DATA SCIENCE & AD:TECH

MIKE SCHUMACHER 2018-11

AGENDA

- Quick Bio
- Ad:Tech 101
- DS Application – Propensity Ranking
- DS Application – Purchase Likelihood
- DS Application – Clustering
- Summary

AGENDA

- **Quick Bio**
- Ad:Tech 101
- DS Application – Propensity Ranking
- DS Application – Purchase Likelihood
- DS Application – Clustering
- Summary

QUICK BIO

Economics + Statistics

MS, Statistics

(Analyst) Predictive Modeling, Design of Experiment, BI

(Director) Managed ~10, Large Scale Predictive Modeling, Causal Inference

(VP) Managed ~25, Machine Learning, Optimization Algo's, Consumer Insights

(VP) Manage ~50, Machine Learning, College Recruiting, Tech/Tools

AGENDA

- Quick Bio
- **Ad:Tech 101**
- DS Application – Propensity Ranking
- DS Application – Purchase Likelihood
- DS Application – Clustering
- Summary

AD:TECH 101

- Enables “free” Social Networking, News, Search, Music, etc.
  - 84% of Google Revenue is Advertising
  - 90% of Facebook Revenue is Advertising
- Ad:Tech Attempts to Connect Consumers With Relevant Ads
  - More effective ads
  - Better consumer experience

MARKETER
Agencies
OmnicomGroup
WPP
PUBLICIS
IPG Interpublic Group
HAVAS
AEGIS
MERKLE
MDC PARTNERS
AKQA
LBi
ROSETTA
icrossing
InnovationINTERACTIVE
camelot
at&t Interactive
ReachLocal
yodle
WebVisible
Agency Trading Desks
Accuen The Trading Desk
XAXIS
MEDIA INNOVATION GROUP
VivaKi
CADREON
RUN
DIGILANT
accordantmedia
IMPAQT
VARICKMEDIA
DSPs
TURN
invitemedia
MediaMath
DataXu
appnexus
EfficientFrontier
theTradeDesk
TRIGGIT
Simplifi
Chango
sitescout
AdBuyer.com
adz
BRANDSCREEN
Exchanges
doubleclick by Google
RIGHTMEDIA
Facebook Exchange
ADECN
OpenX
ZENOVIA
adBrite
adjug
Ad Networks
YAHOO!
Horizontal
Google
Tribal FUSION
Microsoft Media Network
amazon
Aol.
CPX
24/7
audience
Casale
ValueClick
ADVERTISING.COM
Undertone
SOLVE media
Brand.net
AT&T AdWorks
BurstMedia
UNDERDOG
Xertive media
AdDynamix
traffic marketplace
Video / Rich Media
SAY
TREMOR VIDEO
videology
YuMe
BBE
TubeMogul
BrightRoll
spotXchange
adotube
adconion
gumgum
BREAK MEDIA
meebo
ScanScout
Jambo Media
Sharing Data / Social Tools
facebook
ShareThis
AddThis
tynt
gigya
SSPs
Admeld
rubicon project
PubMatic
ReviNet
LIFT DNA
ALTITUDE
sonobi
Vertical / Custom
ravora
LONGBOARD MEDIA
ADIFY
GORILLA NATION
