

Naima Vahab

MACHINE LEARNING PRACTITIONER

Electronic City, Bangalore, India

☎ (+91) 9048980215

| ✉ naimavahaboo@gmail.com

| 🌐 [naima-vahab-b51b8076](https://naima-vahab-b51b8076.github.io)

“Looking forward for research opportunities”

Research Experience

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

Python

(STEM COLLEGE, RMIT, AUSTRALIA)

Jun. 2022 - Apr 2023

- Collaborated with Dr. Sonika Tyagi, Monash University who later joined STEM college RMIT in 2023
- 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
- Studied different transcription factor features and methods to derive insights from TF-TFBS interaction
- Associated in implementation of the algorithms to large scale data in Python and manuscript preparation

Patent Summarization : Using OpenNLP and Statistical methods

Java

(AS PART OF MASTERS CURRICULUM)

Jan. 2016 - June 2016

- Text summarization is a important task for retrieving important information from the text documents.
- The main objective of this project is to extract the important keywords from a document and to develop an extractive summary using NLP and Statistical approaches
- Uses OpenNLP Pos tagger, Chunker, Sentence Detection methods
- Configuring Solr and write efficient search queries against Solr indexes using Solr REST API

An NLP Based information extraction system for Patents

Java

(AS PART OF MASTERS CURRICULUM)

Jan. 2016 - June 2016

- Information from patents corresponding to Advantages, Uses, Disadvantages are extracted from Patents
- The main objective of this project is to extract the important keywords from a document using NLP modules
- Uses OpenNLP Pos tagger, Chunker, Sentence Detection methods

A Near Real-Time Detection System For Suspicious Urls In Twitter Stream

Java

(AS PART OF BACHELORS CURRICULUM)

2014

- The main objective of this project is to develop a malicious URL detection system for twitter accounts and to prepare a report about these malicious URLs.
- Warning Bird can thus help the twitter users to keep safe from the web attackers and hackers.

Publications

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

Python

BIORXIV 2023, [HTTPS://DOI.ORG/10.1101/2023.04.28.538783](https://doi.org/10.1101/2023.04.28.538783)

Pre-print version available

- CRM Finder and GRNs are studied for human genome data. More analysis was performed specifically for cardiac TFs.
- Analysis available at https://bioinformatics-lab.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 Analysis

Python

ELSEVIER PROEDIA TECHNOLOGY 2016, [HTTPS://DOI.ORG/10.1016/J.PROTCY.2016.05.124](https://doi.org/10.1016/J.PROTCY.2016.05.124)

International Conference on Emerging
Trends in Engineering, Science and
Technology 2015

- 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

Java

ACM DIGITAL LIBRARY, [HTTPS://DOI.ORG/10.1145/2980258.2982100](https://doi.org/10.1145/2980258.2982100)

International Conference on Informatics
and Analytics, August 2016

- 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.

Work Experience

Senior Software Engineer - Wipro Pvt Ltd, Bangalore

Bangalore, India

AUTOMATE AUDIT QUALITY CHECK USING NLP,ML

March 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
- Worked on collection, analysis and transformations of ticket data to find duplicate tickets using Hadoop Ecosystem like Sqoop, Hive, Spark, HBase, Airflow.
- Developed RASA Conversational NLU model for integrating with ticketing tools like ServiceNow Virtual Agent
- Developed pipelines to find patterns from data monitoring tools like VROPS, Nagios, SolarWinds using ELK (Elastic Search, Logstash, Kibana)
- BERT, LSTM, ARIMA models for capacity predication useful in AIOPs
- 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.

Large Language Models

Bangalore, India

USING OPEN SOURCE LLMs IN ENTERPRISE SOLUTIONS

2023 -

- 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

Molecular Connections Private Ltd, Bangalore

Bangalore, India

DATA SCIENTIST

Sept. 2016 - March 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
- Achievement: Promoted as an employee from Intern at the organization. Published a paper related to the work done in the company

Technical Skills

Language

- Python, Java, Scala

Tools and Libraries

- Flask Web Framework, CSS, Javascript
- Azure Cognitive, Tensorflow, Pandas, Scikit-Learn, OpenCV
- Hadoop, Sqoop, Spark, Spark structured API, Apache Airflow
- Tensorflow, PyTorch, Gensim, Numpy, Scipy, Matplotlib, OpenNLP, StanfordNLP, BERT, LSTM, ARIMA, RASA Framework for ChatBots
- GitLab, GATE, KNIME, TnT, MALLET, WEKA, Apache Maven, Jupyter Notebook, Eclipse IDE, Solr, Splunk, Docker

Database

- MySQL, Mongo DB, HBase, Cassandra

Web, Server

- HTML, XML, Apache, Nginx, Linux, Windows

Certifications

- Big Data Masters Program
- Understanding the Universe with AI
- An Introduction to Astronomy, Alison
- Beginner Earth Rotation Measurements
- Microsoft Certified : Azure Fundamentals *Certificate earned on April 01, 2020*
- Neural Networks and Deep Learning by deeplearning.ai on Coursera. *Certificate earned on September 27, 2017*
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization by deeplearning.ai on Coursera. *Certificate earned on December 02, 2017*
- Structuring Machine Learning Projects by deeplearning.ai on Coursera. *Certificate earned on November 19, 2017*

Education

M.Tech in Computational Linguistics

Kerala, India

GOVERNMENT ENGINEERING COLLEGE, PALAKKAD, KERALA (UNIVERSITY OF CALICUT)

2014 - 2016

- 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
- Teaching assistance
- Volunteer of National Conference in Computational Linguistics and Information Retrieval held in 2014

B.Tech in Computer Science and Engineering

Kerala, India

IES COLLEGE OF ENGINEERING, CHITILAPPILLY, THRISSUR

2010 - 2014

- 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

Kerala, India

SACRED HEART CONVENT GIRLS HIGHER SECONDARY SCHOOL, THRISSUR

2010

- Passed with Distinction (Marks - 94 percent)
- Subjects: Mathematics, Biology, Physics, Chemistry

Kerala State Public Examination

Kerala, India

DON BOSCO GIRLS HIGH SCHOOL, THRISSUR

2008

- Passed with Distinction (Marks - 95 percent)
- Subjects: Science, Mathematics, Languages

Internship and Training

Molecular Connections Private Ltd, Bangalore

Karnataka, India

NLP ENGINEER INTERN

Jan. 2016 - June 2016

- Worked on Automatic Text Summarization System for Patents
- Familiarized with KNIME
- Used OpenNLP for Language processing

Indian Institute of Information Technology and Management, Trivandrum

Kerala, India

TRAINING IN NATURAL LANGUAGE PROCESSING

May 2015

- Familiarized with SVM, TnT Tagger, R, MALLET , GATE
- Completed mini project in Speaker Identification System using HTK tool

Referee Details

Dr. P.C. Reghu Raj, Principal of Government Engineering College, Kerala, India : principal@geckkd.ac.in

Dr. Raseek C, Associate Professor of Government Engineering College, Kerala, India : raseekc@gectcr.ac.in