



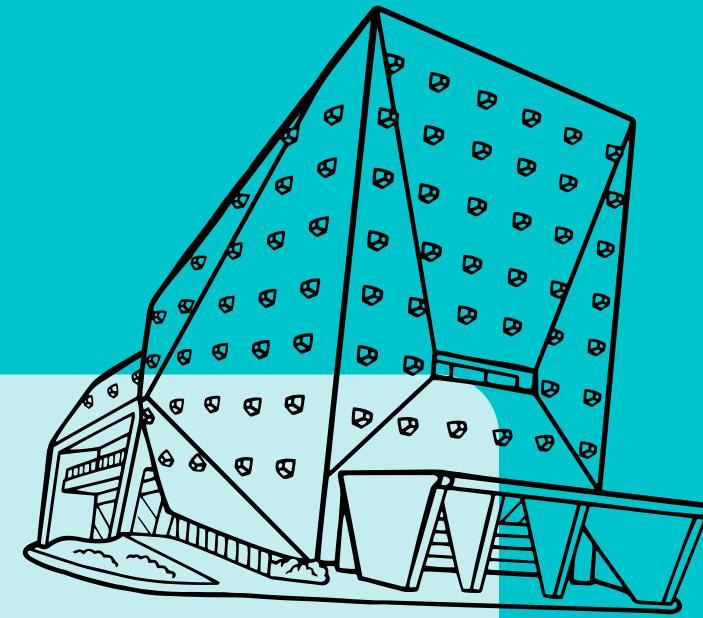
Mini Project NLP



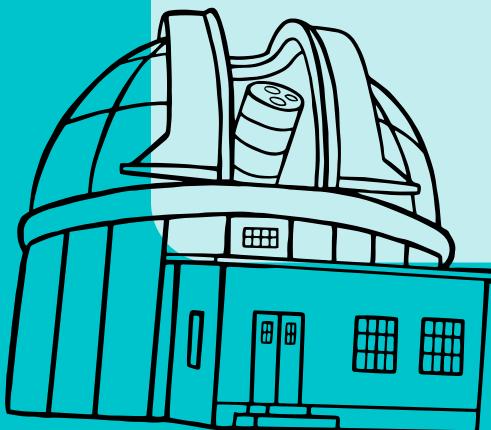
Chat Bot info.bdg

Wisnu Badrani

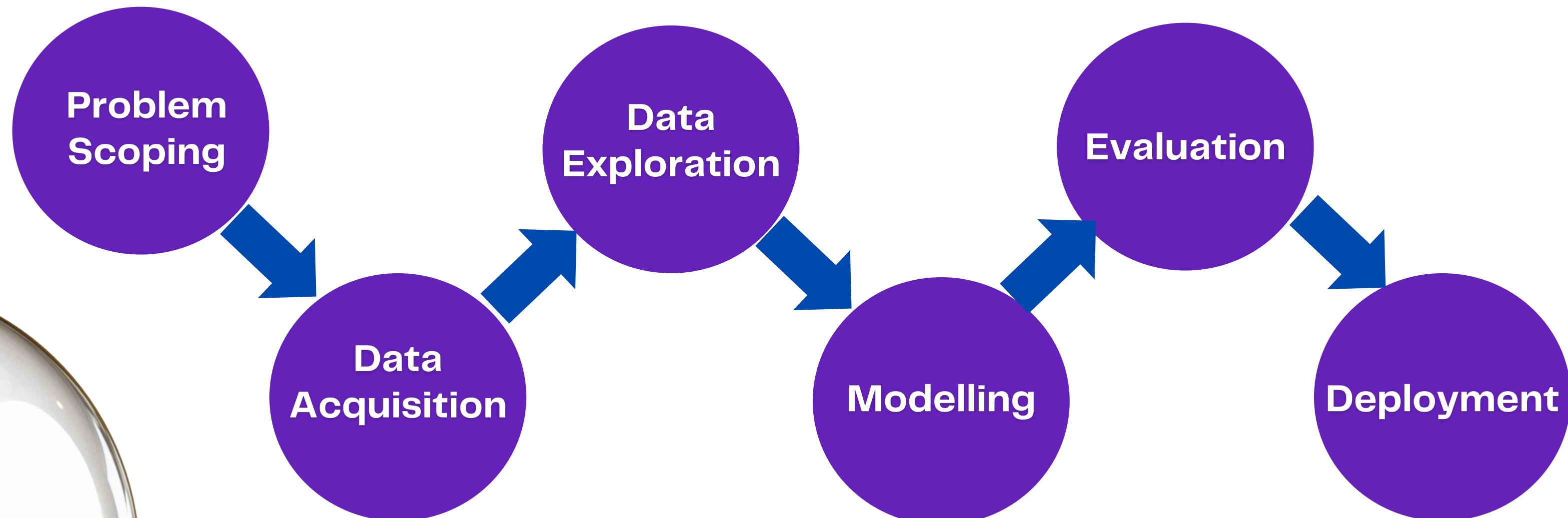
Introduction



Bandung merupakan salah satu kota terbesar di Indonesia. Bandung menjadi salah satu kota favorit sebagai destinasi wisata baik oleh wisatawan lokal maupun manca negara menurut VP of Commercial & Marketing Marketing Pegipegi pada Maret 2022. info.bdg merupakan sebuah aplikasi chatbot yang dibuat sebagai asisten digital yang dapat memahami serta memproses permintaan pengguna, dan memberikan jawaban yang relevan dengan cepat terkait dengan informasi mengenai Kota Bandung. info.bdg dibuat untuk menemani user serta memberi informasi dengan cepat terkait dengan informasi yang berkaitan dengan pemerintahan, pendidikan, kesehatan, transportasi, dan tempat rekreasi di Kota Bandung.



AI Project Cycle



Problem Scoping

Who?

User yang membutuhkan
infromasi terkait Kota
Bandung

Where?

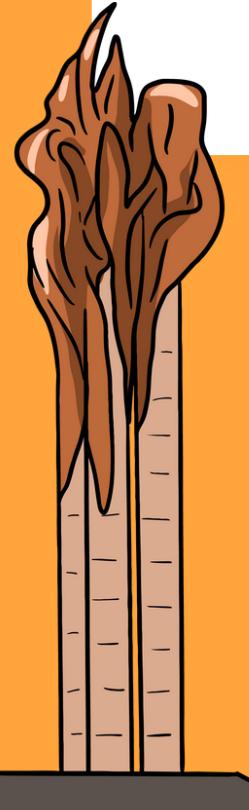
Seluruh wilayah Indonesia

What?

Membantu User
mendapatkan informasi
Kota Bandung dengan cepat

Why?

User membutuhkan
informasi dengan cepat
terakit informasi di Kota
Bandung



Data Acquisition

Metode yang digunakan untuk data acquisition pada project ini adalah dengan Web Scraping yaitu mengumpulkan informasi yang ada di dalam sebuah web. Web yang digunakan sebagai sumber data adalah website resmi Pemerintahan Kota Bandung : <https://www.bandung.go.id/> dengan data yang diambil adalah artikel berkaitan dengan sebagai berikut:



Transportasi