Sportgenic
NetShelter
RGM ALLIANCE
GourmetAds
jumpstart
GLAM MEDIA
Q1Media
martini media
quadrantONE
InflectionPointMedia
IDGTechNetwork
SOJERN
Ideas People
Tone Fuse
FEDERATED MEDIA
Targeted Networks / AMPs
rocketfuel
interclick
collective
dstillery
acuity
choicestream
Taykey
specificmedia
precisiondemand
33across
radiumone
resonate networks
socialvibe
OwnerIQ
NetSeer
Adblade
pulse360
adaramedia
peer39
inadco
AUDIENCESCIENCE
XGRAPH
CONTEXTWEB
Performance
Kontera
epic
MATOMY
Vantage Media
convertmedia
ENGAGE
moxymedia
commission junction
VIBRANT
MadisonLogic
LinkShare
phg
adknowledge
MARCHEX
neverblue
Datran Media
hydra
Tatto Media
TheUseful
MediaTrust
Mobile
admob
pontiflex
Quattro wireless
Jumptap
mojiva
millennialmedia
GREYSTRIPE
rhythm
InMobi
transpera
Publisher Tools
Yieldex
fatTail
ShortTail
isocket
BuySellAds
SHINY ADS
yieldbot
BuzzFeed
DailyMe
scout analytics
daylife
METAMARKETS
Mochila
MAXIFIER
PerfectMarket
PARSE.LY
Outbrain
Jump Time
Taboola
Gravity
Luminate
Ad Servers
doubleclick by Google
24/7 MEDIA
apt
LiveIntent
ADTECH
adgear
ZEDO
OpenX
AdJuggler
Adzerk
aiMatch
PUBLISHER
CONSUMER
Creative Optimization
teracent
TUMRI
dapper
AdExtent
adroit
Impossible Software
struq
MIXPO
paper g
SPONGECELL
STEELHOUSE
ipromote
adacado
snapads
Dynamix
CERTONA
Ad Ready
richrelevance
ADISN
Media Planning and Attribution
MarketShare
COGNITIVE MATCH
NETWORKED INSIGHTS
M FACTOR
visualIQ
pulsepoint
convertro
C3 Metrics
has offers
ThinkVine
adometry
Encore
COVARIO
MARKETING EVOLUTION
Korrelate
DMPs and Data Aggregators
bluekai
[x+1]
aki
Demdex
PARIO
exelate
brilig
quantcast
bizo
LOTAME
JovianDATA
krux
Red Aril
datonics
upfront
LiveRamp
proximic
acerno
PEER39
Admantx
Tag Mgmt
TagMan
ensighten
TEALIUM
BRIGHTTAG
Retargeting
dotomi
criteo
TellApart
STEELHOUSE
AdRoll
sociomantic
MYBUYS
FetchBack
buysight
crosspixelmedia
ReTargeter
MAGNETIC
myThings
mediaFORGE
Ad Servers
doubleclick by Google
atlas
Mediaplex
mediamind
pictela
pointroll
Verification / Privacy
DV DoubleVerify
Integral Ad Science
Ad-Juster
ADXPOSE
Measurement and Analytics
OMNITURE
unica
Coremetrics
tracksimple
webtrends
FLURRY
nielsen
bioscience
comScore
Data Suppliers
Experian
acxiom
QUOVA
datalogix
epsilon
TARGUSinfo
TruSignal
NETEZZA
Where
bluecava
Media Mgmt Systems and Operations

