Rehan Guha

Machine Learning Researcher · Data Scientist

rehanguha29@gmail.com | (+91) 9874223189 | India
rehanguha.github.io | rehanguha | rehanguha| G-Scholar

Summary

Let me start by giving a simple introduction about my work. I am a Senior Machine Learning Researcher presently employed by a product based company. I have worked in several ML projects mainly dealing with Optimization Problems, Explainable AI, Deep Learning, Imbalance Classification, et cetera...

I am also the author of a book **"Machine Learning Cookbook with Python" published by BPB publications**. I have been part of multiple publications through Journals and Open Publications, along with this I have few applied patents to my name.

Recently I have been invited by Arizona State University and to be a part of their team, and we became runners up in **CPS challenge 2020 funded by National Science Foundation(NSF), USA and sponsored by NASA**. I have also conducted multiple seminars/webinars in this field of research for various universities and made contributions to Open Source Software as well.

I have worked on **Open Source contribution for Google TensorFlow Addons Optimizer**. Currently, I am experimenting the behavior of multiple gradient based optimizers.

Work Experience

Pramati Technologies — *Chennai, India*

Senior Machine Learning Researcher — *Oct 2019 - Jun 2021*

i) During this period I worked on Google TensorFlow SIG Addons for pushing a PR **ii)** Studying a behavior on Gradient based Optimizers. **iii)** I have been working on techniques to dewrap a scanned or photographed text document.

Machine Learning Researcher — *Oct 2018 – Sep 2019*

i) I was responsible to integrate Fonduer in our pipeline. **ii)** I also ran a project to De-noising a Natural image using a random weights from a CNN structure.

Imaginea Technologies Inc. — *Chennai, India*

Senior Data Scientist — *Oct 2019 - Jun 2021*

i) Had to perform a quality check for the scanned documents and write an algorithm to dewrap a scanned document. **ii)** Managed the Data Science team of 2 including myself from Imaginea. Actively worked on their clinical dataset for increasing their coverage and accuracy for their existing scripts. **iii)** For another project I was responsible for analyzing the CLV and building a Data Analytics dashboard.

Data Scientist — *Oct 2018 – Sep 2019*

i) Handled Litigation Likelihood (LitLikelihood) module of the project – from data analysis to production deployment. Including Imbalance data classification I had a chance to work on another module where I have used LDA for Topic modeling.

Contributor @ Google TensorFlow — *India*

Added an optimizer to TensorFlow Addons SIG — *Aug 2020 - May 2021*

This optimizer is unique as it is a learning rate free algorithm on SGD which works on a betting framework.
Links:

- GitHub Code: COCOB
- Link of the Paper: https://arxiv.org/abs/1705.07795

Book Author

India

BPB Publications (ISBN-10 : 9389898005 / ISBN-13 : 978-9389898002) (Links: Amazon /BPB Online)

Aug 2019 - Nov 2020

A Cookbook that will help you implement Machine Learning algorithms and techniques by building real-world projects.
Key Features:

- Learn how to handle an entire Machine Learning Pipeline supported with adequate mathematics.
- Create Predictive Models and choose the right model for various types of Datasets.
- Learn the art of tuning a model to improve accuracy as per Business requirements.
- Get familiar with concepts related to Data Analytics with Visualization, Data Science and Machine Learning.

Participant

Part Time

Arizona StateUniversity (ASU) & National ScienceFoundation (NSF)

May 2020 - Jun 2020

2020 CPS Challenge "SoilScope – Mars edition" :
Mars 2020 inspired mission scenario for the 2020 NSF CPS Challenge will be a two-week virtual event, emulating an autonomous probe deployment sciencemission by a rover and drone duo, at the Jezero crater landing site.
Organizer: Arizona State University, NASA

Associate Software Engineer

Chennai, India

Accenture

Jan 2017 - Sep 2018

I was working as a Machine Learning Developer at Accenture for multiple Research projects, and few of the research projects won awards and it has been considered as product by the company.

- Financial Service Development Project (Quartz Development) (Full Time, Managed team size of 2)
- Self Learning Credit Approval System (Part time, 3 months)
- AI Model Based Risk Management (Part time, 4 months)
- Ship and Air Cargo Routing (Part time, 2 months)
- Digital Content and Transaction Management using
  an Artificial Intelligence (AI) based Communication System (Part time, 3 months)
- Routing System for Car Pool and Drop for the Employees (Part time, 3 months)

Bounty Hunter

Part Time

HackerOne

Jan 2017 - Present

Found and Resolved Bugs for:

- Zomato

Creator

India

PixRead

Dec 2016 - Present

One of a kind OCR & 2D Codes (QR Code, Bar Code etc) search engine where the extracted text will be searched over a search engine. This will ease all the users to take a picture (which has text) from a device and search over different search engines.

