ACM Projects
Team White Bread

Meet the Team!

Sia (mentor)   **Keerthi (mentor)**   **Kiran**   **Hikaru**
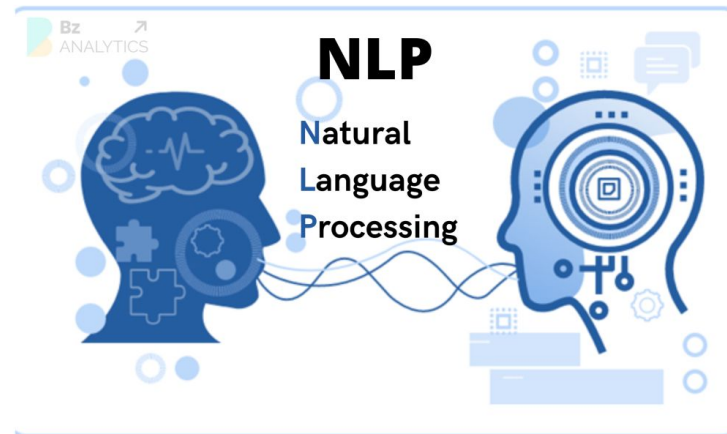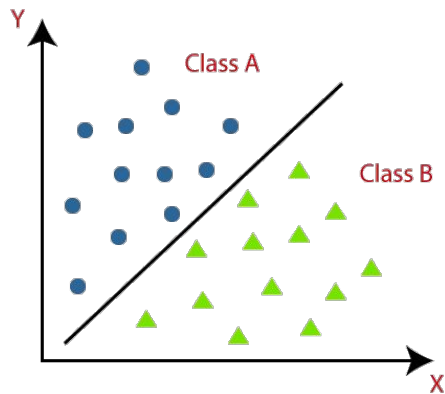
Nicole   **Rahul**   **Jocelyn**   **Rishab**

acm

Brainstorming

SENTIMENT ANALYSIS

POSITIVE

"Great service for an affordable price.
We will definitely be booking again."

NEUTRAL

"Just booked two nights at this hotel."

NEGATIVE

"Horrible services. The room was dirty and unpleasant.
Not worth the money."

Inspiration & Background

Inspiration & Background

Our dataset

51 New Notebook ⬇ Download (847 kB) ▾ ⋮

Customer Support Ticket Dataset

Customer support ticket dataset.

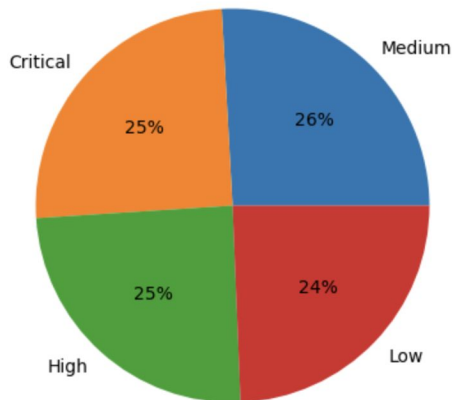| STT	Attribute Name	Meaning
1	Ticket ID	A unique identifier for each ticket
2	Customer Name	The name of the customer who raised the ticket
3	Customer Email	The email address of the customer (Domain name - @example.com is intentional for user data privacy concern)
4	Customer Age	The age of the customer
5	Customer Gender	The gender of the customer
6	Product Purchased	The tech product purchased by the customer
7	Date of Purchase	The date when the product was purchased
8	Ticket Type	The type of ticket (e.g., technical issue, billing inquiry, product inquiry)
9	Ticket Subject	The subject/topic of the ticket
10	Ticket Description	The description of the customer's issue or inquiry
11	Ticket Status	The status of the ticket (e.g., open, closed, pending customer response)
12	Resolution	The resolution or solution provided for closed tickets
13	Ticket Priority	The priority level assigned to the ticket (e.g., low, medium, high, critical)
14	Ticket Channel	The channel through which the ticket was raised (e.g., email, phone, chat, social media)
15	First Response Time	The time taken to provide the first response to the customer
16	Time to Resolution	The time taken to resolve the ticket
17	Customer Satisfaction Rating	The customer's satisfaction rating for closed tickets (on a scale of 1 to 5)

