# Machine Learning: Usage of AI to Check the Truth Value of Political Candidate's Tweets

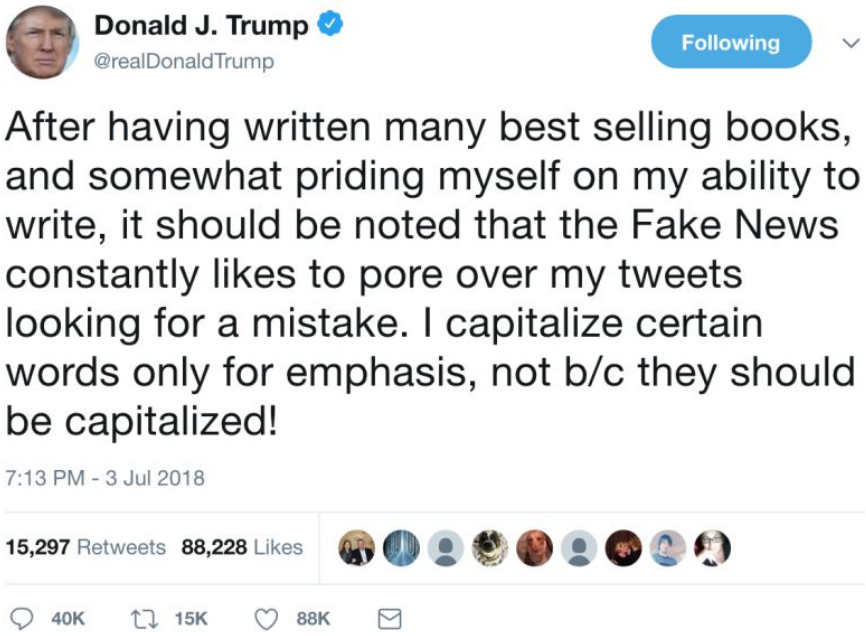
Featuring: Rishik Hombal, Brandon Rudolph, and Jack Zhang

# Social Media and Politics

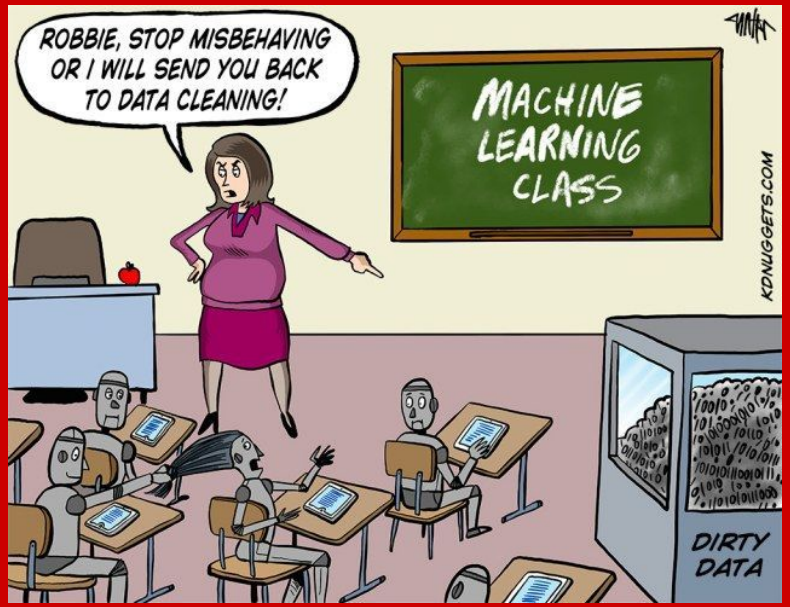
- From December 18th, 2018 to March 11, 2019 the White House did not have a press briefing.
- Over this span, the president averaged nearly ten tweets per day.


