A Project Report on

SLEEPING DISORDER PREDICTION USING MACHINE LEARING

Submitted in partial fulfilment of the requirements for award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

By

K. HARSHITHA

(20U41A0543)

K. PAVAN SAI KUMAR(20U41A0501) **N. MOUNIKA (20U41A0572)**

M.MANIDEEPAK (20U41A0542) **K. BHAGYA SRI (20U41A0574)**

Under the Esteemed guidance of

Mrs. T. SANTHOSHI LAKSHMI

Assistant professor, Department of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DADI INSTITUTE OF ENGINEERING & TECHNOLOGY

(AN AUTONOMOUS INSTITUTE)

(Approved by A.I.C.T.E., Permanently Affiliated to JNTU GV)

Accredited by NAAC with 'A' Grade and Inclusion u/s 2(f) & 12(B) of UGC Act

An ISO 9001:2015, ISO 14001:2015 & ISO 45001:2018 Certified Institute.

NH-16, Anakapalle – 531002, Visakhapatnam, A.P.

(2020-2024)

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the project report entitled **"SLEEPING DISORDER PREDICTION USING MACHINE LEARNING"** submitted by K. HARSHITA (20U41A0543), K. PAVAN SAI KUMAR (20U41A0501), M. MANIDEEPAK (20U41A0542), N. MOUNIKA (20U41A0572), K. BHAGYA SRI (20U41A0574). In partial fulfilment of the requirements for award of the Degree of **Bachelor of Technology** in **COMPUTER SCIENCE & ENGINEERING**, from Dadi Institute of Engineering & Technology(A), Anakapalle affiliated to JNTUGV, accredited by NAAC with 'A' grade is a record of bonafide work carried out by them under my guidance and supervision.

PROJECT GUIDE **HEAD OF THE DEPARTMENT**

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We would like to express our gratitude to our project guide, **T. SANTHOSHI LAXMI**, **Assistant Professor** of the Department of COMPUTER SCIENCE & ENGINEERING, for her guidance and assistance throughout the development of this project. Without her supervision, support, and encouragement, we would not have gained awareness of many new things during our project.

We convey our heartfelt thanks **to Dr. K. SUJATHA, Professor & Head of the Department** of COMPUTER SCIENCE & ENGINEERING, for motivating us to successfully complete the project.

We would like to thank Principal **Dr. R. Vaikunta Rao**, Dadi Institute of Engineering & Technology, for providing the necessary facilities to carry out our project work successfully.

We would like to thank Dean of Academics, **Dr. K. Parvathi**, Dadi Institute of Engineering & Technology, for providing essential facilities for our project.

We express our gratitude to the Teaching and Non-teaching Staff of the Department of COMPUTER SCIENCE & ENGINEERING, who have been directly and indirectly involved in this journey, for their encouragement in completing our project.

We would like express our deep sense of gratitude to **honorable Chairman**, **Sri Dadi Ratnakar**, of Dadi Institute of Engineering & Technology, for providing necessary facilities to carry out our project work successfully**.**

We wish to take this opportunity to express our deep gratitude to all the people who have extended their cooperation in various ways during our project. It is our pleasure to acknowledge the help of all those respected individuals.

PROJECT ASSOCIATES

K. HARSHITHA (20U41A0543)

K. PAVAN SAI KUMAR (20U41A0501)

M.MANIDEEPAK (20U41A0542)

N. MOUNIKA (20U41A0572)

K. BHAGYA SRI (20U41A0574)

DECLARATION

We hereby declare that the project entitled **"SLEEPONG DISORDER PREDICTION USING MACHINE LEARNING "** is submitted in partial fulfilment of the requirements for the award of Bachelor of Technology in **COMPUTER SCIENCE and ENGINEERING** under esteemed supervision of **T.SANTHOSHI LAXMI Asistant professor DEPT OF CSE**. This is a record of work carried out by us and results embodied in this project report have not been submitted to any other university for the award of any Degree.

