Project Report on Course

DATA ANALYSIS USING PYTHON (21CS120)

Bachelor of Technology
In

Computer Science & Artificial Intelligence

By

Name: G SAI CHARAN **Roll Number: 2203A52021**

Under the Guidance of

Dr. DADI RAMESH
Asst. Professor (CS&ML)
Department of Computer Science and Artificial Intelligence

COMPUTER SCIENCE
SCHOOL OF COMPUTER SCIENCE
AND ARTIFICIAL INTELLIGENCE

SR UNIVERSITY, ANANTHASAGAR, WARANGAL
April, 2025.

SCHOOL OF COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE

CERTIFICATE OF COMPLETION

This is to certify that **G SAI CHARAN** bearing Hall Ticket Number **2203A52021**, a student of **CSE-AIML, 3rd Year - 2nd Semester**, has successfully completed the **Data Analysis Using Python** Course and has submitted the following 3 projects as part of the curriculum:

Project Submissions:

- **CSV** Project**: SLEEP CYCLE PRODUCTIVITY**
- **IMAGE** Project**: BABY CRY DETECTION**
- **TEXT** Project**: CROP DETECTION**

Dr. Dadi Ramesh
Asst. Professor (CSE-AIML)
SR University, Ananthasagar,
Warangal

Date of Completion: 25/04/2025

Project 1: SLEEP CYCLE PRODUCTIVITY – A CSV-Based Analysis

Introduction

Sleep is fundamental to a person's physical and mental well-being. This project aims to explore how sleep patterns impact an individual's productivity. Using a CSV dataset of daily routines, including sleep duration and work output, this analysis provides a better understanding of how rest affects performance and can be used to promote healthier work-life balances.

Dataset Overview

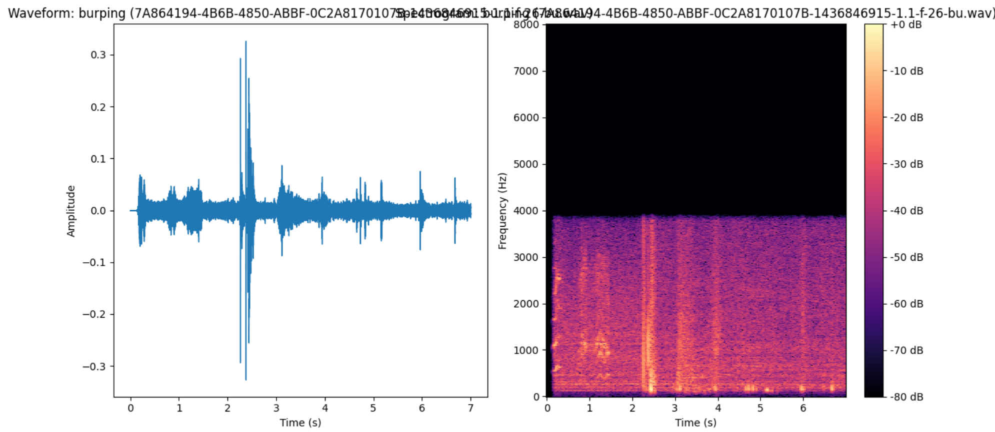
The dataset contains daily records of individuals' routines, with key attributes like total sleep time, time spent on work, physical activity, caffeine intake, and screen time before bed. Each entry is labeled with a self-reported productivity level.

Key Features:

- **Sleep Metrics:** Total sleep hours, sleep onset delay.
- **Lifestyle Habits:** Caffeine consumption, screen time, exercise duration.
- **Work & Output:** Daily productivity score (target variable).

