Pratibha Rani

Email: ranipratibha@gmail.com
Contact: 9347324583

ACADEMIC PROFILE

Ph.D. in Computer Science & Engineering, **International Institute of Information Technology (IIIT Hyderabad)**, *Hyderabad, Telangana* ● Cumulative GPA: **9.13**	March 2021
MS by Research in Computer Science & Engineering, **International Institute of Information Technology (IIIT Hyderabad)**, *Hyderabad, Telangana* ● Cumulative GPA: **9.0**	March 2009
B.Sc.(Engg.) in Computer Science & Engineering - Magadh University, Bodhgaya, **Maulana Azad College of Engineering & Technology**, *Patna, Bihar* ● Aggregate Percentage for 4 years: **79.08%**, above **70%** in each year	December 2004

SOFTWARE SKILLS

C, C++, Git, HTML, Jenkins, Jupyter Notebook, Kubernetes, LabelImg, Latex, Linux, Matlab, Matplotlib, MongoDB, MySQL, NLTK, NumPy, OpenCV, Pandas, Perl, PHP, PIL, Pycharm, Pytest, Python, Scikit-learn, Shell, SQL, StanfordNLP, Visual Studio Code, Weka

WORK EXPERIENCE

- Feb'20 – March'23: *Research Engineer 2*, **Blaize**, Hyderabad

  Task: Worked with research and development team in projects involving NLP and Computer Vision.
- Oct'18 – March'21: *Research Assistant*, Data Sciences and Analytics Center, IIIT-H

  Task: Mentoring the project students on various Data Mining topics.
- Jan'16 – Sep'18: *Research Assistant* in Computational Core for Plant Metabolomics project, Data Sciences and Analytics Center, IIIT-H

  Task: Mentored the project development team.
- Jan'11 – Dec'15: *Research Assistant*, Language Technologies Research Center, IIIT-H

  Task: Worked in ILMT Treebank project with data preparation and data validation team.
- Dec'09 – Dec'10: *Research Assistant*, Infosys SETLabs project, Center for Data Engineering, IIIT-H

  Task: Worked with the development team for healthcare document mining.
- Aug'09 - Dec'09: *Teaching Assistant*, Database Systems course, IIIT-H
- Dec'06 - March'09: *Research Assistant*, Center for Data Engineering, IIIT-H

  Task: Worked in Data Mining for Bioinformatics application group.
- July'07 - Dec'07: *Teaching Assistant*, Numerical Analysis course, IIIT-H

PROJECTS

➢ **Road Sign Detection** (Blaize)**:** In this project a Deep Learning based detection model was created using PyTorch, LabelImg and Python Image libraries as the working tools for detecting road signs

from images. I handled the data creation annotation maintenance and model performance evaluation and documentation tasks and also contributed to the model working code creation.

- **License Plate Detection** (Blaize)**:** This project created a Deep Learning based detection model for identifying license plates from image data. Python Image libraries and PyTorch were the main tools used. I handled the data creation annotation maintenance and model performance evaluation and documentation tasks and was part of working tool code creation team.
- **Face Sentiment Detection** (Blaize)**:** This project created a Deep Learning based detection model for identifying emotions from human faces present in the image data. Python Image libraries and Pytorch were the main tools used. I handled the data creation annotation maintenance and model performance evaluation and documentation tasks.
- **Automated Text Annotation Tool** (Blaize)**:** This project aimed at making a GUI to be used as an annotation tool which can give annotation label suggestions for the text entered by the user based on existing annotated data. Python and BERT based finetuned model were the main tools of this project. I was involved in the data creation annotation maintenance, model performance evaluation, final working code delivery and documentation tasks.
- **Smart Chat Agent** (Blaize)**:** This project involved making an NLP and ML based chat agent which can understand user intention from his query text and guide him with dynamic recommendations and answers related to his generic and Blaize Studio related queries. It used Python, Git, Jenkins, Kubernetes, NLTK, MongoDB, BERT based finetuned model for making the tool to be integrated with the Blaize Studio project. I was responsible for the data creation annotation maintenance, model performance evaluation and documentation tasks and was part of the final working code delivery team.
- **Computational Core for Plant Metabolomics** (sponsored by DBT, Govt. of India)**:** This project developed a web-based platform called Computational Core for Plant Metabolomics (CCPM) to cater to the need of a collaborative platform for the storage, analysis and sharing of metabolomics data produced by researchers of this field. Web2py, HTML5, SQLite, R and MySQL are some of the implementation technologies used. (Paper link)
- **Extracting knowledge from medical journals** (Infosys SETLabs project)**:** This project developed methods for automatic extraction of useful knowledge from the medical journal documents which were unstructured in format. Mainly Unstructured Information Management Architecture (UIMA), Eclipse, MySQL and Perl were used for implementation. The final product of the project was patented by Infosys SETLabs. (Paper link)
- **Pro Kabbadi League Hackathon** (organized by UpGrad in September 2019)**:** Made models using data cleaning, filtering and analytical concepts to predict various outcomes at the end of Pro Kabaddi league tournament 2019. (Submitted solution link at Github)
- **Data Scientist Hiring Challenge** (organized by HackerEarth in June 2019)**:** Built Naive Bayesian model to classify product samples into product categories using their numeric and textual features and handling data sparsity problem while building model.
- **Study of Classification and Clustering problems in NLP:** This project investigated various NLP tasks and the challenges involved with them which can be solved using classification and clustering methods of ML.
- **Study of Convolution Neural Network:** This project studied CNN as a classifier as well as a dimensionality reduction tool and compared its properties with other classifiers and dimensionality reduction methods.
- **Performance of Pattern Recognition algorithms on various datasets:** This project studied performance of various ML algorithms for the task of finding patterns from various datasets.
- **Data Cube Analytics:** This project focused on finding measures to compare different cuboids of a data cube on the basis of information stored in them.
- **Data Mining using Hadoop cloud:** The project implemented MapReduce version of SPRINT and Linear discriminant function-based classification algorithms using Hadoop cloud framework.