Demo

AGENDA

- Quick Bio
- Ad:Tech 101
- **DS Application – Propensity Ranking**
- DS Application – Purchase Likelihood
- DS Application – Clustering
- Summary

APPLIED EXAMPLES - APP INSTALLS

Business Question:

Which Facebook users are most likely to install Doodle Jump?

APPLIED EXAMPLES - APP INSTALLS

Naive Bayes: http://en.wikipedia.org/wiki/Naive_Bayes_classifier

$$p(C_k|x_1,\dots,x_n) = \frac{1}{Z}p(C_k)\prod_{i=1}^{n}p(x_i|C_k)$$

Ck = 1 is "Class 1", meaning the app *was* installed for a randomly selected set of users.

x1 = predictive variable #1 (number of game app's installed)
x2 = predictive variable #2 (age of device owner)

Z = is likelihood of feature set (xi's)

In English: What is the probability of having Doodle Jump installed given your age and number of game apps?

APPLIED EXAMPLES - APP INSTALLS

Let's consider the ratio of the probabilities under each class. In addition:

… take the log (monotonic)

… use Bayes rule (flip the conditional)

… and assume independence (*Naive* Bayes.)

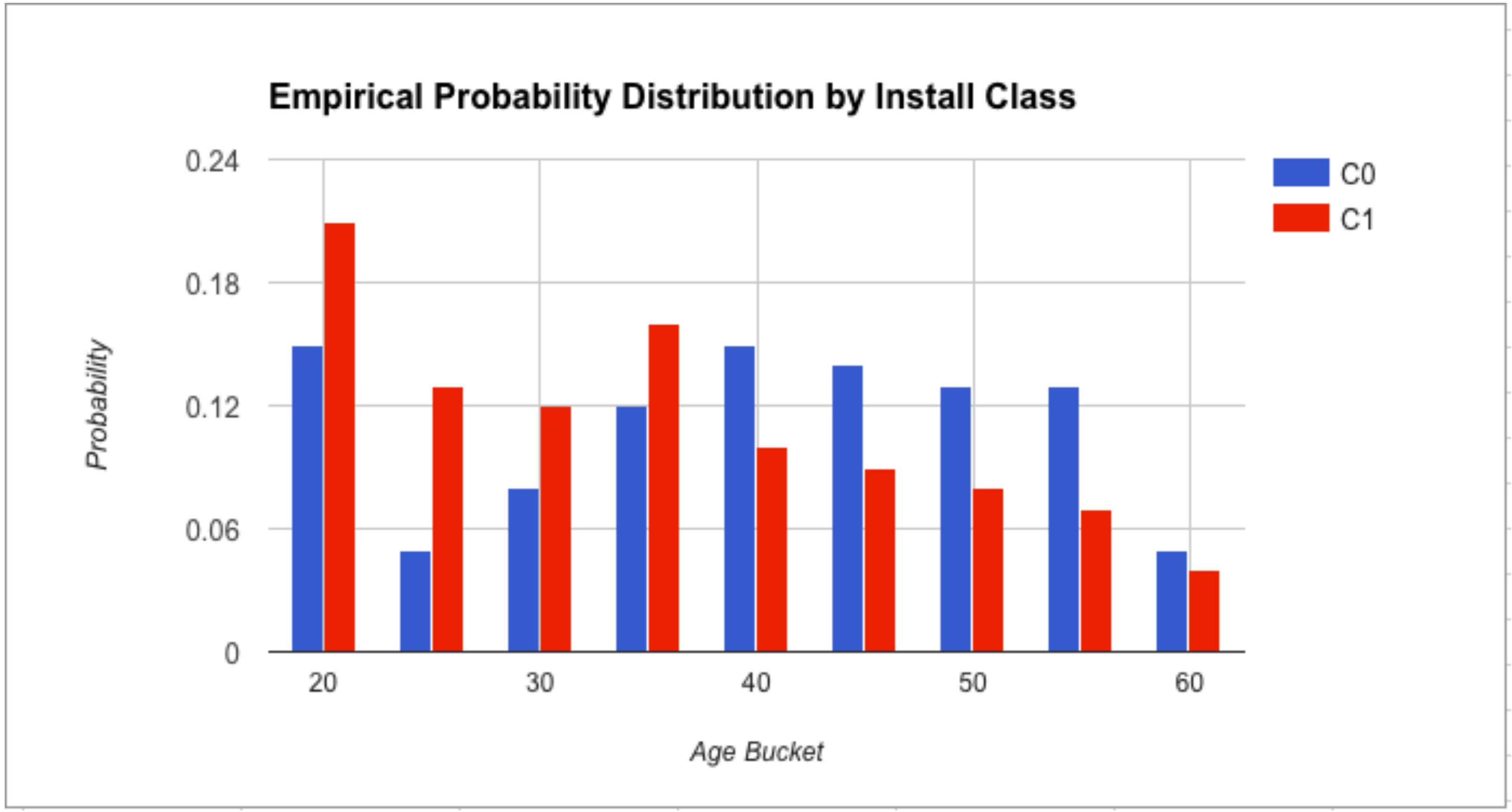
$$ln\left(\frac{P\left(C_1|x_1,x_2\right)}{P\left(C_0|x_1,x_2\right)}\right) \propto \sum_{i=1}^{2} ln\left(\frac{P\left(x_i|C_1\right)}{P\left(x_i|C_0\right)}\right)$$

Now we can sum the log-odds which can be empirically calculated from "big data."