acm

Exploratory data analysis (EDA)

```python
# EDA  of Ticket Priority column
count_labels = df['Ticket Priority'].value_counts()

plt.figure(figsize = (10,5))
plt.pie(count_labels, labels = count_labels.index, autopct='%.0f%%')
plt.title('Distribution of priority labels', loc = 'left', pad = 10, size = 15)
plt.show()
print("The four types of priority labels are: " + str(df['Ticket Priority'].unique()))
```

```python
# EDA of Ticket Channel column
count_ticket_channel = df['Ticket Channel'].value_counts()

plt.figure(figsize = (10,5))
plt.pie(count_ticket_channel, labels = count_ticket_channel.index, autopct='%.0f%%')
plt.title('Distribution of Ticket Channels', loc = 'left', pad = 10, size = 15)
plt.show()

print("Out of " + str(df.shape[0]) + " tickets,")
print(str(df[df['Ticket Channel'] == 'Chat'].shape[0]) + " tickets are through the chat")
```


Distribution of priority labels

Distribution of Ticket Channels

```
The four types of priority labels are: ['Critical' 'Low' 'High' 'Medium']
```
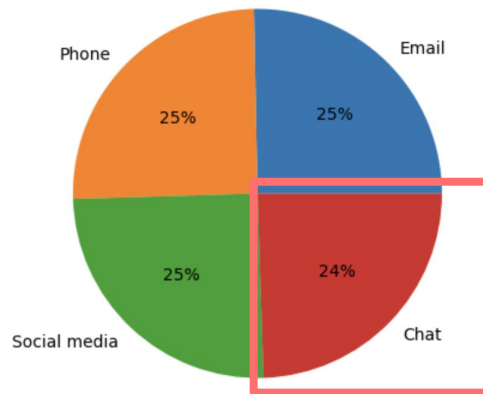
```
Out of 8469 tickets,
2073 tickets are through the chat
```

Data Cleaning

```python
# First 5 rows of relavent rows
cleaned_df = df[df['Ticket Channel'] == 'Chat'][['Ticket Description', 'Ticket Priority']]
cleaned_df
```

	Ticket Description	Ticket Priority
1	I'm having an issue with the {product_purchase...	Critical
11	I'm having an issue with the {product_purchase...	High
12	I'm having an issue with the {product_purchase...	Low
13	I've recently set up my {product_purchased}, b...	Low
14	I'm having an issue with the {product_purchase...	High
...	...	...
8448	I've recently set up my {product_purchased}, b...	Medium
8450	I'm having an issue with the {product_purchase...	Low
8453	The {product_purchased} is unable to establish...	High
8457	I'm having an issue with the {product_purchase...	High
8458	I'm having an issue with the {product_purchase...	Medium

2073 rows × 2 columns

"I'm having an issue with the {product_purchased}. Please assist."

"I'm having an issue with the {product_purchased}. Please assist.\n\nIf you need to change an existing product.\n\nI'm having an issue with the {product_purchased}. Please assist.\n\nIf The issue I'm facing is intermittent. Sometimes it works fine, but other times it acts up unexpectedly."

Data Cleaning

```python
def remove_filler(text):
    return text.replace("I'm having an issue with the {product_purchased}. Please assist.", "")
cleaned_df['Ticket Description'] = cleaned_df['Ticket Description'].apply(remove_filler)
cleaned_df
```

	Ticket Description	Ticket Priority
1	if you need to change an existing product if t...	Critical
11	it is possible that we cannot find some type...	High
12	cqw why didn t i send him the invoice thanks ...	Low
13	i ve recently set up my product purchased but...	Low
14	product name date oct posts i ve recen...	High
...	...	...
8448	i ve recently set up my product purchased but...	Medium
8450	i ve performed a factory reset on my product ...	Low
8453	the product purchased is unable to establish a...	High
8457	for the and shipping please click on "purc...	High
8458	i m having an issue with the product purchas...	Medium

2073 rows × 2 columns

- Cleaning using pandas

- Apply function to column

- Removes the auto-filled sentence

"\n\nIf you need to change an existing product.\n\n\n\nIf The issue I'm facing is intermittent. Sometimes it works fine, but other times it acts up unexpectedly."

