

MGMT 59000 AUD

# Craigslist Platform Improvement Project

---

## Cluster Centers -

- Pulkit Batra
- Esha Kaushal
- Laura Rossi
- Sachin Singhal
- Abhinav Reddy Vangala

# Background

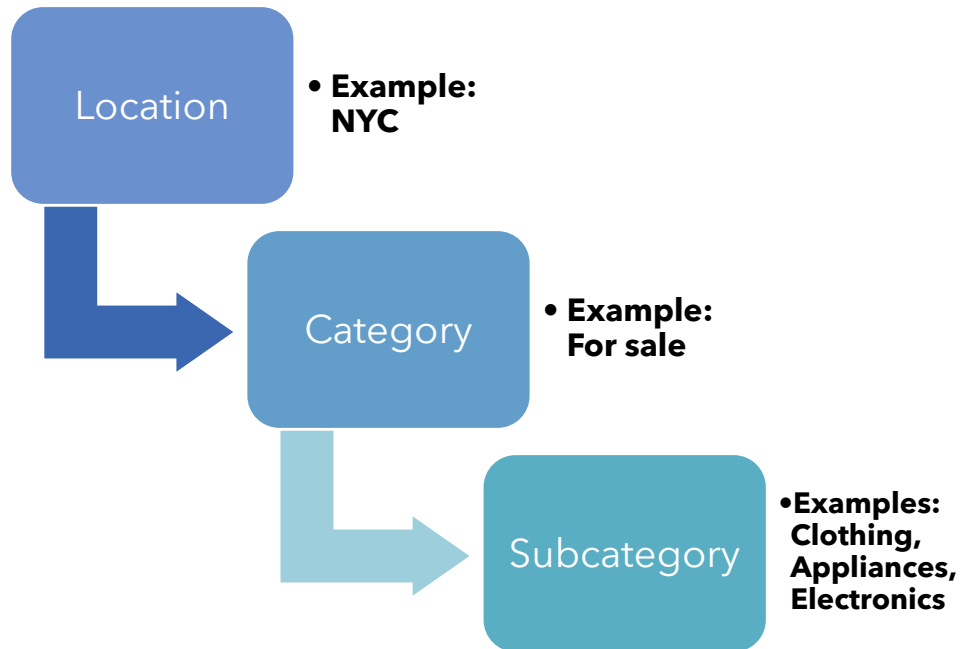
- **Craigslist** (stylized as craigslist) is an American **classified advertisements website**
- The craigslist platform is **advertiser oriented**, providing a simple, **low-structure** interface where advertisers can create their ads with few constraints.
- However, this low-structure creates **additional work for potential buyers** who have to parse through the content to find what they need.

# Business Analysis

## Current State

tippecanoe > for sale > appliances

### 3 tier classification system :



### Cons:

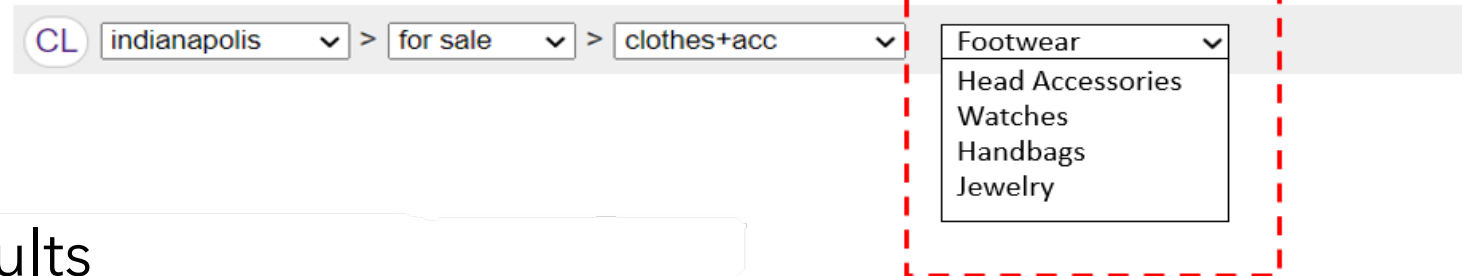
- ✗ Irrelevant results
- ✗ Inability to refine searches
- ✗ Not at par with competitors

