INTERNSHIP REPORT

Name: Maitri Deepak Savla

Email Id: maitrisavla2003@gmail.com

INDEX

INTRODUCTION

The one month Internship as Data Science Developer intern at **NullClass EdTech Pvt. Ltd.** Focuses on exploring and working on real-time industry projects on Image Processing and Computer Vision.

Computer vision and Image processing are two closely related fields that utilizes techniques from Artificial Intelligence and pattern recognition to derive meaningful information from the images, videos and other visual inputs. Image processing algorithm are used to extract information from the images, restore and compresses images and video data, and build new experience in virtual and augmented reality.

Computer Vision uses image processing to recognize and categorize image data. Computer Vision Technology is being used across industries from healthcare and media to gaming and transportation.

This internship report outlines my experience and achievements during the internship, focusing on three distinct machine learning tasks: Child Detection, Voice Gender Detection, and Pedestrian Detection. Each task required a unique set of skills and methodologies, contributing to my overall growth as a machine learning practitioner.

This internship has helped me to build my base strong in Image Processing and Computer Vision. Image processing and computer vision have become integral components of various industries, revolutionizing how businesses operate and interact with visual data. These technologies, fueled by advancements in machine learning and deep learning, enable automation, enhance decision-making processes, and unlock new possibilities across diverse sectors.

It has not only helped me in exploring different Computer Vision algorithm but gave me an opportunity to play with the audio data. It helped me to explore different python library that supports in preprocessing of the audio file. The assignments allotted to me had helped me to dive in-depth about different algorithm and preprocessing techniques used in image and audio processing.

The projects assigned includes popular libraries for image processing such as OpenCV, TensorFlow, Scikit learn library, Numpy, Pandas, librosa, Keras, ElementTree. The journey was not without its challenges. Addressing issues

related to data quality, model optimization and problem-solving skills. Overcoming these challenges was a testament to my adaptability and resilience in the face of complex real-world problems.

The above mentioned task helps me to successfully contribute to the development of cutting-edge solutions within the stipulated timeframe. This internship provided invaluable real-time industry exposure, allowing me to apply theoretical knowledge to practical scenarios and contribute to the development of impactful machine learning solutions.

I express my heartfelt gratitude to the mentors, colleagues and NullClass EdTech Pvt. Ltd. For providing an enriching environment and invaluable support throughout this internship. Through this report, I aim to share my journey and contribute to the ongoing discourse in the ever-evolving field of data science.

