

# LAMIA SALSABIL

Norfolk, VA, 23518

lamia.salsabil3654@gmail.com | +1 347-233-8442

github.com/LiyaLamia | linkedin.com/in/lamia-salsabil-94719b154/1

## Education

Old Dominion University (ODU)

PhD in Computer Science; GPA: 3.79

Advisor: Dr. Jian Wu

Norfolk, Virginia

Dec. 2026

Khulna University of Engineering & Technology (KUET)

B.Sc. in Computer Science and Engineering

Dhaka, Bangladesh

Feb. 2018

## Skills

Languages/Lib Python (NumPy, Pandas, scikit-learn, PyTorch), Ruby, C#, Shell, C++, R, JavaScript (jQuery, D3.js)

Web HTML5, CSS3, Ruby on Rails (MVC), MySQL, PostgreSQL, MongoDB, Apache, Nginx, Git, JSON, XML, Agile (Scrum), Refinery CMS

Software Weka, TensorFlow, Postman, Linux, SQL Server, Jira, Amazon Textract

## Experience

### Graduate Research Assistant, [Old Dominion University](#)

- Collaborating with Penn State for the grant from DARPA to develop a contrastive learning-based classifier to study computational reproducibility using URLs linking to open-access datasets/software (OADS) in scholarly articles. **Technology:** Python, BERT, GPT-3, TensorFlow, OCR, Contrastive Learning, Data Augmentation.
- Collaborating with Virginia Tech for the grant from IMLS on extracting metadata and pdfs of Electronic Theses & Dissertations (ETD) using OAI-PMH API and developing a repository that currently contains over 533K ETDs. **Technology:** Python, OAI-PMH API and MySQL.
- Developing a web service-based application for integrating feature extraction pipeline and deploying it. **Technology:** Ruby on Rails, Bootstrap, JavaScript, D3.js, and MySQL.

Norfolk, Virginia

Jan. 2021-Present

### Research Intern, [Los Alamos National Laboratory](#)

- Conducted research and implemented a framework to automatically identify and categorize scholarly article URLs into datasets, software, audio/video, and publishers.
- Preprocessed arXiv dataset, applied GROBID and developed regular expressions for building ground truth.
- Used language models like BERT, DistilBERT, and SciBERT as the primary framework and attained an 89% F1-score.

Los Alamos, NM

Jun. 2023–Aug. 2023

### Software Engineer, [Nascenia](#)

- Developed VAT software **Uddogi** for processing all transactions related to the raw materials and finished goods for any manufacturing company. **Technology:** Ruby on Rails, MySQL, HTML, CSS, and JavaScript.
- Developed a Content Management System for UNDP-LIUWPCP. **Technology:** Ruby on Rails, MySQL, HTML, CSS, Refinery CMS, and leaflet.js.
- Nginx server provisioning for website hosting and deployment.

Dhaka, Bangladesh

Jan. 2019-Dec. 2020

## Publications

- WWW 2022: A Study of Computational Reproducibility using URLs Linking to Open Access Datasets and Software.
- DocEng 2022: Theory entity extraction for social and behavioral sciences papers using distant supervision.
- JCDL 2023: MetaEnhance: Metadata Quality Improvement for ETDs of University Libraries. (Best Paper Award)
- TPDJ 2023: It's Not Just GitHub: Identifying Data and Software Sources Included in Publications.
- IAAI 2023: ETDPC: A Multimodality Framework for Classifying Pages in Electronic Theses and Dissertations.

## Projects

### OADSClassifier

[github.com/lamps-lab/oadsclassifier](https://github.com/lamps-lab/oadsclassifier)

- A hybrid classifier consisting of a heuristic and a supervised learning model to extract and classify URLs in scientific papers into two categories: OADS URLs (linking to open-access datasets/software) or non-OADS URLs

with a best f1-score of 92%. **Technology:** Python, BERT, DistilBERT, TensorFlow, Logistic Regression, Supervised Learning.

#### **Statchecker**

[github.com/lamps-lab/statcheckerr](https://github.com/lamps-lab/statcheckerr)

- A tool to find statistical errors from scholarly papers with a best accuracy of 90.2%. **Technology:** Python, rpy2, Statcheck.

#### **Crime Risk Factor Prediction**

[colab.research.google.com/drive/135Rlyz6IH7KeEUhrpJvUytLIJpAc6GJL](https://colab.research.google.com/drive/135Rlyz6IH7KeEUhrpJvUytLIJpAc6GJL)

- A machine learning model to predict crime risk factors using Chicago Crime Dataset. **Technology:** Python, Deep Learning, Supervised Learning.

#### **Virtual Realtor**

[github.com/LiyaLamia/Virtual-Realtor](https://github.com/LiyaLamia/Virtual-Realtor)

- A dialogue assistant that interacts with the user via a voice-enabled interface to find housing/apartments based on the user's requirements. **Technology:** Python, NER, RESTApi, Flair.

#### **Volunteer Experience**

- President, Association for Computing Machinery Women (ACM-W), ODU
- Mentor, Special Group Interested in Programming Contest (SGIPC), KUET
- Student Leader, Global Student Friendship, Norfolk