GUJARAT TECHNOLOGICAL UNIVERSITY
(Established under Gujarat Act No. 20 of 2007)
ગુજરાત ટેકનોલોજિકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંક: ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Enrollment no: 210280116082

STUDENT'S WEEKLY RECORD OF INTERNSHIP

NAME OF STUDENT: VINAYAK TUSHARBHAI CHAVDA

DIARY OF THE WEEK: Dt: 17/03/2025 TO 21/03/2025

DEPARTMENT: INFORMATION TECHNOLOGY   SEM: 8th

NAME OF THE ORGANISATION: BUCHANAN TECHNOLOGIES LLP

NAME OF THE PLANT/SECTION/DEPARTMENT: EMERGING TECHNOLOGIES

NAME OF OFFICER INCHARGE OF THE PLANT/SECTION/DEPARTMENT: HARDIK SHETH

DESCRIPTION OF THE WORK DONE IN BRIEF

This week, I focused on improving both the user interface (UI) of my previous projects and addressing specific needs in my current project. I started by exploring **Tailwind CSS**, which helped me tackle the UI issues that had been affecting my earlier work. I learned about various classes Tailwind provides, which enabled me to create more **beautiful** and **responsive UIs** for my applications.

In my current project, which involves extracting data from **PDFs**, I realized that **regular expressions (regex)** would play a critical role in identifying and differentiating data within the documents. To better understand how regex can be applied in this context, I dedicated time to **learning and practicing regular expressions**. This hands-on approach gave me a clearer understanding of how to parse and extract data efficiently.

As the project progressed, I did extensive research on **PDF data extraction** techniques, especially focused on extracting data from **invoices**. I came across multiple methods for extracting data, including using **Tesseract**, **pdfPlumber**, and **easyOCR**. However, the most efficient method I found was leveraging the **Gemini Pro API** (AI-based) for invoice data extraction. I wrote a tool in Python using the Gemini API to extract data from invoices and convert it into **JSON format**, making it easier to process the data further.

To improve the tool's accuracy and flexibility, I spent the next day refining it to generate **invoice-specific JSON data**. After completing this modification, I shifted my focus to learning **Docker**, an essential tool for working on projects with multiple people and systems. Understanding Docker will allow me to containerize applications and streamline development and deployment, especially when collaborating in a team setting.

GUJARAT TECHNOLOGICAL UNIVERSITY
(Established under Gujarat Act No. 20 of 2007)
ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંક: ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

TOTAL HOURS: 40 HOURS

SIGNATURE OF STUDENT

✪ **The above entries are correct and the grading of work done by Trainee is**
EXCELLENT / VERY GOOD / GOOD / FAIR / BELOW AVERAGE / POOR

Signature of Faculty Mentor

Signature of officer-in-charge of Dept. / Section / Plant

Date:

Date:

✪ **Grading of Work, for trainee may be given depending upon your judgement about his Punctuality, Regularity, Sincerity, Interest taken, Work done etc.**