Donald J. Trump
@realDonaldTrump
Follow
Despite the constant negative press covfefe...
10:09 AM - 31 May 2017
74,
me not love him. We are
't agree with everything
ve the right to
Donald J. Trump
@realDonaldTrump
Follow
Snow and freezing weather all over mid-
section of Country. Global warming
specialists better start thinking fast!
9:04 AM
249 Ret
203
Donald J. Trump
@realDonaldTrump
A very big part of the Anger we see today in
our society is caused by the purposely false


Donald J. Trump
@realDonaldTrump
Following
Some people HATE the fact that I got along well with President Putin of Russia. They would rather go to war than see this. It's called Trump Derangement Syndrome!
7:27 AM - 18 Jul 2018
33,652 Retweets 130,857 Likes
55K
34K
131K


Donald J. Trump
@realDonaldTrump
Following
After having written many best selling books, and somewhat priding myself on my ability to write, it should be noted that the Fake News constantly likes to pore over my tweets looking for a mistake. I capitalize certain words only for emphasis, not b/c they should be capitalized!
7:13 PM - 3 Jul 2018
15,297 Retweets 88,228 Likes
40K
15K
88K

# Is Machine Learning an Answer for Fact Checking?




Similar to packages we have used in lab to recommend or suggest whether a particular comment is a troll, we hope to use similar algorithms to determine truth...

# The Data

- Requested permission from Twitter to create a Twitter Developer account (security purposes)
- Wrote code to collect tweets from Presidential  Candidates
- Retweets, emojis, hyperlinks, etc. were removed in cleaning the data so our algorithm could read the tweets

# The Training Data

- William Yang Wang used the now defunct Politifact API to fact check a series of political statements made on social media, television, radio, and other mediums and assign a truthfulness score ranging from “Pants on Fire”, the most untrue statement, and “true”, a completely true statement

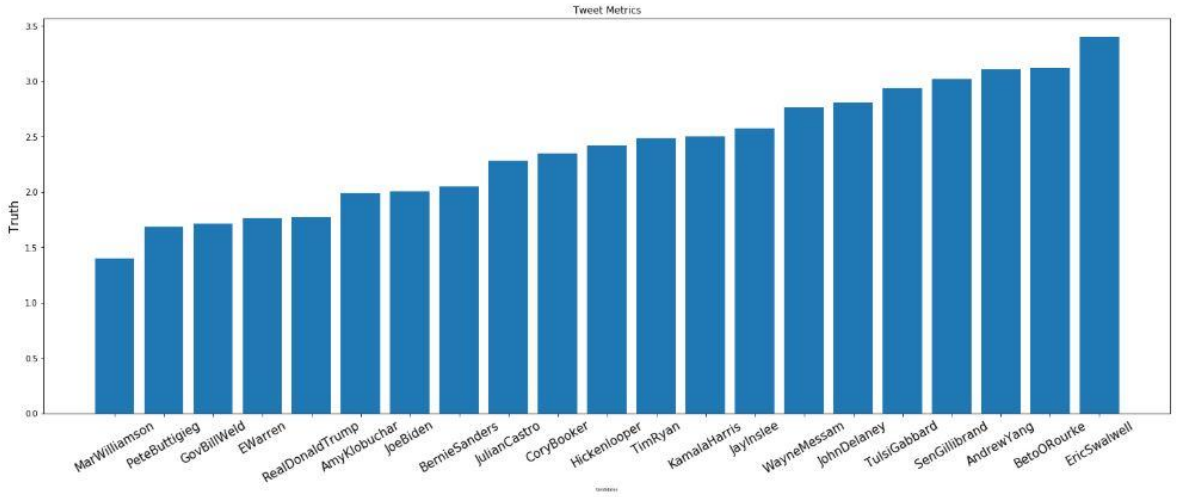
# Intermission

Please wait patiently while we pull up a short presentation of our coding processes. While you're waiting, feel free to sit quickly and patiently, humans are not perfect and neither are you Tommy.

