

Verso l'identificazione automatica di utenti Twitter a rischio binge drinking

Relatore: Prof.ssa Gabriella Pasi

Correlatore: Dott. Marco Viviani

Laureando:

Luca Chiodini

Binge drinking

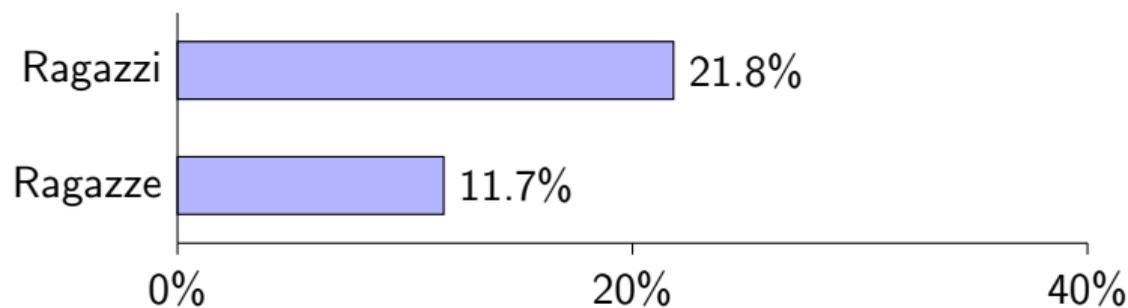
- “Bere alcol in modo estremo”

Binge drinking

- “Bere alcol in modo estremo”
- Oltre 5 UA per gli uomini (4 per le donne)

Binge drinking

- “Bere alcol in modo estremo”
- Oltre 5 UA per gli uomini (4 per le donne)
- Fenomeno anglosassone, diffuso anche in Italia



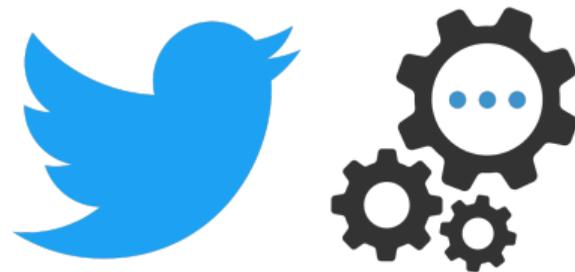
(ISTAT, 2017)

Studi sul binge drinking



Bartoli et al. 2014; Carrà et al. 2015

Studi sul binge drinking



Bartoli et al. 2014; Carrà et al. 2015

Ricerca per 23 hashtag:

| | |
|-----------------|-----------|
| #bingedrinking | #alcohol |
| #nomorealcohol | #cocktail |
| #drinking | #vodka |
| #toomuchalcohol | #hangover |
| #drunk | #rum |
| #drunkasfuck | ... |

Ricerca per 23 hashtag:

| | | | D1 | D2 |
|-----------------|-----------|-------------------|------------|------------|
| #bingedrinking | #alcohol | Data inizio | 29/11/2017 | 16/04/2018 |
| #nomorealcohol | #cocktail | Data fine | 22/03/2018 | 31/05/2018 |
| #drinking | #vodka | Giorni | 114 | 46 |
| #toomuchalcohol | #hangover | Numero di tweet | 452 262 | 194 815 |
| #drunk | #rum | Tweet giorn. medi | 3 967 | 4 235 |
| #drunkasfuck | ... | | | |

Uno sguardo da vicino al dataset. . .



Thanks for a great welcome #ANU. Students were keen yesterday to learn about our campaign. We'll be around for the rest of the year to help reduce the harm to students from #alcohol.

— @MichaelTThorn

Uno sguardo da vicino al dataset. . .

 Thanks for a great welcome #ANU. Students were keen yesterday to learn about our campaign. We'll be around for the rest of the year to help reduce the harm to students from #alcohol.
— @MichaelTThorn

 \$5 #appetizers, \$2 Domestic #beer bottles and well #drinks : 4p.m. - 6p.m. #happyhour
#BOOM
— @CityTapTheAttic

Uno sguardo da vicino al dataset. . .

 Thanks for a great welcome #ANU. Students were keen yesterday to learn about our campaign. We'll be around for the rest of the year to help reduce the harm to students from #alcohol.

— @MichaelTThorn

 \$5 #appetizers, \$2 Domestic #beer bottles and well #drinks : 4p.m. - 6p.m. #happyhour #BOOM

— @CityTapTheAttic

 Read. Eat. Drink. R E D. #feltrinelli #spritz #aperolspritz #drink #alcohol #friends #weekend #saturday #milan <https://t.co/s4eHxXuhDn>

