Gujarat Technological University

NEW L. J. INSTITUTE OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY

ACADEMIC YEAR – 2023-2024

Summer Internship (3170001)

Data Analytics and Machine Learning

PREPARED BY,

STUDENT NAME:
Bokil Neel Milind
ENROLLMENT NUMBER:201430116080

GUIDED BY,
GUIDE NAME: Ms. Roshni Mandli

HOD NAME: Dr. Gayatri Pandi
CSE(AIML)&IT,
NEW LJIET.

Outline

1. Objective
2. Introduction
3. Role and Responsibilities During Internship
4. Daily Task
5. Internship Work
6. Conclusion
7. Future Enhancement
8. Completion Certificate

1. Objective

- To acquire new skills, knowledge and practical experience in the field of data analysis and machine learning.
- To learn to manage time effectively, meet deadlines and prioritize tasks in a professional environment.
- To develop the ability to solve the problems that occur during the internship.
- To gain real-world, hands-on experience related to Data analysis and Machine Learning.
- To gain a better understanding of the company’s culture, values, operations and goals.
- To build a professional relationship with colleagues, supervisors and mentors that can provide guidance, advice and potential future job opportunities.

2. Introduction

- Data analytics is the process of deriving insights about the data and helps us to make decisions based on that data.
- It makes use of various techniques to collect, clean, and transform data into a format that can be analyzed.
- Various statistical and computational techniques are then applied to discover patterns, trends, correlations, and anomalies within the data.
- Data analytics is widely used across industries to optimize operations and improve strategies.
- Machine Learning (ML) is a branch of AI, which enables the machine i.e. the system to learn and improve from experience.
- It is widely used to create algorithms which can be used to analyze and identify patterns in data which then can be used to make predictions and decisions without the need of the human interference.
- Machine Learning models are trained on old data and is used to make predictions on the new data.
- This technology is utilized in many applications such as healthcare diagnosis, image recognition, autonomous vehicles and much more.

3. Role and Responsibilities During Internship

- To collect data and clean the data obtained from various sources.
- To perform exploratory data analysis to uncover insights and understand the data in a better manner.
- Create data visualization in order to represent the data using various charts, graphs and other visual elements so that the data can be understood in a more proper way.
- Applying basic statistical analysis to draw necessary conclusions related to the data.
- Prepare and preprocess the data for the machine learning models.
- Develop and examine the data with machine learning algorithms.
- Train the model, optimize the model and test whether the model is making the accurate predictions or not. Check the accuracy of the model.
- Collaborate and communicate with the supervisor, mentor and colleagues during the internship to improve your overall performance.

4. Daily Task

Days	Date	Activities
Day 1	27/07/23	• Introduction about the Internship and Data Science • Basics of Data Analytics • Types of Data • Understanding Data Cleaning and Preprocessing • Introduction to Dictionary • Concept of API and Requests Package

Days	Date	Activities
Day 2	28/07/23	• API Handling: 1. COVID API 2. ISRO API 3. Bitcoin API 4. Mutual Fund API • API Search
Day 3	31/08/23	• Data Visualization using Matplotlib library • Types of graphs: Bar graph, Multiple Line chart, Multiple Bar graph, Pie chart
Day 4	01/08/23	• Introduction to Data Visualization • API Data Visualization • Graphs plotted on API Data

Days	Date	Activities
Day 5	02/08/23	• Assignment -1 • News API • ISRO Spacecrafts API • ISRO Customers API • Dynamic API – PINCODE API
Day 6	03/08/23	• Introduction to Pandas library • Introduction to NumPy library • How are these libraries useful in Data Science? • Pandas: Data Frames • Pandas: Excel • Pandas: API to CSV • XLRD: IPL Data Analysis
Day 7	04/08/23	• Introduction to Machine Learning • Introduction to Linear Regression • Linear Model: Mathematics • Linear Model: Implementation

Days	Date	Activities
Day 8	07/08/23	• Multiple Linear Regression • Multiple Linear Model: Mathematics • Multiple Linear Model: Implementation
Day 9	08/08/23	• Polynomial Linear Regression • Polynomial Linear Regression model: Implementation • Image Data
Day 10	09/08/23	• Assignment – 2 • OpenCV • Convolutional Neural Network Project
Day 11	10/08/23	• Conclusion and words of appreciation from the supervisor

5. Internship Work

❖ INSHORTS NEWS API

API link: https://inshortsapi.vercel.app/news?category=all

1) How many main keys are there in this API? Extract and print all keys.

