Kapil Pathak

Education

- **Indian Institute of Science, Bangalore** — Karnataka, India
  Master of Technology (Research) - Computer Science; CGPA: 8.0 — *July 2018 - April 2021*
  Advisor: *Prof. Shirish Shevade (Computer Science and Automation, IISc)*
  Research Interests: *Fine-grained Opinion Mining, Natural Language Understanding for Reviews and Dialogue Systems, Sentiment-oriented Relation Extraction for Aspect-Opinion pairs in the Product Reviews, Design of Pre-training approaches to substitute Masked-Language Model (MLM) for various tasks in Conversational Systems*

- **Indian Institute of Technology (IIT), Gandhinagar** — Gujarat, India
  Bachelor of Technology - Electrical Engineering; CGPA: 7.69 — *July 2013 - June 2017*

- **Sir Parashurambhau College, Pune** — Maharashtra, India
  Higher Secondary Certificate (HSC); Percentage: 86.00 — *July 2011 - June 2013*

- **Anant English School, Satara** — Maharashtra, India
  Secondary School Certificate (SSC); Percentage: 97.27 — *July 2005 - June 2011*

Internship and Experience

- **Tata Research Design and Development Center (TRDDC)** — Pune
  Internship Advisor: Arun Bahulkar, Rajendra Gokhale, TCS Research — *May 2019 - Aug 2019*
  - To solve Opening Lead problem in Contract Bridge game bot using past play history and the auction.
  - Design of deep learning based opening lead predictor equivalent to rule based system with less response time.

- **Capgemini Technology Services India Limited**
  Senior Software Engineer — *Sept 2017 - July 2018*
  - Morgan Stanley: Performance Engineer.

- **Royal Melbourne Institute of Technology (RMIT University)** — Melbourne, Australia
  Internship Advisor: Prof. Dinesh Kant Kumar, RMIT University — *May 2016 - July 2016*
  - To detect the abundance of Oxy-haemoglobin and Deoxy-haemoglobin from hyper spectral imaging using constrained ICA.

- **Underwriters Laboratory Inc: An Investigation of Electrically Induced Fire** — Northbrook, USA
  Internship Advisor: Dr. Pravinray Gandhi, Director, Corporate Research, UL. — *July 2015*
  - **Phase 1 & 2**: To prepare an elaborate report on causes, statistical studies and case studies on electrical fire data and design experiments demonstrating specific causes and corresponding effects of electrically induced fire.
  - **Phase 3 @UL lab**: To carry out power data acquisition of electronic ballasts used for fluorescent lights and also conducted IR (thermal) imaging on the ballasts.

Projects

- **Building an Aspect-based Sentiment Analysis Model for Unstructured Reviews in Mattress Domain**
  https://github.com/Kapil-Pathak/ABSA-from-unstructured-reviews
  - To identify different aspects from unstructured product reviews by designing dependency parsing rules and perform clustering on aspects to identify different product categories followed by sentiment analysis with respect to those aspects.
  - Streamlit application deployed on EC2 with the link: http://13.233.35.211:8501/

- **Taxi demand prediction in New York City** — July'19
  Dataset: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
  - Trip data is clustered into 40 clusters using kmeans algorithm based on location data. Further each cluster data is binned into 10 min time span to convert the demand prediction problem into time series problem.
  - XGBoost and Random Forest regressors are used along with double exponential smoothing to solve the demand prediction problem.

- **Question Answering with Reasoning Across Common Knowledge Base** — Jan'19 - April'19
  Course: Natural Language Understanding, Instructor: Prof. Partha Talukdar, IISc Bangalore
  - Retrieving question-related sentences from the common knowledge base (ARC Challenge Dataset).
  - Creation of entity graph from the retrieved sentences, reasoning across retrieved sentences through a relational graph with graph attention network.
  - Report and code available at https://github.com/Kapil-Pathak/QA-GAT