PROJECT ASSOCIATES

K. HARSHITHA (20U41A0543)
K. PAVAN SAI KUMAR(20U41A0501)
M.MANIDEEPAK (20U41A0542)
N. MOUNIKA (20U41A0572)
K. BHAGYA SRI (20U41A0574)

ABSTRACT

Sleep disorders are a major public health problem, affecting millions of people worldwide. They can have a significant impact on an individual's health and wellbeing, leading to fatigue, decreased productivity, and increased risk of accidents and chronic diseases. The main objective of this data science project is to analyze various lifestyle and medical variables of individuals, such as age, BMI, physical activity, sleep duration, blood pressure, etc., and use this information to predict the occurrence and type of sleep disorder they may experience. Sleep disorders, like Insomnia and Sleep Apnea, can have significant impacts on an individual's health and overall well-being. The development of this predictive model has the potential to improve the early detection and treatment of sleep disorders. The potential impact of the project, such as the number of people who could be identified as being at risk of sleep disorders, or the cost savings that could be realized by early detection and treatment. By identifying individuals at risk of sleep disorders, appropriate interventions and treatments can be provided to improve their sleep quality and overall health.

TABLE OF CONTENTS

LIST OF CONTENTS

LIST OF FIGURES

CHAPTER - 1

INTRODUCTION

INTRODUCTION

Sleep, a fundamental pillar of health and well-being, Sleep disorders cast a long shadow over our well-being, affecting millions and often going undiagnosed. Traditional diagnostic methods, like sleep studies, can be a hurdle due to cost and inconvenience. This leaves many suffering in silence, battling the consequences of disrupted sleep. These consequences are far-reaching, impacting our physical and mental health. Chronic sleep disorders can lead to daytime fatigue, difficulty concentrating, and irritability. Over time, the health risks escalate, increasing the likelihood of developing conditions like heart disease, diabetes, and even depression. The economic burden is significant as well, with sleep disorders linked to decreased productivity and increased healthcare costs. Early detection is key to preventing these negative consequences many. Sleep disorders, characterized by disturbances in sleep patterns, affect a significant portion of the population. Unfortunately, traditional diagnostic methods for these disorders often prove to be expensive, time-consuming, and inconvenient, leading to underdiagnosis and delayed treatment. So we came with This project ventures into the exciting realm of Machine Learning (ML) to develop a novel approach for predicting the risk of sleep disorders.

Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create models that enable machines to perform tasks that would otherwise only be possible for humans, such as categorizing images, analyzing data, or predicting price fluctuations. Machine learning (ML), offers a promising avenue for improving sleep disorder prediction. By analyzing diverse data sources, ML models can learn complex patterns associated with different sleep conditions, offering several advantages

Early detection: ML models can potentially identify individuals at risk before symptoms become severe, facilitating early intervention and better disease management.

Accessibility: ML-based tools could provide preliminary assessments, potentially reducing reliance on expensive in-clinic tests.

Personalized insights: ML models can analyze individual factors like demographics, lifestyle habits, and physiological data to provide more personalized risk assessments and treatment recommendations.

Key research areas in this field include:

Developing accurate and reliable prediction models: This involves choosing appropriate algorithms, collecting high-quality data, and addressing challenges like data imbalance and missing values. Identifying important risk factors

ML models can help pinpoint key factors contributing to different sleep disorders, informing preventive strategies and treatment approaches. Integrating diverse data sources: Combining data from surveys, wearables, electronic health records, and other sources can offer a more comprehensive understanding of sleep health.

Ensuring ethical considerations: Data privacy, fairness, and interpretability of ML models are crucial aspects to address when developing these tools for clinical use. Overall, machine learning holds immense potential for revolutionizing sleep disorder prediction and management.

1.1 Project objective

Imagine a project that uses everyday info to predict sleep problems! We'd collect data on how you sleep (trackers or surveys) and consider things like age, weight, and health conditions. We'd even look at your coffee habit and exercise routine!

This project could be a game-changer! By catching sleep disorders early, people could improve their sleep habits and overall health. It could even reduce healthcare costs. Ultimately, this project aims to help people sleep better and live healthier lives!

1.2 Problem Statement

Sleeping disorders are a major public health problem, affecting millions of people worldwide. They can have a significant impact on an individual's health and well-being, leading to fatigue, decreased productivity, and increased risk of accidents and chronic diseases . Sleeping disorders encompass a broad spectrum of conditions, including insomnia, sleep apnea, narcolepsy, restless legs syndrome, and more. These disorders disrupt the natural sleep-wake cycle, resulting in inadequate or poor-quality sleep. . Moreover the consequences of sleeping disorders are far-reaching.

They contribute to a host of health problems, including cardiovascular diseases, diabetes, obesity, and mental health disorders like depression and anxiety. Impaired cognitive function, reduced work productivity, One of the biggest challenges in diagnosing sleep disorders is that the symptoms can be similar to those of other medical conditions. For example, fatigue is a symptom of many different medical conditions, so it is important to rule out other causes before diagnosing a sleep disorder

This problem statement underscores the urgent need for increased awareness, research, and accessible healthcare services to tackle sleeping disorders comprehensively. By doing so, we can mitigate the personal suffering, societal costs, and public health challenges associated with these prevalent conditions.

1.3 Scope

This project aims to develop a machine learning model to predict sleep disorders using easily accessible data.

We'll gather information on sleep patterns (duration, quality, trackers), demographics (age, weight), and lifestyle factors (food intake, exercise). This data will be cleaned and transformed to highlight relevant aspects like sleep efficiency (percentage of actual sleep time) and nighttime awakenings.

Next, various machine learning algorithms will be trained on this data, allowing them to identify patterns that differentiate healthy sleep from disrupted sleep. The best performing model will be chosen based on its accuracy and ability to correctly identify sleep disorders.

By achieving early detection, this project has the potential to empower individuals to improve their sleep hygiene and overall health. Additionally, it could lead to lower healthcare costs associated with undiagnosed sleep issues.

1.4 Methodology

After gathering this data, we'd clean it up to make it clear and consistent. Then comes the magic: transforming this data into clues for a computer program. We'd look at how efficiently you sleep (percentage of time actually sleeping) and how often you wake up at night. We'd even consider how sleepy you feel during the day.

Information from wearable devices, health apps, and medical records will be amalgamated into a rich dataset. This dataset will then undergo meticulous cleaning and preprocessing to ensure data quality and consistency. Subsequently, this curated dataset will serve as the foundation for the development of a predictive modeling algorithm. The model's primary objective will be to discern intricate relationships between predictor variables and the target variable, which, in this case, is the presence and type of sleep disorder. The model's efficacy will be rigorously tested and validated through a robust evaluation process.

Upon successful validation, the predictive model will be integrated into a user-friendly platform, allowing individuals to predict the likelihood and specific type of sleep disorder they may be susceptible to.

This proactive approach promises to empower individuals with valuable insights into their sleep health, enabling early intervention and personalized strategies for improved sleep quality and overall well-being. summary, our proposed solution leverages digital technology and predictive modeling to revolutionize the identification and management of sleeping disorders, offering a pathway to better sleep and enhanced quality of life for countless individuals.

CHAPTER -2

LITERATURE SURVEY

LITERATURE SURVEY

A literature survey for a sleeping disorder project would typically involve reviewing existing research, studies, and literature related to various aspects of sleep disorders

1. **Insomnia**:

    - Morin, C. M., Benca, R., Chronic Insomnia, The Lancet, 379(9821), 1129-1141. (2012)
    - Qaseem, A., Kansagara, D., Forciea, M. A., Cooke, M., Denberg, T. D., & Clinical Guidelines Committee of the American College of Physicians. Management of Chronic Insomnia Disorder in Adults: A Clinical Practice Guideline From the American College of Physicians. Annals of Internal Medicine, 165(2), 125-133. (2016)

2. **Sleep Apnea**:

    - Peppard, P. E., Young, T., Barnet, J. H., Palta, M., Hagen, E. W., & Hla, K. M. Increased prevalence of sleep-disordered breathing in adults. American Journal of Epidemiology, 177(9), 1006-1014. (2013)
    - Weaver, T. E., & Sawyer, A. M. Adherence to continuous positive airway pressure treatment for obstructive sleep apnea: implications for future interventions. Indian Journal of Medical Research, 131, 245–258. (2010)

3. **Restless Legs Syndrome (RLS)**:

