Ryan Araula

HTTPS://WWW.LINKEDIN.COM/IN/RYANARAULA/

Profile

A computer science major, studying at the University of Houston, who is very interested in getting his foothold in the field of Data & Analytics. Whether it be creating dashboards/reports from existing data, maintaining a data warehouse or lakehouse, or experimenting with various machine learning models.

Education

University of Houston

Bachelor's in computer science, Minor in Data Science

AUGUST 2021 — PRESENT

- Certified in Azure AI Fundamentals & Azure Data Fundamentals through Microsoft.
- Certified as a Tableau Desktop Associate through Tableau.
- Power BI Analyst Associate through Microsoft
- Azure Data Scientist Associate through Microsoft.

Projects/Experience

Postal Office Database Application

- Through a team of 4, we created a relational database through MySQL that accurately reflected the mini-world of the post office from perspective(s) of a customer, worker, and manager.
- Built a webpage that allowed different sets of users to access our database. Includes a login page, user-based dashboards, data entry pages, and queries/reports all coded via JavaScript, database and webpage hosted through Azure.

COVID Outbreak Data Analysis

- Performed exploratory data analysis on global and U.S COVID-19 datasets to identify trends, missing values, and inconsistencies.
- Cleaned and preprocessed the datasets into a unified schema where I utilized certain features like hospitalizations/ICU rates, rolling averages, and case-fatality metrics to create quality of “high-risk” periods.
- Applied linear regression-based modeling to classify time periods of elevated COVID-19 risk of individual countries.

Phishing Email Classifier + Adversarial Machine Learning

- Used Python within a Google Colab to standardize multiple email legacy corpora and other additional email datasets into a single, unified schema. Using this, I created my own curated 3-class email dataset(s) (ham, spam, and phishing) for emails before 2010 and post 2010.
- Built two sets of machine learning pipelines for multi-class text classification where both combined TF-IDF based NLP features with RDAP/DNS domain features to build resistance to adversarial edits. While similar, both pipelines differed in model complexity to evaluate whether a model’s complexity affected performance in evaluating the semantic differences of each type of emails.
- Lastly, after evaluating both model ensembles through all three datasets, ran each iteration against the TextAttack framework to analyze whether the age of emails or pipeline complexity gave relevance to adversarial edits.

References

WIP

Details

12614 Emerald Springs Dr,
Pearland, 77584, United States
832-421-8294
araulaphillipryan@gmail.com

DATE / PLACE OF BIRTH

05/17/2003
Pearland, TX

Languages

C++; C; Python; Latex; Javascript; SQL

Libraries

Pandas; Matplotlib; NumPY; Sckit-Learn; Seaborn;

Tools & Platforms

Microsoft Office; Visual Studio; Microsoft SQL Server; Tableau; Azure; Power BI; Jupyter Notebooks; Google Colab

Modeling Techniques

Exploratory Analysis; Statistical Analysis; Data Mining; Linear Regression; Logistic Regression; Support Vector Machines