In English: We can evaluate each variable (age, # of apps) on its own after splitting the data into "Doodle Jump" vs "No DJ"

APPLIED EXAMPLES - APP INSTALLS

Example: Predictive Variable Age

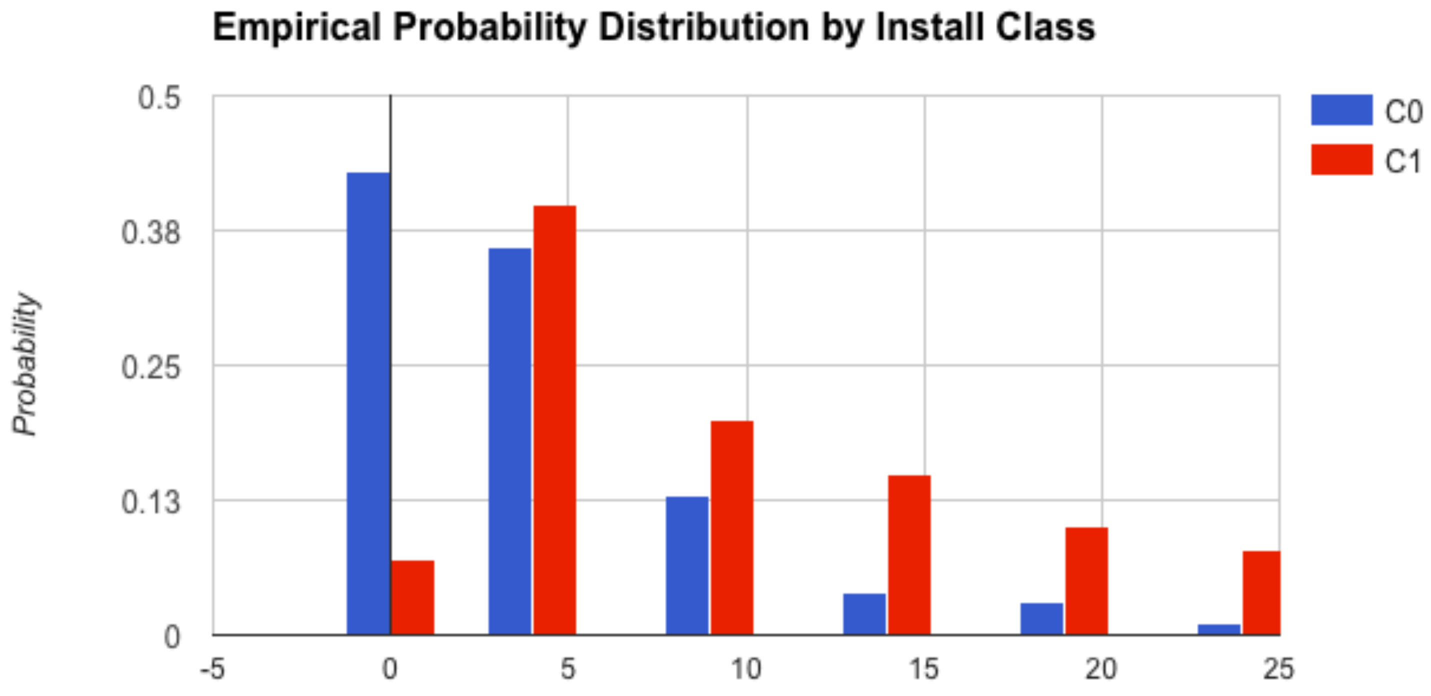


$$ln\left(\frac{P(x_1|C_1)}{P(x_1|C_0)}\right)$$

If “Ron” is in age bucket 25, his ‘term score’ for age is the natural log of 13% divided by 5%, or 0.95.

APPLIED EXAMPLES - APP INSTALLS

Example: Predictive Variable Number of Games Installed



Number of Game Apps Installed

$$ln\left(\frac{P\left(x_2|C_1\right)}{P\left(x_2|C_0\right)}\right)$$

If "Ron" has 25 game apps installed, his 'term score' for "game apps" is the natural log of 8% divided by 1%, or 2.1.

APPLIED EXAMPLES - APP INSTALLS

$$ln\left(\frac{P\left(C_1|x_1,x_2\right)}{P\left(C_0|x_1,x_2\right)}\right) \propto \sum_{i=1}^{2} ln\left(\frac{P\left(x_i|C_1\right)}{P\left(x_i|C_0\right)}\right)$$

Score “Ron” who has 25 games installed and is age 25:

User Score = Age_Term_Score + Game_Apps_Age_Score

User Score = 0.95 + 2.10 = 3.05

Compare this score to peers… select the best ranking users (no probabilities.)

