

# SAMAR SRIVASTAVA

☎ +91-70600 04225 • Github • ✉ samarsrivastava44@gmail.com • LinkedIn • Gurugram, India

## EDUCATION

**Dr. APJ Abdul Kalam Technical University** • B.Tech  
*Computer Science*

August 2015 – May 2019

## TECHNICAL SKILLS

**Programming Languages** • Tools

Python, Elastic Search, Kibana, Apache AirFlow, OOP, Bash, Git

**Packages** • Frameworks

sklearn, pandas, folium, seaborn, nltk, spacy, Flask, FastAPI, Selenium, BeautifulSoup, spacy, nltk

**Cloud** • DevOps

AWS, Docker, Azure

**Domain Expertise** • Domain Knowledge

NLP, Machine Learning, Data Scraping, Text mining, Text analytics, data cleaning, data pre-processing, data visualizations, Statistics, ChatGPT, Rest APIs

## WORK EXPERIENCE

**Data Scientist** – WiseStep  
Gurugram, India

November 2020 – Present

- Spearheaded development of a real-time industry classifier with a Macro F1 Score of 0.67 leading to streamlined candidate experience tagging.
- Innovated an in-house CV parser with a median accuracy of 85%, processing 157 CVs per minute boosting efficiency of consultant hiring process.
  - Parsing of free-form resume to detect and extract candidate name, employers, skills, certifications, employment history;
  - Parsing of job descriptions to detect and extract job title, salary, employer name, assign a standard occupation code;
  - Automated noisy and multi-language data normalization, cleanup, deduplication;
  - Tech stack : Python, NER, Spacy, RegeX, POS Tagging, Kafka, AWS, Flask, Celery
- Developed a lead generation pipeline to source job leads for Recruiting Associates based on keyword search & jobs sourced from LinkedIn, other job boards.
- Managed data duplication of active consultants by identifying duplicate entries, ensuring talent pool uniqueness using Levenshtein distance & Jaro-Winkler distance techniques.
- Day to day task involves data collection to Elasticsearch indices, EDA for job titles, company names normalizing them and performing pre-processing to build knowledge base for data products.
- Engineered Airflow DAGs to automate repetitive tasks such as index maintenance and data pipeline orchestration.
- Generate and manage data dashboards for the Customer Success team and stakeholders.
- **Tech Stack** - python, flask, elasticsearch, kibana, nltk, spacy, transformers, regex, airflow, AWS, chatGPT, LLMs

**Machine Learning Engineer** – Scanta Inc.  
Gurugram, India

April 2019 – October 2020

- Designed data dashboard for analyzing virtual assistant requests, successfully improved anomaly detection and reporting of malicious events.
- 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.
- Experimented with Uber AI's Plug & Play Language Model to introduce personalities in text, boosting customer engagement.
- Experimented with Uber AI's Plug & Play Language Model to introduce personalities in text, boosting customer engagement.
- **Major tech stack** - python3, huggingface tokenizers, nltk, sklearn, Docker, T-SQL, HTML, CSS, JS, spacy

## PERSONAL PROJECTS (AVAILABLE ON GITHUB)

**Classification of Business Licence Status** Kaggle

Source Code | Tech Stack - python, data science, nltk, pandas, regression, ML

**Predicting Stack Overflow Tags** Kaggle

Source Code | Tech Stack - python, data science, nltk, pandas, regression, ML