# Mainak Chain

mainakchain21@gmail.com
https://mainakchain.github.io/ +91-9932349443 | LinkedIn | GitHub | Kaggle

Data Scientist with 1.5 years of experience applying AI/ML techniques to solve real world problems. Kaggle competitions contributor with strong coding and analytical skills. B.Tech and M.Tech from IIT Kharagpur

## Education

**Indian Institute of Technology, Kharagpur** Jul 2015 – Apr 2020
*Dual-Degree (B.Tech + M.Tech), Metallurgical and Materials Engineering, CGPA: 8.36/10, Dept. Rank: 4th*

- Courses: Design & Analysis of Algorithms, Linear Algebra, Software Engineering, Probability & Statistics
- MOOCs: Machine Learning, TensorFlow for Deep Learning Research, Natural Language Processing

## Work Experience

**ANI Technologies Private Limited (Ola) — Data Science Team** Bangalore, KA, India
*Research Engineer/Data Scientist* Sep 2020 - Present

Enhanced *customer* and *partner experience* for 200M+ customers & 2.5M+ partners on the Ola Cabs app

- **Digitization, Automating Data Capturing**: Impact: $2M PA. Laid the foundation for creating a seamless scalable partner on-boarding experience with automation of data creation and validation
  - Developed *scalable* and *fast* API services in **Django** for 25+ documents across 20+ countries
  - Created text and image based document type and sub-type classifiers with **98%+** mean accuracy
  - Built & tuned **transfer-learning** models with **VGG16** for document classification with **99%** precision
  - Optimised key-value retrieval from OCR with improved RegEx and devised *contextual confidence-scores*
- **Allocation Optimisation**: Impact: $0.8M PA. Smarter cab-filtering for allocations
  - Implemented spatio-temporal city-level filters for cab allocation to construct **isochrones**
  - Engineered and optimised a pipeline for **isochrone** creation and storage using *Concaveman*, *HDBSCAN*, *Spark* & *Amazon S3*, handling **1.5 PB** of pings data. Provided truer discovery ETA & reduced API calls
- **Ride Use-case Identification**:
  - Built a pipeline to generate use-case tags for 50+ cities across IN, AU, NZ and UK using in-house data sources & OSM. Created heuristics-based ride level use-cases with spatio-temporal booking patterns
  - Built an **Active Learning** driven pipeline for predicting ride use-cases using hive, modAL, shiny & sf
  - Created interactive dashboards for visual analysis to aid **decision-making** for strategy team
- **Traffic Lights Optimisation**: Minimised wait-time for cars in road-network to mitigate congestion
  - Simulated the traffic scenario using **SUMO** framework & in-house cab-pings data (Baseline established)
  - Devised algorithms to optimise congestion locally and globally with fixed and dynamic traffic lights
  - Developed an RL agent on the sumo environment with various policies to decrease wait-time by **18.3%**
- **Others**: In-house utility-packages (pyutilsds & rutilsds), Drop suggestion API, Conditional LTV model

**ANI Technologies Private Limited (Ola) — Data Science Team** Bangalore, KA, India
*Research Engineer Intern* May 2019 – Jul 2019

- **Scalable Drop Suggestions Model**: Possible Savings: $2M PA, **0.4%** increase in booking-completions
  - Built a Gradient-Boosted Tree based drop-suggestion model adding better engineered time-based features (recency & frequency). Improved on the previous in-production model by reducing error by **18.7%**
  - Spearheaded the development of **one-touch booking system** to facilitate hassle-free bookings
  - Granted a **pre-placement offer** for showcasing excellence in performance and project results

**Innoplexus Consulting Services Private Limited — Innovation Team** Pune, MH, India
*Data Science Intern* May 2018 – Jul 2018

- **Text Summarization of Clinical Trial (CT) documents**:
  - Built a graphical sentence ranking algorithm like TextRank for extractive summarization(ROUGE-L: 31)
  - Designed UMLS based biomedical-tokenizer and encoder in **transformer** model pipeline in TensorFlow

* Awarded a **pre-placement offer** for data scientist position for exceeding expectations in deliverables

**Dipper Technologies Private Limited — Core Technology Team** New Delhi, DL, India
*Data Science Intern* Nov 2017 – Dec 2017

- **ETA Model and Road Logistics Optimization**:
  * Constructed pipeline to fetch GPS-timestamp data for 800+ trucks to engineer features for modelling
  * Constructed a **DNN** model for real-time predictions of ETA on toll booths using historical travel data
  * Analyzed 30+ routes & optimized road logistics for 250+ trucks by stoppage clustering using **DBSCAN**

## Publications and Conferences

**Neural factorization for Offer Recommendation using Knowledge Graph Embeddings**:
G. Chowdhury, M. Srilakshmi, M. Chain, S. Sarkar
*Special Interest Group on Information Retrieval Workshop On eCommerce (SIGIReCom), 2019*

**Decision Support System for Prediction of Occupational Accident: A Case study on Steel Plant**:
S. Sarkar, M. Chain, S. Nayak, J. Maiti
*Advances in Intelligent Systems and Computing series (AISC), Springer 2018*

**Data-driven Decision Support System for Prediction of Occupational Accidents**:
S. Sarkar, M. Chain, S. Nayak, J. Maiti
*Institute of Industrial and Systems Engineers Annual Conference (IISE), 2018, Orlando, Florida, USA*

## Projects

- **Offer recommendation using Retailer Sales data**: B.Tech Project [paper]
  - Built a **neural factorization machine model** to generate score for a given user-item pair using their embeddings from **knowledge graph**. These scores were used to rank the items for recommendation.
  - Improved the model by adding temporal features using **LSTM** with **attention** framework
  - Boosted recall@5 & MRR@5 adding temporal features, beating tree-based methods by **91%** & **93%** resp.
- **Development of an Early Warning System**:
  - Developed a real-time health monitoring app that notifies of imminent threat to any site-worker's health
  - Trained a SVM classifier for workers' health prediction using environment's WBGT & RSPM values and worker's blood oxygen-level & heart-rate as input signals. Tuned hyper-parameters with genetic algorithm
- **Development of a Decision Support System (DSS)**:
  - Built a DSS with Kivy to help decision-maker strengthen occupational safety & alleviate potential hazards
  - Integrated pipeline with feature selection, descriptive analysis and automatic model evaluation & selection

## Technical Skills

**Deep Learning**: NLP, Computer Vision, Time Series Analysis, Training models on GPU/CUDA/TPU
**Modeling**: CNNs, RNNs, Gradient Boosting, Attention/Transformer/BERT, Linear Regression, SVM
**Python Packages**: TensorFlow, PyTorch, scikit-learn, LightGBM/XGBoost, OpenCV, HyperOpt, NLTK
**Experience with**: R, C/C++, SQL, Apache Spark, Apache Hive, AWS, Linux, Django, Apache Solr

## Achievements

- Received the **Spotlight Award** for impact created on business by enhancing customer-experience in Ola app
- Presenting author at IEMIS, 2018. Achieved the **best paper award** for our paper presentation.
- Top 10% in Microsoft code.fun.do 2018, Top 3% - AMEX AnalyzeThis 2017
- Captained my team to win the Fine Arts Cup at SoCult General Championships 2019, IIT Kharagpur
- Privileged as 1 of the 1000 eligible students across India for **Prime Minister's Scholarship Scheme 2016**
- Bagged **81st** position among **2000+ teams** in Amex Analyze This 2017, conducted by American Express
- Procured **multiple gold medals** in Inter-Hall & Open-IIT Fine Arts Competitions at IIT Kharagpur