AGENDA

- Quick Bio
- Ad:Tech 101
- DS Application – Propensity Ranking
- **DS Application – Purchase Likelihood**
- DS Application – Clustering
- Summary

APPLIED EXAMPLES - PURCHASE PREDICTION

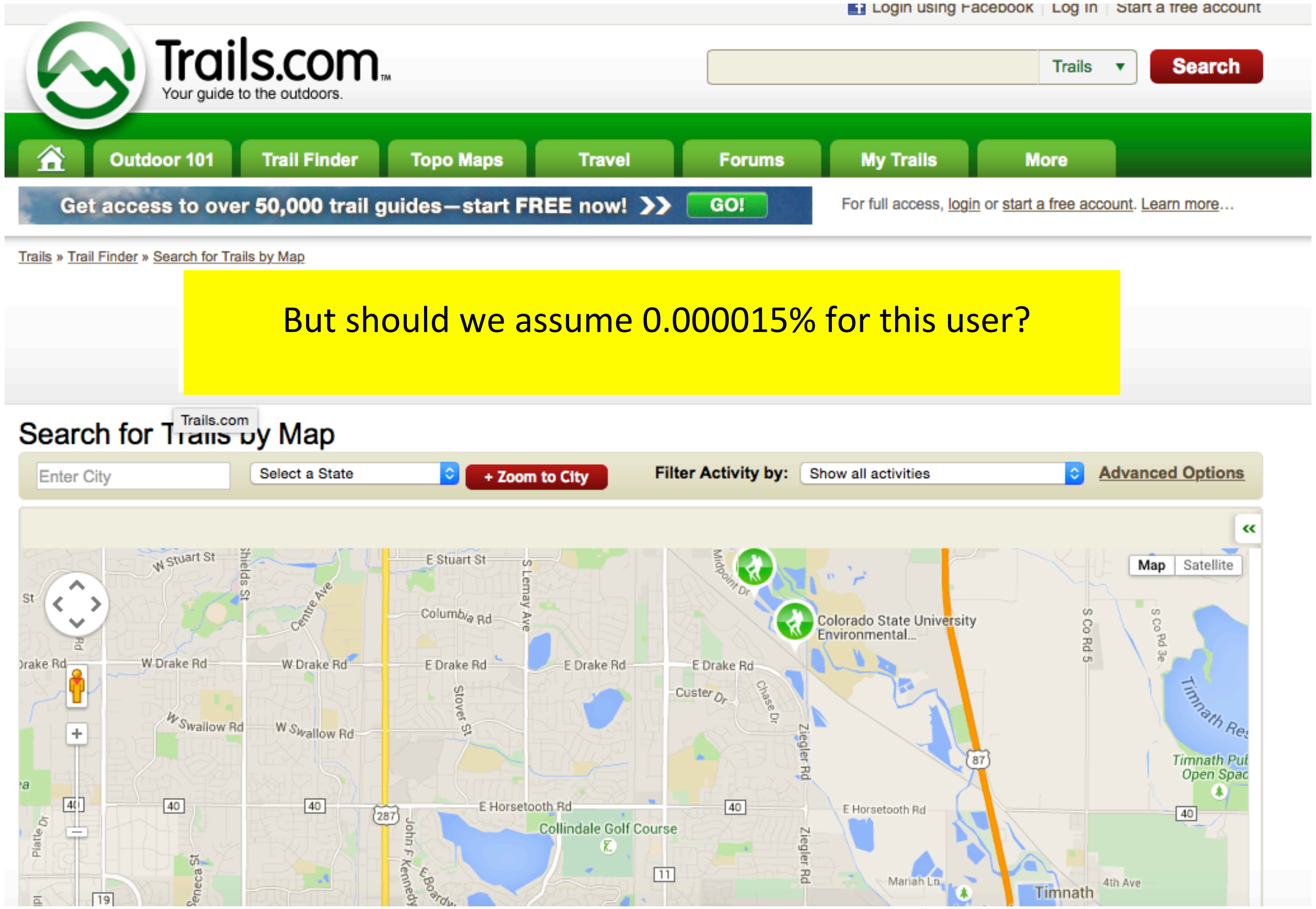
If a random user shows up on a webpage, what is the probability that he/she will buy from REI in the next 30 days?

APPLIED EXAMPLES - PURCHASE PREDICTION

If a random user shows up on a webpage, what is the probability that he/she will buy from REI in the next 30 days?

