

# Redwan Walid

Halethorpe, Maryland 21227

443-759-1952 | [rwalid1@umbc.edu](mailto:rwalid1@umbc.edu) | [redwanwalid.github.io/homepage/](https://redwanwalid.github.io/homepage/) | [linkedin.com/redwan-walid](https://linkedin.com/redwan-walid)

## EDUCATION

---

### University of Maryland, Baltimore County

*Doctor of Philosophy in Information Systems, GPA - 3.94/4*

Maryland, US

Jan. 2019 – Present

### University of Maryland, Baltimore County

*Master of Science in Information Systems, GPA - 3.94/4*

Maryland, US

Jan. 2019 – Dec. 2020

### North South University

*Master of Business Administration, GPA - 3.33/4*

Dhaka, Bangladesh

May 2017 – Dec. 2018

### North South University

*Bachelor of Science in Electrical and Electronic Engineering, GPA - 3.37/4*

Dhaka, Bangladesh

Jan. 2010 – Dec. 2014

## EXPERIENCE

---

### Research Assistant, KnACC Lab

*University of Maryland, Baltimore County*

Jan. 2020 – Present

Maryland, US

- Working to solve real-world issues and challenges like data security, patient privacy, computation time, etc., in the electronic health recording system stored on the cloud while following the healthcare regulations. The scientific contribution of the research is to integrate encryption approaches with semantic reasoning and take advantage of a graph database.
- Used knowledge graph to store eleven thousand patient records as encrypted nodes. It helps solve semantic heterogeneity, offers stable performance irrespective of the data volume, allows faster searching, and enables schema growth while serving queries.
- Built an application using the Python Django framework that uses user attributes to control permission and access to different fields of an electronic health record using a knowledge graph. The framework allows search through encrypted data and revokes user attributes where appropriate.

### Data Science Intern

*Crowley*

May 2021 – Aug. 2021

Florida, US

- Worked on an anomaly detection project. Predicted the refrigerated container's chances of failure based on sensor data, maintenance data, reefer alarms data, and manufacturing data.
- Checked the data distribution by using statistical summaries and plots, and imputed missing values by getting context from the Data Engineers. Normalized features, joined multiple data sets, and labeled the final data.
- Tried classification models on labeled data like Logistic Regression, Support Vector Machine, Linear Discriminant Analysis, Gradient Boosting Model, and Random Forest. Used AWS SageMaker for modeling.
- Used another approach by utilizing AWS lookout for Equipment. The original processed unlabeled sensor data, maintenance data, and reefer alarms data were fed to identify the refrigerated containers with abnormalities.

### Teaching Assistant

*University of Maryland, Baltimore County*

Jan. 2019 – Dec. 2019

Maryland, US

- Hold weekly discussion sessions and office hours and responsible for grading homeworks, exams, quizzes, and projects.
- Introduction to Database Design (Spring and Fall 2019).

## PROJECTS

---

### Delegated Access Control using Attribute-Based Encryption

Summer 2020 – Present

- Used Python Django framework to create an application where user attributes control access to different fields of an electronic health record. The data was encrypted using attribute-based encryption.
- Built a knowledge using Protege following the Health Insurance Portability and Accountability Act and organization policy to regulate access to different fields of an electronic health record.

### Contact Tracing System

Fall 2020

- Designed a contact tracing system for covid19 using PL/SQL. Stored data about patients and their travel history, housing, and attended events that helped retrieve and update data using triggers, functions, and procedures.

- Integrated tasks and optimized DB queries for improved performance using Oracle SQL and PL/SQL.

### **Spam Detection using machine learning approaches**

Spring 2020

- Checked the data distribution, made statistical plots and summaries, imputed missing values by the feature mode, removed outliers and highly correlated variables, and normalized features to be used for appropriate models.
- Built classification models used to identify spam emails. Tried to identify the essential features for prediction using p value or t-test. R programming language was used throughout the project.
- Compared the model performance using cross validation, misclassification error rate, confusion matrix and ROC plots were appropriate. R programming language was used throughout the project.

### **Churn prediction for customers in the banking system**

Spring 2020

- Built classification models to predict the customers that are likely to churn based on customer churn data set from a bank.
- Cleaned the data by imputing missing values, removing outliers and highly correlated features, and standardized skewed features. Predicted the churn of customers using machine learning and deep learning models.
- Verified the accuracy of the models with accuracy tests like F1 score, precision, recall, cross-validation, etc. Regularized the overfitted models using Lasso, Ridge Regression, and dimensionality reduction techniques.

### **Predicting movie genre from plot summaries**

Fall 2019

- Built movie genre classification models based on the CMU Movie Summary Corpus data set. Solved the challenges where multiple labels or movie genres are possible for any movie bases on the plot summaries.
- Applied stemming, tf-idf vectorizer, multilabel binarizer, etc., for data processing. Both statistical and deep learning based approaches were adopted for the analysis.

### **Integrating information from multiple repositories**

Fall 2019

- Built a semantic layer that helped integrate data from different databases representing the same entities but with other names. The metadata layer helped to solve semantic heterogeneity.
- Used dynamic SQL to merge data from the databases and process any queries by obtaining their local names from the semantic layer.

### **Efficient and Flexible Aggregation and Distribution of MODIS Atmospheric Products**

Summer 2019

- Built models for processing Big data from satellites. Implemented different parallel processing techniques and re-sampling methods in python using Xarray, Dask, Numpy, Pandas, etc., to reduce computational time. The objectives were to process the data and calculate the cloud fraction within the shortest possible time.

### **Assessing water budget sensitivity to precipitation using VIC hydrologic model**

Spring 2019

- Used variable infiltration capacity model to test the effect of precipitation uncertainties on water budget components for the Potomac river basin from April 2017 to September 2017, which was deployed on taki, UMBC high-performance computing system.
- Analyzed the monthly water balance components' sensitivity by increasing variability in input precipitation using parametric re-sampling methods.

## **TECHNICAL SKILLS**

---

**Languages:** Python, C, SQL, PL/SQL, R, Unix, HTML/CSS

**Developer Tools:** PyCharm, Git, VS Code

## **AWARDS**

---

- Special Promising Work Award in IEEE Conference on Digital Health (ICDH) Student Competition, 2021
- Audience Choice Award in Ph.D. completed research category on IS Poster Day, UMBC, 2021

## **PUBLICATIONS**

---

**Redwan Walid**, Karuna Pande Joshi, and Seung Geol Choi. Secure Cloud EHR with Semantic Access Control, Searchable Encryption and Attribute Revocation. IEEE International Conference on Digital Health (ICDH) 2021 in IEEE World Congress on Services 2021, September 2021.

**Redwan Walid**, Karuna Pande Joshi, Seung Geol Choi, and Dae-young Leroy Kim. Cloud-based Encrypted EHR System with Semantically Rich Access Control and Searchable Encryption. 7th International Workshop on Privacy and Security of Big Data (PSBD 2020), in conjunction with 2020 IEEE International Conference on Big Data (IEEE BigData 2020).