In [4]:
```
tweets = []
total= 0
for i in os.listdir():
    if i.endswith('.xlsx'):
        tweets.append(i)
for i in range(len(tweets)):
    current = tweets[i]
    app = pd.read_excel(current)
    total += len(app)
total
```

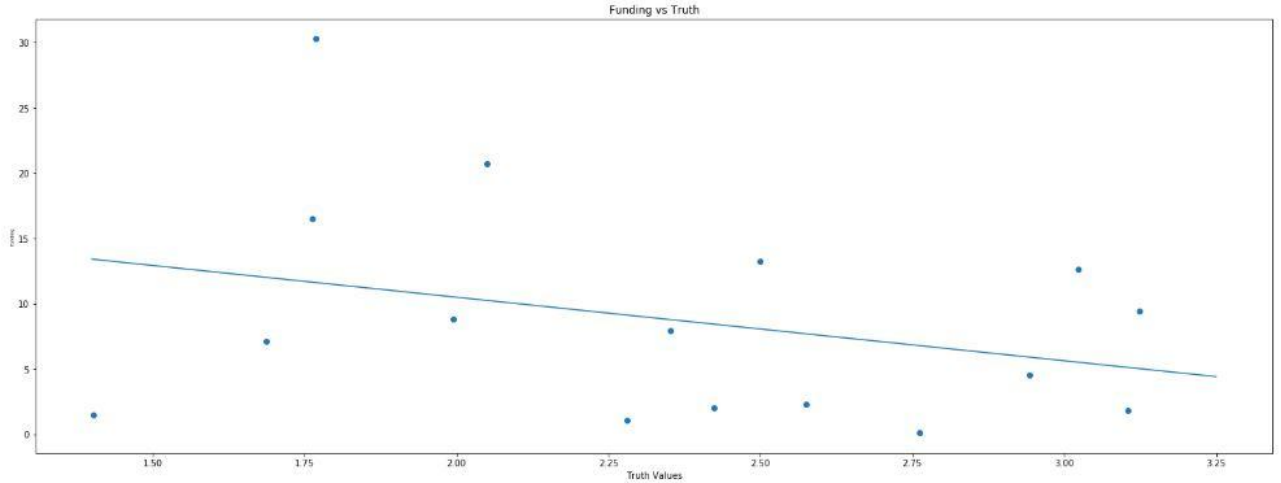
Out[4]: 3281

In [10]:
```
from operator import itemgetter
sortDick = sorted(averageDick.items(), key=itemgetter(1))

sortPeople, sortScores = [], []

for i in range(len(files)):
    sortPeople.append(sortDick[i][0])
    sortScores.append(sortDick[i][1])


index = np.arange(len(files))
plt.bar(index, sortScores)
plt.xlabel('Candidates', fontsize=5)
plt.ylabel('Truth', fontsize=15)
plt.xticks(index, sortPeople, fontsize=15, rotation=30)
plt.title('Tweet Metrics')
plt.rcParams["figure.figsize"] = [25,9]
plt.show()
```



In [18]:
```
plt.scatter(scatterScores, scatterFun)
plt.xlabel('Truth Values', fontsize=10)
plt.ylabel('Funding', fontsize=5)
plt.title('Funding vs Truth')
line = np.linspace(1.4,3.25, 500)
C = (line * model.coef_) + model.intercept_
plt.plot(line, C)
plt.show
```

