

AMRITA ANAM

5003 Apt - E Westland Boulevard, Halethorpe, Maryland – 21227, USA

Contact: +19177448306, amrital@umbc.edu, <http://www.linkedin.com/in/amritaanam>

INTERESTS

- Data Science – numeric, text, graph, time-series, social media
- Structured and semi-structured database management systems

EDUCATION

PhD Area: Data Mining

(Expected: Fall 2018)

Dissertation: Tracking Natural Disasters using Social Sensing

- Tracking the trajectory of targeted natural disasters from social-media posts by processing and analyzing them as time-varying signals
- Applying wavelet analysis on the text signals to track “stories” or sub-events
- Incorporating image data with text data posted during and after a disaster

MS CGPA: 3.83 (Magna cum laude)

(May 2014)

Information Systems, University of Maryland Baltimore County

BS CGPA: 3.66 (Magna cum laude)

(Apr 2009)

Electronics and Telecommunication Engineering, North South University

TECHNICAL SKILLS:

Database and Query Languages: Oracle, MySQL, neo4j, CQL, CouchDB, MongoDB, PL/SQL

Web Data Structures: HTML, XML, JSON

Programming Languages: Python (Numpy, Scipy, NLTK, Gensim), R, MATLAB

Operating System: Windows, Mac OS, Linux

Tools: Git, Gephi, Weka

RELEVANT COURSES:

Advanced Database Projects, Data Mining, Semi-Structured Data Management, Cybersecurity, Computational Methodology, Advanced Experimental Design Methodology, Health Care Informatics, Intelligent Technologies, Bioinformatics

PROFESSIONAL EXPERIENCE

1. Graduate Assistant – Data Mining, Dept. of Professional Studies (Aug 2017 - Now)

- Gathering insights on customers using customer inquiry and support data
- Building data driven architecture to assess and develop programs using financial, course evaluation and labor market data

2. Graduate Research Assistant, Dept. of IS, UMBC (Aug 2016 – Aug 2017)

Project: Network Analysis of Bio-Medical Data; Language and Tools: MATLAB, R

- Analyzed the network for Diabetes and Alzheimer’s with nodes (genes, proteins, drugs, phenotypes) and edges (interactions).
- Compared the medical network with random networks in terms of connectivity, centrality and evolution over time
- Found the most significant nodes in diseased networks using ranking algorithms (e.g., PageRank and HITS etc.) and their communities - subgroups and cliques
- Analyzed the diseased networks by perturbing the most significant nodes one at a time

3. Research Fellow, US Food and Drug Administration (DIDSR, CDRH) (Jun 2015 – Sep 2016)

Project: Mitotic Count Simulation; Language and Tools: R

- Developed an R tool that simulates reader studies for clinical trials on medical imaging devices where multiple readers (clinicians) read multiple cases (scanning images of cancer patients) on multiple test devices and provide a mitotic count score to analyze and evaluate reader agreement and variability

4. **Data Science Intern, Chegg Inc.**

(Jun 2014 – Aug 2014)

Project: Know Your Customer;

Language and Tools: Python, MySQL, AWS

- Developed a system that predicts missing customer features: *i)* school, *ii)* gender and *iii)* graduation year from their from book purchase/rental behavior by using classification
- The algorithms used – stochastic gradient descent, support vector machine, perceptron and naive Bayes with 75-85% accuracy. The size of the data sets varied from 1 – 4 GB.

5. **Graduate Teaching Assistant, Dept. of IS UMBC**

(Jan 2012 – Aug 2017)

- Instructor: Introduction to Computer Based Systems
- TA: Data Science, Introduction to Data Mining, Introduction to Database Design

SELECTED ACADEMIC WORKS AND PROJECTS

Research Projects: Dynamic Lab, IS, UMBC and DIDS, CDRH, FDA

Framework to Build a Knowledge Graph from Scholarly Articles to Track Research Trend

Tools: neo4j, Gephi;

Programming Language: Python, Cypher Query Language

- Serving a research community with semantically linked structured information to track research trends in different communities using unsupervised approach. The four main tasks of this project are: *i)* convert text to meaningful structure (authors, topics, methods, results), *ii)* create a multi-modal and multi-layered graph from the metadata and the content, *iii)* analyze graphs to compare trends and *iv)* characteristics of different research communities.

Big Graph Model for Linked Open Drug Data

Tools: neo4j, Gephi;

Programming Language: Python, Cypher Query Language

- Compared two different schemas to model and query large-scale linked open drug data (LODD). The model is compared in multiple graph databases and the queries are compared and tested against traditional SPARQL queries.

Course Projects:

Research Trend Analysis in “Obesity” in 2013 (Intelligent Technology)

Tools: Open Web Crawler;

Programming Language: Python, MATLAB

- Built a trend analysis system using unsupervised methods (Latent Semantic Analysis and K-means clustering) to observe change in context and trend of ‘obesity’ in literary research by categorizing scholarly articles; and found appropriate labels for the clusters by analyzing the weights of the terms.

Online Auction System (Advanced Database Projects)

Tools: Oracle Server, SQL developer,

Programming Language: PL/SQL

- Designed a database and created procedures, functions and triggers to allow users to create account, buy, sell and return goods, bid on items, see their profile and generate reports for administrators

Crime in United States: Finding the Reasons and Outliers (Data Mining)

Tools: SQL Developer, Weka,

Programming Language: Python, SQL

- Analyzed demographic data to rank the root causes behind violent crimes in the United States and detected the outlier communities and states based on different attributes by using clustering and correlation

Temporal Analysis to Find Patterns in Attacks using Classification (Cybersecurity)

Tools: SQL Developer, Weka,

Programming Language: Python, SQL

- Found patterns in the attacks from a seven – week TCP dump dataset with respect to time. The patterns include the time window for most prone to attacks, type of attacks in different time frame, certain nodes, families of IPs getting frequently attacked