VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-560014, Karnataka

A PROJECT REPORT ON

"EXTRACTION OF DATA FROM IMAGE PDF USING OCR "

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

Submitted By

BASAVARAJA	**(1SV18CS005)**
ENCHARA M	**(1SV18CS014)**
PRAJWAL C	**(1SV18CS032)**
SUSHMA H S	**(1SV18CS042)**

Under the guidance of

Prof. Shanmukaswamy C V B.E., M.Tech.,MISTE
Associate Professor & HOD, Dept. of CSE.

SHRIDEVI EDUCATION

PRINCIPAL
SIET., TUMAKURU

Department of Computer Science and Engineering

SHRIDEVI INSTITUTE OF ENGINEERING AND TECHNOLOGY
(Affiliated To Visvesvaraya Technological University)
Sira Road, Tumakuru – 572 106, Karnataka.

2021-22

Sri Shridevi Charitable Trust (R.)
SHRIDEVI INSTITUTE OF ENGINEERING & TECHNOLOGY
(Recognised by Govt. of Karnataka, Affiliated to VTU, Belagavi and Approved by AICTE, New Delhi)
Sira Road, Tumakuru - 572 106, Karnataka.
Phone: 0816-2210699 | Fax: 0816-2210698 | Email: info@shridevienginering.org | Web: http://www.shridevienginering.org

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that, the project entitled "***EXTRACTION OF DATA FROM IMAGE PDF USING OCR***" has been successfully carried out by Basa[illegible] [1SV18CS005], Enchara M [1SV18CS014], Prajwal C [1SV18CS032], Sushma H S [1SV18CS042], in partial fulfillment for the award of **Bachelor of Engineering** in **Computer Science & Engineering** of the **Visvesvaraya Technological University, Belagavi** during the academic year **2021-22.** It is certified that all the corrections/suggestions indicated for internal assessments have been incorporated in the report. The project report has been approved as it certifies the academic requirements in respect of **Project work-II** prescribed for the Bachelor of Engineering Degree.

Signature of The Guide & HOD

Prof. Shanmukaswamy C V B.E.,M.Tech.
Associate Professor & HOD
Dept. of CSE,
SIET, Tumakuru.

Signature of The Principal

Dr.Narendra Viswanath M.E.,Ph.D.,MIE,MISTE,MIWS,FIV.,
Principal,
SIET, Tumakuru.

PRINCIPAL
SIET., TUMAKURU

Name of the Examiners

1. Shanmukaswamy C V
2. Renukaradhya P C

Signature with date

ABSTRACT

Text data present in images contain useful information for automatic explanation, indexing, and structuring of images. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging in the computer vision research area. The following proposed method are used here in this project to extract text from the desired images, they are OCR. Text recognition in images is a research area which attempts to develop a computer system with the ability to automatically read the text from images. These days there is a huge demand in storing the information available in paper documents format in to a computer storage disk and then later reusing this information by searching process. One simple way to store information from these paper documents in to computer system is to first scan the documents and then store them as images. But to reuse this information it is very difficult to read the individual contents and searching the contents form these documents line-by-line and word-by-word.

PRINCIPAL
SIET., TUMAKURU.