Jay Naresh Dhanwant

+91-9834123803 Email Linkedin Github

Education

IIT (BHU) Varanasi – *Integrated Dual Degree (2022)* **8.51/10**

Skills and Interest

Languages and Tools: Scala, Spark, Linux, Bash, Python, C++, SQL, MS-Excel, Flask, AWS
Areas of Interest: Data Structures, Algorithms, Machine Learning, Computer Vision, NLP

Experience

Data Science Intern | HiLabs Inc. (Pune, India) **January 2022 – Current**

- Worked directly on US-Insurance data to solve data quality issues in **value based care**.
- Developed **association mining algorithm** using **scala**. Engaged domain experts to reduce **false possitives**.
- **Proposed** and wrote **dfs based algorithm** for feature selection making manual workflow efficient.
- Automated model-results compilation in MS-Excel using **apache-poi**. Deployed the scala script on **AWS-EMR** cluster.

Data Scientist | Foudning Team Member | DoubtBuddy (Gurgaon, India) **November 2020 – November 2021**

- Spearheaded data science team to create multiple solutions and experienced a journey from 0 to **45000** users.
- Developed end-to-end **OCR Search** consisting of **google vision, latex parser, text prepossessing, SBERT and Approximated Nearest Neighbours** implemented using **FAISS** on **AWS EC2** running over **1Million+** records.
- Implemented fraud detection based on tutor behaviour and student-tutor conversations.
- Automated QnA extractions from text-book - saved 1000+ hours. [Recommendation]

Data Science Intern | Binhex AI Technologies (Bengaluru, India) **May 2021 – July 2021**

- Ideated, experimented, prototyped **invoice extractor** robust to key-value template variations.
- Developed pipeline with transfer learning on object detection, OCR, NER (auto-annotation) parsing.
- Designed disjoint set union based rectifier for OCR and YOLOv5 increasing successful extractions from **60% to 95%+**. Developed a demonstrable prototype using HTML, Flask NGROK. [Video Demo]

Data Science Intern | Flatfolder Ltd (Helsinki, Finland) **May 2020 – July 2020**

- Utilised scraped ad-data for (1) Selling and Rental modelling (2) Trend Analysis (2017-2020) (3) Unseen appartments' rental and selling price estimation **(r_2:**0.96)*for* **0.4 Million+ data points**.
- Automated ML pipeline for ETL(MySQL) clustering (DBSCAN) and training of tuned Catboost.
- Developed FlatFolder Score of financial health. This mutually serves landlord tenants for effective credit check using tenant's transaction history (Tink API).
- Designed KPIs including RTI, persistence, cash flow, gradients, etc. to calculate tenant's score for a rent budget. The solution bifurcates secondary accounts detects recurrent expenses.

Projects

Remote, Satellite office and HQ allotment - Sterlite Power | Havish M. Consulting

- Automated **HR workflow** using employee-generated labelled form data. Deployed ML algorithms in **MS Excel**.
- Exposure - **xlwings, pearson correlation, f-classif permutation based feature importance.** [Recommendation]

Customer query clustering on CRM data | IT Data Consulting, LLC

- Solved database redundancies by predicting multiple queries from the same person.
- Exposure - Hierarchical clustering, Caverphone, custom distance metric. [Recommendation]

Thesis Project — Pathogen Detection using spectroscopy | Dr. V. Ramanathan

- Designed and trained custom 1D-ResNet on 60,000 spectra of 30 pathogen stains grouped into 7 treatment classes. Observed 99.1% treatment accuracy and 82% stain accuracy.
- Exposure - Data Cleaning Preprocessing, Tensorflow, Bi-LSTMs, 1D-ResNet.

Covid-19 Risk Factor Predictor and Pathology report parser | MediPort.in

- Performed statistical analysis to find significant blood parameters affecting the prognosis under the supervision of Dr. Aman Dev (IMS-BHU). Exposed the service and OCR-based parser using Flask.
- Work featured in various news medium including [Lokmat], [Times of India], [Jagran], [Health Wire] etc.

Relative Grading system | Dr. V. Ramanathan

- Presented drawbacks of current evaluation methods and proposed clustering-based alternative.
- Exposure - Relative Grading, DBSCAN , K-Means clustering, Seaborn. [Kaggle Kernel]

Social Contact Pattern Estimation | Dr. V. Ramanathan

- Designed susceptible-infected-recovered model with custom loss function using SciPy [Public-reviewed paper]

Beer Recommendation | AbInBev Mavericks 2.0 Hackathon

- Created time-variant beer-recommendation-system using whole-saler features purchase history.
- Carried out RFM analysis for cross-sell upsell. Within top 20/2204 teams. [Github]

Position of Responsibilities / Extra-Curricular Activities

Manager — Technex 2020 | IIT(BHU) Varanasi

- Led an 8 member team to execute institute's case-study event with a footfall of **1000+** students.

Educator | Unacademy

- Created online lectures in engineering subjects with **11,400+ views and 700+ followers**.