Samar Srivastava

Gurugram, India | samarsrivastava44@gmail.com | +91-7060004225 | https://samacker77.github.io

linkedin.com/in/samacker771 | github.com/samacker77

Summary

Innovative Data Scientist with 5+ years of experience specializing in NLP, text analytics, and machine learning. Proven track record in developing high-impact solutions that streamline processes and enhance data-driven decision-making. Expert in leveraging cutting-edge technologies and tools to drive efficiency and accuracy in diverse applications.

Education

Dr. APJ Abdul Kalam Technical University

August 2015 - May 2019

• B. Tech, Computer Science

Technical Skills

Programming Languages & Tools

• Python, Tensorflow, Linux, ElasticSearch, PostgreSQL, Kibana, Git, Airflow, Kafka

Packages & Frameworks

Flask, Fast-API, Sklearn, Pandas, Numpy, LLMs, Celery, ChatGPT, LLama

Methods & Data

• Text analytics, EDA, NLP, LLMs, Spacy, NLTK, Machine learning, Neural Networks, RNNs, LSTMs, Deep Learning Data Science, Neural Language Modelling, Classification, Regression, Unstructured data, POS Tagging

Cloud & Deployment

• Amazon Web Services(AWS), Fargate, S3, Dockers

Experience

Data Scientist II, WiseStep - Remote

- Designed a real-time industry classifier with a Macro F1 Score of 0.78 leading to streamlined candidate experience tagging and job-candidate matching;
- Parsing of free-form resume to detect and various candidate details with a median accuracy of 85%, processing 157 CVs per minute boosting efficiency of consultant hiring process;
- Parsing of job descriptions to detect and extract job title, salary, employer name, assign a standard occupation code using NLP & various information retrieval methods;
- Building LLM led features to help in interviewing, & screen candidates faster.

Data Scientist I, WiseStep – Remote

- Optimized usage for Airflow for job orchestration, and scheduling to smoothen complex day-to-day data science activities like report generation, data extraction, etc.
- Performed Data collection to accumulate data into Elasticsearch indices for educational institutes, professional degrees, and then performed pre-processing to build knowledge base for data products.
- Built data sync process to synchronize VMS jobs with WiseStep platform by using NLP, selenium.

Associate Data Scientist, WiseStep - Remote

- Built an automated system for lead generation for recruiting associates, boosting throughput of the team by 30%
- Designed a JD to Boolean generator to enable end-users to quickly create boolean search strings by submitting job descriptions. The feature ended up boosting productivity of recruiters by 3X.

Junior Data Scientist, Scanta - Gurugram, India

• Developed pipeline for basic NLP preprocessing tasks like text cleaning, tokenization, and evaluating n-grams, achieving significant enhancement in text similarity evaluation using cosine similarity.

Nov 2021 - Oct 2023

Nov 2023 - Present

Nov 2020 - Oct 2021

April 2019 - Oct 2020