In [1]:
```
import requests
url = requests.get("https://inshortsapi.vercel.app/news?category=all")
inshorts_data = url.json()
```

In [2]:
```
count=0
for i in inshorts_data:
    count=count+1
print("The total number of main keys in this API are : ",count)
print("All the main keys are as follows : ")
for i in inshorts_data:
    print(i)
```

```
The total number of main keys in this API are :  3
All the main keys are as follows :
category
data
success
```

2) How many news are available in this API?

In [3]:
```
x = len(inshorts_data["data"])
print(x)
```

```
9
```

3) Print all the news in below format:

News content, Author: Author Name, Date: Date of news

News content, Author: Author Name, Date: Date of news

News content, Author: Author Name, Date: Date of news

```
In [10]: for i in range(0, len(inshorts_data["data"])):
             print("News content : ", inshorts_data["data"][i]["content"])
             print("Author : ", inshorts_data["data"][i]["author"])
             print("DATE : ", inshorts_data["data"][i]["date"])
             print(" ")
```

```
News content :  The Thar Desert, which stretches across Rajasthan and parts of Pakistan and is known for its arid landscape, ma
y turn green by century's end due to climate change, a study stated. The mean rainfall in semi-arid northwest regions of India
and Pakistan increased by 10-50% from 1901-2015. Under moderate greenhouse gas scenarios, this rainfall could surge by 50-200%,
researchers said.
Author :  Swati Dubey
DATE :  Thursday, 17 August, 2023

News content :  The rupee weakened 19 paise to close at an all-time low of 83.15 against the US dollar on Thursday as rising US
Treasury Yields and a risk-averse environment weighed on the Indian currency. The rupee had touched its lifetime low of 83.29 o
n October 20, 2022 during intra-day trade. It closed at 82.96 on Monday.
Author :  Srishty Choudhury
DATE :  Thursday, 17 August, 2023

News content :  A group of Australian surfers who went missing at sea were located off the Indonesian coast after 36 hours. The
surfers were aboard wooden longboats between Nias Island and the Pinang Surf resort on Banyak Islands when they ran into bad we
ather. An Indonesian sailor with the group is still missing, while two other Indonesian crew have been rescued.
Author :  Disha Jana
DATE :  Thursday, 17 August, 2023

News content :  Unacademy has removed its teacher named Karan Sangwan who advised students to vote for a politician who is well
-educated. There were calls to boycott Unacademy on social media over his remarks. "Don't elect someone who only knows changing
names," he had said. Several X (Twitter) users accused him of pushing a political agenda.
Author :  Deepika Bhatt
DATE :  Thursday, 17 August, 2023

News content :  Bill Gates-backed UK biotech company Oxitec has engineered "super mosquitoes" that combat disease-spreading cou
nterpart responsible for 6,00,000 annual deaths. Homeowners can purchase these kits to protect their yards. These male mosquito
es carry a gene preventing female survival, curbing malaria transmission. If these modified males mate with wild females, all f
emale offspring allegedly die.
Author :  Swati Dubey
DATE :  Thursday, 17 August, 2023
```

❖ ISRO Spacecrafts API:

API Link: https://isro.vercel.app/api/spacecrafts

1) To print all the main keys and the total number of keys as well.

In [1]:
```
import requests
url = requests.get("https://isro.vercel.app/api/spacecrafts")
isro_data = url.json()
```

In [2]:
```
count=0
for i in isro_data:
    count=count+1
print("The total number of main keys in this API are : ",count)
print("All the main keys are as follows : ")
for i in isro_data:
    print(i)
```

```
The total number of main keys in this API are :  1
All the main keys are as follows :
spacecrafts
```

2) Print all the spacecrafts names.

```
In [3]: for i in range(0, len(isro_data["spacecrafts"])):
            print(isro_data["spacecrafts"][i]["id"], ": ", isro_data["spacecrafts"][i]["name"])
```

```
1 :  Aryabhata
2 :  Bhaskara-I
3 :  Rohini Technology Payload (RTP)
4 :  Rohini Satellite RS-1
5 :  Rohini Satellite RS-D1
6 :  APPLE
7 :  Bhaskara-II
8 :  INSAT-1A
9 :  Rohini Satellite RS-D2
10 :  INSAT-1B
11 :  SROSS-1
12 :  IRS-1A
13 :  SROSS-2
14 :  INSAT-1C
15 :  INSAT-1D
16 :  IRS-1B
17 :  SROSS-C
18 :  INSAT-2A
19 :  INSAT-2B
```

3) Allow user to enter the name of the spacecraft and check whether that spacecraft exists or not.