— @PasiniDebora

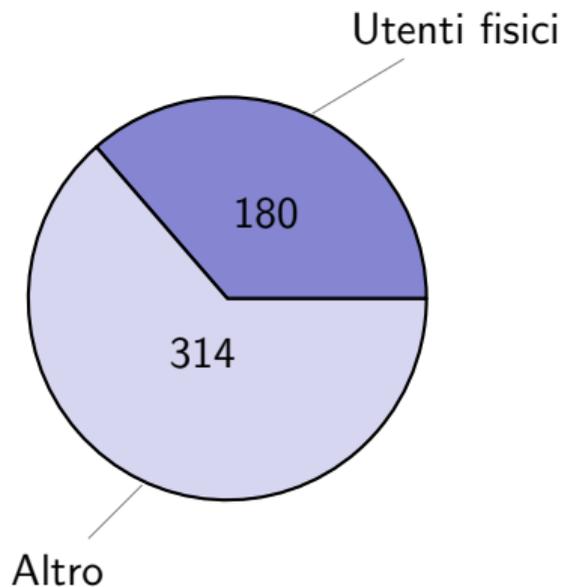
Fissiamo una definizione: vogliamo tenere solo
“persone fisiche che utilizzano l’account senza finalità commerciali”

Fissiamo una definizione: vogliamo tenere solo
“persone fisiche che utilizzano l’account senza finalità commerciali”

Alcune tra le features discriminanti:

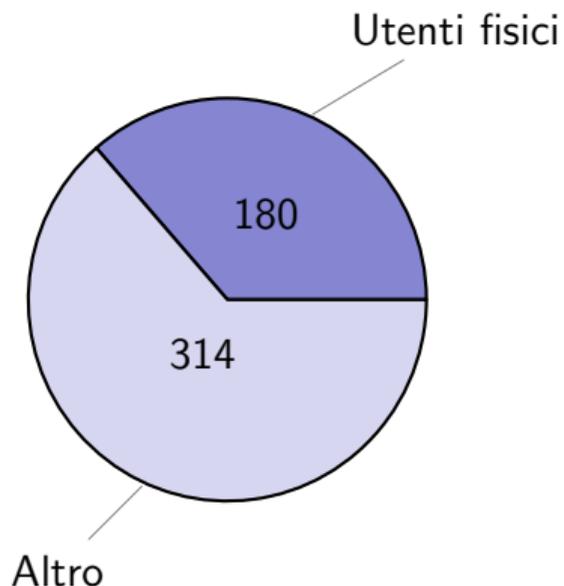
- Numero di tweet
- Numero di # e @ utilizzati in media nei tweet
- Dimensione rete sociale e rapporto followers / followee
- Percentuale di retweet
- Presenza di tweet geolocalizzati

Apprendimento automatico supervisionato



Groundtruth (494 utenti)

Apprendimento automatico supervisionato



Groundtruth (494 utenti)

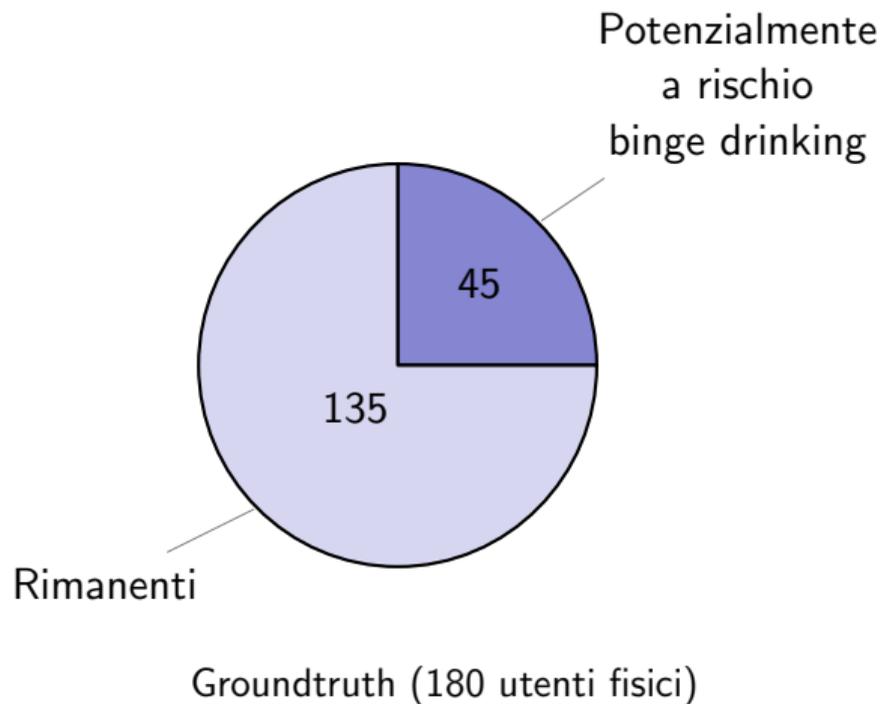
| | Accuracy | F1 score |
|---------------|----------|----------|
| SVMs | 0.76 | 0.66 |
| Random Forest | 0.73 | 0.57 |

Seconda classificazione: rischio binge drinking

Features linguistiche: tweet visti come vettori in uno spazio pluridimensionale, pesati con schema **tf-idf**.