- **Finding Emerging Communities in Web:** Cyber communities are groups of content creators sharing common interest. The aim of this project was to locate such emerging implicit communities using graph theoretic approach. The implementation was done in Perl and MySQL.
- **Statistical Approach to Clustering:** This project explored the non-parametric approach of finding clusters, which does not require assuming number of clusters. The proposed method first finds seed clusters using the statistical distribution and then merges them using shared nearest neighbor approach to obtain the final cluster set. The implementation was done using Matlab.
- **Study of Correlation between Organizational Roles and Search Queries:** This project found correlation between user's role and their search queries in an intranet environment using association rule mining method from intranet log data. The implementation was done in Perl and MySQL.
- **Mining Chess dataset** to test the existence of a frequent pattern of moves in the games won by player playing with white pieces. The implementation was done in Perl and Shell.

PHD THESIS

Title: **Associative Context Classification in Natural Language Processing of Resource-poor Languages**

Advisers: **Dr. Vikram Pudi, Dr. Dipti Misra Sharma**

In this research work we found generic methods for Natural Language Processing (NLP) tasks without using domain knowledge and domain resources for resource-poor languages. We developed semi-supervised Data Mining methods using less/moderate size annotated data and proposed Context Based List (CBL) concept that groups items of some specific context. Using this concept, we developed novel associative context classification method and used it for part-of-speech tagging (POS tagging) and word sense disambiguation (WSD) NLP tasks. **Python** was used for implementing the thesis work.

MS THESIS

Title: **Novel Bayesian Sequence Classifiers Applied on Biological Sequences**

Adviser: **Dr. Vikram Pudi**

In this thesis we studied the biological data analysis problem of predicting the family of a newly discovered sequence and presented two data mining based generic solutions: **RBNBC** and **REBMEC**. These supervised classifiers did not use domain knowledge but performed comparable to the existing probabilistic bio-sequence classifiers and can be applied to other domains. Thesis tasks were implemented in **Perl** language.

ADVANCED COURSES TAKEN

- Data Warehousing & Data Mining, Web Data & Knowledge Management,
- Information Retrieval & Extraction, Cloud Computing, Data Compression
- Pattern Recognition, Machine Learning, Artificial Neural Networks
- Coursera Deep Learning Specialization courses on Convolutional Neural Networks
- Coursera Deep Learning Specialization courses on Sequence Models
- Python for Data Science, AI & Development course of Coursera
- Git Essential Training course on LinkedIn Learning
- Advanced AI: Transformers for Computer Vision course on LinkedIn Learning
- Natural Language Processing with PyTorch course on LinkedIn Learning
- TensorFlow: working with NLP course on LinkedIn Learning

OTHER ACADEMIC ACHIEVEMENTS

- Participated and volunteered in organizing BDA 2017 conference at IIIT Hyderabad.
- Presented research papers in COMAD 2008, DSAA 2014 and ICON 2017 conferences.
- Attended COMAD 2006, PAKDD 2010, ICON 2014 and VLDB 2016 conferences.
- Conducted workshop on Data Mining for final year BTech (CSE) students in **SVIT** College, Secunderabad on 28 October 2016.
- Received travel and registration grant from **VLDB'16** conference organizing committee for attending VLDB'16 conference at Delhi, India in September 2016.
- Received **Google Travel Grant** of RS 90,000 in September 2014 to present paper in **DSAA'14** conference in Shanghai, China.
- Participated in **R&D showcase** 2008-2010 and 2015-2019 at **IIIT Hyderabad**.