- **Distributed Training Strategies for Knowledge Base Completion Model(ConvKB)** Aug'18 - Dec'18
  Course: Machine Learning with Large Datasets, Instructor: Prof. Partha Talukdar, IISc Bangalore
  - Implemented and scaled knowledge graph embedding ConvKB for pre-preocessed KB Freebase, studied different parallel algorithm schemes such as synchronous, asynchronous, stale-synchronous and downpour training strategies on FB15K-237 as well as pre-processed FBase.
  - Working hadoop and distributed tensorflow codes :https://github.com/Kapil-Pathak/Dist-ConvKB

- **Semantic Search Engine for QA Systems** July 2020
  https://www.kaggle.com/stackoverflow/stacksample
  - Built hybrid keyword and semantic similarity based Search Engine for QA using ElasticSearch, Universal Sentence Encoder.
  - Dataset: StackSample: 10% of Stack Overflow QA

- **ICDAR 2019 Robust Reading Challenge on Scanned Receipts OCR and Information Extraction** Now
  Project Collaborator: Mr. Saurabh Zha, Director of Data Science
  - Information Extraction from Form-like Documents using Graph Convolution Network.
  - Dataset: ICDAR 2019 Robust Reading Challenge on Scanned Receipts OCR and Information Extraction.

- **Answer Evidence Extraction using Reinforcement Learning** Aug'19 - Dec'19
  Course: Deep Learning for Natural Language Processing, Instructor: Prof. Shirish Shevade, IISc Bangalore
  - Employed BERT based Question-Answering model for QA task on HotpotQA dataset.
  - Implemented actor critic algorithm to select sentences as evidences given question and a set of passages to support Question-Answering task.

- **Women Apparel Product Recommendation Engine** July 2020
  https://www.kaggle.com/ajaysh/women-apparel-recommendation-engine-amazoncom
  - Implemented a content based recommendation engine using weighted word2vec and tf-idf based text features and CNN based image features.
  - The products are ranked according to cosine-similarity among weighted product representations.

- **Self-driving car application for steering angle prediction** Jun 2020
  https://medium.com/@pathak.kapil/self-driving-car-steering-angle-prediction-304517df69d0
  - Trained a CNN based model regressed with steering angles and created Stream-lit based app to deploy the model.
  - Dataset: https://github.com/SullyChen/Autopilot-TensorFlow

Relevant Coursework

- **IISc Bangalore**: Linear Algebra and Probability, Machine Learning with Large Datasets, Practical Data Science, Machine Learning, Natural Language Understanding, Deep Learning for Natural Language Processing, Computational Methods in Optimization
- **IIT Gandhinagar**: Probability and Random Processes, Digital Signal Processing, Digital Image Processing, Advanced Signal Processing, Signals and Systems

Computer Skills

- **Intermediate Knowledge**: Python, Pytorch, Scikit, NLTK, Keras
- **Basic Knowledge**: TensorFlow, Distributed Tensorflow, Hadoop, C++, MLlib, AWS EC2, Elastic Search

Honors and Awards

- GATE(ECE) 2018: Secured AIR 328 with score of (747/1000)
- JEE Advanced (2013): Secured AIR 3168
- Awarded for appearing in top 20% cumulative performance-wise among 150 participants in 'Summer School in Recent Advances in Computer Vision' at IIIT Hyderabad
- Winner of 2014 UL Engineering Challenge: Investigation of Electrically Induced Fire.
- National Talent Search Examination (NTSE), Year of Selection: 2009

Interest and Activities

- Represented IISc Bangalore and secured 4th Position in Inter IISER Sports Meet (IISM, IISER Pune) (Game: Chess)
- Represented IIT Gandhinagar at Inter IIT Sports Meet 2013, IIT Guwahati (Game: Cricket)
- Table Tennis, Contract Bridge, Travelling.

Personal DATA

- **IISc Email ID**: kapilpathak@iisc.ac.in
- **Gmail ID**: pathak.kapil@btech2013.iitgn.ac.in
- **Skype ID**: live:.cid.6d210e5ce1c7a743
- **Mobile No**: 8329877183, 7249489064
- **LinkedIn**: linkedin.com/in/kapil-pathak-12032882
- **Github**: https://github.com/Kapil-Pathak