In [4]:
```
name_sc = input("Enter the name of the spacecraft: ")
for i in range(0, len(isro_data["spacecrafts"])):
    if(isro_data["spacecrafts"][i]["name"] == name_sc):
        print("Spacecraft is Found")
        break
else:
    print("Spacecraft is NOT Found")
```

```
Enter the name of the spacecraft: INSAT-2D
Spacecraft is Found
```

ISRO Customer API:

API Link: https://isro.vercel.app/api/customer_satellites

To generate a pie chat indicating the percentage of ISRO's own spacecrafts vs customer satellites from above API(s) for analysis of domestic vs foreign customer involvement.

In [2]:
```
import requests
from matplotlib import pyplot as plt
url1 = requests.get("https://isro.vercel.app/api/customer_satellites")
url2 = requests.get("https://isro.vercel.app/api/spacecrafts")
ownsc_data = url2.json()
customersc_data = url1.json()
```

In [3]:
```
x = len(ownsc_data["spacecrafts"])
print(x)
```

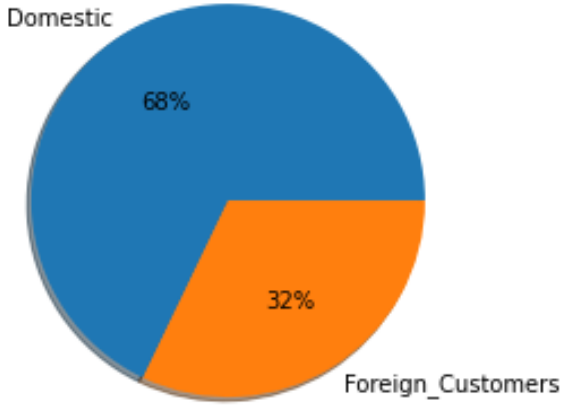
```
112
```

In [4]:
```
y = len(customersc_data["customer_satellites"])
print(y)
```

```
53
```

In [7]:
```
satellites = ['Domestic', 'Foreign_Customers']
num = [x, y]
```

In [13]:
```
plt.pie(num, labels=satellites, shadow=5, autopct="%1.0f%%")
plt.show()
```



Dynamic API:

API: https://api.postalpincode.in/pincode/380001

Allow user to insert pincode, Print name of all areas which fall under that pincode, Pincode entered by user shall be merged in below URL'S XXXXXX part.

```
In [1]: import requests
        user_inp = input("Enter the PINCode: ")
        url = requests.get("https://api.postalpincode.in/pincode/"+ user_inp)
        postal_data = url.json()
        print("Areas under pincode", user_inp, "are:")
        for i in range(0, len(postal_data[0]["PostOffice"])):
            print(postal_data[0]["PostOffice"][i]["Name"])
```

```
Enter the PINCode: 380058
Areas under pincode 380058 are:
Ambli
Bopal
Ghuma
Shela
```

Convolutional Neural Network (CNN):

- These are a type of deep neural network designed for image recognition and processing.
- They use convolutional layers to automatically and adaptively learn spatial hierarchies of features from input images.
- CNNs are effective in tasks like image classification, object detection, and image segmentation due to their ability to capture local patterns and hierarchies of features.
- The architecture typically includes convolutional, pooling and fully connected layers, enabling the network to learn and extract complex visual representations.

- To develop a neural network which could predict the images in the testing folder.(whether it is a laptop or a mobile). We make use of the laptop and mobile images datasets here.

In [1]:
```
import tensorflow as tf
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
import os
import matplotlib.pyplot as plt
import numpy as np
from PIL import Image
```

In [2]:
```
train_dir = 'Image_Data/Training/'
validation_dir = 'Image_Data/Validation/'
test_dir = 'Image_Data/Testing/'
```

In [4]:
```
image_width = 50
image_height = 50
batch_size = 32
epochs = 10
```

In [5]:
```
train_datagen = ImageDataGenerator(
    rescale=1.0/255.0,  # Rescale pixel values to [0, 1]
    rotation_range=40,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

In [6]:
```
validation_datagen = ImageDataGenerator(rescale=1.0/255.0)
```

In [7]:
```
train_generator = train_datagen.flow_from_directory(
    train_dir,
    target_size=(image_width, image_height),
    batch_size=batch_size,
    class_mode='binary'
)
```

```
Found 175 images belonging to 2 classes.
```

In [8]:
```
validation_generator = validation_datagen.flow_from_directory(
    validation_dir,
    target_size=(image_width, image_height),
    batch_size=batch_size,
    class_mode='binary'
)
```