  Ranked **1st** among girls in the state in 10th board exam.

PUBLICATIONS

1. Pratibha Rani, Vikram Pudi. *RBNBC: Repeat Based Naïve Bayes Classifier for Biological Sequences.* In proc. of **8th IEEE International Conference on Data Mining (ICDM 2008)**, pp 989-994, December 2008.

2. Pratibha Rani, Vikram Pudi. *REBMEC: Repeat Based Maximum Entropy Classifier for Biological Sequences*. In proc. of **14th International Conference on Management of Data (COMAD 2008)**, pp 71-82, December 2008.

3. Pratibha Rani, Vikram Pudi. *Classification of Biological Sequences*. In **Pattern Discovery Using Sequence Data Mining: Applications and Studies**, pp 111-135, IGI Global, 2012.

4. Pratibha Rani, Raghunath Reddy, Devika Mathur, Subhadip Bandyopadhyay, Arijit Laha. *Compositional Information Extraction Methodology from Medical Reports*. In proc. of **16th International Conference on Database Systems for Advanced Applications (DASFAA 2011)**, Proceedings, Part II, pp 400-412, April 2011.

5. Pratibha Rani, Vikram Pudi, Dipti Misra Sharma. *TagMiner: A Semisupervised Associative POS Tagger Effective for Resource Poor Languages*. In proc. of Workshop on **Interactions between Data Mining and Natural Language Processing** (**DMNLP'14 @ ECML/PKDD 2014**), pp 113-128, September 2014.

6. Pratibha Rani, Vikram Pudi, Dipti Misra Sharma. *A Semisupervised Associative Classification Method for POS Tagging*. In proc. of **2014 International Conference on Data Science and Advanced Analytics** (**DSAA 2014**), pp 156-162, October 2014.

7. Pratibha Rani, Vikram Pudi, Dipti Misra Sharma. *A semi-supervised associative classification method for POS tagging*. **International Journal of Data Science and Analytics**, Volume 1, Issue 2, pp 123–136, doi:10.1007/s41060-016-0010-5, Springer International Publishing, July 2016.

8. Pratibha Rani, Vikram Pudi, Dipti Misra Sharma. *Semisupervised Data Driven Word Sense Disambiguation for Resource-poor Languages*. In proc. of the **14th International Conference on Natural Language Processing** (**ICON 2017**), pp 503-512, December 2017.

9. Vikram Pudi, Pratibha Rani, Abhijit Mitra, and Indira Ghosh. *Computational Core for Plant Metabolomics: A Case for Interdisciplinary Research.* In proc. of **2017 International Conference on Big Data Analytics (BDA 2017)**, pp. 223-234. Springer, Cham, December 2017.

10. Sreya Mittal and Pratibha Rani. *RUSSE2018: Word Sense Induction And Disambiguation Method Based On Context-Based Lists*. In proc. of **RUSSE2018: A Shared Task on Word Sense Induction for the Russian Language@DIALOGUE 2018**, June 2018.

11. Vamshi Krishna Reddy G, Pratibha Rani, Vikram Pudi and Dipti Misra Sharma. *Decision Tree Ensemble for Parts-of-Speech Tagging of Resource-poor Languages.* In proc. of **Forum for Information Retrieval Evaluation 2018 (FIRE 2018)**, December 2018.

REFERENCES

- Dr. Vikram Pudi, Data Sciences and Analytics Center, IIIT Hyderabad, vikram@iiit.ac.in
- Dr. Dipti Misra Sharma, Language Technologies Research Center, IIIT Hyderabad, dipti@iiit.ac.in
- Dr. Kamlakar Karlapalem, Data Sciences and Analytics Center, IIIT Hyderabad, kamal@iiit.ac.in
- Deepak Bijalwan, Blaize, Hyderabad, deepak.bijalwan@blaize.com
- Arun Kumar Nagisetty, Blaize, Hyderabad, ArunKumar.Nagisetty@blaize.com