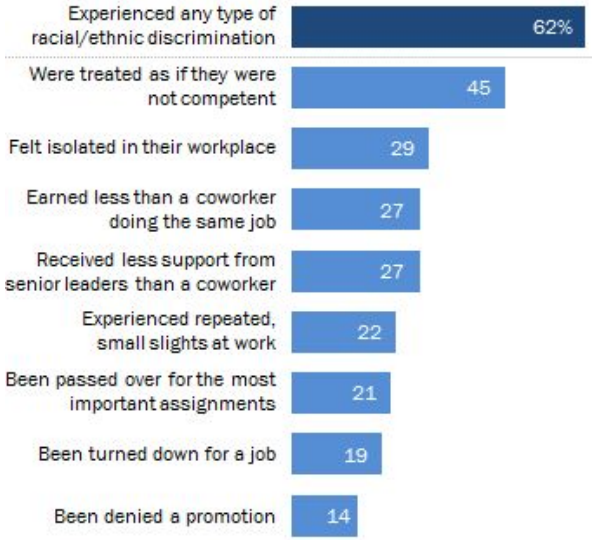
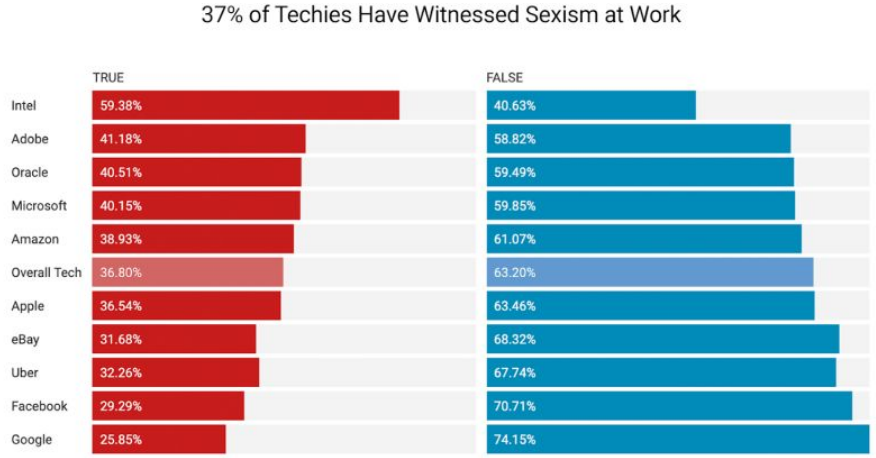
Inappropriate speech detection

Team Mavericks
@fortune @sai @serena @sofia

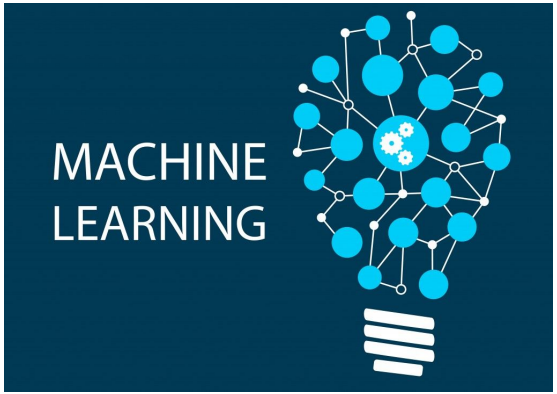
What's the problem we try to solve?

