Somnath Rakshit

+1 (512) 798-3552| somnath@utexas.edu
somnath.pages.dev | linkedin.com/in/somnathrakshit/ | github.com/somnathrakshit/

EDUCATION

Master of Science, Information Studies (Data Science Track), May/2021
The University of Texas at Austin | GPA: 3.88/4 **| Courses taken:** *Mathematical Statistics for Applications, Linear Models, Data Mining, Introduction to Machine Learning, Database Management*
Teaching Assistant: *User Generated Content Analytics* (Fall 2019), *Data Science Principles* (Spring 2020 & 2021), *Advanced Predictive Modeling* (Fall 2020).

Bachelor of Technology, Computer Science and Engineering, May/2018
Jalpaiguri Government Engineering College, India | **GPA**: 8.68/10
Courses taken: *Artificial intelligence, Data Structures, Calculus, Discrete Mathematics, Probability and Statistics, Design and Analysis of Algorithms, Object Oriented Programming*

RELEVANT EXPERIENCE

Research Assistant, Intelligent Data Exploration and Analysis Laboratory, UT Austin, May/2020 – Aug/2020
- Privacy-preserving federated learning using healthcare data (Mentor: Prof. Joydeep Ghosh).
- Generated subphenotypes of COVID-19 with tensor factorization to understand its demographical impacts.

Research Assistant, Centre of New Technologies, University of Warsaw, Jan/2019 – Aug/2019
- Ranked genetic biomarkers for breast cancer subtypes (Mentors: Dr. I. Saha and Dr. D. Plewczyński).
- Developed a novel ranking algorithm using multi-objective genetic algorithms.
- Demonstrated the relation between up/down regulation of genes and survival probability of a population

Software Engineer – Machine Learning, Cyware Labs, July/2018 – Nov/2018
- Built a Convolutional Neural Network to cluster similar articles resulting in 2x no. of articles selected.
- Determined trending keywords using Named Entity Recognition from news articles.
- Deployed models in AWS EC2 using Celery, Gunicorn, and Django.

TECHNICAL SKILLS

Programming Languages: Python, SAS, Java **| Frameworks**: Scikit-learn, Numpy, Pandas, Matplotlib, Spacy, PyTorch, Tensorflow, Keras, Git, Django **| Databases**: SQL, Elasticsearch

RELEVANT PROJECTS

Credit Risk Model and Scorecard, Dec/2020 – Built a data-driven credit risk model using logistic regression after relevant preprocessing and feature engineering in Python to predict the probabilities of loan default and assigned credit scores to potential borrowers; used the co-efficients of the trained model to implement a scorecard to easily calculate credit scores.

Prediction of Interest Rates, Dec/2020 – Predicted the interest rate based on the credit history of a person based on historical loan data that helps investors quantify credit risks using scikit learn.

A GPU-accelerated MRI Sequence Simulation for Differentiable Optimization and Learning, **Mar - Dec/2020 –** Wrote an open source GPU accelerated simulator in PyTorch for the Extended Phase Graph Algorithm for certain sets of sequences. After successfully parallelizing the simulator, we increased its speed by almost 100,000 times. This work was accepted as an abstract in ISMRM virtual conference. Guide: Dr. Jon Tamir

Deep Stock Predictions, Mar - May/2020 – Built a multi-stack Bidirectional LSTM model strengthened by the addition of Attention units. Proposed a data driven approach for optimal selection of window length and multi-step prediction length, and added analyst calls as technical indicators to the model. Guide: Dr. Edison Thomaz

geograpy3, Sep/2018 (Ongoing) – geograpy3 is a Python library that is used to extract place names from a URL or text, and add context to those names, e.g., distinguishing between a country, region or city. It can be installed from PyPi using pip and has over 38,000 downloads.