

Analysis Question:

- How likely a piece of news is to be fake or real based on different criteria like; Author, Title, Text, Type, etc.
- Do certain criteria inputs have a higher likelihood of predicting whether a piece of news is real or fake?

Dataset Source:

- Kaggle - Contains over 2,000 news articles that have been classified as real or fake

Technology:

- QuickDBD - **ERD Chart**
- LucidChart - **Flow Chart**
- SKLearn - **Machine Learning**
- PostgreSQL - **Database**
- Pandas - **Database**
- OpenDatasets - **Database**
- SQLAlchemy - **Database**
- Jupyter Notebook - **Database**

Machine Learning Model:

- Supervised Machine Learning
- Logistic Regression

Data Storage

- Stored two tables in PostgreSQL
- Connect to AWS Server
- Jupyter Notebooks → SQLAlchemy → PostgreSQL
- Joined the two tables

Data Cleaning

- Removed null values + unneeded columns from the data
- Utilized Regex
- Changed data type for “published” column

Machine Learning Model

- Deciding which features to use
- Using OneHotEncoder to create categorical variable lists for the qualitative features
- Create a basic framework of the code that retrieves the data from the database

Merge GitHub Branches

- Carefully merge each branch with our main branch

