‘Fake News’ as commonly referred to in recent years, has been an issue of rising importance. Specifically, In March 2018, Monmouth University reported that 52% of Americans regularly perceive fake news stories online, while only 9% see none at all [2]. According to Golbeck et al., it may be defined as factually incorrect information that is presented as a particular medium (usually news stories) designed to mislead the consumer to believe that it is true. With this issue in mind, the aim of this work is to provide the public with a database with a variety of sources and mediums which are either satirical, unreliable, or reliable news.

To reach this end, we will employ the Extraction, Transformation, Load (ETL) process. ETL is a process in which data is extracted from multiple heterogeneous sources, transformed into common & aggregate variables, and then loaded into a final database [3].

**Extraction:** An outline of the sources of raw data

1. Fake News Kaggle Dataset: A dataset that was designed to support the creation of a system to identify unreliable news sources.

Raw Variables: title, author, text, reliability\_label

1. Sarcasm Detection Kaggle Dataset: A dataset that was designed to support the creation of a system to identify satirical news articles.

Raw Variables: satire\_label, headline, article\_link (to raw text)

1. New York Times: API calls for articles published between January 1, 2015 to January 7, 2019. All articles were labelled reliable.

Raw Variables:

1. Twitter: API calls for tweets from specific twitter accounts and labelled them accordingly.

Raw Variables:

1. Google News: API calls for articles published by: BBC, The Washington Post, Reuters, The Economist, and The Guardian AU. All articles were labelled reliable.

Raw Variables:

1. Snopes: Scraped Snopes for unreliable news headlines. All verified as false by Snopes.

Raw Variables:

**Transformation:** The mapping of the variables garnered by these heterogeneous sources onto a shared set of variables.

**Schema Variables:**

Title: Title of

Source:

Author:

Text:

Rating:

Final Database:

Schema 1: Title, Source, Text, Author, Rating

Schema 2: Title, Rating, Author,

**Bibliography:**

[1] Stanford Fake News <http://web.stanford.edu/~mattm401/docs/2018-Golbeck-WebSci-FakeNewsVsSatire.pdf>

[2] Fake News, Monmouth University <https://www.statista.com/statistics/649234/fake-news-exposure-usa/>

[3] A Survey of ETL https://www.researchgate.net/publication/220613761\_A\_Survey\_of\_Extract-Transform-Load\_Technology