    - Allen, R. P., Picchietti, D. L., Garcia-Borreguero, D., Ondo, W. G., Walters, A. S., Winkelman, J. W., ... & Zucconi, M. Restless legs syndrome/Willis–Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG)

consensus criteria–history, rationale, description, and significance. Sleep Medicine, 15(8), 860-873. (2014)

- Winkelman, J. W., Armstrong, M. J., Allen, R. P., Chaudhuri, K. R., Ondo, W., Trenkwalder, C., ... & Postuma, R. B. Practice guideline summary: Treatment of restless legs syndrome in adults: Report of the Guideline Development, Dissemination, Subcommittee of the American Academy of Neurology. Neurology, 87(24),

4. **Narcolepsy**:

- Thorpy, M. J., & Krieger, A. C. Delayed diagnosis of narcolepsy: characterization and impact. Sleep Medicine, 14(10), 1017-1021. (2013)
- Ruoff, C., Rye, D., & Parker, K. P. Understanding Narcolepsy: What It Is, What Causes It, and How to Treat It. Sleep, 31(7), 777–797. (2008)

5. **Circadian Rhythm Disorders**:

- Sack, R. L., Auckley, D., Auger, R. R., Carskadon, M. A., Wright, K. P., Vitiello, M. V., ... & Twery, M. Circadian rhythm sleep disorders: Part II, advanced sleep phase disorder, delayed sleep phase disorder, free-running disorder, and irregular sleep-wake rhythm. Sleep, 30(11), 1484–1501. (2007)
- Shanahan, T. L., & Czeisler, C. A. Light exposure induces equivalent phase shifts of the endogenous circadian rhythms of circulating plasma melatonin and core body temperature in men. Journal of Clinical Endocrinology & Metabolism, 73(2), 227-235. (1991)

6. **Impact on Health**:

- Cappuccio, F. P., D'Elia, L., Strazzullo, P., & Miller, M. A. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care, 33(2), 414-420. (2010)
- Grandner, M. A., Patel, N. P., Gehrman, P. R., Xie, D., Sha, D., Weaver, T., & Gooneratne, N. S. Who gets the best sleep? Ethnic and socioeconomic factors related to sleep complaints. Sleep Medicine, 11(5), 470–478. (2010)

These studies cover a range of sleep disorders, including insomnia, sleep apnea, restless legs syndrome, narcolepsy, and circadian rhythm disorders, as well as their impact on health and treatment approaches. They provide valuable insights into the current state of knowledge in the field of sleep medicine.

Junoh, A.K.; Ahmad, F.K.; Mohsen, M.F.M.; Alazaidah, R. Open research directions for multi label learning. In Proceedings of the 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 28–29 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 125–128.

Anbarasi, L.J.; Jawahar, M.; Ravi, V.; Cherian, S.M.; Shreenidhi, S.; Sharen, H. Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown. Health Technol. 2022, 12, 825–838.

CHAPTER -3.

EXISTING SYSTEM

EXISTING SYSTEM

3.1 Insomnia Diagnosis and Management:

Clinical Evaluation: The gold standard for diagnosing insomnia is a comprehensive evaluation by a sleep specialist. This typically involves a detailed sleep history, physical examination, and potentially sleep questionnaires. In some cases, a polysomnography (PSG) test might be necessary, which involves monitoring brain waves, muscle activity, breathing, and eye movements during sleep.

Cognitive Behavioral Therapy for Insomnia (CBT-I): This is the first-line treatment for chronic insomnia. CBT-I focuses on identifying and modifying negative thoughts and behaviors that contribute to sleep problems. It can include techniques such as sleep restriction, stimulus control, relaxation training, and cognitive restructuring.

Medications: Prescription sleep medications might be considered for short-term relief of insomnia, particularly when CBT-I is not effective. However, their use should be monitored by a healthcare professional due to potential side effects and dependency risks.

3.2 Sleep Apnea Diagnosis and Management:

Polysomnography (PSG): This is the definitive test for diagnosing sleep apnea. PSG monitors various physiological parameters during sleep, allowing doctors to identify episodes of breathing cessation or shallow breathing characteristic of sleep apnea.

