CompSci216 Spring '15 Everything Data

Global Tastes in Music

Global Superstars:
1. Sam Smith
2. Nicki Minaj
3. Ellie Goulding
4. Rihanna
5. Bruno Mars
6. Ed Sheeran
7. Calvin Harris
8. Ariana Grande
9. Maroon 5

World Spotify
Valid Cou

Language   Income   Crime

Yoon Choi, M

14

Quantifying the Rivalry

Team.txt: Arihant Jain, Jennie Ju, Arun Karottu, and Devin Solank

1

What can I play for you today?

Tell me about yourself.

How are you feeling?   Chill
What era are you feeling?   1950's

SUBMIT

SmartTunes
Music For How You Feel

Made with ♥ by Group 4

4

Team 13: #FoodPorn

- Problem: People spend more time on social media platforms than on Yelp. Are Twitter and Instagram viable restaurant review sources?
- Gathered relevant Twitter and Instagram posts and Yelp reviews on Austin restaurants.
- Ran sentiment analysis algorithm on posts to predict Yelp rating.
- Conclusions:
  - Yelp reviews are great predictors of Yelp ratings (87% explained variability).
  - Twitter and Instagram are very poor predictors (24% explained variability); especially lacking in negative dining experiences.

#FoodPorn

13

"Examining relationships between income, food access, and health"

Munger Beans:
Elizabeth, Janet, Julian, & Abigail

12

Summary: Health Hackers

Problem	Given a patient profile and a drug, predict the most likely adverse effects
Dataset	OpenFDA
Methodology	• Modified Naive Bayes prediction algorithm • Relative Likelihood Index • Cross-validation • plot.ly visualization tool
Case Study	Lipitor

11

Current News and Trending Chews:
The Effect of Current Events on Food Industry Trends

10

---

yelp 2.0

Group 9   Alice Ni
Arvind Viswanathan
Jimmy Fang
Leanna Zhan

WHERE TO EAT

Input one restaurant you **love**, we'll recommend **more**

- ✓ Ask user to select a **restaurant** of choice, with **distance range**
- ✓ Output **top 10 characteristics** of restaurant based on content of reviews
- ✓ User **indicates own preferences** of characteristics with sliding
- ✓ Recommend **other restaurants** with similar characteristics

9

FOOD Girls Generation

yelp

8

The Evolution of Music

- Processed song data from Million Song Dataset
  — Focused on Billboard Hot 100
- Processed lyrics from musiXmatch Dataset
  — TFIDF to measure saliency of lyrics

MILLION SONG DATASET

musiXmatch
not only words

Billboard HOT 100

7

College to NBA:
Predicting Performance and Salary

Group 6:
Jinzi, Matt, Victor, Vincent, Willy

6

Analyzing Presidential Campaign Data

■ $$$MONEY$$$ ■ TRUE VALUE

5

Statistics on COURSE EVALUATIONS

17,000 class sections have ratings available on SACES

These classes have been taught by over **2,000** instructors

75% of classes with available ratings have instructor ratings of 4.22 or above

Department Ratings:
Highest: Theater Studies, Russian, Dance, English
Lowest: Physics, Engineering, Mech. Engineering

Class Rating
Average: 4.34/5

Instructor Rating
Average: 4.46/5

Instructors can decide which course to have put into SACES but cannot select or deselect individual items.
Matt Serra
Director, A&S Office of Assessment

Duke (

3

---

yelp

College towns vs. Non-college Towns

Alvin Kang
Paul Kim
Grace Park
Scotty Shaw

2