Beethika Tripathi

Indian Institute of Technology Madras

📞 +91-9199113825 • ✉ beethik@cse.iitm.ac.in

Education

Program	Institution	%/CGPA	Year of Completion
M.S. (Computer Science and Engg.)	Indian Institute of Technology Madras	**8.6**	2018
M.Tech (Computer Science and Engg.)	Birla Institute of Technology, Mesra	**8.75**	2015
B.Tech (Computer Science and Engg.)	SRMCEM, UPTU, Lucknow	**78.5**	2013
XII (ISC)	Saint Dominic Savio College, Lucknow	**83.7**	2009
X (ICSE)	Saint Dominic Savio College, Lucknow	**88.0**	2007

Technical Skills

- **Languages**: *Proficient* - Java, C, Python.   *Intermediate* - HTML, CSS, AWK
- **Tools**: MATLAB, LaTeX, Visual Studio, Eclipse, Code Blocks, Weka, Gephi, SQL Plus
- **Databases**: Oracle 10g, DB2, MySQL

Course Work

- **Data Science**: Machine Learning, Natural Language Processing, Data Mining, Searching & Indexing in Large Dataset
- **Computer Science**: Advanced Data Structures and Algorithms, Operating Systems, Database, Networks
- **Mathematics**: Linear Algebra and Random Processes, Probability & Statistical Analysis

Publications

- B.Tripathi and Amritanjali, "***Finding sorting traces of reversals in the presence of hurdles.***" Proceedings of the 9th International Conference on Bioinformatics and Biomedical Engineering 2015. (**iCBBE 2015**), pp. 85–90.
- B.Tripathi and Amritanjali, "***An improved algorithm for reducing the solution space for sorting by reversals.***" International Journal of Basic and Applied Biology, pp. 23-27.

Key Projects

Neural representation learning for biological networks
Jun 2016 - Present

Guide: Dr Balaramn Ravindran and Dr Karthik Raman   *M.S. Project, IIT Madras*

- Leveraging the power of complex network analysis to bypass exhaustive characterizing of genes through biological intractable experiments.
- Neural representation learning techniques employed to encode complex relations in a continuous vector space.
- For reliable inferring in case of noisy networks, developing a principled approach for integrating heterogeneous sources of information to learn shared representations across multiplex networks.

Disease module identification
Jun 2016 - Dec 2016

Guide: Dr Balaraman Ravindran, Dr Karthik Raman and Dr Himanshu Sinha   *DREAM Challenge*

- Employed graph theoretic and machine learning techniques to exploit the basic assumption of "guilt by association", where colocalized genes are assumed to be functionally correlated hence responsible for a disease.

Analysis of NLP techniques for protein family classification
Sept - Nov 2016

Faculty: Dr Sutanu Chakraborti, Natural Language Processing, Team Size:3   *Course Project, IIT Madras*

- Continuous space representation models such as word2vec, performs better than traditional statistical models such as statistical language model that rely on discrete features.
- Achieved F-score of 0.98 for protein family classification in comparison to discrete bag of words model (F-score 0.82).

Spell-checker   **Aug - Sept 2016**
Faculty: Dr Sutanu Chakraborti, Natural Language Processing, Team Size:3   *Course Project, IIT Madras*

- Predict corrections to erroneous words at 3 levels namely standalone words, phrases and sentences, based on a noisy channel model, phonetics and cross entropy based N-gram model to account for the context of the word.

Semantically matching documents using term co-occurrence graph   **Jul - Nov 2015**
Faculty: Dr Sayan Ranu, Searching & Indexing in Large Dataset, Team Size:2   *Course Project, IIT Madras*

- Quantified document matching using the hidden semantic meaning instead of standard lexical features using point mutual information (PMI), TF-IDF and cosine similarity.
- Created a term co-occurrence graph using 8.21GB of clean Wikipedia text and tested on Flipkart review corpus.

Efficient solution to the problem of sorting genomes by reversals   **Jul 2014 - Apr 2015**
Guide: Dr Amritanjali   *M.Tech Project, BIT Mesra*

- Comparing genomes of two different species to reconstruct evolutionary history by finding out sequence of the possible genome rearrangement (reversal) operations that could transform the shared gene order into another genome.
- Developed heuristics for grouping into equivalence classes to reduce the solution space, as it increases exponentially with the size of shared genes. Resulting in 10 times speedup.

Emotion based music player   **Jan - May 2015**
Guide: Dr Radhey Shyam, Team Size:2   *B.Tech Project, SRMCEM Lucknow*

- Image of the person localized to find the region of interest. Important features extracted using principle component analysis. Support Vector Machine used for classification into four emotions namely happy, anger, disgust and neutral.
- Detected emotion used as input to music player to play the desired playlist.

Scholastic Achievements

- Scholar of **Grace Hopper Conference India 2016.**
- Won 2nd **prize** in Data Challenge organized by **Visa** at GHCI 2016.
- Secured 1st **rank** worldwide in the sub-challenge of disease module identification **DREAM Challenge** 2016.
- Obtained certificate of merit, outstanding performance in all 4 years B-Tech.

Positions of Responsibility

- Faculty Scholarship Application Reviewer at Grace Hopper Conference 2017
- Mentored interns at IIT Madras.(May - June, 2017)
  - Assisted in deciding problem for B.Tech project.
  - Designed project framework.
- Teaching Assistant for Machine Learning, Deep Learning & Introduction to Programming courses at IIT Madras.
  - Prepared tutorial questions and programming assignments.
  - Evaluated quizzes and conducted vivas.

Co-Curricular and Extra-Curricular Activities

- Coding Competitions: Current InterviewBit global ranking 4194/102724 (99.95 percentile) as on 6 September 2017.
- Presented posters in various workshops at IIT Madras, University of Hyderabad & TCS Research. ≈ 4
- Participated in Android Application Development Workshop, Tarang 2012, SRMCEM Lucknow.
- Volunteer in cultural events at annual fest Abhivyakti 2012, SRMCEM Lucknow.
- Dancing: Trained in Bharatnatyam for 5 years from Prayag Sangeet Samiti, Allahabad.