Continuous Positive Airway Pressure (CPAP): CPAP is the most effective treatment for obstructive sleep apnea, the most common type. A CPAP machine delivers constant and positive air pressure through a mask worn during sleep, preventing airway collapse.

Auto-CPAP: For some patients, an auto-CPAP machine can be used. It automatically adjusts airway pressure throughout the night to maintain an open airway.

3.3 Restless Legs Syndrome (RLS) Diagnosis and Management:

Clinical Evaluation: Similar to insomnia, RLS diagnosis involves a thorough medical history and physical examination. Additionally, specific criteria established by the International Restless Legs Syndrome Study Group (IRLSSG) are used to confirm the diagnosis.

Dopamine Agonists: Medications that mimic the effects of dopamine in the brain are the primary treatment for RLS. Several types are available, and a doctor will determine the most suitable option based on individual needs and side effects.

Iron Therapy: Iron deficiency can sometimes contribute to RLS. In such cases, iron supplementation might be recommended.

3.4 Narcolepsy Diagnosis and Management:

Multiple Sleep Latency Test (MSLT): This test measures daytime sleepiness by recording sleep patterns during multiple naps scheduled throughout the day. Excessive sleepiness and rapid eye movement (REM) sleep onset during the MSLT are indicative of narcolepsy.

Sodium Oxybate: This medication helps regulate sleep-wake cycles and reduce daytime sleepiness in narcolepsy patients. Other medications like stimulants can also be used to manage symptoms.

Lifestyle Modifications: Maintaining a consistent sleep schedule, avoiding caffeine and alcohol before bed, and engaging in regular exercise can significantly improve sleep quality for people with narcolepsy.

3.5 Limitations of Existing systems

1.Accessibility and Cost:

Specialist Availability: Consulting a sleep specialist can be challenging due to limited availability, especially in remote areas. This can lead to delays in diagnosis and treatment.

Cost of Diagnostics: Sleep studies like PSG are expensive, and insurance coverage might not always be comprehensive. This can prevent some individuals from accessing necessary diagnostic tools.

Treatment Costs: Medications and CPAP machines can be costly, posing a financial barrier for some patients.

2. Accuracy and Limitations:

Subjectivity in Diagnosis: Diagnoses often rely on clinical evaluation and patient history, which can be subjective. This can lead to misdiagnosis or delayed diagnosis, especially for complex sleep disorders.

Limited Scope of Diagnostics: Sleep studies like PSG capture a snapshot of sleep, and some sleep disorders might not manifest during the study. Additionally, these tests might not be suitable for everyone due to discomfort or claustrophobia**.**

One-Size-Fits-All Approach: Current treatment options often take a "one-size-fits-all" approach. Individual responses to medications and therapies can vary significantly.

3. Long-Term Management:

Treatment Adherence: Long-term adherence to therapies like CPAP or medications can be challenging due to side effects or lifestyle disruptions.

Lack of Personalized Support: Existing systems often lack long-term support programs to help patients manage their sleep disorders effectively and adjust treatments as needed.

Limited Focus on Preventative Measures: The focus is often on treating existing sleep disorders rather than addressing underlying factors and promoting healthy sleep habits to prevent their occurrence in the first place.

4. Technological Limitations:

Accuracy of Wearables and Apps: While wearable devices and sleep tracking apps are becoming more popular, their accuracy in diagnosing specific sleep disorders is still under development.

Limited Integration: Existing systems might not seamlessly integrate with new technologies, potentially hindering data sharing and comprehensive sleep management.

CHAPTER -4

PROPOSED SYSTEM

Our proposed solution aims to revolutionize the identification and management of sleeping disorders by developing an advanced predictive model.