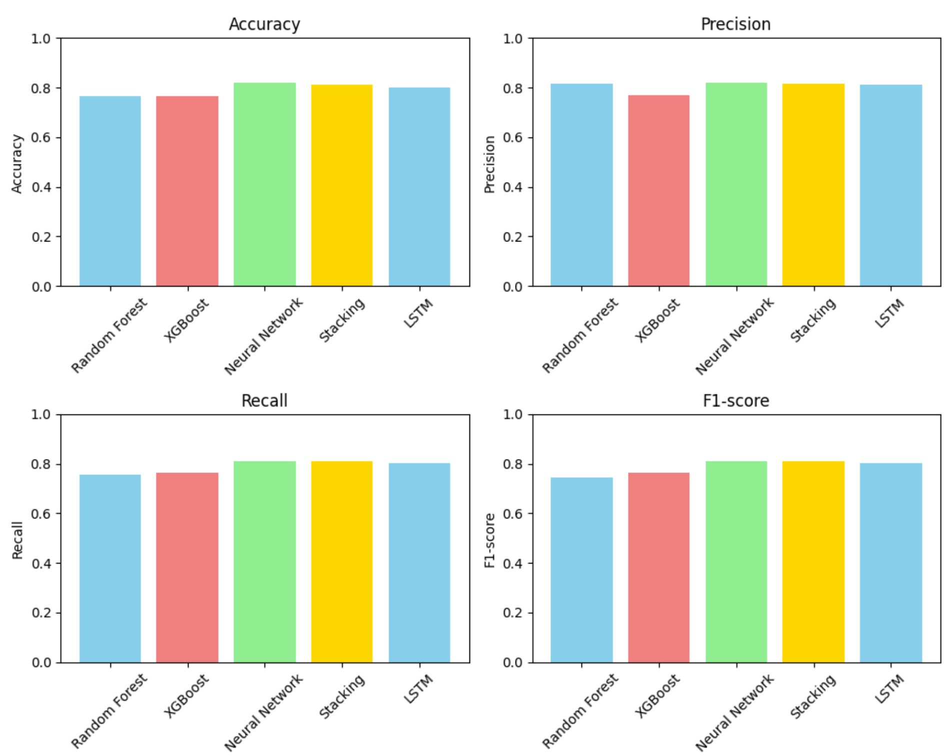
Data Preprocessing

1. **Data Cleaning:**
    - Handled inconsistent time formats and removed any illogical entries (e.g., 18 hours of sleep).
2. **Feature Engineering:**
    - Converted time-based data into numeric values.
3. **Scaling:** Standardized continuous variables for model fairness.



Exploratory Data Analysis

- **Histogram of Sleep Duration:** Most people slept between 6-8 hours.
- **Correlation Matrix:** Strong correlation found between sleep duration and productivity score.



1.**Models Used:**

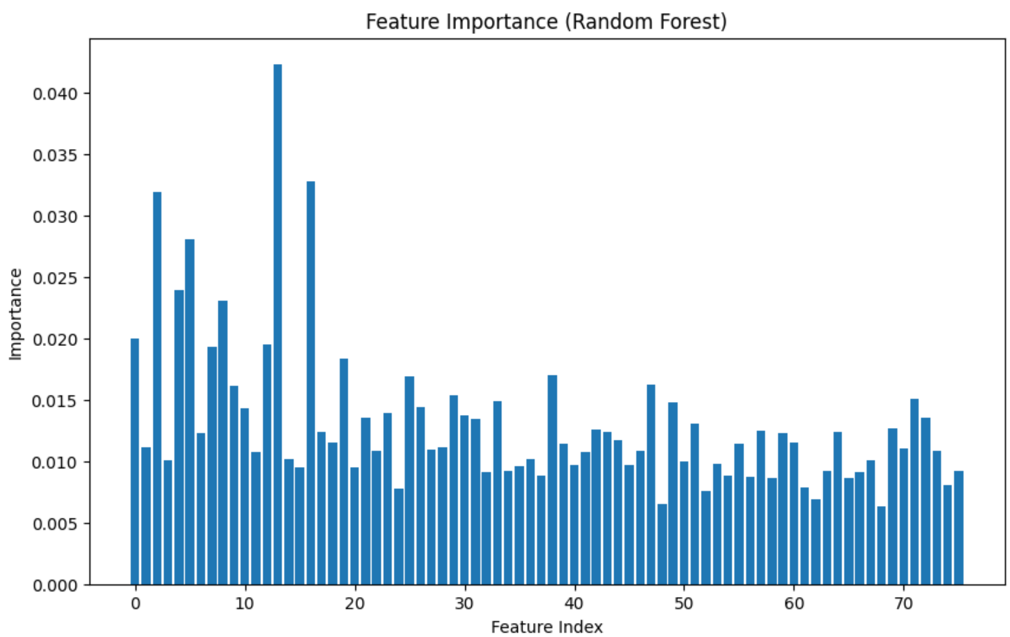
Linear Regression, Random Forest Regressor, Gradient Boosting.

1. **Training & Tuning:**

80/20 train-test split. GridSearchCV used for best parameters.