ANS: 0.000015%

APPLIED EXAMPLES - PURCHASE PREDICTION



But should we assume 0.000015% for this user?

APPLIED EXAMPLES - PURCHASE PREDICTION

Decision Tree Splitting Criteria - Information Value

$$WOE_i = ln\left(\frac{BuyPerc_i}{NoBuyPerc_i}\right) \; i = 1, 2, ..., n$$

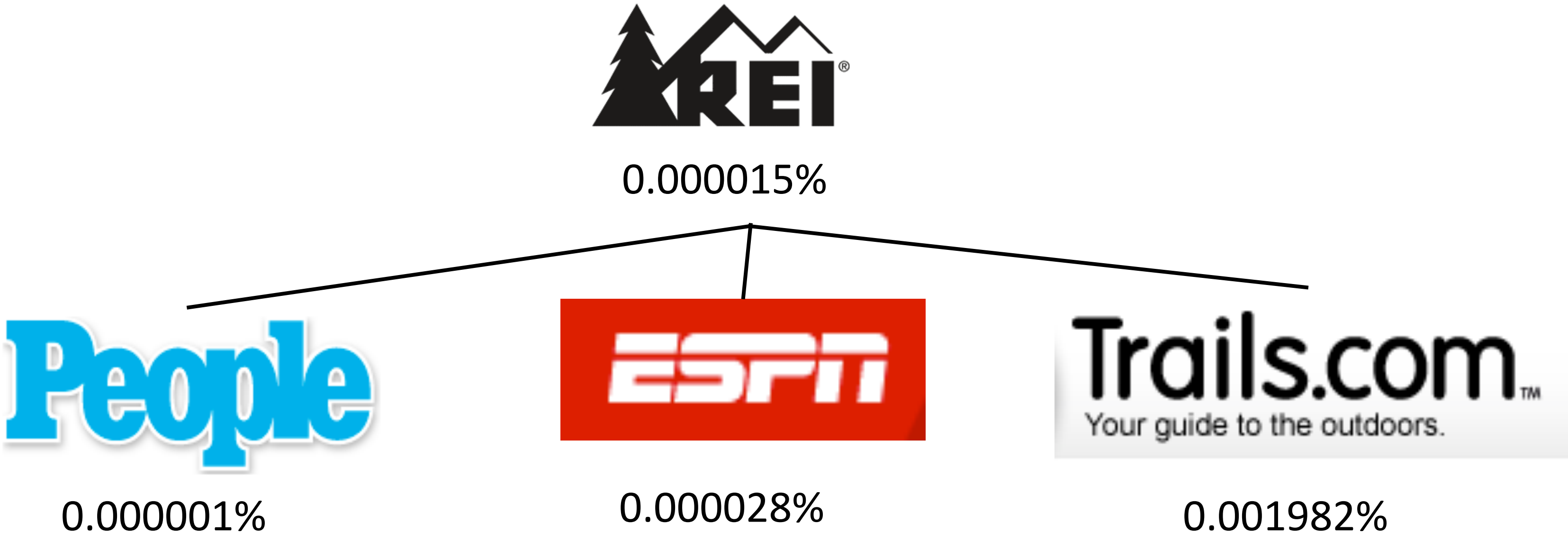
$$IV = \sum_{i=1}^{n}((BuyPerc_i - NoBuyPerc_i) * WOE_i)$$

URL	Perc_Buy	Perc_NoBuy	Woe	IV
people	4%	8%	-0.693	0.028
trails	7%	2%	1.253	0.063
espn	3%	2%	0.405	0.004
Other	60%	88%	-0.383	0.107
				0.202