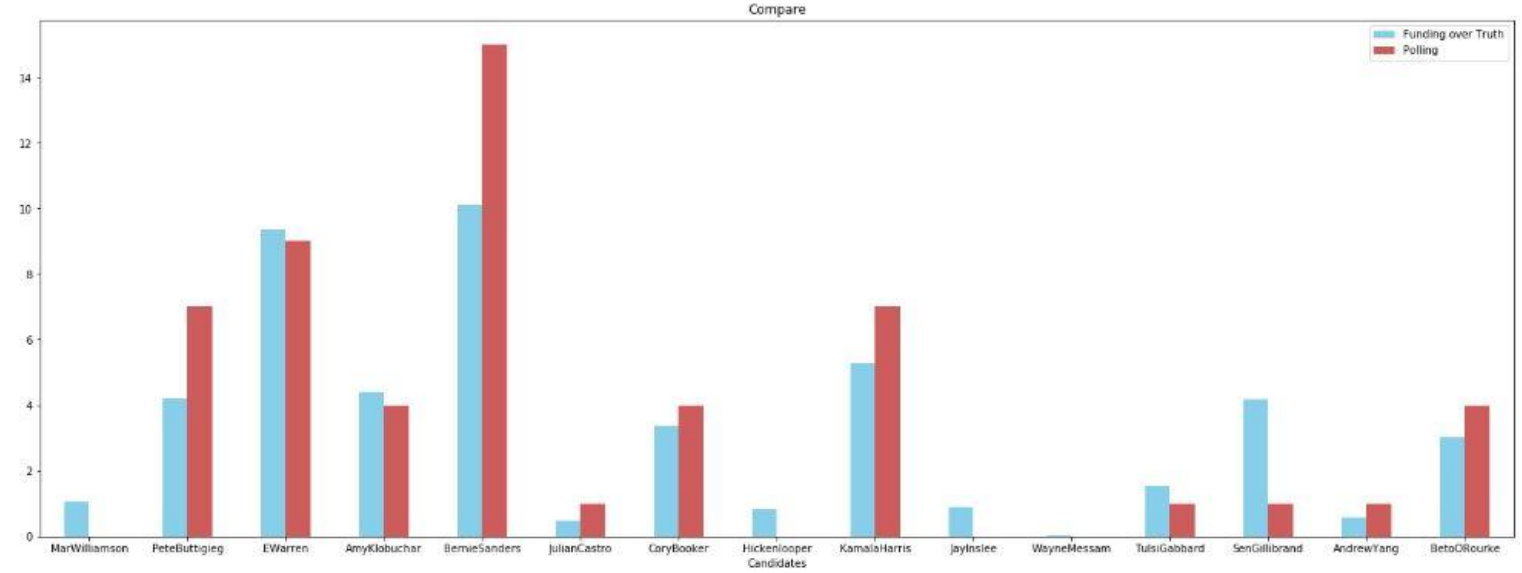
Out[18]: <function matplotlib.pyplot.show(*args, **kw)>




In [19]:
```
model.score(X,y)
```

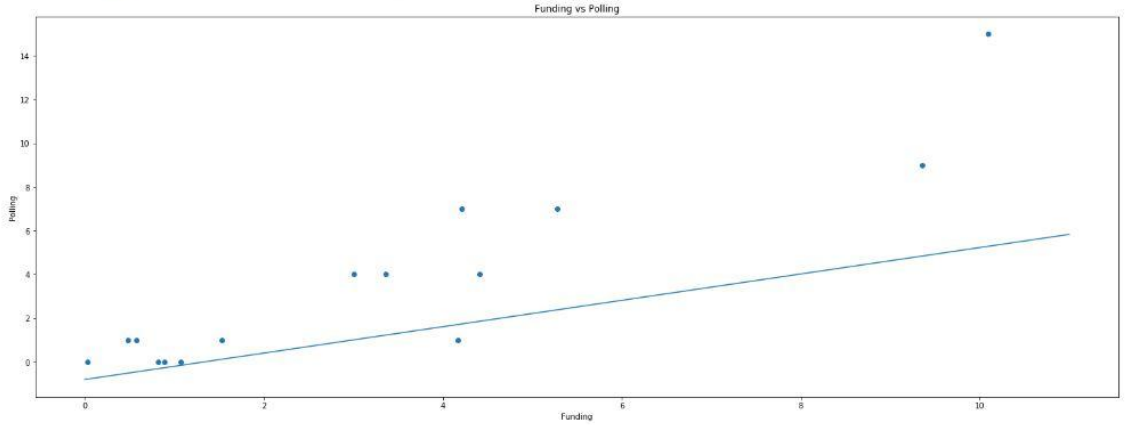
Out[19]: 0.10093438464069338

In [27]:
```
df = pd.DataFrame({"Funding over Truth":ratioList,"Polling":iowa})
ax = df.plot.bar(color=["SkyBlue","IndianRed"], rot=0, title="Compare")
ax.set_xlabel("Candidates")
ax.xaxis.set_major_formatter(plt.FixedFormatter(scatterPeople))
```



