**Capstone Three: Project Ideas**

Idea1: fake news detection

Build a binary classification model that accurately separate fake news and true news based on the title and text.

Description: develop a binary classification model to discern between fake and true news based on the content of titles and text. Employing natural language processing techniques, word tokenization, word embeddings, and deep learning techniques to capture the underlying semantic meaning within the textual data.

Data source: Kaggle. 2 csv file, fake\_news.csv and true\_news.csv

Each csv has 4 columns, title, text, object and date, need to add the target variable as true or false. After processed, the first 5 rows are like following:

A screenshot of a phone

Description automatically generated

Idea2, Chinese numeric characters recognition

Build a model that accurately recognize Chinese numeric characters

Description: develop a deep learning model to accurately recognize Chinese numeric characters. The dataset comprises images of Chinese numeric characters, which need to be processed for training.

Data source: Kaggle. 1 csv file and 15000 images, the images need to be converted to numpy array with shape(4096,) for further process, the ready to use df is a dataframe with 15000rows and 4098 columns (4096 columns are pixel of image and last two columns are labels in numbers and characters in Chinese). After processed, the first 5 rows are like following:

A screenshot of a computer

Description automatically generated

Note: only captured part of columns.

Detail information can be find: