

2022

MGMT 59000

Analyzing Unstructured Data

TEAM: AUD Ones Out
MILESTONE 1



**ENHANCING THE USER EXPERIENCE OF
ELECTRONICS CATEGORY IN CRAIGSLIST**

Submitted By - Amrit Singh
Anusha Reddy
Deepa Narayanan
Mrinmoy Dalal
Sanchit Aggarwal
Vedanti Gulalkari

PROPOSAL DEVELOPMENT

Business Scenario:

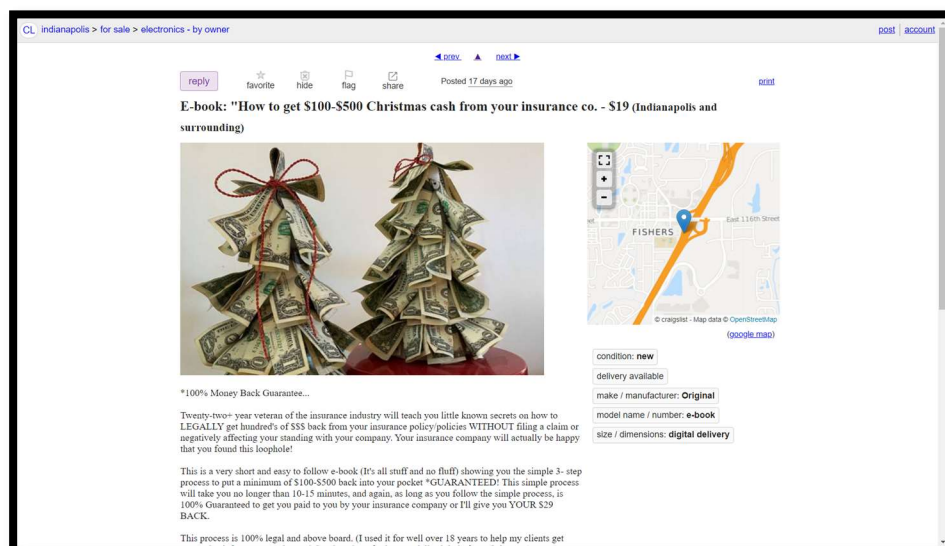
Attracting 250 million website visits every month as of March 2022, Craigslist is one of the most highly visited websites on the planet with about 75% of the visits originating from the US alone. Craigslist acts as one of the critical hubs for classifieds, enabling individuals and businesses to exchange goods and services on the internet. The website is currently hosting ~3500 listings under electronics category across Chicago and Indianapolis. However, the site often becomes prey to fraudsters levying the exposed vulnerabilities and their targets include highly expensive items like electronics and name-brand goods. The project aims to curb the rise fraudulent activities by analysing the listings of electronic goods focussed on these 2 regions and flag them out.

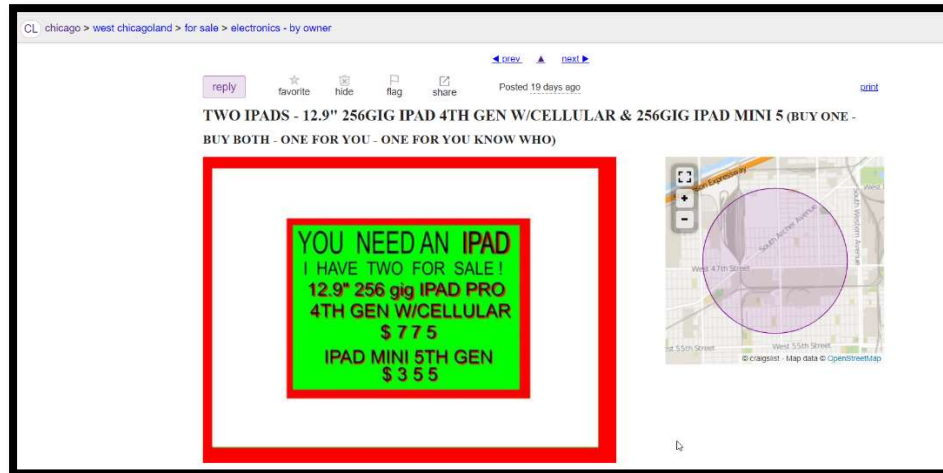
Problem Statement:

The project spans across 2 phases:

- **Improved classification for products under electronic category:** The ads popping in electronics category are highly unstructured making it difficult for user to search for a desired product, and a lack of efficient filtering mechanism to narrow down a search for finding a potential match of product offering for user
- **Spam Detection of fraudulent listing:** Presence of duplicate and misleading ads with the aim of deceiving the customer, that potentially require constant moderation

Example depicting fraudulent sale on Craigslist for Electronics category:





Proposed Feature Enhancement:

- Provide users the privilege to narrow down their search with filtering options via Topic Modelling
- Our tool uses ensemble of powerful Natural Language Processing and prediction models to find any type of misclassified adverts
- Spam detection by performing comparative analysis of the product on ecommerce website to recapitalize the customer reliability on the popular products and their trusted sellers

Description of data requirement and collection of the data

1) Data Requirement:

- Date Posted
- Product Name
- Post ID
- Description
- Region
- Topic Headline
- Product Reviews from competitive website

2) Data Collection:

- Scraping data from Craigslist Website – Electronics Category
- Scraping data from other comparable ecommerce websites in the market for comparison with competitor offerings
- Leverage pre-trained models