PROJECT

Problem statement

↓

Data extraction

↓

EDA

↓

Preprocessing

↓

What if unlabelled data?

$\downarrow$

Multiple language data

$\downarrow$

Detect language

Translate

$\downarrow$

Clustering $\rightarrow$ Count vectorizer

$\rightarrow$ TFIDF

$\rightarrow$ Word2Vec

$\downarrow$

Evaluate $\rightarrow$ Silhouette Score

$\rightarrow$ Silhouette visualizer

$\downarrow$

Training data

$\downarrow$

Root cause analysis

Problem statement :

"We were trying to build a model that will find out if customers are happy or not with the insurance product. If not then what are the reasons behind it."

combination of 2 things

① Sentiment analysis   ② Root cause analysis

" We are having client with chain of retailers (rsc: Lifestyle, shoppers stop). They observed that number of customers are visiting their store but very less no. of people buying products from them."

Reasons → Rate↑

→ Bad customer service

Domain $\longrightarrow$ Bank Services $\rightarrow$ credit card

$\rightarrow$ internet services

$\rightarrow$ ATM services

Human Resources domain (HR)

$\rightarrow$ Reasons of employee churning

Retail industry

Hospitality domain

Health domain

Supply chain management

Telecom domain

"We had a customer they were famous for their main course initially. Recently they added snacks and desserts so they found out that sales of main course has been drastically decreased."

Two types of Projects,

In house Project   Client based Project

Client $\longrightarrow$ (Problem)

$\downarrow$

Business analyst

$\downarrow$

Manager

$\downarrow$

Team lead

$\downarrow$

1, 2, 3, - - - .

Problem statement

↓

Architecture

↓

Present

↓

Client Satisfied

↓

Ask for specific data

↓

Analyse the data

↓

Decide flow of Project

Architecture

```
┌─────────────┐          ┌──────────────┐
│ Data Source │   ───►   │  NLP engine  │   ───►   Modelling          {Reason - Depart-
└─────────────┘          └──────────────┘              │                    - ment}
      ▲                                                 ▼                      │
      │                                               Reasons  ───►  Static   ▼
      │  +                                               │           mapping
      │                                                  ▼              │
      │                                            Notification         ▼
      │                                            will trigger  ◄───  Script
     DB    ◄───  Sharepoint  ◄───                  to respected
                 access                            team / department
                    │
                    ▼
              { Review - Reason }
```

Problems

1. Time consuming process
2. lack of communication
3. Vehicle repair shop
4. App / Website

Data

① Quality

② Quantity

③ Not continuous flow of data

④ Exact data not available for our use case

$\downarrow$

Public options

$\downarrow$

Feedback form / google form / voice calls

Q What problems you faced while extracting data?
→ Private data / Secure data

Q what is difference between structured and unstructured data?

Q. Quantity of data → > 1 GB

→ 10 lacs ⇒ Rows

Data $\rightarrow$ cycles

    $\rightarrow$ ✓ Monthly $\longrightarrow$ $\underbrace{1^{\checkmark} \quad 2^{\times}}$ $\underbrace{3^{\checkmark} \quad 4^{\times}}$

    $\longrightarrow$ Quarterly

    $\longrightarrow$ Yearly

$$\Downarrow$$

Data

$$\Downarrow$$

Initial Preprocessing
- Language detection
- Language translation

$\downarrow$

EDA $\longrightarrow$ Ngrams

$\longrightarrow$ Word cloud

$\rightarrow$ Keyphrase extraction