

SAHIL GADGE

Boston, MA | 8574379167 | gadge.s@northeastern.edu | <https://www.linkedin.com/in/sahil-gadge/>

EDUCATION

Northeastern University, Boston, MA September 2021 - May 2023
Master of Science in Information Systems GPA - 4.0/4.0

- Relevant Courses - Data Science Engineering and Methods, Deep learning and Reinforcement learning in Game Engineering

Savitribai Phule Pune University, Pune, India July 2015 - May 2019

Bachelor of Science in Computer Engineering

- Relevant Courses – Artificial Intelligence and Robotics, Data Mining and Warehouse, Machine Learning, Data Analytics

TECHNICAL KNOWLEDGE

- **Languages:** Python (PyTorch, TensorFlow, NLTK, Pandas, Scikit-Learn, Matplotlib, NumPy Flask, Keras, etc.), R
- **Other Languages and Framework:** Java, C, C++, HTML, CSS, JavaScript, PHP
- **Databases:** MySQL, PostgreSQL, Mongo DB, NoSQL
- **Tools:** Anaconda, AWS SageMaker, Jupyter, Tableau, MS Excel, RStudio, Power BI, SQL Server Management, AWS

WORK EXPERIENCE

KHELIYA TOYS, Pune, India August 2020 - July 2021

Data Analyst and Python Developer

- Discerned insights from data gathered by 22 shops located across Maharashtra State to increase sales
- Generated following reports. Report to estimate re-order quantity of each product based on sales and stock availability, report of classified fast-moving products, and others. Reports lead to fast decision making
- Designed a web application on Digital Ocean cloud as a hosting platform and Python (Flask), HTML, CSS, JavaScript coding languages. Therefore, shops were able to generate various reports remotely faster
- Developed auto mailer to send reports to shops via mail. Reduced 2 hours of daily work required to format and mail reports
- Collaborated with team to verify correct implementation of web app and accuracy of reports created by python script

TATA CONSULTANCY SERVICES, Pune, India May 2020 - August 2020

Data Science Intern

- Built a Convolutional Neural Network model to identify handwritten text from an image. Used Keras and TensorFlow libraries to build Convolutional Recurrent Neural Network. Training and testing of models were done using IAM Handwritten dataset
- Incorporated loss functions and various explanation techniques to compare performance metrics
- Presented findings to peers, illustrating progress made during various steps. And validated correctness of model

PROJECTS

TEXT FROM IMAGE (LANGUAGES- ENGLISH, HINDI, AND MARATHI) July 2020 - August 2020

- Developed a web application for extracting text from an image. Text in following languages English, Hindi, and Marathi can be detected by web application. Web application results into faster predictions on input images
- Designed AI Algorithm it takes image as input and return string of text detected in given image
- Constructed web application on Heroku Cloud using Python, Flask web framework as a backend and tailored front end in HTML, CSS, JavaScript languages. Also applied OpenCV to perform cleaning steps on images from user

SCIKIT-LEARN VS AUTOML ON CATEGORICAL DATA October 2021 - November 2021

- Compared AutoML algorithms including H2O, Light AutoML, and Scikit-Learn Logistic Regression and Random Forest algorithms based on how accurately each algorithm performed in classifying target variable
- Trained and tested ML models using Cat-in-dat Kaggle competitions categorical dataset. Conducted data cleaning, feature engineering steps such as encoding, mapping. Examined results of encoding data with One Hot encoding and Target encoding
- Improved accuracy of SK-Learn Logistic Regression and Random Forest algorithm by tuning hyperparameters. Scikit-Learn's Logistic Regression model outperformed both H2O, Light AutoML models

VARIOUS WAYS TO HANDLE IMBALANCED DATASETS November 2021 - December 2021

- Assessed various techniques to handle imbalanced datasets. Imbalanced Datasets such as, credit card fraud, Earthquake prediction, Disease detection dataset creates issue of biased models
- Completed analysis by scaling imbalanced dataset with various oversampling and under sampling techniques. With AWS SageMaker trained and tested ML models in python and Scikit-learn
- Concluded oversampling is better than under sampling dataset, as it removes valuable data. Oversampling increased the accuracy of model from 86% to 99%

CERTIFICATIONS

- **IBM Data Science Specialization**

Topics Include - Python, Databases and SQL, Data visualization, Data analysis and Machine learning

- **Machine Learning** (by Stanford University)

Topics Include - Supervised and unsupervised learning algorithms, Machine learning and statistical pattern recognition