In [34]:
```
X = np.reshape(scatterFun, (-1,1))
y = iowa
model = apple.LinearRegression().fit(X, y)
model.score(X,y)
plt.scatter(ratioList, iowa)
plt.xlabel('Funding', fontsize=10)
plt.ylabel('Polling', fontsize=10)
plt.title('Funding vs Polling')
line = np.linspace(0,11, 1000)
C = (line * model.coef_) + model.intercept_
plt.plot(line, C)
plt.show
```

Out[34]:
```
<function matplotlib.pyplot.show(*args, **kw)>
```

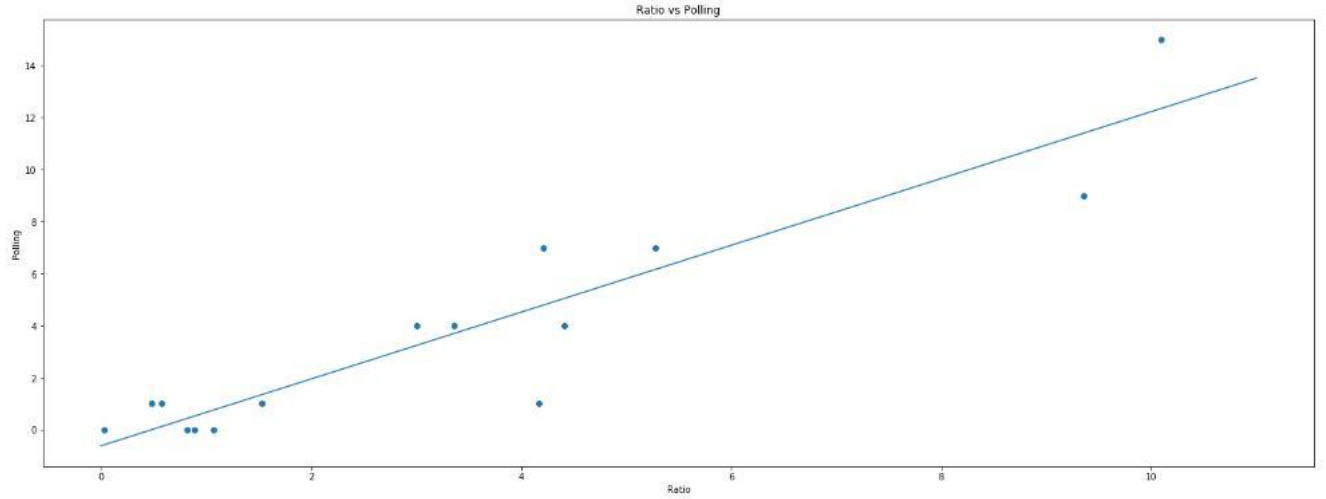



In [35]:
```
model.score(X,y)
```

Out[35]:
```
0.7638249941458373
```

In [30]:
```
plt.scatter(ratioList, iowa)
plt.xlabel('Ratio', fontsize=10)
plt.ylabel('Polling', fontsize=10)
plt.title('Ratio vs Polling')
line = np.linspace(0,11, 1000)
C = (line * model.coef_) + model.intercept_
plt.plot(line, C)
plt.show
```

