

Shikhar Agrawal

+91 98881 93338 | me@shikharagrawal.com | [linkedin.com/in/shikhar](https://www.linkedin.com/in/shikhar) | github.com/shikhar | www.shikharagrawal.com

EDUCATION

Indian Institute of Technology, Hyderabad

Hyderabad, IN

Bachelor of Technology in Computer Science and Engineering with Honors - CGPA: 7.07

Aug. 2015 – May 2019

Relevant Coursework: Data Structures and Algorithms, Database Management, Applied Machine Learning, Operating Systems, Data Mining, Information Retrieval, Principles of Programming Languages, Compilers

TECHNICAL SKILLS

Languages: Python, C++, Golang, Java, MySQL, PostgreSQL, MongoDB, Haskell, LaTeX, HTML, CSS, JS

Frameworks: Django, PySpark, WordPress, Swagger-UI, FastAPI

Developer Tools: AWS, Git, SVN, Selenium, Vim, Kubernetes, Google Cloud Platform, PyCharm, Eclipse

ML Tools, Frameworks, and Libraries: TensorFlow, Keras, ElasticSearch, Pandas, NumPy, Matplotlib, Sci-kit Learn, MLFlow

EXPERIENCE

Associate

July 2021 – Present

Global Stock Selection, AQR Capital

Bengaluru, IN

- **Batch API:** implemented batch API that acts as a wrapper for AWS Batch distributed framework. This enabled every user to submit multiple batch tasks with each task running a specific function with different argument. The API combines the result and stores it in **Redis/S3**. The user can override batch parameters, can give custom code paths to override production code. This API was significant work in batch migration.
- **Systematic Weighting:** implemented and deployed the systematic weighting framework that **dynamically assigns weights** to factors. Traditionally this weights were assigned manually. This new process **improved the accuracy** of model generation along with returns and decreased the turn around time from a **week to a day**. Implemented regression framework that compares the changes in code/data with respect to production and **flags** the models that **fails** regression.
- **Views/Returns Loader:** designed and developed new framework to get factor view and returns. Traditional process utilised SQL server to get factor views/returns and had **scalability** issues. The new architecture utilised **DaaS**(Data as a Service) which was built internally that utilised AWS S3/Redis/RDS as data layers. It resulted in speed up of view/returns retrieval by over **200%** and could handle more than **10 times** the previous concurrent requests.
- **Cluster API:** Cluster API is a wrapper developed around HT Condor for **distributed computing**. I worked on several improvements in the API from ability to submit and **distribute tasks** on **AWS EC2 instances** while keeping the user interaction same, moved all shared drive data to S3, and added ability to view **live logs** in the Cluster UI.
- **Graph Compute:** Graph Compute is one of the most important process that generates daily **model views**. The code was part of huge monolithic application shared with other team and worked only in python 2. I worked on **modularising** the monolithic application into multiple small modules while converting all the code to python 3. **Found and fixed multiple bugs** in the original codebase. **Implemented a regression framework** that compares the model views and flags models that differ in views above a threshold.
- **NLP/ML Signals:** worked on NLP model averaging and ML Combo jobs that enhances factors. ML combo use data driven approach and **ML methods** to find best way to combine factors. NLP model averaging jobs **process text data** from multiple sources and **enrich the factor views**. I also migrated these jobs to batch utilising the batch API developed by me that significantly reduced the daily run time from **6hrs to 2hrs**.

Data Scientist / Software Developer

July 2019 – 2021

Zenlabs (R&D), Zensar

Pune, IN

- **News influence on product sales: extracted and processed** news articles for the products in Ecuador's supermarket chain. Extracted **key events** from the news headlines. Implemented a **relevancy module** that retrieves relevant news articles using by generating keywords from **148 million** Amazon reviews. A **Hybrid Attention Network (HAN)** was used to get feature vectors. These vectors were then used in **framenet, deepglo, and MALSTM** for multimodal sales forecasting. Achieved **better sMAPE** values for few categories.
- **ZEVA (Zensar's Enterprise Virtual Assistant):** Zeva is Zensar's voice-based enterprise personal assistant. **Feedback based continuous learning:** designed and implemented a feedback loop for improving the answers by ZEVA. Feedback is provided through a **like/dislike** button. Trained a **ML model** to continuously learn from feedback. **Unstructured text retrieval:** implemented a custom fine-tuned model based on **BERT** to retrieve answers from unstructured text. **Improved accuracy** by approximately **48%** compared to previous IR model.
- **Address Matching Solution:** built a scalable solution for a client to match the names and addresses of customers from disparate sources. **Achieved 90% overall accuracy** and **95% precision** in one of the categories by leveraging Natural Language Processing(NLP) techniques to replace the current rule-based manual approach.
- **Social Distancing Tracker:** implemented a social distancing tracker using **Computer Vision (CV)** to track the violation of social distancing norms. It requires a **one-time calibration step**. IDs of people violating the norms get **logged in the SQL database**. This solution can be integrated with existing CCTV surveillance in retail and manufacturing workplaces.

- **Image Similarity Solution:** implemented an image similarity solution for a major **UK based retail client**. Multiple algorithms were built, tested, and compared based on **accuracy, response time, and match rate** for various image transformations.

SELECTED PROJECTS

Fake News Detection | *Python, NLP*

Jan 2019 – Apr 2020

- Collected data on rumours in social media sites such as Twitter. Classified the tweets following a rumour by its stance into one of **supporting, denying, questioning or commenting**. Used the **Hawkes process (HP)** approach to decide the veracity of the rumour.

Finding Active Expert Users in CQA System | *Python, Django, NLP*

Aug 2018 – Nov 2018

- Used topic modeling techniques such as the **LDA** to find expert users for a given topic. Used **time-series analysis** to find the **active users** in the system. Used **ensembling techniques** to find active expert users in Community Question Answering (CQA) system such as Stack Exchange.

P2P File Sharing Application | *GoLang*

Aug 2017 – Nov 2017

- Built a **file-sharing application** where files can be **downloaded/uploaded** by multiple people on the server.

PATENTS

- **Application Number - 201921052195:** A system and a method for facilitating meeting content for attending a meeting.
- **Application Number - 202021021207:** A system for alerting an user to prevent free fall of an object.
- **Application Number - 202021036190:** Responding to customer's query while identifying topic drift in an ongoing conversation along with customer agents optimization.
- **Application Number - 202021038710:** System and method for determining influence of online content on products using Artificial Intelligence (AI).

ACHIEVEMENTS

- Achieved **All India Rank(AIR) of 964** in JEE Advanced out of **13,56,000 candidates** (2015)