Semester Project III
Abstract

1. **Project Title:** Automated Object Recognition and Speech Synthesis
2. **Domain / Technology:** Machine Learning, Data Analytics, Agriculture Technology
3. **Application:** Interactive learning tool for children to identify objects with speech output. Helps improve vocabulary, observation, and engagement through fun technology.
4. **Dataset :**COCO.ImageNet – 14M images across 20k categories. Pascal VOC – 20 object classes with bounding box annotations. Open Images – 9M images with rich object labels.
5. **Expected Algorithm /Processing**: **Data Collection:** Preprocessing – Resize, normalize, and augment image data to improve model generalization. Object Detection Model Training – Train deep learning models (YOLO, Faster R-CNN) for object recognition. Object Label Extraction – Retrieve predicted object names with confidence scores.
6. **Expected Output from Project:**
   - Detected Object List – Display names of identified objects in images or video frames.
   - Real-Time Speech Output – Verbal description of recognized objects.
   - Confidence Scores – Show prediction accuracy for each detected object.
   - Visual Annotations – Draw bounding boxes and labels on detected objects.
7. **Abstract in your own words :**

   Children are naturally curious and learn best through interaction with their surroundings. One effective way to support their learning is by providing tools that help them detect, recognize, and hear the names of objects around them in real time. This project is specifically designed as a learning and interactive system for kids, aiming to make object recognition both educational and entertaining. By integrating object detection with speech synthesis, the project goes beyond traditional learning methods, creating an innovative solution that bridges technology and education. Ultimately, this project aims to make education more dynamic, enjoyable, and accessible for children by harnessing the power of artificial intelligence and speech technology. we are developing a system. Text-to-speech converters like YOLOv8 can help achieve this. Keywords: YOLOv8 assistive technologies, deep learning, object detection, and Mask R-CNN

Team Members:

1. Pradnya Chhotu Patil. (TY-DS-12-231106013)
2. Prerna Motilal Patil. (TY-DS-13-231106014)
3. Divya Jagdish Khairnar. (TY-DS-63-241206004)
4. Ashwini Yogesh Deore. (TY-DS-65-241206006)