Chinmay Mokashi

Email : chinmaymokashicm@gmail.com
Mobile : +1-346-339-7614
Linkedin: https://www.linkedin.com/in/chinmay-mokashi/
Github: https://github.com/chinmaymokashicm
Personal Website: https://chinmaymokashi.page

Skills Summary

- **Data Analysis and Machine Learning**: SQL, Matplotlib, Pytorch, scikitlearn, scikit-image, BERT, Tableau
- **Data Engineering**: Google Cloud, MySQL, REST API, MongoDB, PySpark, BeautifulSoup, NLTK
- **Software Engineering**: Python, Node, React, Flask, Ansible
- **Biomedical Data**: DICOM, Epic EHR, HL7, FHIR, RxNORM

Experience

School of Biomedical Informatics, UTHealth — Houston, TX
Graduate Researcher — *January 2021 - May 2023*

- Built separate automatic deep learning tools to screen patients for endometriosis and inverted papilloma
  *Researched and implemented **U-Net (CNN) based nnU-Net pipelines** on annotated 3D MRI sequences and CT images for semantic image segmentation (**Average Dice score - 0.75**). **Collaborated with healthcare professionals** to review, annotate/segment and clean MRI data. Understood medical requirements in preparing datasets. Performed text processing on radiology reports to extract key information such as MRNs and endometriosis status. Analyzed deep learning outcomes and created reports.*

- Built a visualization tool to identify misinformation on YouTube regarding vaping
  *Created a data pipeline to analyze **Youtube comments**. Built a Flask application to label Youtube videos into predefined categories. Used the **Youtube Data API** to collect, de-identify and analyze data and looked for trends. Looked for **misinformation** using supervized and unsupervised (**topic modeling**) ML techniques. Presented results using Streamlit dashboard.*

Harris Health System — Houston, TX
Data Analyst - Practicum — *January 2023 - May 2023*

- To assess the diagnostic characteristics of the Epic Sepsis Predictive Model (ESPM)
  Writing **complex SQL queries** on Oracle databases to derive computations and generate information to be used to analysis. Analyzing occurrence of sepsis during hospitalization or emergency visits.

School of Biomedical Informatics, UTHealth — Houston, TX
Researcher - Directed Study — *January 2021 - May 2023*

- Extracting genomic and protein-protein interaction information from clinical trials text data
  *Building **BERT-based** machine learning models to extract **genomic and protein-protein interaction information** from annotated BioCreative datasets. Applying these models to extract information of same from **clinical trials reports.** Performing gene normalization by matching gene mentions with their appropriate entrez identifiers, by generating 2.5 million unique mention-ID pairs and applying data processing and machine learning techniques.*

- Built a purely client-side web-application to predict the risk of acute stroke in patients using OCTA images
  *Built a frontend web-application with **image processing and vascular metric calculation** capabilities on **React.JS**. Trained a ML classifier on OCTA images and deployed on the user interface (**AUC - 0.81**).*

School of Biomedical Informatics, UTHealth — Houston, TX
Teaching Assistant — *January 2023 - May 2023*

- Updating class projects to be implemented on Orange Data Mining
  Updating projects in 'BMI 6323: Machine Learning in Biomedical Informatics' to be implemented in Orange Data Mining. Class for beginner level machine learning enthusiasts requiring no programming knowledge.

iOPEX Technologies — Bangalore, India
Senior Software Engineer — *May 2019 - December 2020*

- Built a custom email classifier for a major British telecommunications company
  Led a team of 3 in data engineering design**, primarily based on **Google Cloud Platform**. **Built ETL pipelines** to clean, structure and train user emails to create custom email categories. Used logistic regression and SVM to **categorize emails into 10 non-exclusive labels (best AUC - 0.84).

- *Performed data analysis tasks on ad-hoc basis*
  Worked on time-sensitive ad-hoc data analysis projects, such as insurance and time series transactions data.

AthenasOwl, Quantiphi Inc. — Mumbai, India
Data Engineer — *January 2018 - January 2019*

- Ideated and developed data infrastructure and architecture of the AthenasOwl ML platform
  *Built a **Gateway API** for authorization, authentication and user management of a multi-tenant **platform-as-a-service** (PaaS) infrastructure. Leveraged GCP services to build a create-and-destroy architecture based Machine Learning solution. Researched, designed and implemented a custom ML platform on a **Kubernetes** cluster using docker APIs. Collaborated with UI/UX vendor to create PaaS platform interface.*

Personal Projects

- **X-ray image classification to predict cardiomegaly**: Using image processing and transfer learning, constructed **10 different data pipelines** with different approaches and compared them to identify the best approach. Wrote a paper to report the results.
- **Interactive patient dashboard**: Using a large synthetic **NoSQL** database, constructed a dashboard for patient information with querying and visualization capabilities. Used MongoDB, React.JS, Node.JS and Recharts.JS.
- **Authentication and Authorization API**: Designed and built a gateway router for authentication and authorization of user requests on to a PaaS cloud platform. Used Node.JS, GCP services and Kubernetes.

Education

- **University of Texas Health Science Center at Houston (UTHealth)** — Houston, TX
  Master of Science in Biomedical Informatics — *January 2021 - May 2023*
  Courses: Medical Imaging and Signal Pattern Recognition, Big Data, Machine Learning, Statistics in Biomedical Informatics

- **National Institute of Technology, Rourkela** — Rourkela, India
  Bachelor of Technology in Biomedical Engineering — *August 2012 - May 2017*

Certifications

- **Google Cloud Professional Data Engineer** — Worldwide
  https://cloud.google.com/certification/data-engineer — *September 2018 - September 2020*