This model will harness the power of digital gadgets and smart devices to collect comprehensive data on individuals' lifestyles and medical variables.To initiate this transformative process, researchers will embark on data collection, focusing on a wide array of parameters relevant to sleep health. Summary,

our proposed solution leverages digital technology and predictive modeling to revolutionize the identification and management of sleeping disorders, offering a pathway to better sleep and enhanced quality of life for countless individuals

The core component of the proposed system is the sleep disorder prediction module, which utilizes machine learning algorithms to predict the sleep disorder based on attributes

4.1 Demographic attributes

such as:

Column Name	Description
Person_ID	Unique ID assigned to each person
Gender	The gender of the person (Male/Female)
Age	Age of the person in years

Column Name	Description
Occupation	The occupation of the person
Sleep_duration	The duration of sleep of the person in hours
Quality_of_sleep	A subjective rating of the quality of sleep, ranging from 1 to 10
Physical_activity	The level of physical activity of the person (Low/Medium/High)
Stress Level	A subjective rating of the stress level, ranging from 1 to 10
BMI_category	The BMI category of the person (Underweight/Normal/Overweight/Obesity)
Blood_pressure	The blood pressure of the person in mmHg
Heart_rate	The heart rate of the person in beats per minute
Daily Steps	The number of steps taken by the person per day
Sleep_disorder	The presence or absence of a sleep disorder in the person (None, Insomnia, Sleep Apnea)

4.2 Dataset Description:

The dataset used for this project is called the "**Sleep Health and Lifestyle Dataset**." It consists of 400 rows (individuals) and 13 columns (variables) that cover a wide range of information related to sleep patterns and daily habits. The dataset includes the following key features:

Comprehensive Sleep Metrics: This section allows exploring various sleep-related metrics such as sleep duration, quality of sleep, and factors influencing sleep patterns.

Lifestyle Factors: The dataset provides insights into lifestyle factors such as physical activity levels, stress levels, and BMI categories, which may have an impact on an individual's sleep health.

Cardiovascular Health: The dataset includes measurements of blood pressure and heart rate, which are crucial indicators of an individual's cardiovascular health and may have a correlation with sleep disorders.

Sleep Disorder Analysis: The primary focus of this project is to identify the presence or absence of sleep disorders in individuals. The dataset labels individuals with three categories in the "Sleep Disorder" column:

- None: Individuals who do not exhibit any specific sleep disorder.
- Insomnia: Individuals who experience difficulty falling asleep or staying asleep, leading to inadequate or poor-quality sleep.
- Sleep Apnea: Individuals who suffer from pauses in breathing during sleep, resulting in disrupted sleep patterns and potential health risk

4.3 Work flow of Proposed System

After cleaning and preparing the data, a model is trained on 80% of it, learning patterns that correlate with different sleep disorders. The remaining 20% is used for testing, ensuring the model works well with unseen data. Rigorous evaluation techniques confirm the model's accuracy, and comparisons with traditional diagnostics like sleep studies further validate its reliability.

The key benefit lies in early detection, allowing for prompt intervention and personalized treatment plans. However, limitations exist. The model's accuracy depends on high-quality data, and its decision-making process might be difficult to interpret. While not a replacement for healthcare professionals, this model can be a valuable tool to assist in diagnosis and promote better sleep health overall.

Future advancements could involve incorporating data from wearable devices and improving the model's transparency for even more effective sleep disorder prediction.

4.4 Advantages of Proposed System

Early Detection: ML models can potentially identify individuals at risk for developing sleep disorders before symptoms become severe. This allows for earlier intervention and treatment, potentially leading to better long-term outcomes.

Improved Diagnosis: ML models can act as a valuable tool to assist healthcare professionals in the diagnostic process. By analyzing large amounts of data, they can highlight potential sleep disorders that might be missed in traditional evaluations

Personalized Care: By identifying individuals at risk for specific sleep disorders, ML can help tailor treatment plans to individual needs. This can lead to more effective and targeted therapies.

Cost Reduction: Early detection and intervention of sleep disorders can potentially reduce healthcare costs associated with managing chronic and severe sleep issues.

Resource Optimization: ML models can help healthcare systems allocate resources more efficiently by identifying individuals most likely to benefit from sleep studies or specialist consultations.

Large-Scale Screening: ML models can analyze data from large populations, potentially enabling proactive screening efforts to identify individuals at risk for sleep disorders, especially in underserved communities.

CHAPTER -5

SYSTEM REQUIREMNTS

5.1 Functional Requirements:

1. Pandas:

Pandas works with many different types of data sets such as comma-separated values (CSV) files, Excel files, extensible markup language (XML) files, JavaScript object notation (JSON) files and relational database tables.

