Shubham Thakur

sythakur@usfca.edu • 6282242735 • San Francisco, CA • linkedin.com/in/sythakur/• **github.com/shubhamthakur01**

EDUCATION

University of San Francisco, San Francisco, USA — Jun *2022 (expected)*
MS in Data Science
Courses: Probability and Statistics, Machine Learning, SQL, Distributed Computing (Spark), Time Series Analysis

Indian Institute of Technology, Guwahati, Guwahati, India — *Jul 2015 - Jun 2019*
B.Tech in Chemical Engineering
Courses: Pattern Recognition and Machine Learning, Optimization Techniques, Linear Algebra

WORK EXPERIENCE

LexisNexis, San Francisco, USA — *Nov 2021 – Present*
Data Science Intern - NLP

- Working on a facts extraction model to get the semantic information like age, gender, and occupation of the plaintiff from a corpus of millions of legal briefs
- Built an ensemble of Coreference Resolution and Bert Question Answering model to achieve a document level F1-score of 0.8 for gender extraction

EXL Analytics, Ahmedabad, India — Sep 2019 - Jun 2021
Data Science Consultant II/Lead Assistant Manager

- Used high transaction volumes and personally identifiable data to develop risk mitigation strategies across different business lines of a leading US banking client
- Identified the loss exposure of over $10+ MM by early fraud detection in credit and deposit portfolio
- Developed a satellite-imagery based building damage assessment web application using the ResUNet framework to automate the First Notice Of Loss(FNOL) during calamities

University of Vienna - IAESTE trainee, Vienna, Austria — *Jun 2019 – Aug. 2019*
Research Intern, Deep Learning

- Worked on the development of explainable AI models to detect the substructures(Tox-Alerts) of the molecule responsible for toxicity
- Improved the performance of the LIME model by extending its utility for non-linear decision boundaries

EXL Analytics, Gurgaon, India — *May 2018 – Jul 2018*
Product Development Intern

- Developed a word-level handwritten text recognition model by stacking CNN layers and a Bi-directional LSTM layer on a CTC scoring loss function

PROJECTS

Exoplanet Detection using Machine Learning

- Detected the existence of an exoplanet around a given star from its radiated flux data using extracted features from the Box Least Square method
- Used ensemble of multiple regression algorithm which achieved the F1-score of 0.92 on the test dataset

SKILLS

Programming: Python (Scikit-Learn, Pandas, Numpy, SciPy, PyTorch, Keras, TensorFlow, Matplotlib, Seaborn, Spacy, NLTK), SQL, AWS (EC2, EMR, S3, Sagemaker, Lambda), PySpark, No-SQL (MongoDB), Linux, APIs
Skills: Machine Learning, Deep Learning, Data Manipulation, A/B Testing, Data Analysis, Git/Github, NLP, ETL

AWARDS

Promising Newcomer Award FY-2020, EXL Analytics