

Choosing a Hospital and Improving Hospital Quality Care in California

Clients: Patients and Hospitals

Questions and Deliverables:

- 1) For Patients: How do we choose a hospital with the best available medical care?
We will suggest a hospital to patients based on patient location (per zip code) and patient required medical treatment using hospital Yelp reviews in California.
- 2) For Hospitals: How can we improve hospital quality care?
We will suggest to hospitals in what areas they need to make improvements based on medical professions, procedures, conditions and specialties associated with negative Yelp reviews in California.

Dataset: Collect patient hospital reviews from Yelp.com and national hospital ranking from US News (best hospitals in California <http://health.usnews.com/best-hospitals/area/ca/>).

Description of dataset: Reviews and rankings are textual data.

Approach:

- 1) Get hospital names, hospital national ratings, hospital specialties, procedures and conditions per hospital, number of doctors per specialty, hospital address in state of California from US News website.
- 2) Use sentence **sentiment classifier** to calculate hospital ratings from Yelp textual reviews.
- 3) Predict hospital ratings from Yelp reviews using **naïve Bayes**.
- 4) How US news hospital national ratings associate with patient Yelp hospital ratings? We can use **linear or logistic regression** to find associations. **Linear regression** can be used to find association between numerical national rankings and numerical sentiments. **Logistic regression** can be used to find association between numerical national rankings and categorical sentiments (positive or negative).
- 5) How positive and negative sentiments associate with hospital specialties, hospital procedures and conditions? Hospital procedures, conditions and specialties will be found from Yelp reviews and also will be taken from US news data. We can use **logistic regression** to find association between categorical sentiments and hospital procedures, conditions and specialties. In addition, we can do **clustering** to separate groups with different hospital procedures, conditions and specialties. What groups will have positive or negative sentiments?
- 6) How positive and negative sentiments associate with professions (clinicians, surgeons, physicians or nurses) who carried out medical treatments? Professions will be found from Yelp reviews. Number of doctors will be taken from US news data. We can use **logistic regression** to find association between categorical sentiments and medical professions. In addition, we can do **clustering** to separate groups with different medical professions. What groups will have positive or negative sentiments?