# Business Analysis

## Our Solution

### 4<sup>th</sup> tier of **categorization**

- Users can **refine** search results
- **No additional effort** on behalf of advertisers
- Categories are always **relevant** to advertised products



# Data Analysis

1

Data Acquisition

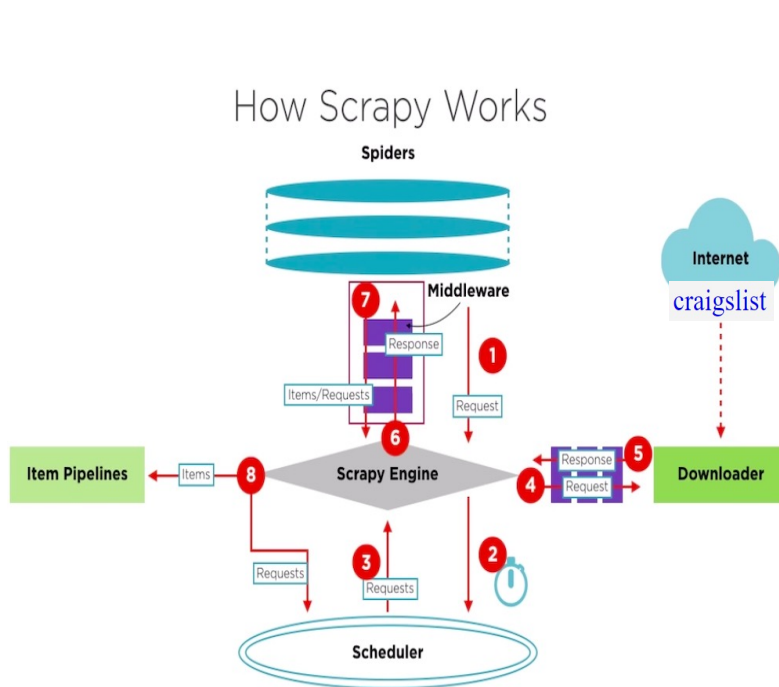
2

Data Cleaning

3

Topic Modeling

How Scrapy Works



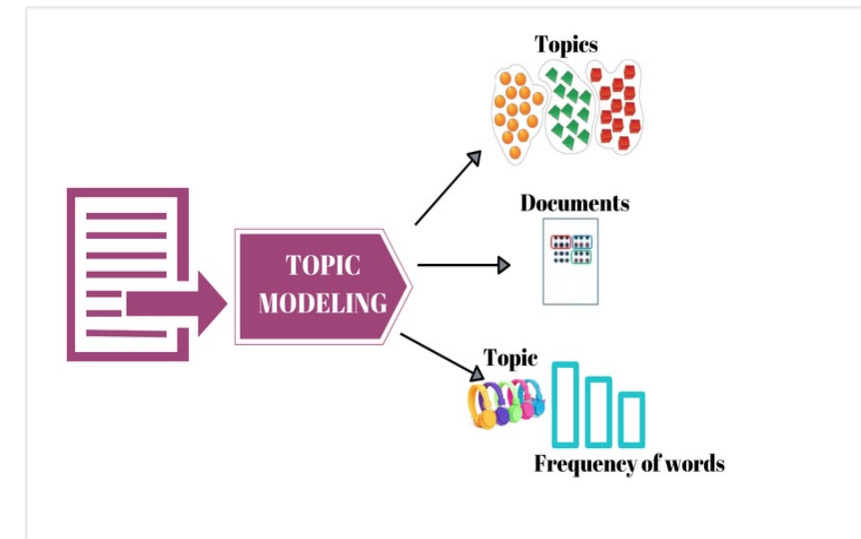
White space

Lemmatization

Punctuation

Capitalization

Vectorization

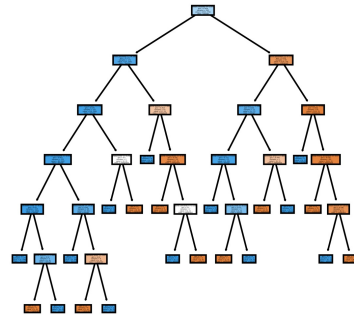
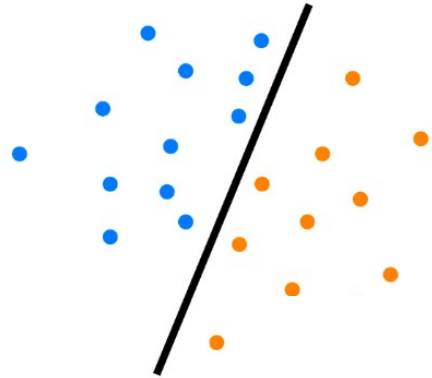


# Data Analysis

## 4 Topics to Categories

|    |   |   |   |   |     |
|----|---|---|---|---|-----|
| T1 | 1 | 2 | 3 | 4 | ... |
| T2 | 1 | 2 | 3 | 4 | ... |
| T3 | 1 | 2 | 3 | 4 | ... |
| T4 | 1 | 2 | 3 | 4 | ... |