BACKGROUND

- Child Detection: The child detection project aimed to develop a solution that can predict whether the uploaded image is of Child or not. The most basic approach is to predict the decision using the age of human. The objective was to create a robust model capable of recognizing and distinguishing minors in online environments. After splitting the data into training and testing we are designing the CNN model to train the data. We have also introduced pooling layer to the model. The pooling layers are used to reduce the dimensions of the feature maps. It reduces the number of parameters to learn and the amount of Computation performed in the network. The CNN model created is:

```
-    Model: "model"
-    _________________________________________________________________
     _____________
-     Layer (type)                Output Shape              Param #   Connected to
-    ==================================================================================
     ============
-     input_1 (InputLayer)        [(None, 48, 48, 3)]       0         []
-
-     conv2d (Conv2D)             (None, 48, 48, 32)        896       ['input_1[0][0]']
-
-     dropout (Dropout)           (None, 48, 48, 32)        0         ['conv2d[0][0]']

-     activation (Activation)     (None, 48, 48, 32)        0         ['dropout[0][0]']
-
```

```
-    max_pooling2d (MaxPooling2  (None, 24, 24, 32)          0         ['activation[0][0]
    ']
-    D)
-
-    conv2d_1 (Conv2D)          (None, 24, 24, 64)          18496     ['max_pooling2d[0]
    [0]']
-
-    dropout_1 (Dropout)        (None, 24, 24, 64)          0         ['conv2d_1[0][0]']
-
-    activation_1 (Activation)  (None, 24, 24, 64)          0         ['dropout_1[0][0]'
    ]
-
-    max_pooling2d_1 (MaxPoolin  (None, 12, 12, 64)          0         ['activation_1[0][
    0]']
-    g2D)
-
-    conv2d_2 (Conv2D)          (None, 12, 12, 128)         73856     ['max_pooling2d_1[
    0][0]']
-
-    dropout_2 (Dropout)        (None, 12, 12, 128)         0         ['conv2d_2[0][0]']
-
-    activation_2 (Activation)  (None, 12, 12, 128)         0         ['dropout_2[0][0]'
    ]
-
-    max_pooling2d_2 (MaxPoolin  (None, 6, 6, 128)          0         ['activation_2[0][
    0]']
-    g2D)
-
-    conv2d_3 (Conv2D)          (None, 6, 6, 256)           295168    ['max_pooling2d_2[
    0][0]']
-
-    dropout_3 (Dropout)        (None, 6, 6, 256)           0         ['conv2d_3[0][0]']
-
-    activation_3 (Activation)  (None, 6, 6, 256)           0         ['dropout_3[0][0]'
    ]
-
-    max_pooling2d_3 (MaxPoolin  (None, 3, 3, 256)          0         ['activation_3[0][
    0]']
-    g2D)
-
-    flatten (Flatten)          (None, 2304)                0         ['max_pooling2d_3[
    0][0]']
-
-    dense (Dense)              (None, 64)                  147520    ['flatten[0][0]']
-
-    dense_1 (Dense)            (None, 64)                  147520    ['flatten[0][0]']
-
-    dropout_4 (Dropout)        (None, 64)                  0         ['dense[0][0]']
-
-    dropout_5 (Dropout)        (None, 64)                  0         ['dense_1[0][0]']
-
-    sex_out (Dense)            (None, 1)                   65        ['dropout_4[0][0]'
    ]
-
-    age_out (Dense)            (None, 1)                   65        ['dropout_5[0][0]'
    ]
-
```

```
=================================================================================
============
Total params: 683586 (2.61 MB)
Trainable params: 683586 (2.61 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________
```

- Voice Gender Prediction: The voice-activated systems and virtual assistants become integral parts of daily life. Voice Gender Detection addressed this need by focusing on accurately predicting the gender of the speaker based on different voice parameters. In this task, we are using the csv file of preprocessed audio file to train the model. For this task a simple sequential model with Dense and Dropout layer produces a descent accuracy. To test the model, we can use the librosa library to extract all important features of an audio (.wav) file and then use it for evaluation of the model.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 32)                4128

 dropout (Dropout)           (None, 32)                0

 dense_1 (Dense)             (None, 64)                2112

 dropout_1 (Dropout)         (None, 64)                0

 dense_2 (Dense)             (None, 128)               8320

 dropout_2 (Dropout)         (None, 128)               0

 dense_3 (Dense)             (None, 128)               16512

 dropout_3 (Dropout)         (None, 128)               0

 dense_4 (Dense)             (None, 1)                 129

=================================================================
Total params: 31201 (121.88 KB)
Trainable params: 31201 (121.88 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

- Pedestrian Prediction: Pedestrian Detection played a pivotal role in addressing challenges related to the advancement of autonomous

vehicles and surveillance systems. This project focuses on creating a reliable model capable of recognizing pedestrians in diverse scenarios. Sometimes the ML model gets confused among the pedestrian and the article which look like pedestrian. The sequential model is designed and trained to differentiate among the pedestrian or the article that look like a pedestrian. The sequential model is the combination of convolution layer and the pooling layer. The pooling layer summarizes the features present in a region of the feature map generated by a convolution layer.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 198, 198, 16)      448

 max_pooling2d (MaxPooling2  (None, 99, 99, 16)        0
 D)

 conv2d_1 (Conv2D)           (None, 97, 97, 32)        4640

 max_pooling2d_1 (MaxPoolin  (None, 48, 48, 32)        0
 g2D)

 conv2d_2 (Conv2D)           (None, 46, 46, 64)        18496

 flatten (Flatten)           (None, 135424)            0

 dense (Dense)               (None, 128)               17334400

 dense_1 (Dense)             (None, 2)                 258

=================================================================
Total params: 17358242 (66.22 MB)
Trainable params: 17358242 (66.22 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

