

Mehul Gupta

Data Scientist @ Tata 1mg

📍 Nagda, MP
📞 8103795345
✉ mehulgupta2016154@gmail.com
🌐 www.linkedin.com/in/mehulgupta7991
📧 medium.com/@mehulgupta_7991

Summary

Data Scientist with 2+ years of professional experience looking out to solve complex problems using AI; Experienced in developing data-driven solutions for automating digitalization of health documents reducing manual efforts by 100% for lab reports & 50% for prescriptions; Competent in developing models for diverse problems using Reinforcement Learning, Time Series, NLP, Computer Vision, and Data Mining as well.

Work experience

Data Scientist

July'20 - Ongoing

Tata 1mg

- Formulated **Lab Labeler**, an algorithm for information extraction from lab report pdfs by detecting important regions and extracting text sequences digitizing 4k reports daily & already parsed 3 Million historic reports covering 90% of total reports received
- Devised **GV-T transformation** algorithm for arranging OCR text output according to image orientation (using coordinate geometry) for information retrieval from image lab reports.
- Established pipeline for Lab reports digitization using Airflow & AWS Lambda Function parsing reports at 1 report/sec.
- Ideated & developed **patient summaries** using users' digital health data increasing user footfall on the Health Records section of Tata 1mg app by 25% (in progress)

Data Science Intern

May'19 - June'20

Tata 1mg

- Concocted algorithms for medicine names extraction Printed & Handwritten prescription images namely '**C-Cube**' & '**3 Step Filtering**' using rule-based filtering and clustering reducing time taken by digitizers for manual verification by 50%. Nearly 10% of all orders are completely automated as well.
- Trained **NIMA**(Neural Image Assessment) for scoring Prescription images on a scale of 1-10. Using TF-Lite, the model size was reduced to nearly 1/10 of the original size for deployment purpose
- Experimented to recognize different regions of interest in a prescription image using YOLO Object Detection

Founder & Editor

2019 - Ongoing

Data Science in your pocket

A [medium publication](#) covering different aspects of Data Science with 10k views/month & 13k read minutes/month

Publications

- **Author:** [Algorithms for rapid digitization of prescriptions](#), Visual Informatics (Elsevier, 2021)
- **Co-Author:** [Attended robotic process automation of prescriptions' digitization](#), Smart Health (Elsevier, 2021)
- 60+ [Data Science blogs](#) ranging from Reinforcement Learning, Time Series, GANs, Transformers in NLP, etc.

Core Skills

- Data Science
- RDBMS, Cassandra & AWS Athena
- Python, Java & C++
- Tableau, Metabase
- Data management using AWS S3
- Apache Airflow
- Comet ML

Relevant Projects

Image Translation using CycleGANs

Developed season translation (Summer-->Winter, Winter-->Summer) using CycleGANs

Vehicle Number Plate Detection

Vehicle number plate detection & extraction of Registration number using YOLOv3 object detection and Microsoft Vision API with an mAP = 98%

Off Policy Q Learning(Reinforcement learning) for Maze Game

The aim was to find a way out given random obstacles. A 4 X 36 Q Table was used with 36 states. No back step is allowed as well. Valid moves were up, down, left, and right.

ARIMA Forecasting Model for Time Series

Forecasted future Mean Global Temperature Using stationary Time Series with ARIMA Model where parameters were determined using autocorrelation and partial autocorrelation plots.

Cuisine Prediction

Determining Cuisine from the given list of ingredients by applying TF-IDF over ingredients and using Voting Classifier for predictions.

Education

B.Tech. PDPM Indian Institute Of Information Technology , Design and Manufacturing Jabalpur 8.1 CPI (CSE)	2016 - 2020
Senior Secondary Bharatiya Vidya Bhavan's K.D.K.V.M 94.2% (C.B.S.E)	2015 - 2016
High School Bharatiya Vidya Bhavan's K.D.K.V.M 10 CGPA (C.B.S.E.)	2013 - 2014

Certificates

- Data Scientist Career Track (Data Camp)
- Data Analyst Career Track
- Statistical Thinking In Python

[Link to Certificates](#)

Achievements

- Top Writer in Artificial Intelligence, Medium
- 1.2k+ reputation on [stack overflow](#)
- TCS HUMAIN finalist (AI competition organized by TCS, 2019)
- 25 Kaggle discussions medals.
- Secured 96th rank out of 8195 participants in Capgemini Data Science Tech Challenge 2018
- Shikshan Bharati Kulapati K.M. Munshi Award in Mathematics(2014)
- Bharatiya Vidya Bhavan Shri C. Subramaniam Award for excellence in character(2008-2009,2011-2012)
- Certificate of Merit(Proficiency in Co-curricular activities) for Declamation and Extempore