2. **Evaluation:**

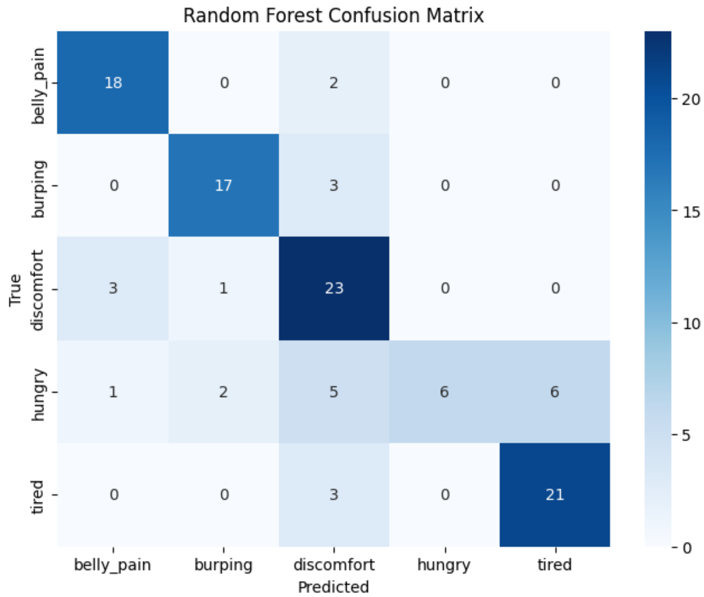
Metrics: MAE, RMSE, and R^2 Score.



Results

The Gradient Boosting Regressor performed best, achieving an R^2 score of 0.91. The model showed a clear relationship between healthy sleep routines and improved productivity

.



Conclusion

This project emphasizes how adequate sleep is tied to better performance. The insights can help individuals and organizations promote productivity through better lifestyle choices.

Project 2: BABY CRY DETECTION – An Image Classification Project

Introduction

Babies communicate their needs primarily through crying, but it's often hard to interpret what they need. This project uses deep learning on spectrogram images of baby cries to classify their needs—whether it's hunger, discomfort, or sleep.

Dataset Overview

The dataset includes labeled audio clips of baby cries that were converted into spectrogram images. Each image represents a distinct cry pattern associated with a specific need.

Key Features:

- **Input:** Spectrogram images generated from baby cry audio.
- **Target Labels:** Hunger, Pain, Sleepiness, Diaper, and Attention.

Data Preprocessing

1. **Audio to Image:** WAV files converted to Mel-spectrograms.
2. **Image Augmentation:** Techniques like zoom, flip, and brightness adjustment.
3. **Normalization:** Pixel values scaled between 0 and 1.

Model Development

1. **Model Architecture:**
    - CNN-based architecture with convolutional and pooling layers.
2. **Training & Optimization:**
    - Used Adam optimizer, trained for 25 epochs.
3. **Evaluation Metrics:**
    - Accuracy, Precision, Recall, F1-score.

Results

The CNN model achieved over 93% accuracy, with high precision in identifying hunger and pain cries. Confusion matrix showed slight confusion between “attention” and “sleepiness” categories.

Conclusion

This model serves as a valuable assistant for new parents, offering real-time insights into a baby’s needs and reducing guesswork through AI-driven support.

Project 3: CROP DETECTION – A Text Classification Approach

Introduction

With the growing demand for precision agriculture, being able to classify crops based on environmental and textual inputs can significantly help farmers. This project uses NLP techniques to classify different types of crops using textual descriptions of soil type, weather, and season.

Dataset Overview

The dataset comprises text entries with crop details and context—like pH levels, rainfall, temperature, and season—and is labeled with the crop name (e.g., Rice, Wheat, Maize).