```
Found 27 images belonging to 2 classes.
```

In [9]:
```
model = Sequential([
    Conv2D(32, (3, 3), activation='relu', input_shape=(image_width, image_height, 3)),
    MaxPooling2D(2, 2),
    Conv2D(64, (3, 3), activation='relu'),
    MaxPooling2D(2, 2),
    Conv2D(128, (3, 3), activation='relu'),
    MaxPooling2D(2, 2),
    Flatten(),
    Dense(512, activation='relu'),
    Dropout(0.5),
    Dense(1, activation='sigmoid')
])
```

In [10]:
```
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

In [11]:
```
history = model.fit(
    train_generator,
    steps_per_epoch=train_generator.samples // batch_size,
    epochs=epochs,
    validation_data=validation_generator,
    validation_steps=validation_generator.samples // batch_size
)
```

```
Epoch 1/10
5/5 [==============================] - 10s 654ms/step - loss: 0.7074 - accuracy: 0.4755
Epoch 2/10
5/5 [==============================] - 1s 168ms/step - loss: 0.7080 - accuracy: 0.4685
Epoch 3/10
5/5 [==============================] - 1s 189ms/step - loss: 0.6951 - accuracy: 0.4895
Epoch 4/10
5/5 [==============================] - 1s 189ms/step - loss: 0.6981 - accuracy: 0.4825
Epoch 5/10
5/5 [==============================] - 1s 168ms/step - loss: 0.6791 - accuracy: 0.5664
Epoch 6/10
5/5 [==============================] - 1s 167ms/step - loss: 0.6775 - accuracy: 0.5944
Epoch 7/10
5/5 [==============================] - 1s 164ms/step - loss: 0.6668 - accuracy: 0.6014
Epoch 8/10
5/5 [==============================] - 1s 194ms/step - loss: 0.6404 - accuracy: 0.6434
Epoch 9/10
5/5 [==============================] - 1s 157ms/step - loss: 0.6080 - accuracy: 0.6573
Epoch 10/10
5/5 [==============================] - 1s 168ms/step - loss: 0.5943 - accuracy: 0.6573
```

In [12]:
```
test_datagen = ImageDataGenerator(rescale=1.0/255.0)
```

In [13]:
```
class_labels = train_generator.class_indices
class_labels = {v: k for k, v in class_labels.items()}
```

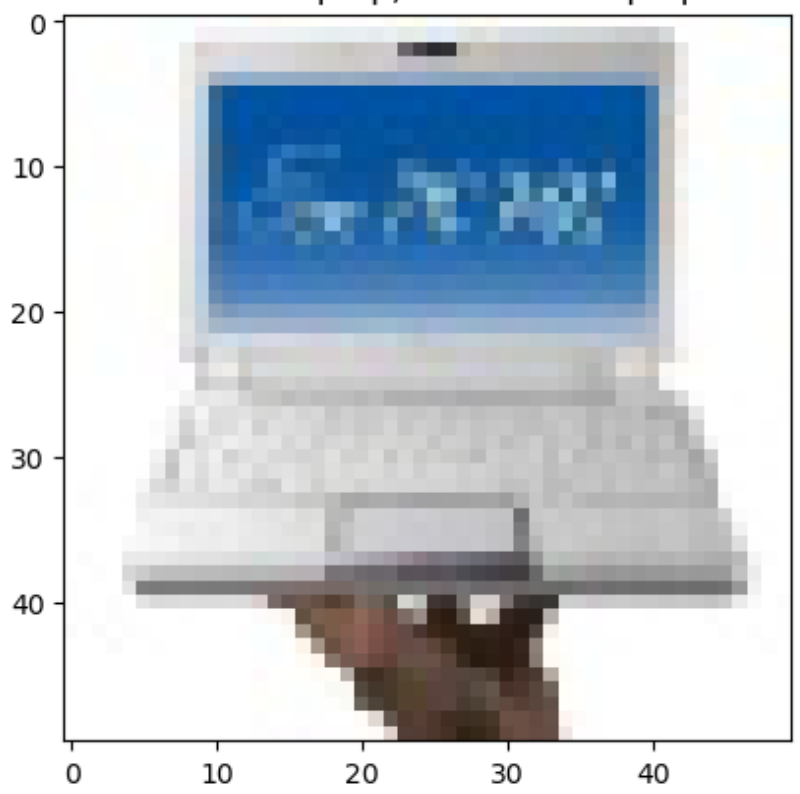
In [14]:
```
test_generator = test_datagen.flow_from_directory(
    test_dir,
    target_size=(image_width, image_height),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=False
)
```

```
Found 91 images belonging to 2 classes.
```