LEARNING OBJECTIVES

- The specific goals set for the internship aimed at equipping the intern with practical skills and a deep understanding of the challenges inherent in real-world machine learning applications.
- **Hands-on Experience in Machine Learning:** The primary objective of this internship is to immerse a hands-on experience with different machine learning algorithm. It focuses on understanding the entire machine learning pipeline.

- **Development of Audio Data Preprocessing:** The internship focuses to enhance my skills in audio data processing. This involves delving into the complexities of handling audio data, exploring feature extraction techniques such as MFCC, and developing a comprehensive understanding about voice-related data.
- **Development of Computer Vision Skills:** The project emphasizes on development of Computer Vision skills. This includes gaining proficiency in image processing, understanding various computer vision algorithm and dealing with challenges such as scale, perspective and occlusion.
- **Integration of Deep Learning Architectures:** Another key goal is to gain exposure to deep learning architectures such as Convolution Neural Network. The objective is to understand how these advanced architectures can be applied to image and audio data, providing a deeper insight into the complexities of deep learning.
- **Ethical considerations in Machine Learning:** The internship seeks to instill an awareness of ethical considerations in machine learning applications. This involves understanding the implications of bias, fairness and transparency in algorithmic decision-making, fostering a responsible and ethical approach to deploying machine learning solutions.
- **Documentation and Communication Skills:** In addition to technical skills, the internship places emphasis on developing strong documentation and communication skills. This involves documenting project processes, and creating comprehensive report to convey findings and insights to both technical and non-technical stakeholders.
- **Adaptability to Diverse Datasets:** The internship also aims to enhance adaptability by working with diverse datasets. This includes understanding how to preprocess and manipulate data from different domains, adapting machine learning models to various data characteristics and gaining experience in handling the intricacies of varied datasets.

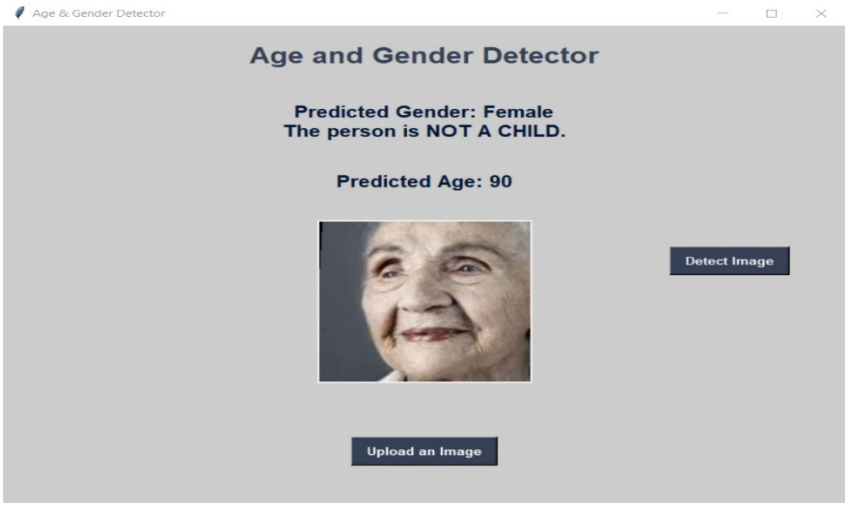
ACTIVITIES AND TASKS

- **Task 1- Child Detection:** The task assigned is Detecting whether the image is of child or not. Here, we are using the dataset of different screenshots of human faces. After training the model, it is able to detect

whether the given image of human is of a child or not. This assignment mainly focus on exploring and understand basics of OpenCV and CNN network.