| T5 | 1 | 2 | 3 | 4 | ... |
| T6 | 1 | 2 | 3 | 4 | ... |

## 5 Classification



## 6 Implementation

CL indianapolis > for sale > clothes+acc > Footwear  
Head Accessories  
Watches  
Handbags  
Jewelry

Our solution: Level 4 categorizing

# Classification Process

## Craigslist Ads title dataset

| S.No | Title  |
|------|--|
| 1    | Two Cutter & Buck men's shirts, size XL..... |
| 2    | Men's Wool Suit Pin Stripes, 33 wai....      |
| 3    | Selling a pair of shoes that....             |
| 4    | 3 Men's size 15 1/2 short sleeve.....        |
| 5    | Real snake Sandals.....                      |

**LDA**

### Topics

Shirt  
 Suit  
 Sleeve } Dress  
 Shoe  
 Sandal } Footwear

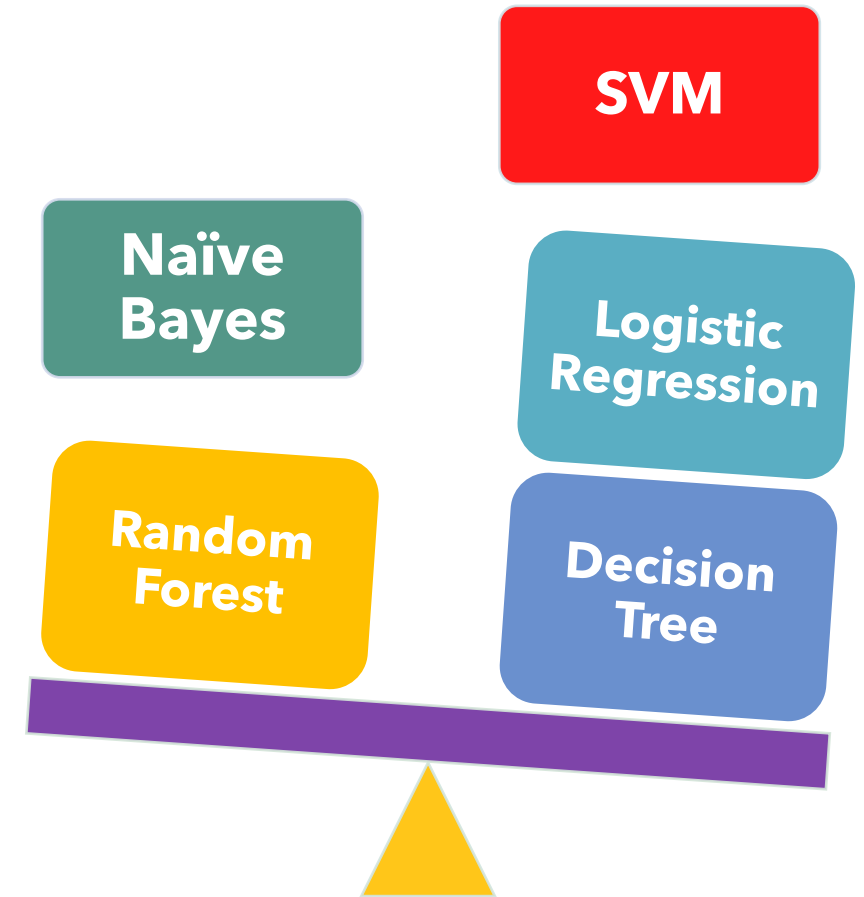
**Manual intervention to merge similar topics**