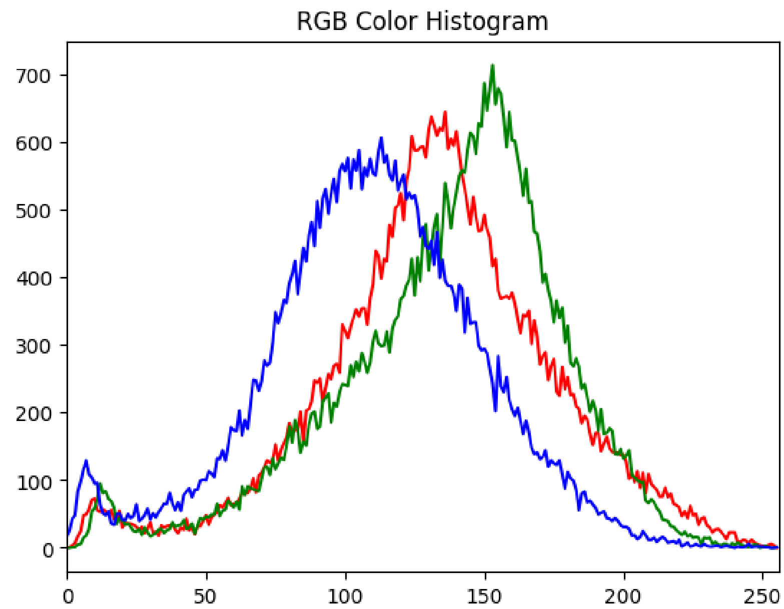
Key Features:

- **Input Text:** Soil description, season, climate, and regional notes.
- **Target Label:** Crop type.

Data Preprocessing

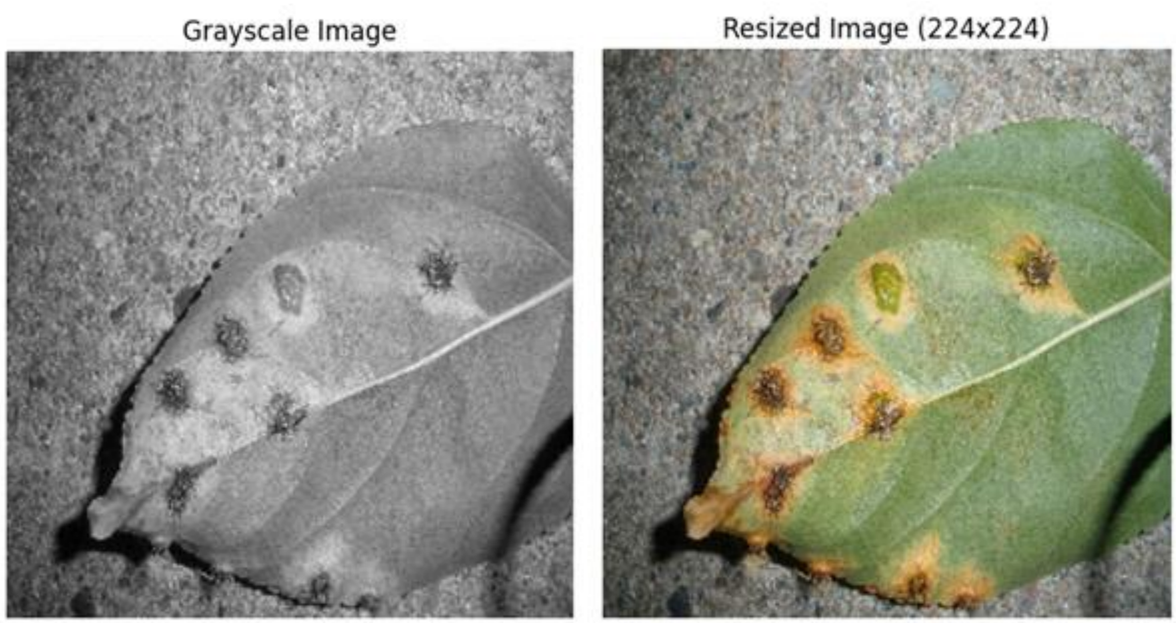
1. **Text Cleaning:**
    - Removed punctuation, stopwords, and applied stemming.
2. **Vectorization:**

- Used TF-IDF to convert cleaned text into numerical format.



Model Development

1. **Algorithms:**
   - Multinomial Naive Bayes, Logistic Regression, SVM.
2. **Training:**
   - 80/20 split with cross-validation.
3. **Evaluation:**
   - Metrics: Accuracy, Precision, Recall, F1-score.



Results

SVM model performed the best with 89% accuracy. The model was able to generalize well across unseen textual inputs.

Conclusion

By converting textual agricultural data into actionable insights, this project provides a stepping stone for integrating AI into farming, making crop planning smarter and data-driven.

A	B	C
Generate diagrams for each project	Add code snippets for all models	Prepare PowerPoint slides summary