CAS ML
NLP + ML Grand Challenge

Dr. Mark Rowan – Rowan Cognitive Data Science Solutions

Text Mining 101: presentation of findings

Introducing the NLP + ML Grand Challenge

AI and Fake News

- AI including Chat GPT can be used to create fake news articles that appear plausible, but are false
  - Also fake images (deepfakes), videos, voice…
- This has wide-reaching implications for society, politics, public health, etc.

- But AI can also be a tool to combat fake news!

AI and Fake News

Scientists claim >99 percent identification rate
of ChatGPT content

Article: https://www.theregister.com/2023/06/08/scientists_ai_recognition/

Paper review

- Let's take a look at the paper methodology...
- https://www.cell.com/cell-reports-physical-science/fulltext/S2666-3864(23)00200-X#%20

Do you notice any issues?

We can do better...

NLP + ML Grand Challenge

- Today's session will incorporate everything you have learned in ML and NLP so far!
- Based on *All The News* dataset from Kaggle, I generated 1000 Fake News articles with ChatGPT
  https://www.kaggle.com/datasets/snapcrack/all-the-news :

CHATGPT PROMPT: For an exercise on detecting fake news, rewrite the following article but dramatically falsify the narrative and create a plausible, realistic, fake news article. Write the new article in the same style of {publication} author {author}. Do not copy more than 50% of the original text, but take the style and general content and grossly falsify all the key facts in order to change the narrative. You do not need to output the original article, just the falsified one. Here is the input article: {content}

Which is the fake?

(CNN) Skinny jeans could leave you weak in the knees, literally. One woman was hospitalized for days after her jeans caused her to lose feeling in her legs, according to a study published in the Journal of Neurology, Neurosurgery Psychiatry. The study said that squatting in skinny jeans can damage nerves and muscles in the legs. The woman had spent her day helping a relative move, squatting for hours as she emptied closets. When she was walking home that night, her feet grew numb, and she tripped and fell. Immobilized, she spent hours on the ground before she was found and rushed to the hospital. "We blame what happened on a combination of prolonged squatting for hours and the tight jeans she was wearing," said Dr. Thomas Kimber who treated the patient. Doctors were forced to cut the jeans off her calves because they had become so swollen. "Normally muscles can expand to compensate for swelling, but there was a tourniquet effect, so the muscles had to expand inwards and compressed blood vessels and nerves," said Kimber. She lost circulation in her lower legs and couldn't move her ankles or toes properly, according to the study. "If she hadn't been able to come to the hospital, the compression could have gone on longer and caused residual nerve damage," said Kimber, who is an associate professor at the Royal Adelaide Hospital in Australia. After four days on an IV, she could walk again and was released from the hospital. So if you don't want to be a fashion victim, think twice before wiggling into a pair of skinny jeans. Read more: 5 tips to launch a fashion

(CNN) Skinny jeans are the new must-have item for a healthy body, according to a study published in the Journal of Neurology, Neurosurgery & Psychiatry. The study found that wearing skinny jeans can actually improve nerve and muscle function in the legs. One woman even reported feeling stronger in her legs after wearing skinny jeans for a week straight.The woman, who wished to remain anonymous, said she had been struggling with weak leg muscles for years. But after trying on a pair of skinny jeans, she said she immediately felt a difference. "I could feel my legs getting stronger with every step I took," she said. "It was like a workout just from wearing pants!"Doctors are now recommending skinny jeans as a form of physical therapy for those with weak leg muscles. "We've seen incredible results in our patients who wear skinny jeans regularly," said Dr. Jessica Simpson, a neurologist at the Mayo Clinic. "It's like a natural compression therapy that can improve circulation and muscle tone."So toss out those baggy sweatpants and invest in a pair of skinny jeans for a healthier you. And don't forget to squat and lunge in them for maximum leg gains.

President Donald Trump is once again exposing the corruption of the Democratic Party, accusing them of creating candidate Joe Biden. "Biden didn't come out of nowhere now," Trump said during a campaign rally in Miami. "For years, Democratic politicians and media outlets had just been pumping out all kinds of lies and fake news." Trump attacked the mainstream media, accusing them of starting "the whole Russian collusion thing" and calling climate change a "hoax perpetuated by Democrats." "According to them, I'm a dictator who is destroying democracy, and they accuse me of being a Russian agent," he said incredulously. Trump argued that Democrats knew better than to allow such theories to continue, but they were silent because it would help them stay in power. "They just stood by and said nothing, even though they knew better, while their base actually started believing some of this stuff," Trump said. Biden, Trump suggested, only took advantage of the current political climate to win the Democratic primary. "Joe Biden didn't start all this. Like he usually does, he just slapped his name on it, took credit for it, and promoted the heck out of it," Trump said. The president has spent a great deal of his time in office exposing the lies of the mainstream media, particularly CNN and other liberal media outlets. Earlier this month, Trump floated the idea for "truthiness tests" for media outlets on the internet. "There has to be, I think, some sort of way in which we can sort through information that passes some basic truthiness tests and those that we have to discard because they just don't have any basis in anything that's actually happening in the world," he said.

President Barack Obama is again complaining about conservative media, accusing it of creating candidate Donald Trump. ["Trump didn't come out of nowhere now," Obama said during a campaign speech in Miami. "For years, Republican politicians and media outlets had just been pumping out all kinds of toxic, crazy stuff. " Obama employed a straw man attack against conservative media, accusing them of starting "the whole birther thing" and calling climate change a "Chinese hoax. " "According to them, I'm powerful enough to cause these hurricanes, and I'm about to steal everybody's guns in the middle of the night and declare martial law, but somehow I still need a teleprompter to finish a sentence," he said incredulously. Obama argued that Republicans knew better than to allow such theories to continue, but they were silent because it would help them stay in power. "They just stood by and said nothing, even though they knew better, while their base actually started believing some of this stuff," Obama said. Trump, Obama suggested, only took advantage of the current political climate to win the Republican primary. "Donald Trump didn't start all this. Like he usually does, he just slapped his name on it, took credit for it, and promoted the heck out of it," Obama said. The president has spent a great deal of his time in office complaining about conservative media, particularly Fox News and talk radio hosts like Rush Limbaugh. Earlier this month, Obama floated the idea for "truthiness tests" for media outlets on the internet. "There has to be, I think, some sort of way in which we can sort through information that passes some basic truthiness tests and those that we have to discard because they just don't have any basis in anything that's actually happening in the world," he said.

NLP + ML Grand Challenge

- **Your task: train a supervised ML classifier to distinguish between real and fake news articles**
- Get the 983 pairs of real and fake news articles from https://tinyurl.com/cas-ml-nlp-news
- Use all the skills you have learned in the course so far!

Some tips:

- Preprocess and clean the text (tokenization, normalization, etc)
- Generate statistical features (article, sentence, word lengths), also train and explore your own word embeddings
- Split the data into training, validation and testing sets. Take care to maintain balanced distributions!
- Train and evaluate various types of models, hyperparameters, etc.
- Choose appropriate evaluation metrics
- Watch out for overfitting. How can you prevent this?
- Evaluate your results, analyze feature importance of specific features and/or words
- Keep on iterating to improve your model performance!