Md Danish Kalim

Staff Machine Learning Engineer, Sharechat || B. Tech. IIT Guwahati

+91 9085 433 444 @ danishkalim14@gmail.com LinkedIn Bangalore, India

Education

- **Indian Institute Of Technology** Guwahati, India
  Bachelor of Technology - Electronics and Communication Engineering | CGPA: 7.63 *2012-2016*

Experience

- **Staff Machine Learning Engineer | Sharechat** | Bangalore *Dec 2019 - Oct 2023*
  - Leading the Ranker System Team in building an AI first feed recommendation system for **180 million** MAU
  - Built realtime system to serve **12k+** peak ranking rps using accelerator hardware.
- **Senior Data Scientist | Paytmlabs, Paytm** | New Delhi *Jul 2016 - Dec 2019*
  - Built end to end machine learning solution at scale for **450 million+** customers
  - Built feature store comprising of **12k+ features**
  - Built personalisation, customer ratings, credit scoring and churn prediction models
- **AI Intern, HOLMES Platform | Wipro Technologies Limited** | Bangalore *May 2015 - Jul 2015*
  - The main goal was to perform the **human level intelligent operations** and reduce the manual effort mainly in the repetitive activities, build its **own decision making capabilities** like a human being and increase the throughput. **Deep Learning** was applied for validating the action performed and **reinforcement learning** was applied for intelligence.

    Language: **C++, Python** Library: **OpenCV, Win32, Keras**

Skills

- **Programming**: Python, Scala, SQL, Bash, LaTeX, C/C++
- **ML Framework**: TensorFlow/Keras, Spark MLlib/GraphX, Intel BigDL/TensorFlowOnSpark/DL4j, Catboost
- **Big Data Framework**: Apache Spark/Spark SQL/Spark Streaming/GraphX, BigQuery/Hive, Hadoop, BigTable
- **Production Tool**: Kubernates/EMR, Docker, Airflow, AWS/GCP
- **Soft Skills**: Leadership, Team Management

Projects

- **Re-Ranking module in Feed Ranking | Sharechat**
  - Improved the Value model setup from tedious manual online AB tuning to offline optimised Learnt Value Model. **Reduced time** to shipment from **3 months to 1 month** for addition of new actions.

    Algorithm: **GPR/ Bayesian Optimisation** Library Used: **BoTorch, Ax**
- **Ranking module in Feed Ranking | Sharechat**
  - Improved the feed ranking model using Multigate Mixture of Expert (MMoE). **Increased** the pointwise AUC in offline results by **10%** and retention by **50 bps** (the most significant gain witnessed in last 2 years at Sharechat). Currently under review for presentation at the FIRE conference

    Algorithm: **MMOE** Library Used: **Tensorflow Recomenders, TPU Embedding**
- **Early Post Life Cycle Recall model in Feed Ranking | Sharechat**
  - Developed enhanced semantic representation learning by integrating behavioural and content data to create a robust representation suitable for the early post-life cycle. **Increased** the **ctr** of early posts by **32 %**. Poster accepted at Stanford Graph Workshop 2023.

    Algorithm: **GraphSAGE, Streaming Factorisation Machine** Library Used: **DGL**
- **Recall model in Feed Ranking based on Knowledge Graph | Sharechat**
  - Built a Knowledge Graph of Sharechat ecosystem capturing user's preferences/aversions. Performed node/edge representation learning. Missing link prediction was used as feed candidates. **Increased** the **retention** of consumer by **15 bps**

    Algorithm: **ComplEx** Library Used: **Pytorch BigGraph, SCANN**
- **Multi-model Representation learning for Video | Sharechat**
  - Inspired by the SimCLR paper by Google Research, applied a similar approach for visual representation learning in video. Created a bipartite graph having user and post. Performed early fusion of all modality and applied unsupervised Graph Convolution Network.

    Algorithm: **SimCLR, PinSAGE/GraphSAGE** Library: **DGL**
- **Actor Metadata Extraction in Video Post | Sharechat**
  - Built model to extract gender, face embedding of all humans present in video. Identified presence of major influencers in video. **Increased** the **accuracy** of the multimodal tag prediction by **7%**.

    Algorithm: **Retina Face** Library: **InsightFace, Hecate**
- **Identifying Associated Accounts | Paytm**

- Generated unique user id for customers having multiple Paytm accounts using semi-supervised learning and graph algorithm. **Reduced the cost** of campaigns by **30%** by deduplicating the customer base in social media reactivation.

  Algorithm: **Random Forest, Graph** Library: **GraphX/Spark ML**

- **Customer Affinity Generation | Paytm**
  - Generated affinity of customers towards L2 and L4 levels of catalog **Increased** click-through rate (**CTR**) of icon rail by **17%** Used to run improved targeted campaigns for lapsed customers by providing discounts offers.

    Algorithm: **Word2Vec, LSTM** Library: **Intel BigDL/DL4j, Spark ML**

- **Credit Score for Paytm Postpaid | Paytm**
  - Built credit score model in collaboration with ICICI bank using more than 100 transactional features of a customer Used by Paytm Postpaid service to provide credit to low risk base.

    Algorithm: **Logistic Regression** Library: **Spark ML**

Invited Talks

- **Enhanced Semantic Representation Learning** — Stanford University
  Poster presentation at prestigious Stanford Graph Learning Workshop 2023 — *Oct 2023*
- **Knowledge Graph based Feed Recommendation System** — IIIT Delhi
  Invited by Prof. Raghava Mutharaju for guest lecture in course Semantic Web — *Mar 2022*
- **Machine Learning in Industry** — CIC, University of Delhi
  Invited by Prof. Shobha Bagai for guest lecture. — *Oct 2018*

Academic Projects

- **Diagnosis of Diabetic Retinopathy Using Deep Learning** — B. Tech. Project
  Prof. Amit Sethi, IIT Bombay — *Jul 2015 - Apr 2016*
  - The project aims to **detect abnormalities** in the blood vessels of the retinal tissue. The system works on a method for **segmenting** and **recognizing** blood vessels in high resolution fundus images of the retina. It employs a pre-processing (PCA, CLAHE and bilateral filter) followed by **deep learning** based blood vessel detection. Finally the local connectivity property of blood vessels was used for noise eradication.
  - Performance achieved was **Accuracy = 94.3** and **Sensitivity = 89.9** (superior to the prevailing state of the art)
  - Algorithm: **PCA, CLAHE, Bilateral Filter, CNN** Library: **OpenCV, Theano**

- **Facial Key point Detection Using Deep Learning** — *Oct 2015*
  - The project aims to predict key point location on face images. The system employs **deep learning** to detect features and correspondingly locate key points. Due to imbalance of data for each class, the system employed a **specialist CNN** for each class. **Dropout** and **data augmentation** is performed for better generalization.
  - Algorithm: **CNN** Library: **Lasagne**

Achievements

- **Entrance exam**: Secured **AIR-1934 (99.6 %ile)** & **AIR-1996(99.82 %ile)** in IIT-JEE & AIEEE 2012 respectively. Secured state rank **110 (99.93 %ile)** in WBJEE 2012.
- **School Level**: **Gold Medal** for Academic Excellence issued by St. Anthony's High School, Kolkata, 2010
- **Aptitude Test**: Award recipient from National Level Talent Search Examination, 2012

Volunteer Experience

Fellowship & Conferences

- International fellow of **FastAI** and **NurtureAI** programs.
  Attended Fifth Elephant Conference, 2017

Community Experience

- Volunteer of **Saarathi**, an NGO aimed towards prevention of **Human Trafficking** and helping victims of human trafficking. Active member of **National Social Service** since 2012.