

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 analysing virtual assistants requests and response to detect anomaly in conversations.
- Setup the pipeline for basic NLP preprocessing like text cleaning, tokenization, generating bag of words, evaluating n-grams. Evaluate similarity between text using cosine similarity, and much more.
- Calculating tf-idf values over conversations, performing stemming, lemmatization.
- Working closely on production level deployment.
- Used Kafka nodes as message brokers, encapsulating modules into docker containers.
- Used SQL Server as local database and redis as in memory database for superfast data fetching.
- Experimented with Uber AI's Plug Play Language Model to induce personalities in text.
- Responsible for development of a paraphrasing tool using Transformers.
- Deployed 3 products on AWS using various services like EC2, API gateway, and AWS Lambda.
-
- Responsible for end to end engineering on NLP products pipelines from data mining, data cleaning , to modelling and deployment on cloud.
- Packages used - nltk, scikit-learn, stanford NLP, and more.

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.

Neural Paraphraser Scanta Inc 2019
Natural Language Processing

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

House Price Predicion - Advanced Regresison Techniques Kaggle Kernel 2019
Kaggle Competition

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

TECHNICAL SKILLS

- Programming languages: Python, SQL, C, sqlite3
- Other framework experience: AWS, Flask, Numpy, Pandas, Scikit-learn, Keras, PyTorch
- Deployment experience: Kafka, Docker, Logger, Redis