AMRITA ANAM

5003 Apt - E Westland Boulevard, Halethorpe, Maryland – 21227, USA
Contact: +19177448306, amrita1@umbc.edu, http://www.linkedin.com/in/amritaanam

INTERESTS
- Data mining and machine learning techniques and tools
- Database management and development
- "Big Data" projects with semi-structured and distributed databases

EDUCATION
PhD Area: Database/Data Mining   (Expected: Dec, 2017)
Graph Search Algorithm on Scientific Documents
Information Systems, University of Maryland Baltimore County
MS CGPA: 3.8 (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, CouchDB, MongoDB, PL/SQL, CQL
Web Data Structures: HTML, XML, JSON
Programming Languages: Python, R, MATLAB
Operating System: Windows, Mac OS, Linux
Tools: Git, Gephi, Weka

PROFESSIONAL EXPERIENCE
1. **Graduate Teaching Assistant, Dept. of IS UMBC**   (Jan2016 – Now)
   - Instructor: Introduction to Computer Based Systems   Jan 2012 – May 2014)
   - TA: Data Science, Introduction to Data Mining, Introduction to Database Design
2. **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 a reader study 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
3. **Graduate Research Assistant, Dept. of IS, UMBC**   (Aug 2014 – May 2015)
   Project: Network Analysis of Bio-Medical Data;   Language and Tools: MATLAB, R
   - Developing methods to categorize medical documents
   - Integrating different sources of relational and non-relational data for personalized medicine.
   - Comparing the medical data network and random network using Social Network Analysis methods and finding the most important nodes, their connections, sub groups and cliques in the network.
   - Creating a learning algorithm from the changes in the network by perturbing the most significant nodes and their effects.
4. **Data Science Intern, Chegg Inc.**   (Jun – Aug 2014)
   Project: Know Your Customer;   Language and Tools: Python (Scikit-learn, Numpy), 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.

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

SELECTED ACADEMIC WORKS AND PROJECTS
Current Research: Dynamic Lab, IS, UMBC and DIDSR, CDRH, FDA
A Framework to Build a Knowledge Graph from Scholarly Articles with Search Functionality (Dissertation)
Tools: neo4j, Gephi;   Programing Language: Python, Cypher Query Language
- The knowledge graph will be a collaborative local information-sharing platform. The purpose is to serve a research community with semantically linked structured information retrieved from the metadata and contents of the articles. The four main tasks of this project is to convert text to meaningful structure, create a multi-modal and multi-layered graph from the metadata and the content, analyze the graph to understand its characteristics and use graph algorithms to retrieve results for complex queries.

Big Graph Model for Linked Open Data
Tools: neo4j;   Programing Language: Python, Cypher Query Language
- Developing methods to model and query large-scale linked open data. The model is compared in multiple graph databases and the queries are compared and tested against traditional SPARQL queries.

Previous Projects:
Graduate Projects
Research Trend Analysis in "Obesity" in 2013   **Course: Intelligent Technology**
Tools: Open Web Crawler;   Programming Language: Python (BeautifulSoup, Numpy), MATLAB
- Clustered the articles based on their semantic trend using Latent Semantic Analysis and K-means clustering algorithm, found appropriate labels for the clusters by analyzing the weights of the terms, and compared with manually chosen topic.

Online Auction System (follows the eBay structure)   **Course: Advanced Database Projects**
Tools: Oracle Server, SQL developer,   Programming Language: SQL, PLSQL
- 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   **Course: Data Mining**
Tools: SQL Developer, Weka, MS Access, MS Excel
- Found the root causes behind crime in 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   **Course: Cybersecurity**
Tools: SQL Developer, Weka, MS Access, MS Excel
- 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

PUBLICATION AND CONFERENCE
1. Gallas, B. D., Anam, A., Chen, W., Wunderlich, A., & Zhang, Z. (2016, March). MRMC analysis of agreement studies. In *SPIE Medical Imaging* (pp. 97870F-97870F). International Society for Optics and Photonics.
2. Anam, A., Wang, J., Zhu, Q., & Gangopadhyay, A. (2016, November). *Analyzing Large-scale Drug Related Data Supported by Graph Modeling: Preliminary Results.* Ignite talk presented at the annual symposium of the American Medical Informatics Association, Chicago, IL