Technical Associate

Kolkata, India

Saltlake Institute of Engineering & Management

Aug 2015 - Feb 2015

Website Developer & SEO Professional

Publications

Multilevel Database Model -A form of Secured stored password

DOI: 10.13140/RG.2.2.12017.94561

Cryptography, Architecture *Apr 2019*

This paper dealt with database architecture for storing password. Many authentication schemes depend on secret passwords. Unfortunately, the length and entropy of user-chosen passwords remain constant over time. In contrast, hardware improvements constantly give attackers increasing computational power. This paper presents and focuses on a novel database architecture and password scheme which will help the users to keep the same password for a long time making it more secure. In the traditional system, the user entered text passwords will be converted into password hash[2] and stored into the database which will be later compared for validation. In the new system, user's text password will be hashed in multiple layers with a different Database architecture and password scheme. So, the attacker won't get hold of the original text password given by the user which will ensure the integrity as well as the security of the system.

Digital Content and Transaction Management using an Artificial Intelligence (AI) based Communication System

Patent: US20200034842A1

Machine Learning, Credit Risk Scoring *Aug 2018*

A system for predicting a non-fraud dispute using an artificial intelligence (AI) based communications system is disclosed. The system may comprise a data access interface to receive instructions historical transaction and disputes data from at least one data source associated with an account issuer. The data access interface may also receive incoming transaction data associated with a transaction from at least one data source associated with an account holder. The system may comprise a processor to predict a likelihood of a non-fraud dispute associated with the transaction by: examining the historical transaction and disputes data; retrieving non-fraud dispute attributes; parsing the incoming transaction data; applying predictive analytics to the incoming transaction data to yield a prediction value; determining that the prediction value meets a predetermined threshold; and generating a predictio for the likelihood of a non-fraud dispute associated with the transaction associated with the account holder to be outputted, via an output interface to a user device.

Prime Numbers – 1st One Lac

Dataset: Link

Data, Numerical *Jun 2016*

Contains 1000 files with 100 prime numbers in each file.

Grid Searching – Novel way of Searching 2D Array

Journal: Link

Algorithm, Research, Searching *Jan 2016*

Linear/Sequential searching is the basic search algorithm used in data structures. Linear search is used to find a particular element in a 2D array. It is not compulsory to arrange an array in any order (Ascending or Descending) as in case of 2D binary search. In this paper, I present a unique searching algorithm named Grid Search, which helps to search an unsorted 2D Array/Matrix with least time complexity and iteration. We also have compared the Grid searching algorithm with Linear Search Algorithm. We used C++ for implementation and analysis of CPU time taken by both the algorithms. Results have shown that Grid Searching Algorithm is working well for all input values and it takes lesser time than Sequential Searching in all aspects.

Biometric Ticketing System

Patent

Electronics, Architecture, Railways *Mar 2014*

To introduce fingerprint ticketing system in metro railway system by replacing RFID tickets.

Biometric Voting Machine And System

Patent

Electronics, Architecture, Election *Dec 2014*

This machinery function and system is more scientific and real alternative to EVM or Ballot paper with stamp system. In this system biometric registration of the voters, voting by biometric identity cards only, no need to make registered voter list before each election, identity checking with fingerprint matching, voting with biometric fingerprint impression, auto-locking of each fingerprint identity after casting vote, one person can cast only one and his own vote, flexibility of votes to cast his/her vote in different booths (under certain conditions), centralise vote counting opportunity are the key functions.

Data Security- Multi-Layer Folder Lock Hiding

Patent: Page-8379

Architeture, Algorithm *Apr 2013*

The invention is a computer program, method and process in which the ID-Data is locked with 5 tier protection system and the ID-Data also protected from rouge software.

Skills

Machine Learning, Data Science, Imbalance Classification, Statistical Analysis, ML Architecture, Deep Learning, Explainable AI, Financial Services, Credit Risk, Product Development, Algorithms, Data Analysis

Honors & Awards

Jul 2020	**Runners-Up**, In this challenge, we use a drone to search for and locate a soil probe. The drone picks up the probe and takes it to a drop-off location. Then, the drone must autonomously return to a rover and land in its trunk. This was accomplished using Rtabmap, a SLAM package used to estimate the position and velocity states of the rover with visual odometry and our own tailor-made Kalman Filter conflated with GPS based state estimation and the visual odometry based states.	*National Science Foundation (NSF)*
Jul 2019	**Finalist**, Infinity	*Pramati Technologies*
Jul 2019	**Finalist**, Infinity	*Pramati Technologies*
Aug 2018	**Accenture Celebrates Excellence**, Category - Innovation; Award type - Team	*Accenture*
Mar 2018	**Asset Harvest Contest**, 1st Pace	*Accenture*
May 2018	**Accenture Celebrates Excellence**, Category - Innovation; Award type - Individual	*Accenture*
Nov 2017	**Accenture Celebrates Excellence**, Category - Client and Customer; Award type - Team	*Accenture*
May 2015	**Student Performance Award**, Received the award for excellence innovation and performace.	*I.E.M., Kolkata*
Nov 2013	**Data Security- Multi Layer Folder Lock Hiding**, Demonstration	*Indian Army -Bengal Area Sig Coy*

Projects

Bank of America (Quartz Development)

Python, Angular JS

Accenture

Problem: We had multiple tools of different technologies, and we require proficient resource to use the tools. Project was people driven.
Action: I made a platform for Environment Management team where all the automated scripts (for multiple technologies) & report generation tools will be placed & used from a single integrated platform with a simple UI.
Result: Saved a lot of manual effort, anybody with the access permission can use the tool can execute it using a simple UI with less or no knowledge of the technology, Project became Process driven.

