Gaurav **Kumar**

+1-619-568-0949 | gkumar@ucsd.edu | gaurav5590.github.io | gaurav5590 | gauravkumar5590

Education

University of California, San Diego (UCSD) — *San Diego, CA*

Master of Science in Computer Science (Specialization in AI) — *Expected: Jun. 2022*

Indian Institute of Technology, Kanpur — *Kanpur, India*

Bachelor of Technology in Electrical Engineering, Minor: Machine Learning (**GPA:** $9.0/10$) — *Jun. 2018*

Skills

Computer Languages	C, C++, Python, R, Matlab, Java, SQL
Software Tools	MS Office, Git, Vim, Linux, Docker, React, Google Colab, Ampss, Php, Jupyter, VSCode, Android Studio
Machine Learning Tools	Pandas, Keras, TensorFlow, TfLite, PyTorch, Scikit-Learn, Opencv

Coursework

- **Computer Science:** Natural Language Processing, Machine Learning, Image Modeling Techniques, Image Processing, Data Structure & Algorithms, Design & analysis of algorithms, Database Systems, Data Mining, Convex Optimization, Fundamentals of Computing
- **Mathematics and Others:** Probability, Statistics, Differential Equations, Linear Algebra, Game Theory, Complex Variables, Calculus

Professional Experience

Samsung Research Institute | Machine Learning Engineer, Bixby Assistant Team — *Jul. 2018 - Jan. 2021* | *Bangalore*

- Wrote a light-weight, *transformer* based domain classifier for text data in Pytorch. Deployed the trained models onto the production server using Java and C++. Utilized ML methods such as label smoothing, residual layering, incremental learning, and knowledge distillation to boost performance.
- Designed a Dialogue Management System using a multi-stream deep learning architecture. Leveraged deep learning techniques like Graph Convolution Networks, Memory Networks and BERT. Incorporated external knowledge in system through knowledge graphs. Work published at LREC 2020.
- Handled data processing pipeline including the regular training, testing and release of a machine learning based intent classifier in Python.

University of Tuebingen | Research Intern, CIN Lab — *May. 2018 - Jul. 2018* | *Germany*

- Worked on a training pipeline to estimate material properties (texture, stiffness) from cloth video data. Experimented with CNN+LSTM networks, 3D CNNs, 2 stream, multi loss networks along with the optical flow for video feature extraction. Proposed triplet loss based optimization for model training.
- Built a web interface (HTML, CSS, PHP, JS) to gather user data and understand the correlation between human perception and neural networks.

Samsung Research Institute | Summer SDE Intern, Knox and Security Team — *May. 2017 - Jul. 2017* | *Bangalore*

- Objective was to identify confidential content by reading text inside a document image and build an image classifier to judge the degree of risk.
- Used Recurrent Neural Network to model sequential text image features. Deployed the Tensorflow model efficiently using an Android App.

Projects

Handwritten Mathematical equations to LaTeX | Course project, Image Modeling Techniques — *Spring '18* | *IIT Kanpur*

- Developed a deep learning pipeline to convert handwritten mathematical equation images to latex. Incorporated edge detectors, Hough Transform and image segmentation during pre-processing and built a novel tree structure based algorithm to identify multilevel superscripts and subscripts
- Trained multiple classifiers including SVM, CNN, Random Forests for symbol recognition with 98% accuracy even on complex trigonometric symbols

Detecting Semantically Similar Question Pairs | Course Project, Natural Language Processing — *Spring '18* | *IIT Kanpur*

- Executed Siamese LSTM, 1D CNN, and attention-based methods (python, keras) to detect duplicate question pairs on Quora Question Dataset.
- Experimented with embedding methods (Word2vec, Glove, LSA) to improve the feature space and utilized ensemble models to generate 82% accuracy

Captcha Breaking | Course Project, Machine Learning Techniques — *Fall '17* | *IIT Kanpur*

- Utilized feature engineering methods such as clustering and dominating color based segmentation to denoise SquirrelMail captcha images
- Implemented CNN in Python (Pytorch) for character recognition and reached 98% accuracy. Clocked 85% accuracy for extremely noisy captchas

Auto Insurance Claim Prediction | Data Mining Project — *Spring '18* | *IIT Kanpur*

- Built a machine learning model to determine the probability of a driver initiating an auto insurance claim. Performed data cleaning in Pandas.
- Implemented machine learning methods like Random Forests, SVM, XgBoost, neural networks, K means Clustering. Achieved 86% test accuracy.

Publications

- **Gaurav Kumar**, Rishabh Joshi, Jaspreet Singh, Promod Yenigalla. "AMUSED: A Multi-Stream Vector Representation Method for Use in Natural Dialogue." International Conference on Language Resources and Evaluation (LREC) 2020
- Bi, Wenyan, **Gaurav Kumar**, Hendrikje Nienborg, and Bei Xiao. "Understanding Information Processing Mechanisms for Estimating Material Properties of Cloth in Deep Neural Networks." Journal of Vision 19, no. 10 (2019): 297c-297c

Honors & Awards

- Received the prestigious J.N. Tata Endowment Scholarship for higher education of Indians to pursue my studies at UCSD
- Awarded Samsung Citizen Award for exceptional contribution and excellence in work at Samsung Research Institute in 2019 and 2020
- Received Academic Excellence Award for exceptional academic performance in 2015-16 session at IIT Kanpur
- Secured All India Rank 519 in IIT-JEE Advanced and achieved 99.91 percentile in JEE Mains Examination 2014 out of 1.3 million candidates