"I did not have sexual relations with that woman... Miss Lewinsky"

Verbal Deception Detection with spaCy

By Victor Palacios

Goals

1. Introduce a verbal deception detection study

2. Introduce spaCy to elevate that study

Where to begin?

- Where's the data?

The Internet!

- politifact.com

Donald Trump

stated on November 22, 2020 in a tweet:

"In certain swing states, there were more votes than people who voted, and in big numbers."

PANTS ON FIRE!
POLITIFACT TRUTH-O-METER

40 Sample Points

Original Data Format:

Statement	Veracity	Source
"Since the Affordable Care Act passed, 90 percent of job growth has been in full-time positions."	TRUE	Alan Krueger

ALAN KRUEGER
1960-2019
CNBC

Count / Gender Addition:

Statement	Veracity	Source	Nouns	Adjectives	Verbs	Gender
"Since the Affordable Care Act passed, 90 percent of job growth has been in full-time positions."	TRUE	Alan Krueger	3	2	2	M

How do I scale?
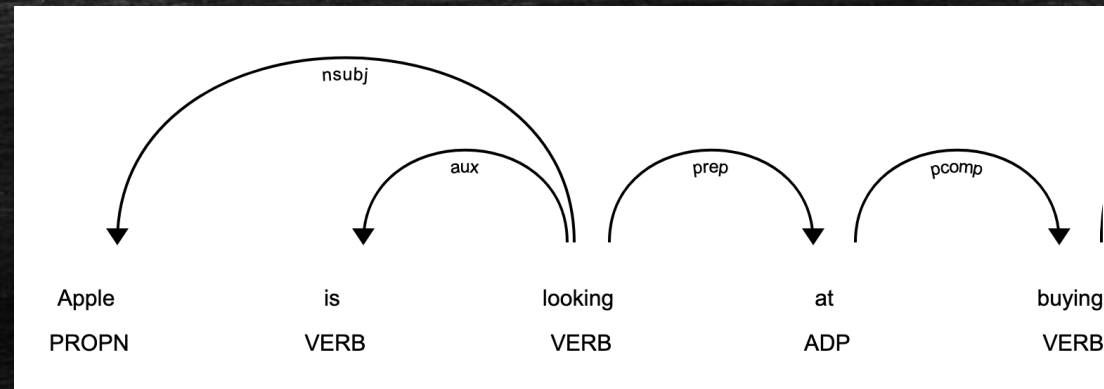
- Labeling 10,000+ data points by hand would be time-consuming and error prone

Scaling with spaCy

- spaCy's Part-of-Speech Tagging


```
I           PRON        PRP         pronoun, personal
hated       VERB        VBD         verb, past tense
it          PRON        PRP         pronoun, personal
in          ADP         IN          conjunction, subordinating or preposition
my          ADJ         PRP$        pronoun, possessive
childhood   NOUN        NN          noun, singular or mass
though      ADP         IN          conjunction, subordinating or preposition
```

Take Away

- Deception detection is multifaceted: data, validity, how to analyze

- AI labeling (spaCy) can augment data labeling for large scale studies