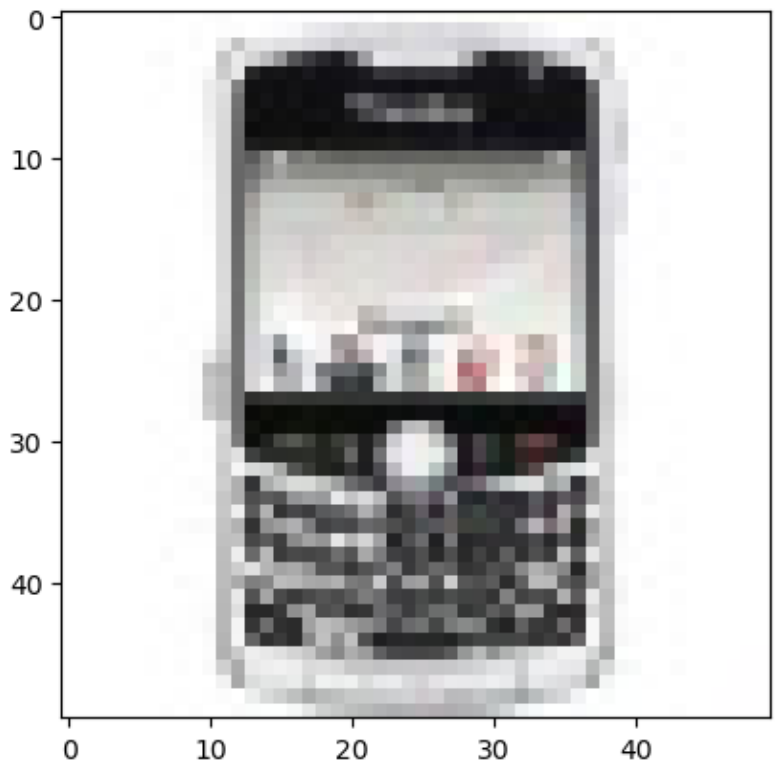
In [15]:
```
num_samples = len(test_generator)
for i in range(num_samples):
    image, actual_label = test_generator[i]
    actual_label = int(actual_label[0])  # Convert one-hot encoded label to integer

    predicted_prob = model.predict(image)[0][0]  # Predicted probability
    predicted_label = 1 if predicted_prob > 0.5 else 0  # Convert probability to label

    actual_class = class_labels[actual_label]
    predicted_class = class_labels[predicted_label]

    image_array = np.asarray(image[0])  # Convert image to array
    image_pil = Image.fromarray(np.uint8(image_array*255))

    plt.imshow(image_pil)
    plt.title(f'Actual: {actual_class}, Predicted: {predicted_class}')
    plt.show()
```

Output:

```
1/1 [==============================] - 0s 418ms/step
```




```
1/1 [==============================] - 0s 187ms/step
```



6. Conclusion

- In conclusion, this internship journey has been very productive and great which aided me to learn new things such as API data handling, creating better data visualization of large and dynamic data like APIs, and the machine learning concepts like regression and its types and machine learning model generation.
- This internship assisted me to polish my skills, enhance my performance and ingest new concepts regarding Data Analytics and Machine Learning.
- With the internship drawing to a close, the skills obtained and insights gained establish a precious base for ongoing advancement within the technology sector. The incorporation of theories and practical experience underscores the tangible applicability of technology in real-world contexts.

7. Future Enhancement

- With this remarkable learning journey behind me and a genuine attraction to Data Analytics and Machine Learning, my future course involves thorough exploration followed by the application of these concepts in project endeavors.
- My future aims related to this field will mainly include projects which would solely focus on the Data Analysis concepts and models made using Machine Learning concepts and algorithms.

8. Completion Certificate

INFOLABZ IT SERVICES PVT. LTD.
WEB DEVELOPMENT | APP DEVELOPMENT | DATA SCIENCE | IOT

COMPLETION CERTIFICATE OF SUMMER INTERNSHIP

Date: 12-08-2023

Enrollment: 201430116080
Semester: 7th, Information Technology
New Lj Institute Of Engineering And Technology

To whom it may concern,

We are delighted to provide this certificate for the successful completion of the requirements and work performed during the two-week free internship (27 July 2023 to 10 August 2023) by Bokil Neel Milind.

In this internship tenure, we have covered the fundamentals of Data Analytics and Machine Learning. In the data analytics part, we have worked on API data and covered the basics of analysis using pandas and data visualization using matplotlib. In machine learning, we have implemented elementary regression models.

We wish Bokil Neel Milind all the best for future endeavors.

Ms. Twinkle Shah
Internship Coordinator
InfoLabz, Ahmedabad


THANK
YOU!