Rehan **Guha**

MACHINE LEARNING RESEARCHER · DATA SCIENTIST

India

(+91) 98-7422-3189 | rehanguha29@gmail.com | rehanguha.github.io | rehanguha | rehanguha

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**. Along with this I have been part of multiple publications through Journals and Open Publications, and applied a few patents to my name.

Recently I was invited by Arizona State University to be a part of their team participating 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.

Currently, I am working on **Open Source contribution for Google Tensorflow Addons** and studying the behavior of multiple gradient based optimizers.

Work Experience

Vodafone
Pune, India

LEAD DATA SCIENTIST (MANAGER, LEAD-BUSINESS INTELLIGENCE)
July 2021 - Present

Vertical:: BIBDA (BI Big Data and Analytics)
- Handling multiple projects from Machine Learning / Data Science perspective
- Working on FBProphet, LSTM, SKTime
- Using Google Cloud Platform for all the projects
- Handing around 7 ML/DL projects

Pramati Technologies
Chennai, India

SENIOR MACHINE LEARNING RESEARCHER
Oct 2019 - Jun 2021

R'n'D
- Contributed an optimizer COCOB to Google TensorFlow SIG Addons
- Worked on how to transform a wrapped page and de-wrap it
- Researching on Gradient based Optimizers

Client Projects
Healthy Paws
- Role: Need to do a quality check for the scanned documents and write an algorithm to dewrap a scanned document
- We successfully developed an algorithm to check quality of the scanned document. We are in the process of developing an algorithm to dewrap an document.

VERANA
- Role: 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.
- We worked on some of their existing code and increased their accuracy and coverage percentage. We developed an address extraction script from a given document using NLP.

BAIN
- Role: Responsible for analyzing the CLV and building a Data Analytics dashboard.
- The project involves analyzing the survey data and figuring out the CLV. It also entails creating a Tableau Dashboard that analyzes all the features in a single dashboard.

R'n'D
- Worked on Fonduer to incorporate in your pipeline
- Natural Image Deniosing using a random pixel image

Client Projects
RPX Corporation
- Role: Handled Litigation Likelihood (LitLikelihood) module of the project – from data analysis to production deployment.
- In this project, the data (12M records) requires a lot of transformation where some major errors have been observed leading to a drastic improvement in the prediction accuracy. Therefore, LightGBM has been used here and the model is accepted by the client as well. Other projects under RPX, Imbalance Learning with Probabilistic Scoring with Binary Classification, LDA Topic Clustering.

Contributor @ Google TensorFlow   *India*

ADDED COCOB OPTIMIZER TO TENSORFLOW SIG ADDONS   *Aug 2020 - May 2021*

This optimizer is unique as it is a learning rate free algorithm on SGD which works on a betting framework.
- GitHub Code: COCOB Optimizer Link
- Link of the Paper: https://arxiv.org/abs/1705.07795

Book Author   *India*

BPB PUBLICATIONS   *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 STATE UNIVERSITY (ASU) & NATIONAL SCIENCE FOUNDATION (NSF)   *May 2020 - Jun 2020*

2020 CPS Challenge "Soil Scope – 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 science mission 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 am 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 Carpool 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

SALTLAKE INSTITUTE OF ENGINEERING & MANAGEMENT

Website Developer & SEO Professional

Publications

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

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), centralize vote counting opportunity are the key functions.

Data Security- Multi-Layer Folder Lock Hiding

Patent

ARCHITECTURE, 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

ML & DL   Imbalance Classification, Statistical Analysis, ML/DL Architecture, Explainable AI, Optimization
Domain   Financial -Credit Risk, Telecom -Roaming Services

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) & Arizona State University (ASU)*
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.	*Institute of Engineering & Management, Kolkata*

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 Deep Learning 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 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 have designed a machine learning model which predicts both the dispute and the dispute category of a financial transaction given to the model 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 CarPool 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 a 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 were a handful number of 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 that knowledge for their business with no knowledge of Audio Processing.

2D to 3D - Floor Plan

Python, Machine Learning, OpenCV,
AR/VR

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, Kharagpur

Deep Learning (AICTE approved FDP course)

Jul 2019 – Oct 2019

Institute of Engineering & Management, Kolkata

Kolkata, India

Bachelor of Computer Application (B.C.A.) in Computer Applications

2013 – 2016

GPA: 8.5

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 Modern World and Expose them to a new technology which they may 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

- Nationality: Indian

- Allergies: None

- Language skills: English, Bengali, Hindi

- Philanthropy service: Volunteer @ World Food Program by United Nations

- Hobbies: Pencil Sketching, Playing Table Tennis, Trekking, Cooking

Declaration:
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.