

Problem Statement:

Classification of digital advertisements on the basis of its textual data

Dataset

About the dataset:

Input

Textual description of advertisements

Output

Categories

Multiclass

- Banking
- Rent – Apartment
- Sell – House
- Jobs – IT
- Jobs – Retail
- Gossip – Politics

Process

- Build a scraper
- Scrap data
- Label data using prodi.gy tool
- Text cleaning
- Text preprocessing
- Model building
- Model validation
- Build classification API

Team

Name	Programmer	Technology	
Kushal	Python Dev	AI	4

User Stories

As a User, I should be able to add text, so that the system could categorize it.

Use Case

User Enters Data

Input

2 rooms (1 big & 1 small) available for rent located near Everett CHA. Clouds to bus route 109, 110 and 104. Utilities and Wi-Fi included.

Price negotiable

Category

Rent – Apartment

Scrapping

Team

Name	Programming	
Siddhartha	Python	2
Jeevan	NoSQL Database	4
Anish	Python	2

Sites

Website		
https://www.sajha.com/		
https://us.sulekha.com/		
https://dallas.craigslist.org/		

Tool

Scrapy

Run it as a module within the Nelp Framework.

Labeling

- Progi.gy tool - Kushal
- ?

Team Lumbini

15 people

7 – Interns – Python Developers

Timeline

Siddhartha	Write a scrapper to read text data and export it to a CSV file	Three different websites	xpath, scrapy
Shushant	Look into the NLP		

Shushant	Research Algorithm		