The steps involved in this task are:

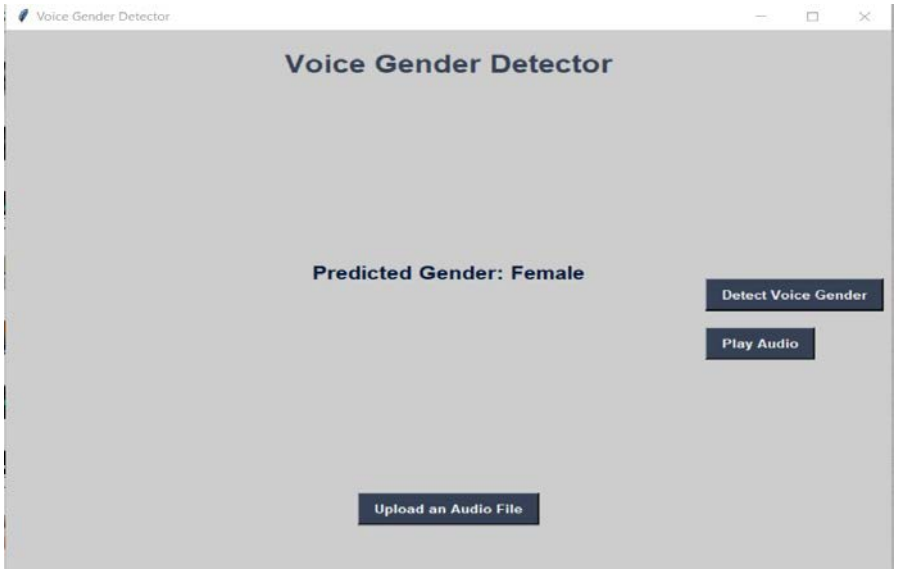
- Extracting the required data from the dataset. The link for the dataset is here .
- We need to create a separate list of Age and Gender from the dataset.
- Save images array, ages array and genders array in the system for future use.
- Performing EDA on dataset is a crucial step to understand the dataset more efficiently. The graph plotted gives the reader and programmer the in-depth understanding about the dataset and different algorithm that can be applied to the dataset.
- Splitting the dataset for training and testing phase.
- Y_test and Y_train contains two element, one representing age and other representing the gender. For better understanding it is to convert the 1-D Y_train and Y_test into 2-D array, one representing age and the other representing the gender.
- Designing and defining the model is the most important part of the process. We are creating a **CNN model** to accomplish our task.
- The activation function used are **ReLU** and **sigmoid**. A rectifier linear unit (ReLU) is an activation function that introduces the property of non-linearity to a deep learning model and solves the vanishing gradient issue. Sigmoidal activation function binds the input values between 0 and 1.
- Initializing and running the model.
- You evaluate the model and test the prediction made by the model.
- Check if any hyper-parameter tuning is possible or not.
- Making the report of the model and analyzing the confusion_matrix for the prediction made.
- Designing the basic GUI using the tkinter libaray in python. The GUI looks like the following images.



