing irrelevant content and tokenizing text.
- Utilized **Word2Vec** to generate word embeddings, capturing semantic information from the text.
- Deployed the **LSTM NER model** as a **Flask API** endpoint, facilitating seamless integration with other applications for **real-time** entity extraction.
- Developed a **user interface** (UI) for iNext using Flask API to retrieve results from the LSTM models.
- By integrating the **new LSTM model and intuitive UI**, insurance operation teams doubled their application processing capacity, experiencing a **100**% increase in efficiency.

Education

University of Petroleum and Energy Studies | Dehradun, Uttrakhund
Applied Petroleum Engineering | 05/2018
- Graduated with — 3.23+ GPA

Skills

SQL, PostgreSQL, Python, Deep Learning, Flask, React, MongoDB, JavaScript, Redux, Machine learning, Python, AWS