

SAMAR SRIVASTAVA

☎ +91-945-7068-769 • ✉ samarsrivastava44@gmail.com • in www.linkedin.com/in/samacker771

EDUCATION

Dr. APJ Abdul Kalam Technical University • Lucknow, Uttar Pradesh August 2015 – May 2019
Bachelor of Technology • Computer Science • Percentage: 71.1

Central Board Of Secondary Education • Delhi March 2014 – March 2015
Higher Secondary School • Physics, Chemistry & Mathematics Major

WORK EXPERIENCE

Machine Learning Engineer – Scanta Inc. April 2019 – Present
Gurugram, Haryana

- Working on various Data Science & NLP projects for data dashboard generation, and data augmentation over various data sources.
- Applied several Natural Language Processing techniques for extracting information from data, like removing non-ASCII characters, stop words, punctuation from documents, tokenization, generating bag of words, finding bigrams and trigrams important for the document.
- Calculating tf-idf values over sentences, performing stemming, and lemmatization.
- Automated tasks like tokenization, lemmatization, and various other NLP preprocessing tasks by setting up a flexible sequential system for performing the above mentioned tasks that earlier need to be done manually one after another over a stream of data, based on requirements.
- Working closely on deployment side to take care of setting up the product on cloud infrastructures like AWS and Azure by binding the models into a Flask Web Application or by creating REST APIs for inferencing to the models.
- Utilize Git, Python, Scikit-Learn, NLTK, StanfordNLP to transform research principles into working models.

SELECTED PROJECTS

Covid 19 India EDA Kernel Kaggle 2020
Personal Project

A notebook dedicated to data visualization and geospatial analysis of COVID19 Pandemic in India. The notebook visualizes the current location of COVID19 patients in India, and the effects of COVID19 pandemic in India to help understand the effect of the outbreak demographically. The Exploratory Data Analysis showed some really interesting insights from the data.

Text Data Preprocessing Pipeline Scanta Inc 2020
Natural Language Processing

A pipeline to automate tasks like tokenization, lemmatization, and various other NLP preprocessing tasks by setting up a flexible sequential system for performing the above mentioned tasks that earlier need to be done manually one after another, based on requirements.

Personality Engine Uber PPLM 2019
Language Models

Web based demo based on Uber AI's research paper named Plug and Play Language Models implemented using Transformers package by huggingface. Deployed on AWS as a flask web-app.

House Price Predicison - Advanced Regresison Techniques Kaggle Competition 2020
Global Rank - 3000

A notebook for Kaggle Competition of "House Prices:Advanced Regression Techniques". Lots of visualizations and Feature Engineering on the data.

Neural Paraphraser Scanta Inc 2019
Natural Language Processing

Detects sentence polarity and converts negative sentences to positive sentences by modifying embeddings and processing negations.

TECHNICAL SKILLS

-
- Programming languages: Python, SQL, C
 - Other framework experience: AWS, Flask, Numpy, Pandas, Scikit-learn