Out[30]: <function matplotlib.pyplot.show(*args, **kw)>



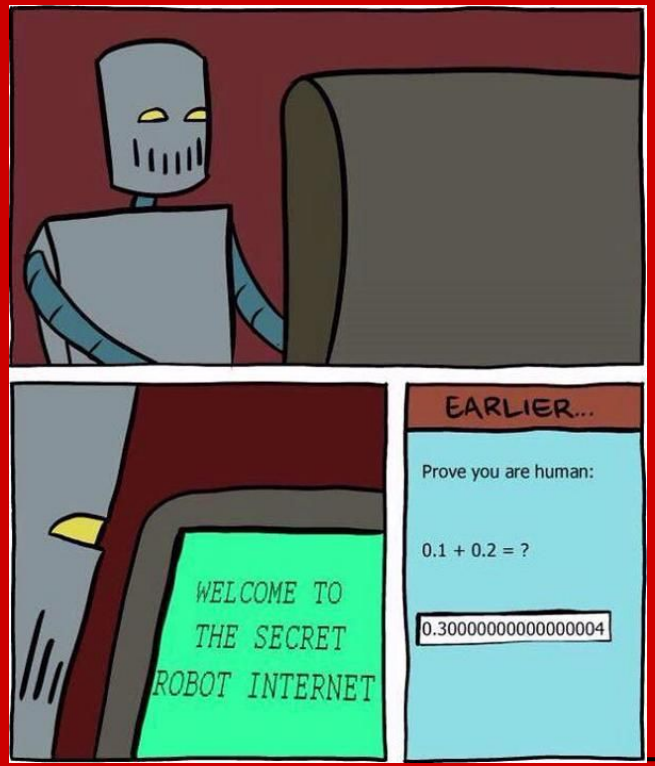

In [31]:
```
model.score(X,y)
```

Out[31]: 0.8571993106393145

# What Did We Learn

- Our methods did not do so well. Why?
- Even top research institutions have struggled to tackle the challenge of “fact checking’
- The best model accurately labeled news outlets with low, medium, or high factuality just 65% of the time.
- Our model had an accuracy of roughly 30%

# What is True?

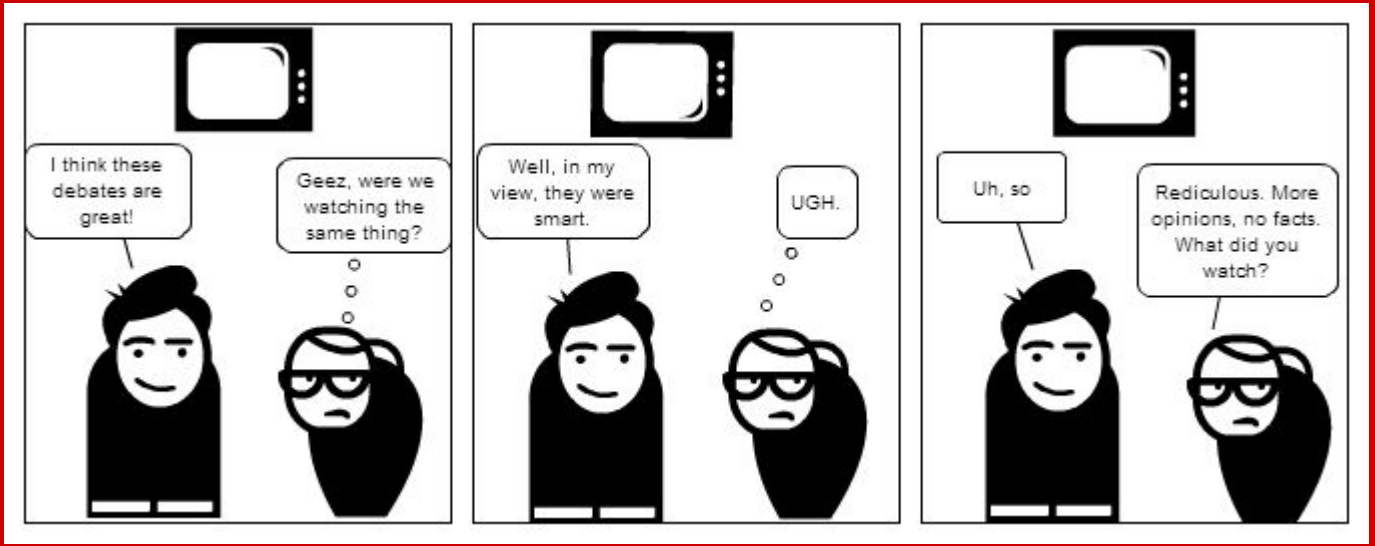



“The United States must expand medicare to all Americans.”

- Politicians often speak in a passive manner, making their statements, and tweets, lack binary-type phrasing.
- The AI excelled at pure True/False type premises, but not all tweets will have a clear premise.
- Argument vs. Opinion vs. Fact

# What Might Improve Results?

- Perhaps a better metric than a True vs False scale for rating the validity of tweets would be an Objective vs Subjective scale.
  - This would categorize the data in a manner which could be used to further break down.



# The Prosecution Rests.