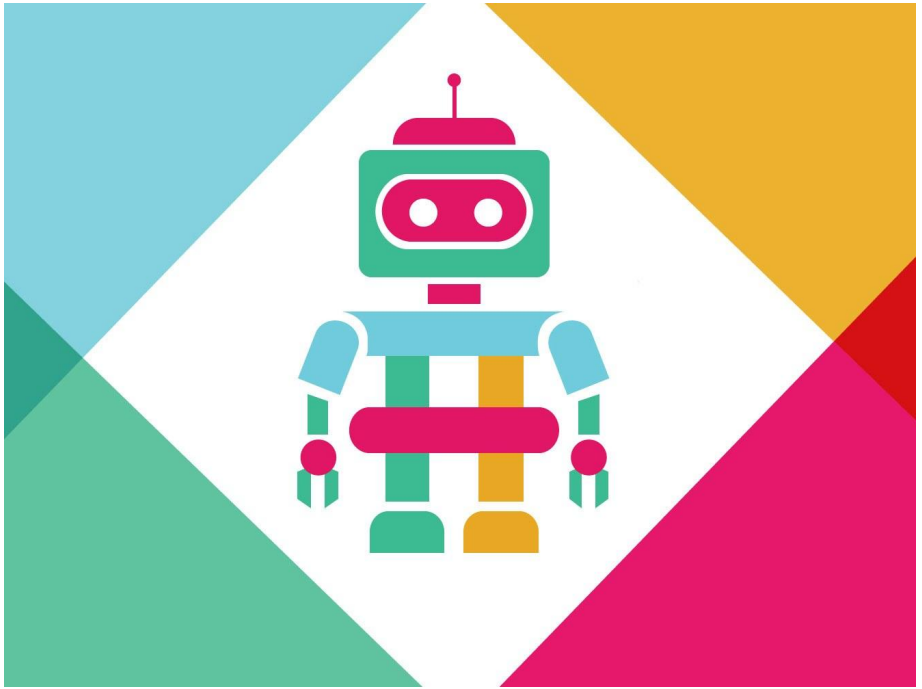
Inappropriate Comments





Solution



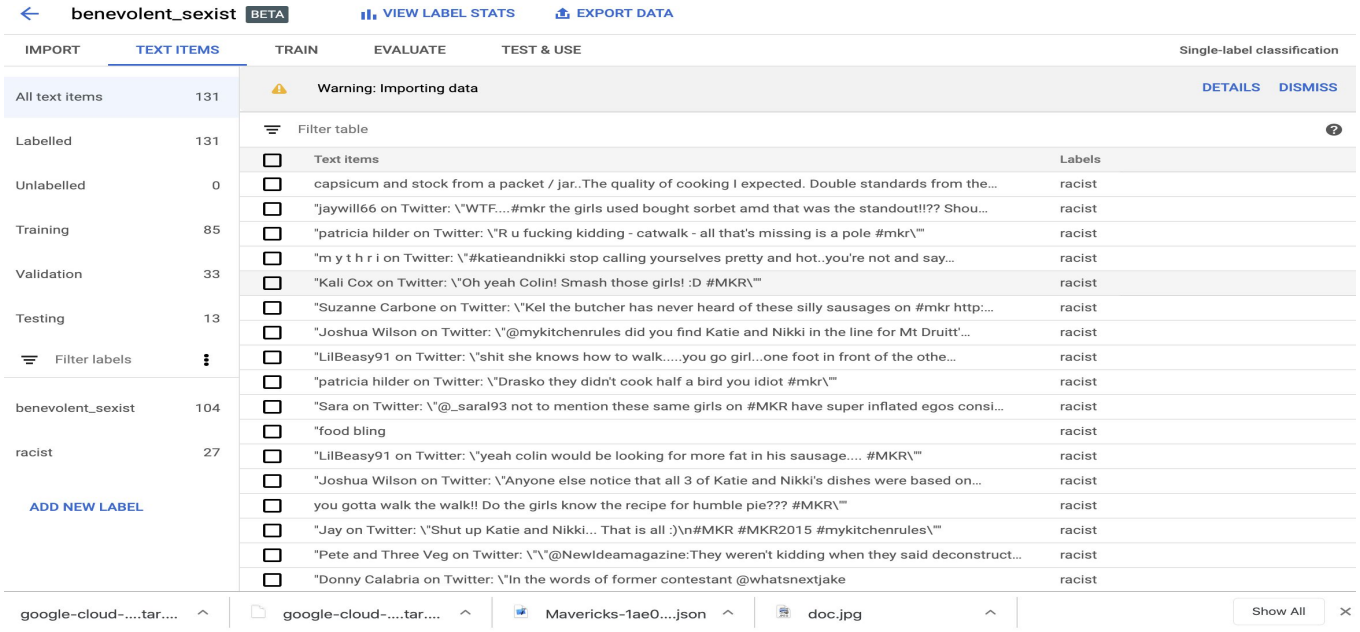


How? (Data set)

Datasets of annotated tweets for racism & sexism

- https://github.com/AkshitaJha/NLP_CSS_2017
- https://github.com/ZeerakW/hatespeech

How? (Training)

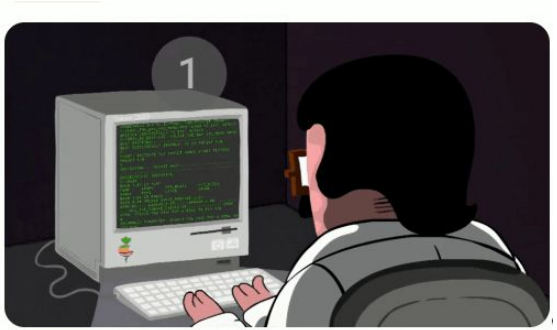


How? (Detection)

Speech and Text

How? (Education - Dialog flow)

- Explain why comments are inappropriate
- Train Dialog-bot to help frame better sentences.



How? (Calls)

- Google Speech to Text for parsing voice

DEMO TIME!!

Challenges

- Training of model
- Dialog-flow requires a lot of examples
- Twitter != workplace communications

Next steps

- Train model with more data
- Report for repetitive comments
- Productionize (Google Run for our API)
- More platforms (Google Assistant, Viber, Telegram,...)

Best part

- Everybody learned and contributed
- We had a lot of fun :)