



RAJALAKSHMI ENGINEERING COLLEGE

(AN AUTONOMOUS INSTITUTION)

RAJALAKSHMI NAGAR, THANDALAM- 602105

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BUILDING A MODEL TO PREDICT THE STAY OF PATIENTS USING MACHINE LEARNING

CS23421 – INTERNSHIP

Internship Report submitted by

REGISTRATION NUMBER	230701177
STUDENT NAME	MANOHARAN K
YEAR	2023–2027
INTERNSHIP COMPANY	TATA CONSULTANCY SERVICES
TRAINER NAME	MANIKANDAN S
INTERNSHIP PERIOD	16.06.2025 - 01.07.2025
INTERNSHIP DURATION	14 DAYS

(Approved / Not
Approved)

EXAMINER 1

(Approved / Not
Approved)

EXAMINER 2

(Approved / Not
Approved)

Deputy HoD

(Approved / Not
Approved)

HoD / CSE

Work Carried In Industry:

During my internship at Tata Consultancy Services (TCS), I worked on a project which used machine learning techniques to predict patient hospital stay duration. The objective involved processing hospital records for model development to estimate patient hospital stay durations through diagnosis age and other health-related features.

The internship included learning about real-world healthcare data and applying data science principles for generating insights through visualizations and models. I was responsible for defining the problem and performing data preprocessing and exploratory data analysis and selecting and training models before evaluating them and creating business presentations for the results.

My workflow followed an industry-standard data science project pipeline structure. Through this experience I gained practical knowledge of applying ML techniques to address actual problems while learning how large enterprises like TCS manage their professional data workflows.

Technologies And Tools Used:

Programming and Environment:

- **Python** – Used as the primary language for all stages of development.
- **Jupyter Notebook** – For writing and testing code in an interactive format.

Data Preprocessing and Analysis:

- **Pandas** – For data cleaning, transformation, and handling missing values.
- **NumPy** – For numerical computations and array operations.

Visualization:

- **Matplotlib** and **Seaborn** – Used to create visual representations such as histograms, box plots, and correlation heatmaps to understand data patterns.

Machine Learning:

- **Scikit-learn** – For model selection, training, and evaluation. Algorithms included Decision Trees, Random Forests, and Gradient Boosting.
- **Evaluation Metrics** – Accuracy, Precision, Recall, F1-Score, and ROC-AUC were used to assess model performance.

Project Overview

Project Title: **Building a model to Predict the Stay of Patients using Machine Learning.**

Objective:

To develop a machine learning model that predicts the duration of a patient's hospital stay based on key parameters such as diagnosis, age, and other clinical features.

Key Features:

- Data preprocessing to handle missing values and encode categorical variables
- Exploratory Data Analysis (EDA) to identify patterns and relationships
- Model training using suitable classification algorithms
- Evaluation of model performance using appropriate metrics
- Visualization and presentation of results in a business context

Technologies Used:

- Python, Jupyter Notebook
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn

Outcome:

Successfully built and evaluated a predictive model that estimated patient hospital stay durations with reliable accuracy. Gained hands-on experience in applying data science techniques to a real-world healthcare problem and developed a structured understanding of machine learning project workflows in an enterprise setting.

Conclusion and Reflection:

My internship at Tata Consultancy Services (TCS) was a valuable learning experience that deepened my understanding of data science and machine learning. Working on a healthcare-based prediction model gave me practical exposure to real-world data challenges and industry expectations.

I was able to apply theoretical concepts to a live problem statement, enhance my coding and analytical skills, and gain confidence in handling structured workflows commonly followed in enterprise environments.

Key Learnings:

- Understood the complete machine learning pipeline from data preprocessing to model evaluation
- Learned to use Python libraries like Pandas, NumPy, Scikit-learn, and Seaborn for data analysis and visualization
- Gained experience in selecting suitable algorithms for classification tasks
- Developed an ability to interpret model performance using key evaluation metrics
- Improved problem-solving, time management, and documentation skills through structured task execution

Final Thoughts:

This internship provided me with a solid base to start my data science career. Through this experience I was able to combine theoretical understanding with practical industry knowledge. The TCS team provided essential guidance during my internship that motivated me to continue exploring machine learning and healthcare analytics projects.

Internship Certificate:



Date: July 1, 2025

Internship Project Completion Certificate

To Whom it May Concern

This is to certify that **MANOIHARAN K (230701177)** Student of Rajalakshmi Engineering College, Thandalam undergoing "B.E Computer Science and Engineering" has successfully completed his internship program with the project titled "**Building a model to predict the stay of patients using Machine Learning**" at Tata Consultancy Services for a period of 2 weeks from 16-Jun-2025 to 01-July-2025.

We appreciate the hard work and dedication towards the same.

For TATA Consultancy Services Limited

A handwritten signature in black ink, appearing to read 'M.P. Lakshman Prakash', is written over a horizontal line.

M P Lakshman Prakash
General Manager – Human Resources