Data read from these sources are returned as Pandas data types known as DataFrame and Series. Dealing with the same data types across the board is convenient because it allows us to read data from one data source, manipulate it and insert it into another data source without worrying about the data's syntax.

2. NumPy:

NumPy is a fundamental library for numerical computing in Python. It supports large, multi- dimensional arrays and matrices, along with a collection of mathematical functions for efficient computations.

3. Matplotlib:

Matplotlib plot is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc. In Matplotlib plot various states are preserved across function calls, so that it keeps track of things like the current figure and plotting area, and the plotting functions are directed 2 6 to the current axes (please note that "axes" here and in most places in the documentation refers to the axes part of a figure and not the strict mathematical term for more than one axis)

4. Pickle:

Pickle is a Python module used for object serialization, allowing the saving and loading of Python objects to/from disk.

5. Scikit-learn:

Scikit-learn (sklearn) is a popular machine-learning library in Python that provides a diverse set of tools and algorithms for tasks such as classification, regression, clustering, and more

6. Seaborn

Python provides a numerous number of libraries for data visualization, we have already seen the Matplotlib library in this article we will know about Seaborn Library. Seaborn is an amazing visualization library for statistical graphics plotting in Python.

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data

5.2 Non Functional Requirements:

Non-functional requirements specify the qualities or characteristics of the system that are not directly related to its functionality but are essential for its overall performance, usability, and maintainability.

Here are some non-functional requirements for sleep disorder prediction model

1. Performance
2. Accuracy
3. Security
4. Reliability

5. Usability
6. Maintainability
7. Compliance

5.3 Hardware Requirements:

Hardware requirements are essential for ensuring optimal performance, reliability, scalability, compatibility, and security of the system. By meeting these requirements, organizations can effectively support their software applications and deliver a positive user experience.

- System version: Windows 7 or above
- Processor: Intel i3 or above/ Ryzen
- Hard Disk: 512GB HDD/SSD or above
- RAM: 4GB or above

5.4 Software Requirements:

software requirements are essential for guiding the development process, aligning with stakeholder needs, managing risks, ensuring quality, facilitating communication, supporting change management, and achieving regulatory compliance.

By capturing and documenting these requirements effectively, organizations can increase the likelihood of project success and deliver high-quality software solutions.

- Technology/Language: Python
- Operating System: Windows/Mac/Linux/Ubuntu
- IDE: Jupyter, Google colab
- Libraries: Numpy, Pandas, Matplotlib, Seaborn

CHAPTER -6

TECHNOLOGY ADOPTED

TECHNOLOGY ADOPTED

Machine Learning (ML) offers a promising alternative, with the potential to predict sleep disorders based on a variety of data points. This technology can empower individuals to take a more proactive approach to their sleep health and assist healthcare professionals in early detection and intervention.

6.1 Machine Learning for Sleep Prediction

1.Pattern Recognition: ML excels at identifying patterns in large datasets. In sleep disorder prediction, these datasets can include information about individuals with and without sleep disorders. By analyzing this data, ML models can learn to recognize patterns associated with specific sleep disorders, such as insomnia, sleep apnea, or restless leg syndrome.

2.Predictive Modeling: Once trained on historical data, ML models can be used to predict the likelihood of an individual developing a sleep disorder based on their own characteristics and habits. This allows for early intervention and personalized treatment plans.

3.Supervised Learning: This dominant approach involves training models on labeled data, where each data point is categorized as having a specific sleep disorder or healthy sleep. Common algorithms include:

Logistic Regression: Identifies the relationship between factors like age and BMI with the presence or absence of a sleep disorder.

Support Vector Machines (SVMs): Create clear separations between data points belonging to different categories (sleep disorder vs healthy sleep).

Decision Trees: Classify data based on a series of sequential decisions, offering interpretability for understanding key risk factors.

Random Forests: Combine multiple decision trees for improved accuracy and reduced overfitting.