Pendidikan



Kesehatan



Pemerintahan



Tempat wisata

Data Exploration

Membuat Korpus

```
# Import Library
import bs4 as bs
import urllib.request

# Scraping data
bdg_data = urllib.request.urlopen('https://www.bandung.go.id/etalase/4/transportasi').read() #pemerintahan
bdg_data1 = urllib.request.urlopen('https://www.bandung.go.id/etalase/2/pendidikan-dan-kebudayaan').read() #pendidikan
bdg_data2 = urllib.request.urlopen('https://www.bandung.go.id/etalase/3/kesehatan').read() #kesehatan
bdg_data3 = urllib.request.urlopen('https://www.bandung.go.id/etalase/4/transportasi').read() #transportasi
bdg_data4 = urllib.request.urlopen('https://www.bandung.go.id/etalase/6/tempat-wisata-dan-rekreasi').read() #tempat rekreasi

# Menemukan keseluruhan paragraf html dari laman web
bdg_data_paragraphs = bs.BeautifulSoup(bdg_data, 'lxml').find_all('p')
bdg_data_paragraphs1 = bs.BeautifulSoup(bdg_data1, 'lxml').find_all('p')
bdg_data_paragraphs2 = bs.BeautifulSoup(bdg_data2, 'lxml').find_all('p')
bdg_data_paragraphs3 = bs.BeautifulSoup(bdg_data3, 'lxml').find_all('p')
bdg_data_paragraphs4 = bs.BeautifulSoup(bdg_data4, 'lxml').find_all('p')

# Membuat korpus dari keseluruhan paragraf artikel dari situs web
bdg_text = ''
# Membuat korpus teks huruf kecil pada paragraf
for p in bdg_data_paragraphs:
    bdg_text += p.text.lower()
for p in bdg_data_paragraphs1:
    bdg_text += p.text.lower()
for p in bdg_data_paragraphs2:
    bdg_text += p.text.lower()
for p in bdg_data_paragraphs3:
    bdg_text += p.text.lower()
for p in bdg_data_paragraphs4:
    bdg_text += p.text.lower()
```

