Sagar Daswani

617.308.7317 | sagardaswani1703@gmail.com

LinkedIn: linkedin.com/in/sagar-daswani **GitHub:** github.com/Sagar401 **Medium:** medium.com/@sagardaswani1703

TECHNICAL SKILLS

- **Proficient:** Python, TensorFlow, Keras, Scikit-learn, Pandas, Numpy, Matplotlib, Seaborn, R, SQL, JavaScript, PostgreSQL, MongoDB, AWS, Azure, Spark, PySpark, Hadoop, Scala, Hive, Databricks, PowerBI, Tableau, Git
- **Exposure:** PyTorch, Flask, Kafka, D3.js, Oracle, BeautifulSoup, Selenium, C/C++, Java

EDUCATION

M.S. Data Science & Analytics, *Northeastern University* Jun 2020

- Relevant Coursework: Foundations and Applications of AI, Predictive Analytics, Advanced Data Mining, Big Data and Machine Learning, Data Warehousing in SQL, Data Structures and Algorithms

B.Eng. Computer Engineering, *Dharamsinh Desai University* May 2018

WORK EXPERIENCE

Tamarin Health | Data Scientist, Intern | Boston, MA Jul 2019 - Dec 2019

- Redesigned warehouse architecture and relational databases design to accelerate stability for +250 customers utilizing Microsoft Azure services (Data Lake, Databricks, PowerBI) and Apache Spark.
- Collaborated in data analysis and preprocessing utilizing Databricks backed by Apache Spark framework to provide analytics to physician group practices.
- Designed supervised machine learning pipeline using PySpark and SparkML on Microsoft Azure, reducing computation time by +15% for physician analytics platform with caching technique.
- Devised XGBoost and Naive Bayes machine learning models for predicting patient readmission rate with over 75% accuracy utilizing Python, PySpark, and SparkML, boosting profit margins by 10 - 15%.
- Communicated analytical solutions to stakeholders and implemented interactive dashboards in PowerBI.

Northeastern University | Big Data Graduate Teaching Assistant | Boston, MA Jul 2019 - Nov 2019

- Taught 90+ students big data concepts (map-reduce, big query) leveraging Spark, Hadoop, Scala, and Hive.
- Assisted Northeastern IT Services with setting up Azure Lab services and Virtual Environment for classrooms.
- Hosted NASA Space App Hackathon and conducted workshops on real time analytics using streaming data.

Physical Research Laboratory - ISRO | Data Science Research, Intern | Ahmedabad, India Jan 2018 - Apr 2018

- Engineered Monte Carlo simulation tool in Java and Swing and performed hypothesis testing to detect behavior and lifetime of objects in space when exposed to high energy solar winds.
- Integrated Tableau dashboards to accelerate analysis on results to communicate and validate with scientists.
- Incorporated open source tools using object-oriented approach with Java and Swing for interactive support.

PROJECT WORK

Disaster Tweet Classification | Machine Learning Engineer | github Mar 2020

Leveraged tweets to develop sentiment analysis model for classifying tweets aimed to report disasters in a community.

- Integrated Tokenization, TF-IDF and built Recurrent Neural Network (RNN) with +85% accuracy in detecting disasters from Tweets using Python, TensorFlow, and Numpy.
- Extended architecture by adding LSTM, Bi-LSTM, GRU and pre-trained word embedding vectors (GloVe and fastText) utilizing Python and TensorFlow.

Historical Document Digitization | Machine Learning Engineer | presentation | github Nov 2019

Devised pipeline to evaluate handwritten documents and extract data using OCR with an average of 85% accuracy.

- Implemented Convolutional Neural Network (CNN) in Python and TensorFlow for classification of documents from different decades or handwritten/typewriter with +91% accuracy.
- Incorporated bounding box algorithm for image segmentation and carried out image pre-processing using Python and OpenCV specific for each segment of image, which boosted accuracy by 10%

LEADERSHIP + CERTIFICATES

Deep Learning Specialization, *deeplearning.ai | Certificate* Jan 2019

Founding Member, *AI Skunkworks at Northeastern University* Mar 2019 - Mar 2020

- Conducted a workshop on Convolutional Neural Network and Backpropagation