Self Learning Credit Approval System

Python, ML, H2o.ai, Scikit-learn

Accenture

Problem: Rule based decision for approval of Credit which lead to loss of business.
Action: I was solely responsible for the development of the Machine Learning Algorithm.
Result: Accenture took this as a product and demonstrated to the client for identifying bad loans or missed opportunities, enable the bank to mitigate potential risk and opportunities

---

AI Model Based Risk Management

Azure, Python, H2o.ai

Accenture

Problem: High level of manual effort from Compliance officer for Risk Management to detect valid and type of alerts raised.
Action: I made an AI based Model using DeepLearning to detect the Alert and type of alert which is performed by Compliance officer.
Result: Dependency on the compliance officer will decrease, model will be capable to figure out new patterns in the generated alert.

Ship and Air Cargo Routing *Python, Machine Learning*

Accenture

Problem: Ship and Air Cargo Routing for Logistics with high traffic and and more than expected loading time.
Action: Used key concepts of algorithm and process Design to optimize the whole system.
Result: Decreased the time for processing more than 20 seconds which improved the performance and saved huge money.

Digital Content and Transaction Management using an Artificial Intelligence (AI) based Communication System *Python, Machine Learning, H2o.ai, Scikit-learn*

Accenture

Problem: There were no existing intelligent model to predict Disputes of a transaction of a financial institute.
Action: I had designed a machine learning model which predicts both the dispute and the dispute category of a financial transaction given to the model using using a life dataset.
Result: Accuracy of the model with the live dataset came to be around 92%. The project is applied for patent by Accenture and will be sold as a product to the clients. This offering is expected to Reduce Annual Billing Disputes Volume by 30% i.e. 330K, lowering Ops Cost by USD 15.3 million and enhancing Customer Experience by guarding them from such dispute prone transactions.

Routing System for Car Pool and Drop for the Employees *Python, Machine Learning, H2o.ai, Scikit-learn*

Accenture

Problem: There were no existing intelligent model to predict Disputes of a transaction of a financial institute.
Action:: Created an software using Google Map and some grouping algorithm with considering total cars, total no of seat in each car, last drop should be a male for women security.
Result: Reduced the entire manual effort form sorting and assigning of the cabs

---

PixRead *Python*

Link

An Innovative Research Project which helps the users to extract the text from an Image using OCR and other image processing algorithms and search through different Popular Search Engines. Technologies used - OpenCV, Google Tesseract OCR, etc...

AudioAnalytica *Python, Machine Learning*

Link

AudioAnalytica is a no-code platform and audio analytics dashboard. This is one of a kind All-in-One Audio analytics platform. There are no to less solutions who offer an audio analysis and analytics platform for general and not specific to one vertical. Dashboard for individuals looking to get insight from audio files and use those knowledge for their business with no knowledge of Audio Processing.

2D to 3D - Floor Plan *Python, Machine Learning, OpenCV, AR/VR*

Pramati Technologies

2D to 3D - Floor Plan is a tool to convert 2D floor plan to 3D. Later using the output from Python and OpenCV, a virtual reality is created that can be interacted by a user. This project easily creates an AR/VR environment without the huge knowledge of 3D model designing.

Education

Indian Institute of Technology(IIT), Kharagpur

Deep Learning (AICTE approved FDP course) *Jul 2019 – Oct 2019*

Institute of Engineering & Management (I.E.M.), Kolkata *Kolkata, India*

Bachelor of Computer Application (B.C.A.) in Computer Applications *2013 – 2016*

GPA: 8.35

Delhi Public School, Ruby Park *Kolkata, India*

All India Senior School Certificate Examination in Computer Science *2011 – 2013*

Volunteering Experience

Guest Lecturer *Nov 2018 – Present*

I give Machine Learning & Data Science related Workshops to University students to bring them up the pace with the Moden World and Expose them to a new technology which they pursue in Future. Top Universities & Colleges visited so far

- KL University (Top 50 University in India)
- B.S. Abdur Rahman Crescent Institute of Science and Technology
- VELS University

Other Information

Father's Name:	Ranjan Guha
Nationlality:	Indian
Allergries:	None
Language skills:	English, Bengali, Hindi
Philantrophy service:	Volunteer @ World Food Program by United Nations
Hobbies:	Pencil Sketching, Playing Table Tennis, Trekking

Decalaration:
I hereby declare that all the information contained in this resume is in accordance with facts or truths to my knowledge. I take full responsibility for the correctness of the said information.

Name: Rehan Guha

Place: Kolkata, India