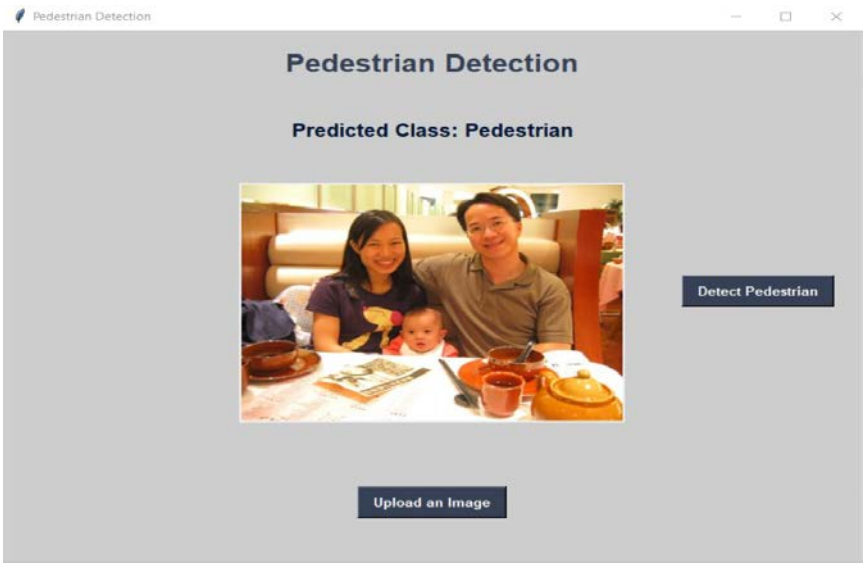
- **Task 2-Voice Gender Detection**: In this task we have to develop a machine learning model that is capable of predicting the gender of a speaker based on a short audio clip. The task involves handling audio data, preprocessing, feature extraction and applying the appropriate machine learning techniques to achieve reliable predictions.
  The steps involved in this task are:
  - Download the dataset. Click here to download the dataset.
  - Importing CSV file of preprocessed audio file.
  - Loading the dataset and extracting all important features from it.
  - The features of the preprocessed audio file are store in **features.npy** array and the labels are stored in **labels.npy** array.
  - Performing EDA on dataset is a crucial step to understand the dataset more efficiently. The graph plotted gives the reader and programmer the in-depth understanding about the dataset and different algorithm that can be applied to the dataset.
  - Building the model. We have designed the sequential model with Dense and Dropout layers. The activation function used are **ReLU** and **sigmoid**.
  - The loss function used is **Binary cross entropy** function. It is a loss function that quantifies the dissimilarity between probability distributions, aiding model training by penalizing inaccurate predictions.
  - We are evaluating and predicting the model outcome.
  - Librosa library in python supports the processing of an audio file in python. The features we are extracting from audio(.wav) file are MFCC, Chroma, MEL Spectrogram Frequency, Tonnetz and Contrast

MFCC	MFCC is a feature extraction technique widely used in speech and audio processing. It are used to represent the spectral characteristics of sound in a way that is well-suited for various Machine Learning tasks.
Chroma	It describes the angle of pitch rotation as it traverses the helix.
MEL Spectrogram Frequency	melSpectrogram applies a frequency-domain filter bank to audio signals that are windowed in time. It aim to mimic non-linear human ear perception of sound.
Tonnetz	The Tonal Centroid features (Tonnetz) representation uses the method to project the chroma features onto a 6-dimensional basis representing the perfect fifth, minor third, and major third each as two-dimensional coordinates.
Contrast	Audio Contrast is the difference in the sound levels, quality or style between two or more audio elements in a video project. It can be used t create a sense of contrast between different locations, time period, moods or perspective.

- Creating the report of the model and analyzing the confusion matrix of the predicted output.
- Designing the basic GUI using tkinter library in Python. The GUI looks like the following images.



- **Task 3-Pedestrian Detection**: In this task the data set consist of images of human and the images which look like humans. The assignment involves developing different machine learning model that can effectively detect and recognize the Pedestrian in the images.
  The steps involved are:
  - Extracting the required data from the dataset. The link for the dataset is here.
  - Reading the dataset and loading the dataset in their respective folder as Train, Test and Validation.
  - Perform EDA on the loaded dataset. It is an important step to understand the given dataset in-depth. The graph plotted gives the visualization about the data in the dataset.
  - We are designing and defining the **sequential CNN model**. The CNN models is designed using the **convolution layer**, **pooling layer** and **the dense layer**.
  - The activation function used in this model are **ReLU** function. We have also used **Adam optimizer** while compiling the model. The Adaptive Moment Estimation (ADAM) is an iterative optimization algorithm used to minimize the loss function during the training of neural networks.
  - Evaluating and Predicting the output obtained from the model designed.
  - Creating report of the model and analyzing the predicted output.
  - We have also designed the basic GUI using the tkinter library of python. The GUI looks like the following images.