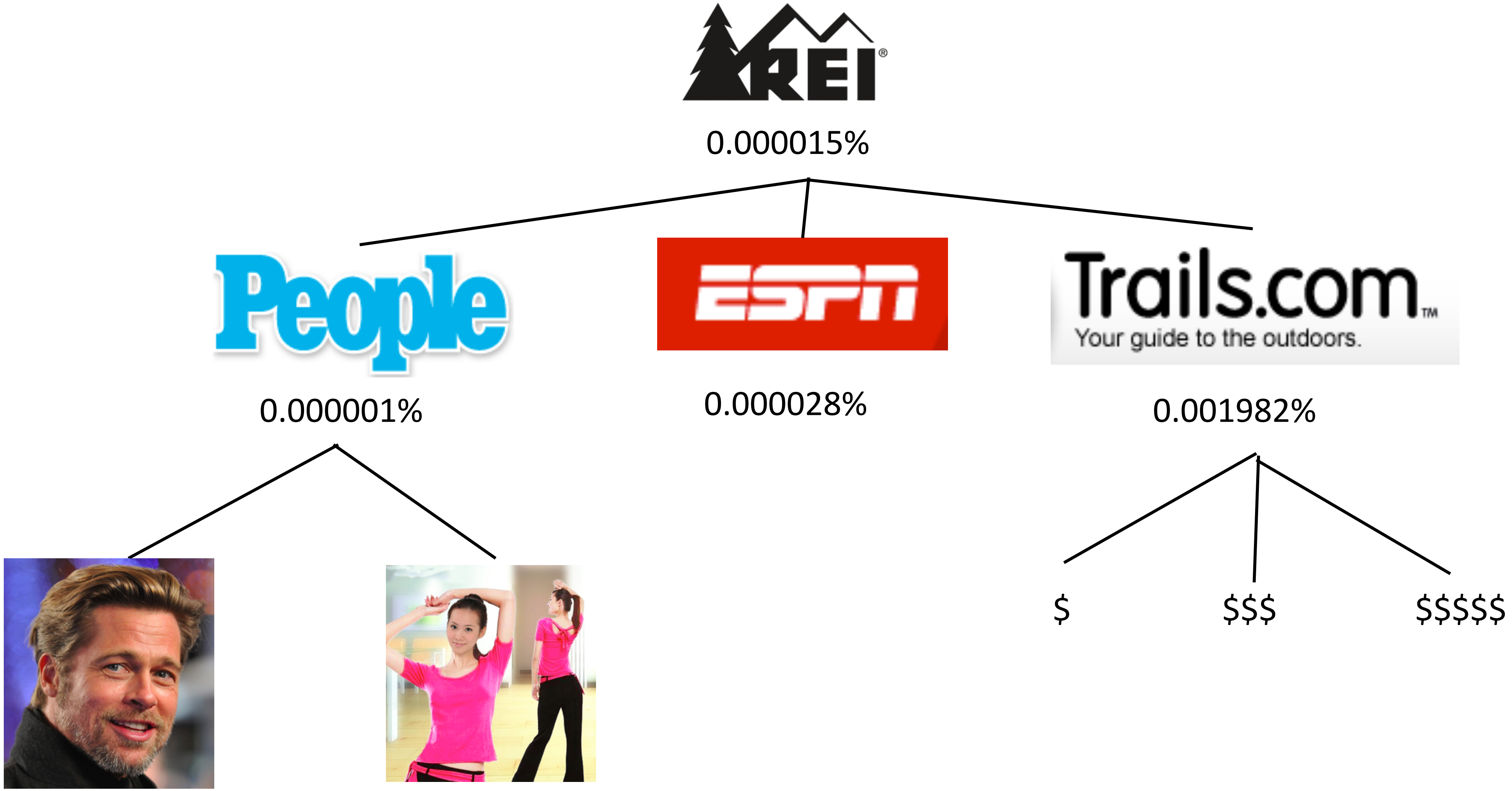
Income	Perc_Buy	Perc_NoBuy	Woe	IV
$	20%	35%	-0.560	0.084
$$$	60%	50%	0.182	0.018
$$$$$	20%	15%	0.288	0.014
				0.117

Calculate the IV for each explanatory variable and determine the best predictor.

APPLIED EXAMPLES - PURCHASE PREDICTION



APPLIED EXAMPLES - PURCHASE PREDICTION



AGENDA

- Quick Bio
- Ad:Tech 101
- DS Application – Propensity Ranking
- DS Application – Purchase Likelihood
- **DS Application – Clustering**
- Summary

APPLIED EXAMPLES - CLUSTERING

AGENDA

- Quick Bio
- Ad:Tech 101
- DS Application – Propensity Ranking
- DS Application – Purchase Likelihood
- DS Application – Clustering
- **Summary**

THANK YOU!

Mike.Schumacher@gmail.com

https://www.linkedin.com/in/analyticmike/