Exposing Privacy Risks on Social Media

Isabela Figueira, Archana Godavarthy, Prof. Yi Fang
Computer Engineering Department

Santa Clara University
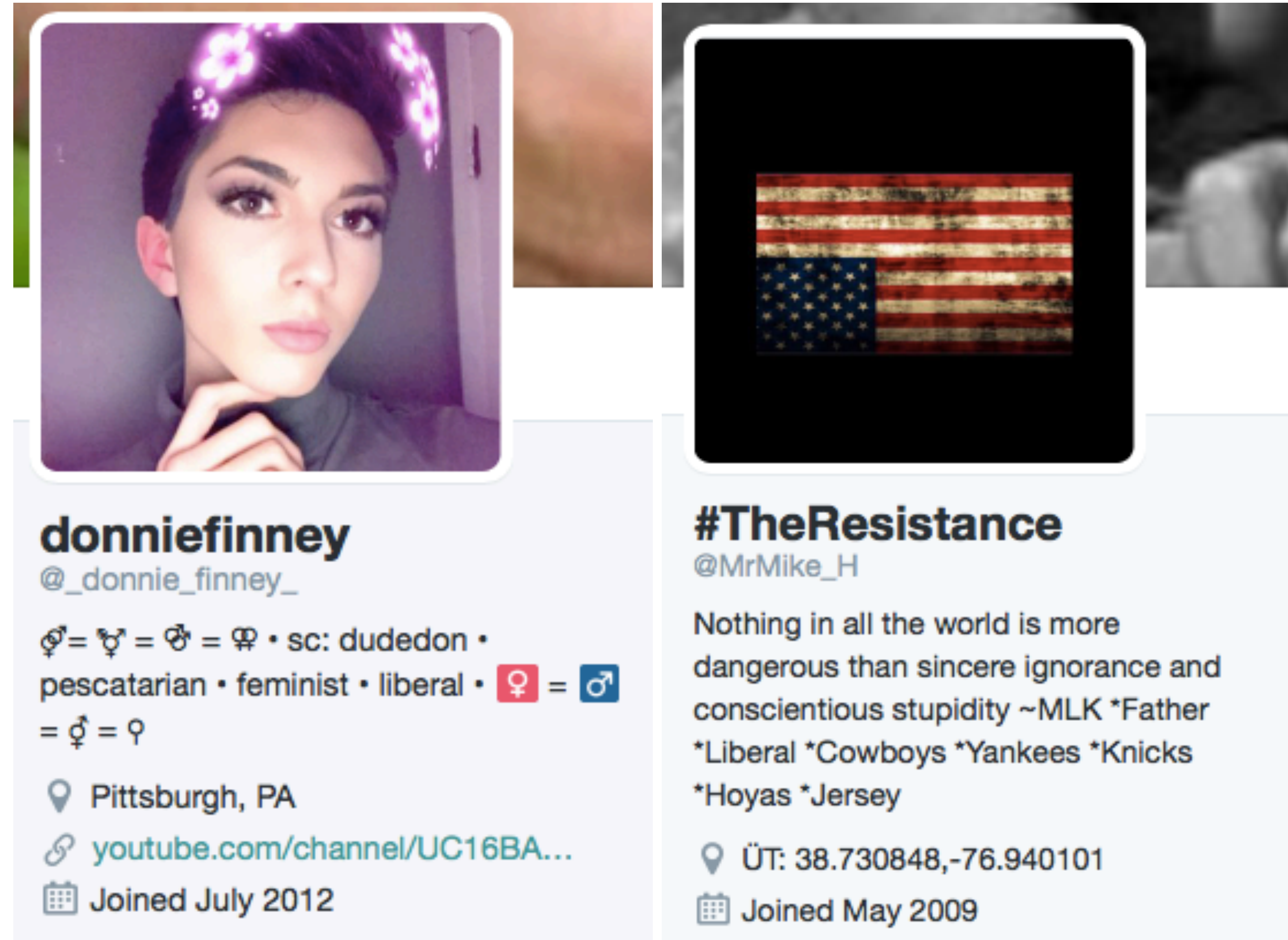
School of Engineering

Introduction:
The info social media users post can be used to learn about the user's personal lives.

Objectives:
Use machine learning to learn user traits from their tweets in order to uncover social media privacy risks.

Why bother?
Now, it doesn't seem like such a threat. But give it time!

Machine learning will be able to be used to parse the vast amounts of data we volunteer about ourselves and learn about things we don't even know we don't want shared with the world.

Who knows what future threats to our data there are.

donniefinney
@_donnie_finney_

♀= ⚥ = ⚨ = ⚢ • sc: dudedon • pescatarian • feminist • liberal • ♀ = ♂ = ⚨ = ♀

📍 Pittsburgh, PA

🔗 youtube.com/channel/UC16BA...

📅 Joined July 2012

#TheResistance
@MrMike_H

Nothing in all the world is more dangerous than sincere ignorance and conscientious stupidity ~MLK *Father *Liberal *Cowboys *Yankees *Knicks *Hoyas *Jersey

📍 ÜT: 38.730848,-76.940101

📅 Joined May 2009

The Process:
• Scrape Twitter stream
• Get users meeting criteria
• Preprocess their tweets
• Train an algorithm that maps a small selection of users to their positive or negative trait
• Test the algorithm on more users we know the traits of and analyze results

Next Step:
After training the data to uncover the relationship, we want to find a way to change the data so the algorithm no longer works to protect people's privacy

Traits in question:
Married (1) vs unmarried (0)
Democrat (1) and Republican (0)

Collect and preprocess data → Train and analyze data → Change the tweets so that the trained algorithm is now ineffective