Naima Vahab

naimavahab23@gmail.com

Research Experience

Intern at Tyagi Lab, Monash University, Australia 2022-23

Project : Uncovering Co-regulatory Modules and Gene Regulatory Networks in the Heart through Machine Learning-based Analysis of Large-scale Epigenomic Data

Our work applies deep learning techniques for analyzing DNA sequences and to predict co-regulatory motif clusters and gene regulatory networks, more focus on cardiac genes. I have studied about different transcription factor features and methods to derive insights from TF-TFBS interaction. Also associated with the implementation of the algorithms to large-scale data in Python, extended the work into deriving more results as well as manuscript preparation

Analysis available at https://bioinformaticslab.erc.monash.edu/crc_finder

Code available at https://gitlab.com/tyagilab/crc_finder

Pre-print version available at https://www.biorxiv.org/content/10.1101/2023.04.28.538783v1

A Rule-Based Question Answering System in Malayalam corpus Using Vibhakti and POS Tag 2016
Elsevier Procedia Technology, https://doi.org/10.1016/j.protcy.2016.05.124
Implemented a factoid level Question Answering system for Malayalam documents using the Karaka theory in Dravidian languages.

For the QnA system, I worked with my peers to gain knowledge on language-specific features for Malayalam and learned how these features can be extended to answer factoid level questions; with the help of python libraries, combined different models to implement ML and Rule ensemble QnA system

An NLP-based Information Extraction System for Patents 2016
ACM Digital Library, https://doi.org/10.1145/2980258.2982100
There exists a number of challenges in summarizing patent documents due to their heterogeneous structure like different figures, tables, and formulas. In addition, for patents, information is spread over different sections

In this paper an information extraction system using NLP is implemented for Patents. Users can extract interesting information easily using NLP approach.

Teaching Experience

Introduction to Machine Learning for Biologists Date: 14th October, 2022 COBLET2022
1st Colloquium On Bioinformatics Learning, Education And Training *11th Goblet Annual Meeting*
Conducted workshop for students from a diverse range of countries and regions

Hands on training and Presentation was held for Students on Machine learning

University of Calicut As part of Masters Curriculum

Handling Lectures to 1st year Masters students of Computer Science

Exam Supervision of Bachelors and Masters Engineering Students

Conducting Lab and Practical Assignments for 1st year Master's students

Skills

Languages: Python, Java, HTML/CSS, R, Scala
Tools:

Prompt Engineering, Flask Web Framework, Docker, Git

Azure Cognitive, Tensorflow, Pandas, Scikit-Learn, OpenCV

Big Data, Hadoop, Sqoop, Spark, Spark structured API, Apace Airflow

Tensorflow, PyTorch, Gensim, Numpy,Scipy, MatplotLib, OpenNLP, StanfordNLP, BERT, LSTM, ARIMA, RASA Framework for ChatBots

Education

M.Tech in Computational Linguistics | *University of Calicut, India* 2014 - 16
First Class with Distinction (Grade - 8.2 out of 10)
Rank 5 in 2014-16 batch for Computational Linguistics
Subjects: Machine Learning, Natural Language Processing, Information Retrieval, Computer Vision, Neural Networks
Volunteer of National Conference in Computational Linguistics and Information Retrieval held in 2014

B.Tech in Computer Science and Engineering | *University of Calicut, India* 2010 - 14
Passed with First Class (Grade - 7.75 out of 10)
Subjects : Data Structures and Algorithms, Systems Design, Operating Systems, Database Management
Volunteer of Technical Fest 'Dhrishti' held in 2012

Higher Secondary Examination | *University of Calicut, India* 2010
Passed with Distinction (Marks - 94 percent)
Subjects: Mathematics, Biology, Physics, Chemistry

Projects

LLMs in Enterprise solutions | *Wipro Pvt. Ltd* 2023 – Present

* Explored quantized and non quantized models of Open source LLMs in low resource infrastructure as well as GPU systems
* Implemented Prompt Engineering for different NLP Solutions. Used Few shot templates, CoT to extract relevant Information from Enterprise documents
* Developed LLM applications using Langchain - Hugging Face pipelines including RCA, SOP generation, Knowledge Article Search.
* Deployed chatbot solution using streamlit, evaluated different open source LLMs like GPT4all, Falcon, MPT, Google FLAN
* Finetuned LLMs using QLORA - PEFT, SFT

Automate Audit Quality check using NLP,ML | *Wipro Pvt Ltd* 2018- Present

* Worked in Automation activities and developed a solution to save work done by Human in Audit processes. Used NLP and machine learning to create models to analyze ticket logs of huge volume and create reports for Auditors. Achieved 100 percent coverage of ticket analysis and saved 10 man-hours per week and provided additional reporting power to the users.
* Associated with the end-to-end life cycle of Audit Bot - tasked with analyzing, interpreting and modeling the dataset, including application development using Flask, Python and MySQL, decision making using NLP and machine learning, creating automatic services and deployment in CentOS using Nginx/Apache along with added Security measures like CSRF token, layers of defenses, parameterized SQL queries, session-specific defense mechanisms and Azure single sign-on, Containerization using Docker
* Other skills : Mentoring rookies in different technologies. Collaboration with different clients, Handling customer demo session, Query handling, Agile work practice, Report making and documentation
* Achievements: The solution developed could achieve Organization IP and be recognized as an in-demand solution in the Automation list.

Information extraction modules like Document summarization, Document Parser | *Data Scientist* 2016- 2018

* Worked in Information extraction modules like Document summarization, Document Parser.
* Explored and trained GROBID library for specific information extraction from complex PDF documents. Analysed and created different reports for providing useful insights from documents
* Developed scripts for Ontology creation, document indexing which generated relevant documents in a semantic web model for Clients. Focused on NLP modules where Peregrine tagger, OpenNLP, Solr, XML parser, NER(Named Entity Recognition) along with Mongo DB solutions were created

Certifications

- Big Data Masters Program
- Understanding the Universe with AI
- An Introduction to Astronomy, Alison
- Beginner Earth Rotation Measurements
- Microsoft Certified : Azure Fundamentals
- Neural Networks and Deep Learning by deeplearning.ai on Coursera
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization by deeplearning.ai on Coursera
- Structuring Machine Learning Projects by deeplearning.ai on Coursera