Data Cleaning

```python
#Ticket Description cleaning
def clean_text(text):
    text = str(text).lower()
    text = re.sub('[({})?/$#|=]', '', text)
    text = re.sub('\n', '', text)
    text = re.sub('[.]', ' ', text)
    text = re.sub('[->]', ' ', text)
    text = re.sub('[:]', ' ', text)
    text = re.sub('[_]', ' ', text)
    text = re.sub('[,]', ' ', text)
    text = re.sub('[-]', ' ', text)
    text = re.sub('[\']', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('\w*\d\w*', '', text)
    return text

cleaned_df['Ticket Description'] = cleaned_df['Ticket Description'].apply(clean_text)
cleaned_df.head()
```

	Ticket Description	Ticket Priority
1	if you need to change an existing product if t...	Critical
11	it is possible that we cannot find some type...	High
12	cqw why didn t i send him the invoice thanks ...	Low
13	i ve recently set up my product purchased but...	Low
14	product name date oct posts i ve recen...	High

- Cleaning using regex

- Apply function to column

- Removes unnecessary symbols, digits, text, etc.

'if you need to change an existing product if the issue im facing is intermittent sometimes it works fine but other times it acts up unexpectedly '

Data Cleaning

"I'm having an issue with the {product_purchased}. Please assist.\n\nIf you need to change an existing product.\n\nI'm having an issue with the {product_purchased}. Please assist.\n\nIf The issue I'm facing is intermittent. Sometimes it works fine, but other times it acts up unexpectedly."

'if you need to change an existing product if the issue im facing is intermittent sometimes it works fine but other times it acts up unexpectedly '

Features

What features did you want to implement and what have you achieved?

✓ Train model using kaggle dataset

○ Test model using IT department tickets

○ Create website/app that is able to test other ticket datasets

Technical Process

- Brainstorming document to outline ideas, potential challenges, and our final goal

- Planning document to list models, methods, datasets

- Used GitHub to manage meetings notes and updates
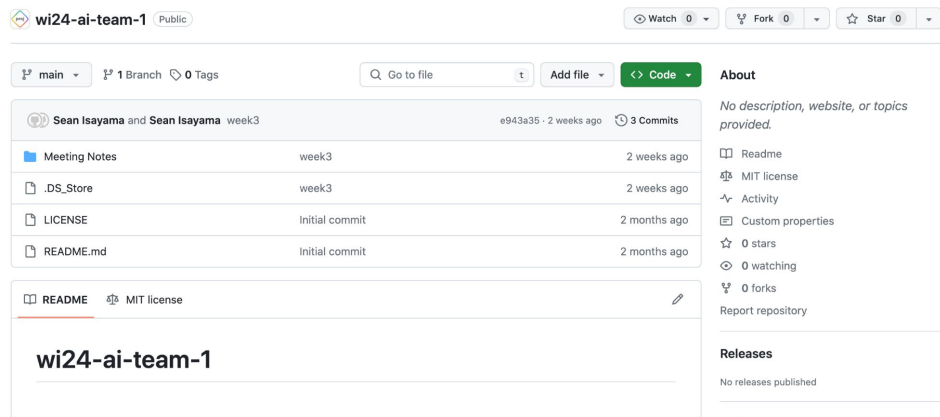
Look at our tech stack !!

Challenges

Coordinating meeting times — Each member has their own schedule, so finding a meeting time that works for everyone is difficult

Dataset issues — At first, our first dataset did not have a strong correlation between the text and its urgency rating

Model