SKILLS AND COMPETENCIES

Skills:

- **Deep Learning model development**
    - Developing a machine learning model for predicting speaker gender from audio data involves utilizing deep learning techniques, possibly leveraging neural networks for effective feature learning.
    - For detecting pedestrians in street images or videos and for child detection task, deep learning models, particularly convolutional neural networks (CNNs), are commonly employed to learn hierarchical features from visual data.
- **Encompassing Data preprocessing**
    - Data preprocessing is crucial for preparing the dataset to fit properly in the model designed. It ensures that the data is appropriately formatted and cleaned for model trainging.
    - Audio data preprocessing is integral to extracting meaningful features, such as MFCC coefficients, from raw audio signals.
- **Feature extraction**
    - Feature extraction is one of the important steps before applying any machine learning algorithm. You select only essential features and exclude the features that has least contribution or is irrelevant for achieving the desired accuracy from the model.
- **Computer Vision technique**
    - Proficiency in Computer Vision techniques is essential for developing models that can effectively detect or recognize the requirements considering challenges like scale, perspective and occlusion.

Competencies:

- **Problem-solving:** All tasks involve addressing unique challenges. These tasks has helps me to enhance my problem solving skills and has increased my eager in exploring different techniques can be used to resolve the problem.

- **Critical thinking:** Critical thinking is crucial in making decisions about the choice of models, preprocessing steps, and parameter tuning to ensure optimal performance in each task.
- **Debugging:** Debugging skills are essential when troubleshooting issues related to model training, data preprocessing, or feature extraction to ensure the models operate as intended.
- **Attention to Detail:** Detecting nuances in voice patterns for Voice Gender Detection and handling variations in image data for Child Detection and Pedestrian detection requires a keen attention to details to achieve accurate results.
- **Ethical Consideration:** Ethical considerations are crucial in tasks such as child detection, where privacy and responsible use of technology are paramount. Ethical considerations also play a role in developing models for voice gender detection.

FEEDBACK AND EVIDENCE

Feedback:

- It was because of training session given before internship, I was able to design and test different model to receive high accuracy to complete the assigned task.
- This internship also gives exposure to the real-life industry level experience.
- The project is not only limited to model designing, but it also train us to build a basic GUI for showcasing your output in a presentable manner.
- Helps us understanding actual importance of Version Control system.
- This internship opportunity gave me an opportunity to explore certain python libraries that is beyond image processing and computer vision.
- It gave me an opportunity to study in depth about librosa library that is used for audio processing.
- Daily report system by the NullClass EdTEch Pvt. Ltd. is very organized approach. It helps me to set the daily target and daily work on it.

Evidence:

- Model Evaluation Metrics: After designing of the model for each task, there is model evaluation metrics included. This encompassed accuracy

rates, precision, recall, F1 score providing a comprehensive understanding about the model's performance.
- Even after analyzing the confusion matrix, we are testing the model by passing certain test images or audio file through the model.
- Successful demonstration of the machine learning models through GUI further validates their practical applicability.

CHALLENGES AND SOLUTIONS:

- In Voice Gender Prediction Task, variations in voice pitch and tone among speakers of the same gender presented a challenge in Voice Gender Detection.

  SOLUTION: Advanced Feature Extraction Methods
  To overcome this challenge, explored advanced feature extraction methods, with a focus on leveraging Mel-frequency cepstral coefficients (MFCC) and other spectrogram-based techniques. This allowed the model to extract more nuanced information from voice samples, improving the accuracy of gender predictions.

