EXTRACTION OF TEXT FROM IMAGES

PROJECT REPORT

Submitted By

POOJA P S (RegNO: MCT20MCA-2030)

THE APJ Abdul Kalam Technological University

In partial fulfillment of the requirements for the award of the

Degree of

MASTER OF COMPUTER APPLICATIONS

DEPARTMENT OF COMPUTER APPLICATIONS

MOHANDAS COLLEGE OF ENGINEERING & TECHNOLOGY

Anad, Nedumangadu, Thiruvananthapuram

695544

2021-22

DEPARTMENT OF COMPUTER APPLICATION

MOHANDAS COLLEGE OF ENGINEERING & TECHNOLOGY

Anad, Nedumangadu, Thiruvananthapuram-695 544

CERTIFICATE

This is to certify that the report entitled **"EXTRACTION OF TEXT FROM IMAGES"** submitted by **Ms. POOJA P S (Register No: MCT20MCA2030)** to **APJ Abdul Kalam Technological University** in partial fulfillment of the requirements for the award of the degree **MASTER OF COMPUTER APPLICATIONS** is a bonafide record of the project work carried out by her under my guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor(s) Project Coordinator

Head of the Department External Examiner

DECLARATION

I hereby declare that the project report "**EXTRACTION OF TEXT FROM IMAGES**", submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Prof Mrs. JAYANTHI T. I have adequately and accurately cited and referenced the original source. I also declare that I have adhered to ethics of academic honestly and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be cause for disciplinary action by the institute and or the University and can also evoke penal action from the source which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of the degree, diploma or similar title of any other University

Place: Trivandrum

Date:

POOJA P S

ACKNOWLEDGEMENT

At the outset, I thank God Almighty for standing by me throughout the project and making it possible for me to complete the project within the stipulated time.

I wish to record my deep sense of gratitude to our Director, **Dr. ASHALATHA THAMPURAN**, for her extensive support and guidance throughout the course of my project.

I wish to record my deep sense of gratitude to our Principal, **Dr. S. SHEELA**, for her extensive support and guidance throughout the course of my project.

I am greatly thankful to our **HOD Mrs. Sreeja. K**, (Department of Computer Applications) for her kind co-operation and guidance throughout the course of my project.

I am also thankful to our Staff Advisor **Mrs. JEEJA. G .S**, (Department of Computer Applications) for her kind co-operation and guidance throughout the course of my project.

I am also thankful to our Project Guide **Mrs. JAYANTHI. T** (Department of Computer Applications) for her kind co-operation and guidance throughout the course of my project.

I also extend my sincere thanks to all other faculty members of Department of Computer Applications and our friends for their co-operation and encouragements.

BY,

POOJA P S

TABLE OF CONTENTS

ABSTRACT

The aim of this project is to extract text from images. The language used is python. Tesseract is an open-source engine for optical character recognition (OCR). It efficiently reads text from images. We Import all the required libraries (tkinter, tesseract, opencv). Provide the location of the tesseract.exe file. In this project, we use Pytesseract as the language used is python to extract text from the image uploaded. Pytesseract stands for python-tesseract. Tkinter provides GUI functionalities: open an image dialog box so user can upload an image from the file directory so as to extract text from the image uploaded. This project also uses OpenCV for image processing. This is a project which automatically detects and extracts text from images very efficiently.