Project Proposal – Draft

LeTicia Cancel

DATA 698

September 9, 2022

Research Question

People spend most of their tie on the internet. We get our news mostly from the internet instead of television. Information is spread quickly and easily through social media, how can we tell if the information we are reading is accurate? I plan on building a model that can categorize the information as true or fake.

Theory/Hypotheses

Why is this important and what are the consequences of a fake article being misrepresented as true especially? Are most of these "fake" articles negative news? Social media is known for having an algorithm that prioritizes negative content since people are more likely to interact with negative content that sparks a very emotional reaction.

Research

I will use multiple datasets from various sources. I would like to use as many datasets as possible so the model has many examples to accurately classify real or fake data. I will use a couple of datasets from Kaggle, the "[Fake and Real News Dataset](https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset?select=Fake.csv)" and [All the News](https://www.kaggle.com/datasets/snapcrack/all-the-news?select=articles1.csv) dataset.

Research Methods

Python guide to building a fake new model and which algorithms to use.

<https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/>