

# Bibhash Chandra Mitra

Male, 23 years

## Contact

Phone No: +919635337441

Email Id : bibhashm220896@gmail.com

## OVERVIEW

Associate Data Scientist with a passion in the field of Artificial Intelligence and Algorithms. Graduated with a Major in Aerospace Engineering and a Minor in Computer Science Engineering. Currently working on critical projects like Novel Drug discovery with AI and Clustering Graph Networks for Entity Normalisation. Professionally skilled in OCR Engines, Deep Learning, data exploration and visualization, predictive modelling and optimization.

## SUMMARY

- Currently working on Novel Drug discovery using AI techniques like GVAE, JTVAE and RL
- Worked on Table detection and extraction using FRCNN and Image processing
- Worked on converting PDF into a categorically normalized structured DataFrame
- Built products and proofs-of-concept in the areas of Graph Clustering for Entity Normalisation
- Got a rating of 5/5 for "Outstanding" performance

## WORK EXPERIENCE

### Drug Discovery using AI

*Innoplexus Consulting Services, Pvt. Ltd. ; September 2019 - Present*

- Used different Variational Autoencoder networks (VAE, GVAE and JTVAE) with Reinforcement Learning for creation of new compounds which show similar properties as the target ligands but are novel.
- Worked on new Junction Tree models for encoding the molecules
- Made custom reward functions based on Self Organising Maps (SOMs) for reinforcing the model to sample from specific parts of the latent space of the VAE
- Integrated custom tokenizers for encoding the SMILES structure
- Benchmarked the models using Zinc dataset and MOSES benchmarking tool

### Hierarchy using Graphs

*Innoplexus Consulting Services Pvt. Ltd. ; September 2019 - November 2019*

- Created different Proof of Concepts to find the hierarchy of different nodes using Graphs
- Used Neo4j for finding a hierarchy of nodes in a graph using a custom function built using Louvain Modularity and Centrality algorithms (like Page Rank)
- Used the graphs to create clusters of similar nodes and used NLP found the relationship between two nodes which do not share a Graph edge directly but are related to each other
- Used LDA for finding keywords and essential nodes for the graph

### PDF Extraction

*Innoplexus Consulting Services Pvt. Ltd. ; June 2019 - September 2019*

- Segmentation of paragraphs in a scanned PDF using rule-based Image Processing techniques
- Used Tesseract OCR on these segments to get the contents and segregated different sections based on the type of content
- Used PDF encodings to convert a digitally generated PDF to HTML
- Recorded style information and segregated different contents in PDF based on measurable style change
- Found different information like headers, footers and page numbers to refine the final output of the PDF
- Analyzed the structured PDF data and using Machine Learning for finding information like Author, Title, Affiliation etc. to map the PDF and the author
- Differential analytics was done on top of the extraction for information which shows the consistency of the contents of the PDFs mainly for PDFs which are prescriptions of patients
- Created NLP based post-processing scripts for refinement of the output and merging unnecessarily separated contents

### Table Extraction

*Innoplexus Consulting Services Pvt. Ltd. ; June 2019 - August 2019*

- Detection of a table using both Rule-based and Deep Learning modules
- Used IOU approach to maximize the coverage and accuracy of table detection
- To generalize the table extraction module, created a custom Table-Maker function that converts any table (different background colours, no grids etc.) into a standard table which can be used for extracting the contents of the table
- Used Tesseract-OCR to get the textual information in the generalized table and mapped it onto a preferred file type

## AREA OF EXPERTISE

### Knowledge of:

- **Programming Languages:** Python, C#, C, C++,
- **Deep Learning:** Faster RCNN, YOLO, Inception and ResNet Architecture, GCN, VAE, GAN
- **OCR Engines:** Tesseract OCR Engine, Asprise OCR
- **Image Processing:** OpenCV, Skimage
- **NLP:** Stanford Core NLP, Spacy, NLTK
- **Libraries:** Keras, Tensorflow, Pytorch
- **Modules:** TensorRT, Rapids, Kubernetes
- Docker, Virtual Environments, Anaconda

### Hands-on experience on:

- DGX-1
- Tesla V100

## PROJECTS

### Optimal Simultaneous Rendezvous of Fixed-Wing UAVs in 3D

*Dr Sikha Hota | July 2017 - April 2019*

**Selected at AIAA SciTech Forum 2019, San Diego, California, USA**

- Finding an optimal path in an environment filled with both static and dynamic obstacles which keep a safe distance from the obstacle and considers the turning constraints for the UAV
- Using circular approximations to estimate the path when turning so that there is a smooth change in the direction and minimum change in overall distance
- Using Fortune's Algorithm for creation of Voronoi Diagram across different boundaries in Real-time which changes along with the moving obstacles

## INTERNSHIPS

### Innoplexus Consulting Services, Pvt. Ltd.

May 2018 - July 2018

- Using Machine Learning and Image Processing methods to segment an image of a document into Title, Author, Affiliation and Abstract and map it onto a CSV file
- Detection of a table from an image of a document and converting that into CSV format using deep learning and rule-based mechanisms
- Contributing to Tesseract OCR Engine, a 20-year long project of HP and Google; modifying wrappers for Asprise OCR Engine and Tabula

### Invention Labs

May 2017 - July 2017

- Worked on Machine Learning involving Decision Trees based algorithms for processing Text
- Our objective was to make a language-specific Engine capable enough to process sentences into a tree structure and in turn be converted into Universal Natural Language (UNL)

### Teknuance Info Solutions

December 2016 - January 2017

- Contributed to designing a Database Management System and also optimizing the same for easy access
- Made a UI for accessing the custom DBMS and defined parameters for optimized data retrieval

## COMPETITIONS

### Inter IIT Tech Meet 2017

#### UAV Design

- Designed a UAV with 58 seconds glide time
- Secured the fourth position in the event and achieved an overall silver at inter-IIT

### IBM Hackcamp

#### Machine Learning and Data Analytics

- Participated in #IBMinclude Career Challenge, a machine learning and data analytics competition organized by IBM
- Cleared two rounds and was in top 100.

### Laws of Motion

#### Aeromodelling Competition organized by Boeing

- Designed an RC plane with 43 seconds glide time
- Finished 32nd in the first round

### Intelligent Game Design

#### Artificial Intelligence

- Learnt Unity 3D and how to incorporate Artificial Intelligence in different elements of a game such as NPC, environment etc.
- In this process, made my own game called as "Beyond Infinity", which was about space and physics simulator

## EXTRA-CURRICULAR ACTIVITIES

### General Secretary Technology

LLR Hall of Residence | July 2016 - April 2017

- Responsible for the performance of our Hall in the event of making an Ad campaign about a product or app
- Incharge of the video editing and computer graphics part of the video

### Autonomous Ground Vehicle

January 2015 - August 2015

- Contributed to designing an autonomous vehicle that can detect obstacles in a defined path and can avoid them
- Was responsible for developing the E-stop and remote control mechanism of the vehicle

### General Championship

August 2015 - March 2019

- Participated in various events on behalf of my Hall in the General Championship
- Sports: Football and Volleyball
- Technology: Ad Design, Product Design, Hardware Modelling
- Social and Cultural: Dramatics, Elocution

## ACADEMIC BACKGROUND

2014-2019	Dual Degree	Major: Aerospace Engineering Minor: Computer Science Engineering	Indian Institute of Technology, Kharagpur Indian Institute of Technology, Kharagpur	8.45 / 10 8.20 / 10
2013-2014	AISSCE		DAV Public School, Kota	92.4%
2012	SSCE		Kendriya Vidyalaya Burdwan	10 / 10