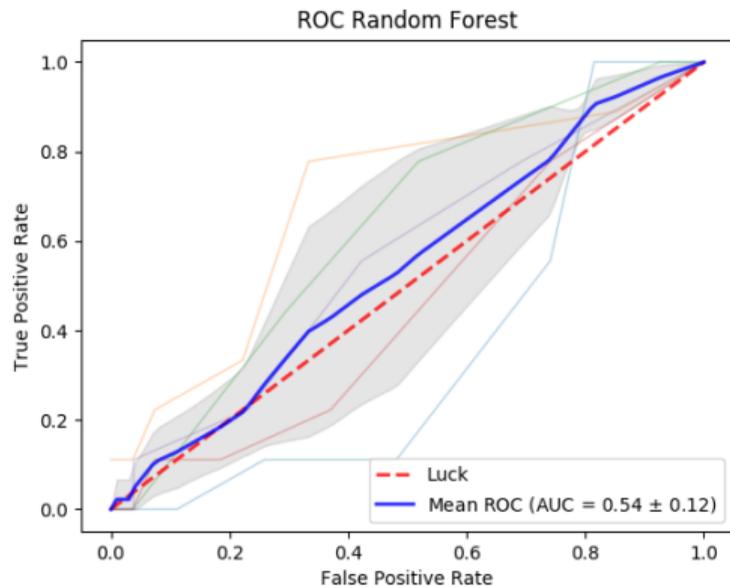
Seconda classificazione: rischio binge drinking

Features linguistiche: tweet visti come vettori in uno spazio pluridimensionale, pesati con schema **tf-idf**.



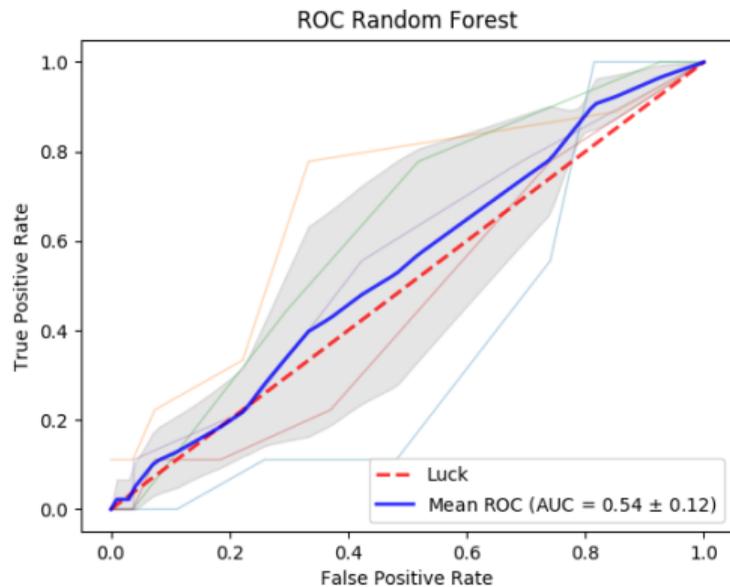
Seconda classificazione: risultati

Dataset iniziale:
~2 tweet per utente

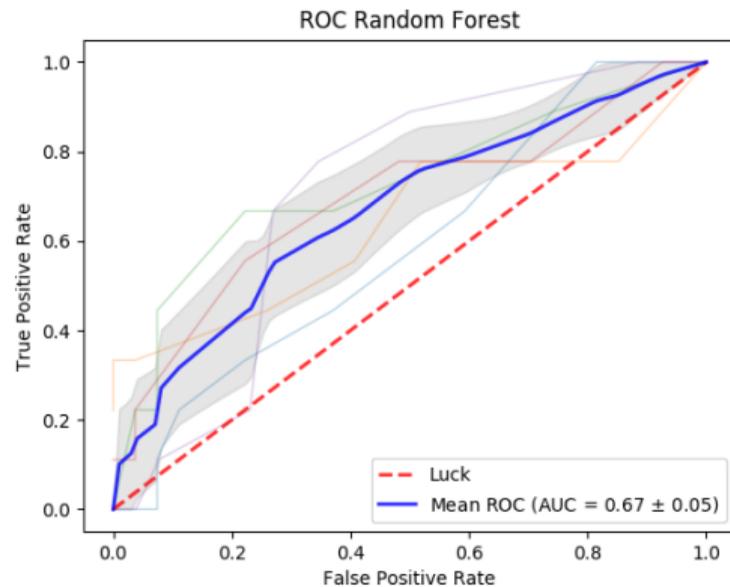


Seconda classificazione: risultati

Dataset iniziale:
~2 tweet per utente



Recrawling intera cronologia:
~1 900 tweet per utente



- Problema difficile: uno dei primi approcci in assoluto in letteratura
- Classificazione degli utenti fisici: livello buono
- Classificazione potenziali binge drinker: da migliorare

- Problema difficile: uno dei primi approcci in assoluto in letteratura
- Classificazione degli utenti fisici: livello buono
- Classificazione potenziali binge drinker: da migliorare
- Sviluppo futuro: esplorare altri social network (foto da Instagram)