## Final Labelled Dataset

| S.No | Title   | Label    |
|------|---|----------|
| 1    | Two Cutter & Buck men's shirts, size XL.....  | Shirt    |
| 2    | Men's Wool Dress Suit Pin Stripes, 33 wai.... | Dress    |
| 3    | Selling a pair of shoes that....              | Footwear |
| 4    | 3 Men's size 15 1/2 short sleeve.....         | Dress    |
| 5    | Real snake Sandals.....                       | Footwear |

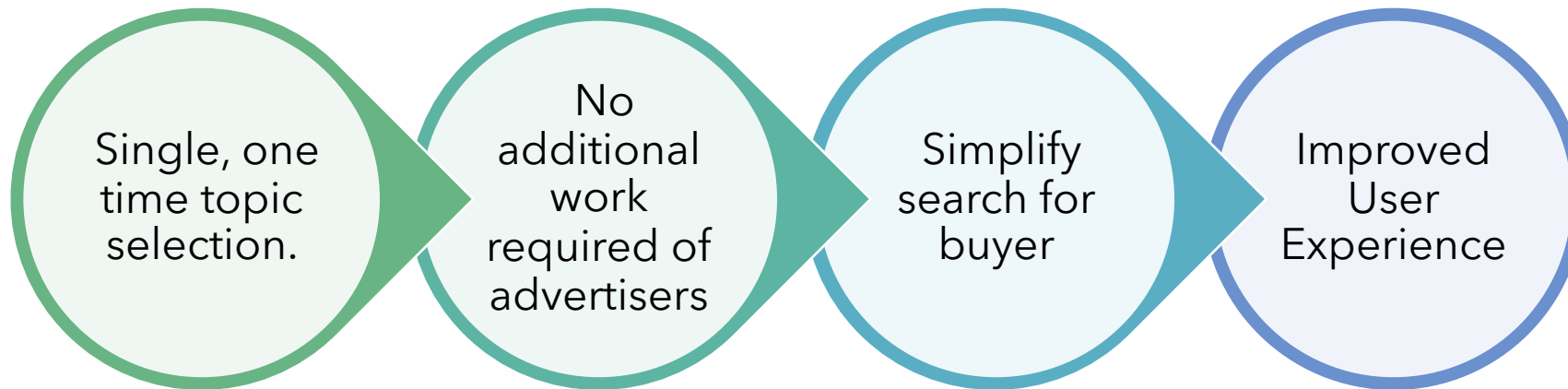
# Validation

|                        | Furniture<br>Classification | Appliances<br>Classification | Clothing and<br>Accessories<br>classification |
|------------------------|-----------------------------|------------------------------|---|
| SVM                    | 98.48%                      | 99.62%                       | 99.9%   |
| Logistic<br>Regression | 96.88%                      | 98.60%                       | 99.04%  |
| Decision Tree          | 94.48%                      | 97.96%                       | 95.49%  |
| Naïve Bayes            | 79.17%                      | 90.57%                       | 87.52%  |
| Random Forest          | 66.69%                      | 81.53%                       | 80.33%  |





# Value add for Business



# Conclusion



Simple algorithm provides easy classification



Some manual intervention required in selecting best topics



Room to grow into more categories





# Thank you

---