

# ENHANCING CRAIGSLIST GIGS WITH NLP

**Client:** Craigslist

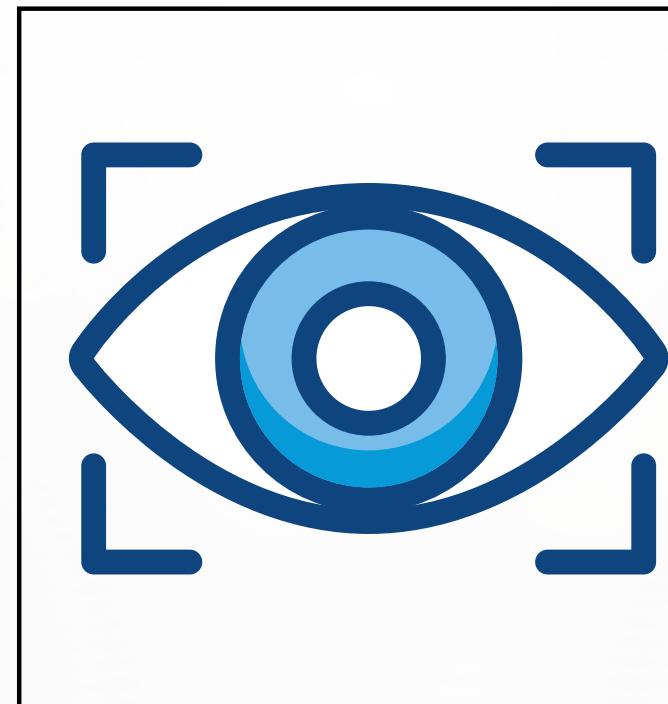
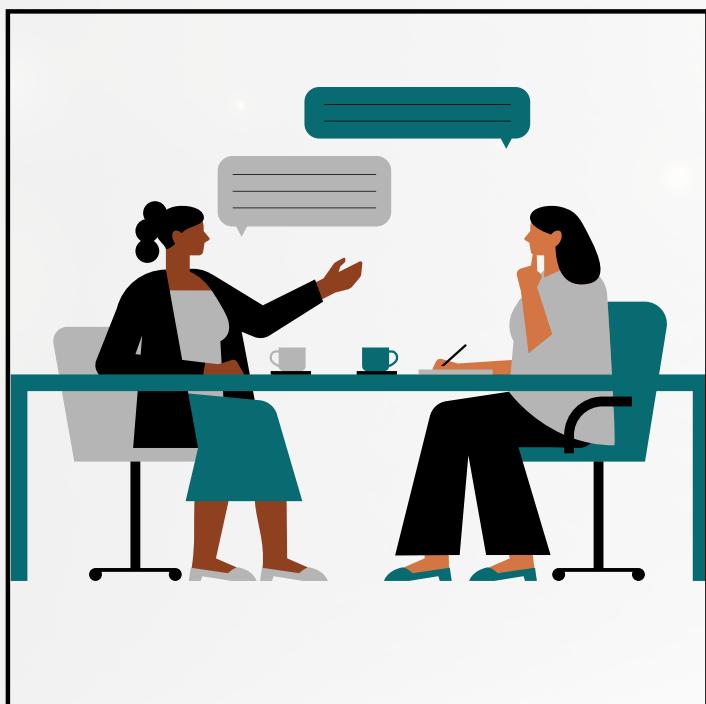
**Team:** Team 12

**Team Members:** Ashwa Ursani, Dipita Biswas, Krisha Vira,  
Parul Gupta, Tanjina Moon, Zohaib Jamil

# BACKGROUND

---

**Client:** Craigslist – a legacy platform for classifieds with millions of user-generated posts.



**Focus Area:** The Gigs section – home to freelance, part-time, and urgent jobs.

**Problem:**

- The Gigs section is text-heavy and unstructured.
- Users struggle to search by task type or emotional tone.
- Craigslist lacks tools for understanding, classifying, or organizing these posts.

# BUSINESS ANALYSIS

---



## **Objective:**

- Organize gig posts into thematic categories using unsupervised learning.
- Predict and label the sentiment of each post.
- Enable smarter navigation for users & automated moderation for managers.



## **Business Value:**

- Helps users find the right gigs faster.
- Flags potentially suspicious listings.
- Enhances Craigslist's platform intelligence without changing its open-posting model.

# WE USED NLP, CLUSTERING, AND CLASSIFICATION TO STRUCTURE GIG ADS

**Web Scraping -**  
Scraped 939 Boston gig posts  
Collected title, description,  
location

**Feature Engineering -**  
TF-IDF (1-2 grams, min\_df=3)  
Converted text to vectors

**Sentiment Labeling -**  
Used VADER to tag posts as  
Positive, Neutral, or Negative

**Data Cleaning -**  
Lowercased text  
Removed stop words, lemmatized

**Clustering -**  
KMeans (k=4)  
Identified 4 themes: Delivery,  
Events, Research, Labor

**Modeling & Evaluation -**  
Models: NB, LR, SVM, RF, DT

# DATA ANALYSIS

---

**Data Source:**

939 Craigslist Boston gig ads

**Descriptive Model:**

Applied TF-IDF (1–2 grams, min\_df=3)

Used KMeans clustering (k=4) to group similar gigs:

Delivery & Logistics

Research / Testing

Events & Promo

Manual Labor

**Predictive Model:**

VADER – label sentiment as Positive/ Neutral/ Negative

**Trained 5 classifiers:**

Naive Bayes,

Logistic Regression

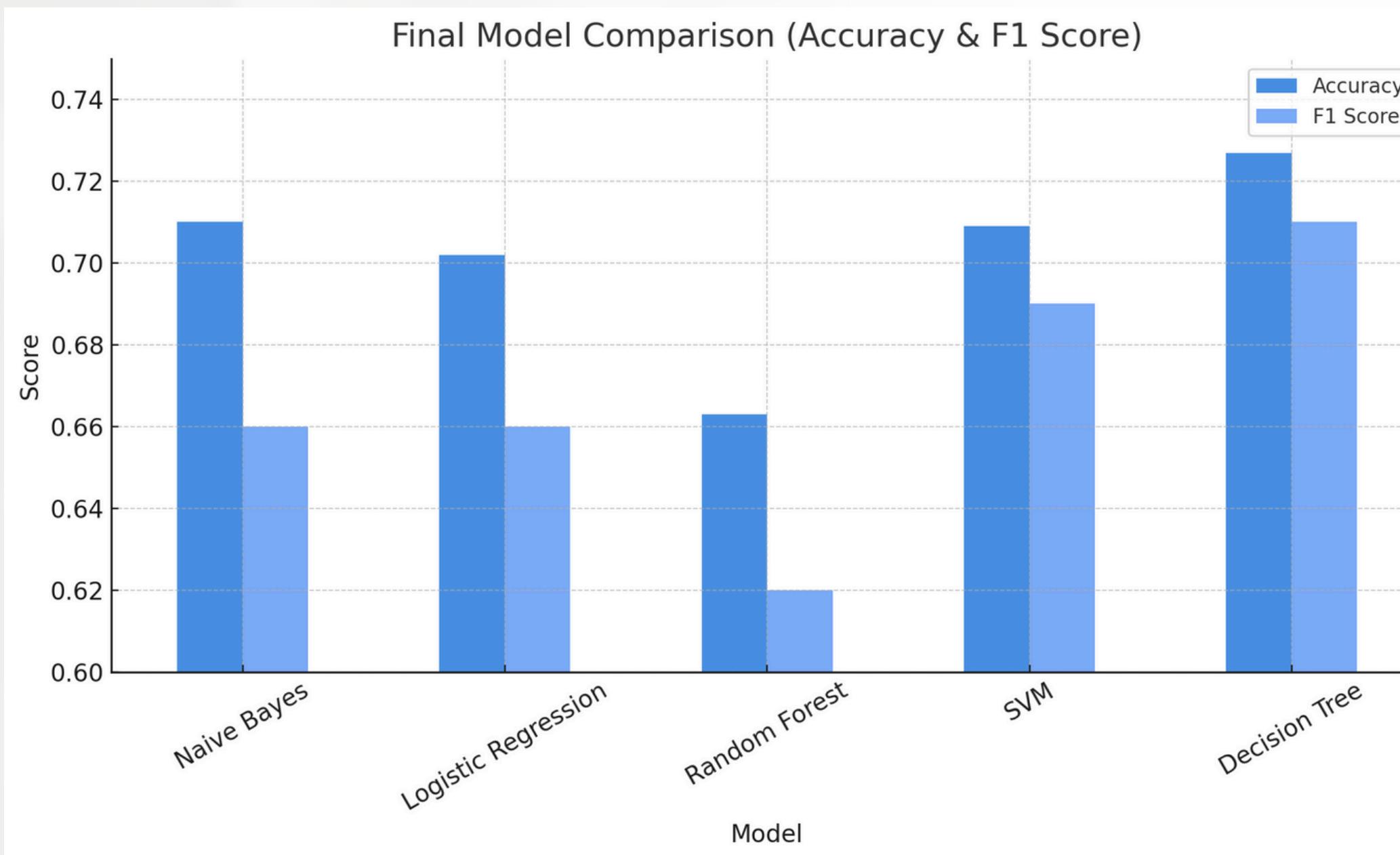
Random Forest

SVM

Decision Tree Model

# VALIDATION

---



- **Train-Test Split:** 70/30 (stratified)
- **Metrics:** Accuracy, F1-Score (Weighted Average)
- **Best Model:** Decision Tree (72.7%)
- SVM: 70.9%, Naive Bayes: 71.0%, Logistic Regression: 70.2%, Random Forest: 66.3%
- **Sentiment Distribution:**  
Positive: 62%, Neutral: 28%, Negative: 10%

# CONCLUSION

---

- Clustered 900+ gig posts into 4 meaningful categories
  - Predicted emotional tone using supervised models
  - Identified Decision Tree as the best-performing model (73% accuracy)
- 
- Improved user experience through smarter categorization
  - Enabled moderation teams to detect high-risk or emotional postings
  - Scalable solution without disrupting Craigslist's open-posting format
- 
- Extend classification to other categories (Jobs, Housing, Services)
  - Build filters based on gig category and emotional tone for end users
  - Explore deep learning models for potential performance boost

1

Key Wins

2

Business Values

3

Future Steps



# **THANK YOU**