Sarthak Singh

asarthaks@gmail.com

asarthaks

+49-15566271040

Education

- **Universität Stuttgart** — Stuttgart, Germany
  M.Sc. Computational Linguistics — *October. 2023 – Present*
  - Courses: Machine Perception and Learning, Deep Learning, Modelling Semantic Plausability
- **Guru Gobind Singh Indraprastha University** — New Delhi, India
  B.Tech. Computer Science and Engineering — *July. 2015 – June. 2019*
  - Courses: Machine Learning, Artificial Intelligence, Applied Linear Algebra, Data Structures & Algorithms

Skills Summary

- **Languages**: Python, C, SQL, Unix & Shell Scripting, Latex, JavaScript
- **Tools**: Numpy, Pandas, PyTorch, Transformers, Langchain, Scikit-Learn, Matplotlib, XGBoost, OpenCV, Huggingface, Git, Pytest, NLTK, SpaCy, Flask, Kubernetes, Docker, GIT, Airflow, AWS, GCP, Django, PostGreSQL
- **MLOPs**: Airbyte, Great Expectations, Weights & Biases, MLFlow, DVC, Sagemaker, Lambda, ECR, ECS, Docker, FastAPI, Streamlit, Gradio, AirFlow, KubeFlow, Kafka, Apache, Gunicorn
- **Technical Skills**: Machine Learning, Deep Learning, Natural Language Processing, Model Deployment, Model APIs

Experience

- **Gigaforce Private Limited** — Noida, India
  Senior Software Engineer - AI/ML — *October 2022 - September 2023*
  - **SubroGPT**: Designed and curated a Custom Dataset using **Falcon 40B** for Subrogation Domain. **Instruct fine-tuned** an open source LLM with 7B parameters using **PEFT** and **LoRA** to perform **transfer learning**. Applied it to several downstream tasks like **email generation, chain of thought answering, and context understanding**.
  - **QA Retrieval Agent**: Engineered a robust QA Retrieval agent by leveraging **Langchain, FastChat5B, and ChromaDB Vector Store**. Analyzed accident descriptions and traffic laws to identify law infringements, streamlining liability determination alongside SubroGPT. Increased claim closure rate by 500%.
  - **Data Extraction from Police Report**: Utilized **OpenCV** and **Tesseract** to identify and extract structured data from Police Reports. This automated data extraction resulted in a decrease of claim processing time by 1 hour/report.
  - **Subro Potential Application**: Created a production-ready application that could successfully analyze a claim and determine its subrogation potential. Used **TFiDF, NER, BERT Embeddings and sentence similarity** to understand the data, gather features and trained on historical data. **Reduced the claim processing time from 100 claims/day to 10,000 claims/hour** while **increasing per claim revenue by 40%**. Used **Dockers** to deploy the application and **Airflow** for scheduling and monitoring. Daily processing 1M+ documents to predict subrogation potential
  - **AI Enabled Claim Document Parser**: Developed an Algorithm to identify patterns and extract data from an unseen pdf while maintaining its structure. Went from POC to deployment in 1 month. This allowed streamlined text extraction resulting in significant time reduction and huge savings for clients. 250K+ documents processed till date
  - **Impact**: Successfully took several AI/ML projects from proof-of-concept to production. Developed and applied advanced algorithms in the domain of **subrogation opportunity identification, liability analysis, and text extraction** leading to increased subrogation efficiency and revenue for clients. Identified need for Airflow, **WandB** incorporation, set up and managed airflow and WandB to automate, schedule, monitor, and debug AI applications and artifacts
- **Navia Life Care** — Gurgaon, India
  Machine Learning Engineer — *August 2019 - September 2022*
  - **Handwriting Annotation and Recognition System**: Built an end-to-end system for collecting and labeling doctor's handwritten prescriptions, Leveraged collected data to enhance **handwriting and character recognition on doctor's prescriptions** through the utilization of **CNN, RNN, and CTC techniques** achieving a **90% accuracy rate in offline recognition**
  - **Clinical Decision Support System**: Implemented personalized **recommendation engine** for 20k+ doctors based on patient's information, symptom details and usage data
  - **ETL Pipelines and Reporting System**: Led the development of an ETL pipeline using PostgreSQL and Django, in conjunction with a reporting system featuring HTML, CSS, and JavaScript visualizations. Automated data pipelines and reporting for extensive doctor and patient records, alongside the development of APIs for seamless integration.

Achievements

- Kaggle Expert
- Achieved top 40 ranking out of 6000 students in the ZS Data Science Challenge 2018.