- One of the major challenges faced in Child Detection and Pedestrian Detection is that the model cannot predict the output properly is there is lot of background disturbance in the image.

  SOLUTION: We can consider this to be one of the limitation and ask user to upload or test images that is less background disturbed. This ensures that the model is able to predict accurate results for majority of the input.

- The other challenges that I have faced during Voice- Gender detection is that the processing of the audio files depends on the type of audio file that is given as input. We cannot generalize the feature extracting function for all types of audio.

  SOLUTION: I have restricted my feature extracting function to extract the features of only wav type audio file. The model cannot predict the

output for file other than .wav audio file. It might throw error or give incorrect output for other type audio file.

OUTCOMES AND IMPACT:

The internship resulted in tangible outcomes, including high-performing model for each task.

CHILD DETECTION

OUTCOME: The child detection yielded a high-performing model, showcasing notable accuracy rates in detecting the presence of a child in images or videos.

IMPACT:

- The model's accuracy contributes to the reliability of child detection, which can have significant implications in various domains such as child safety, surveillance and automated content filtering.
- Applications in childcare monitoring, public spaces and image classification benefit from the robust child detection capabilities, ensuring a safer and more efficient use of visual data.

VOICE GENDER DETECTION

OUTCOME: The voice Gender Detection task resulted in a model with remarkable accuracy, effectively predicting the gender of a speaker across a diverse range of voice samples.

IMPACT:

- The accuracy achieved in predicting speaker gender enhances the usability of voice-based applications, including voice assistants, voice authentication systems and customer service applications.
- This outcome contributes to advancements in voice-related technologies and ensures more personalized and accurate user interactions, especially in applications where user gender information in relevant.

PEDESTRIAN DETECTION

OUTCOME: The pedestrian Detection task demonstrated reliable recognition in street images, showcasing a model capable of effectively detecting and recognizing pedestrians.

IMPACT:

- The model's reliability in pedestrian detection has significant implications for the development of safer autonomous vehicles and advanced surveillance systems.
- In the context of autonomous vehicles, accurate pedestrian detection is crucial for ensuring pedestrian safety, thereby contributing to the advancement of intelligent transportation systems.
- In surveillance systems, the model's capabilities enhances the effectiveness of monitoring and security, especially in crowded urban environments.

CONCLUSION

In reflection, the internship has been a transformative and enriching experience, providing a comprehensive immersion into the realm of machine learning applications. The skills and competencies cultivated during this period extend beyond the confines of the specific tasks undertaken, forming a robust foundation for future endeavors in the dynamic field of artificial intelligence.

Beyond the technical aspects of machine learning, the journey underscored the significance of confronting and surmounting real-world challenges. The acquired skills encompass not only the intricacies of deep learning model development, data preprocessing, feature extraction, and computer vision techniques but also extend to the broader context of ethical considerations, interdisciplinary collaboration, and effective communication.

The internship journey underscored the significance of confronting and surmounting real-world challenges in the realm of machine learning. The acquired skills are not solely to the technical aspects of machine learning; they extend to the broader context of ethical consideration, interdisciplinary collaboration and effective communication.

As the field of artificial intelligence continues to evolve, the insights gained from this internship serve as a compass for navigating the complexities and ethical considerations inherent in the development and deployment of machine learning solutions. The experiences garnered, the challenges overcome, and the positive feedback received collectively validate ones growth and contributions.

The high-performing models developed during the internship for child detection, voice gender detection, and pedestrian detection contribute tangibly to the advancement of machine learning applications, validating not only technical proficiency but also the ability to translate theoretical knowledge into practical, real-world solutions. In conclusion, the internship has been a catalyst for personal and professional development, providing a multifaceted understanding of the intricacies of machine learning and preparing me for the challenges and opportunities that lie ahead in the dynamic field of artificial intelligence.

---------- END OF REPORT ----------