Membersihkan teks

```
# Regular Expressions
import re

bdg_text = re.sub(r'\s+', ' ', re.sub(r'\[[0-9]*\]', ' ', bdg_text))
```

Tokenisasi

```
# Tokenisasi nltk
import nltk
# Unduh package punkt dalam nltk
nltk.download('punkt')

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
True
```

Modelling

Dalam project chatbot ini digunakan berbasis aturan (Rule Based Chatbot) dan menggunakan Algoritma sebagai berikut:

- TF-IDF (Term Frequency - Inverse Document Frequency) yang digunakan untuk ekstraksi fitur teks.
- Cosine Similarity untuk menentukan kemiripan teks yang akan mencari jawaban atau respon bot.

```
# TF-IDF Algorithm
from sklearn.feature_extraction.text import TfidfVectorizer
# Cosine Similartity
from sklearn.metrics.pairwise import cosine_similarity

def chatbot_response(user_query):
    # Menambahkan kueri ke sentences list atau daftar kalimat
    bdg_sentences.append(user_query)
    # Membuat basis vektor kalimat ke dalam list
    vectorizer = TfidfVectorizer()
    sentences_vectors = vectorizer.fit_transform(bdg_sentences)
    vector_values = cosine_similarity(sentences_vectors[-1], sentences_vectors)
    answer = bdg_sentences[vector_values.argsort()[0][-2]]
    input_check = vector_values.flatten()
    input_check.sort()

    if input_check[-2] == 0:
        return "Mohon maaf saya belum bisa menjawab"
    else:
        return answer
```

Evaluation

Chatbot Info.bdg karena merupakan *Rule Based Chatbot* maka mesin akan mengikuti intruksi aturan dengan masukan teks yang diinputkan oleh user. Chatbot ini memiliki kelemahan dimana tidak dapat menjawab pertanyaan yang kompleks.



Deployment



info.bdg nantinya akan deploy menjadi sebuah aplikasi berbasis web yang dapat digunakan oleh user baik menggunakan perangkat komputer ataupun smartphone

Output Program

Halo, selamat datang di Info Bandung. Ada pertanyaan terkait Kota Bandung?:

Apa julukan kota bandung?

info.bdg ?: selain dikenal sebagai paris van java dan kota kembang, bandung pun dijuluki the most european city in the east indies, bandung excelsior, intelectuele centrum van indie, europe in the tropen, kota permai, kota pendidikan, kota kreatif hingga kota kuliner.

Bagaimana fasilitas kesehatan di kota bandung?

info.bdg ?: sebagai ibu kota provinsi jawa barat, kota bandung memiliki sarana pelayanan kesehatan yang paling lengkap di provinsi ini.

Jumlah unit kesehatan di kota bandung?

info.bdg ?: sampai tahun 2007, kota bandung telah memiliki 30 unit rumah sakit dan 70 unit puskesmas yang tersebar di kota ini, di mana dari 17 unit rumah sakit

apa transportasi dalam kota yang biasa digunakan di kota bandung?

info.bdg ?: moda transportasi tersebut dilayani oleh angkutan kota (angkot), bus kota (damri), trans metro bandung (tmb), maupun kereta rel diesel (krd) angkutan kota dan bus kota untuk transportasi di dalam kota, masyarakat bandung biasanya menggunakan angkutan kota atau yang lebih akrab disebut angkot.