4.1 Unsupervised Learning: This approach can be used for data exploration and anomaly detection, identifying unusual sleep patterns that might indicate a potential sleep disorder. Techniques include:

Clustering: Group similar sleep patterns together, potentially revealing clusters associated with specific sleep disorders.The choice of algorithm depends on the specific goals of the prediction task, the nature of the data available, and computational resources.

6.2 Data Acquisition

Building the Foundation the success of any ML model for sleep prediction hinges on data quality. Data sources for sleep disorders can include:

Electronic Health Records (EHRs): Clinical databases containing patient information, sleep study reports, and diagnoses.

Wearable Devices: Smartwatches and fitness trackers that record sleep data like duration, sleep stages, and heart rate variability.

Smartphone Apps: Sleep monitoring apps that track sleep patterns and collect user-reported sleep quality data.

6.3 Evaluation Metrics:

Evaluating the model's effectiveness is crucial to ensure its reliability. Common metrics used include:

Accuracy: The proportion of correctly classified cases (individuals with and without the sleep disorder).

Precision: The proportion of positive predictions that are actually true positives (individuals with the sleep disorder).

Recall: The proportion of actual cases with a sleep disorder that are correctly identified by the model.

F1 Score: A harmonic mean of precision and recall, providing a balanced view of the model's performance.

CHAPTER -7

FEASIBILITY STUDY

1.Data Availability:

Access to diverse and high-quality datasets is paramount for training accurate and robust machine learning (ML) models for predicting sleep disorders. However, obtaining annotated datasets that encompass a wide range of demographic characteristics and sleep disorder phenotypes may pose challenges. Collaboration with healthcare institutions, sleep clinics, and research organizations can facilitate access to relevant data sources. Additionally, efforts should be made to ensure data represent various age groups, genders, ethnicities, and comorbidities commonly associated with sleep disorders.

2.Computational Resources:

ML model training and validation often require significant computational resources, including high-performance computing clusters or cloud platforms. The feasibility of deploying ML-based sleep disorder prediction systems hinges on the availability of such resources. Organizations must assess their computational infrastructure and budgetary constraints to determine the feasibility of implementing ML models for sleep disorder prediction. Optimization techniques, such as model compression, parallelization, and distributed computing, can help mitigate resource requirements and enhance scalability.

3.Model Interpretability:

Ensuring the interpretability of ML models is crucial for gaining insights into prediction outcomes and fostering trust among end-users, including healthcare providers and patients. While complex ML algorithms may achieve high predictive performance, their black-box nature can hinder understanding of the underlying decision-making process.

3.Ethical and Regulatory Considerations:

Compliance with data privacy regulations (e.g., GDPR, HIPAA) and adherence to ethical guidelines for human subjects research are paramount when developing ML-based sleep disorder prediction systems. Organizations must establish robust data governance frameworks to safeguard sensitive health information and ensure data security and confidentiality. Additionally, obtaining informed consent from participants and obtaining institutional review board (IRB) approval for research involving human subjects are essential ethical considerations. Organizations should conduct thorough risk assessments and implement appropriate safeguards to mitigate potential ethical and legal risks associated with ML-based sleep disorder prediction.

4.Clinical Validation:

Validating the clinical utility and effectiveness of ML-based sleep disorder prediction models is critical before widespread deployment in clinical settings. Organizations must conduct rigorous clinical validation studies to assess the accuracy, sensitivity, specificity, and generalizability of ML models compared to standard diagnostic methods, such as polysomnography and clinical assessments by sleep specialists. Collaborating with healthcare providers and sleep medicine experts is essential for designing validation studies, recruiting participants, and interpreting results. Robust clinical validation is necessary to demonstrate the clinical value of ML-based sleep disorder prediction systems and gain acceptance from healthcare stakeholders.

5. Cost-Benefit Analysis:

Conducting a comprehensive cost-benefit analysis is essential to evaluate the economic feasibility of implementing ML-based sleep disorder prediction systems. Organizations must consider the upfront costs associated with data acquisition, model development, computational infrastructure, and personnel training, as well as ongoing maintenance and operational expenses. Comparing the potential benefits, such as improved diagnostic accuracy, reduced healthcare costs, and enhanced patient outcomes, against the investment required can help organizations make informed decisions about the feasibility of adopting ML-based sleep disorder prediction systems.

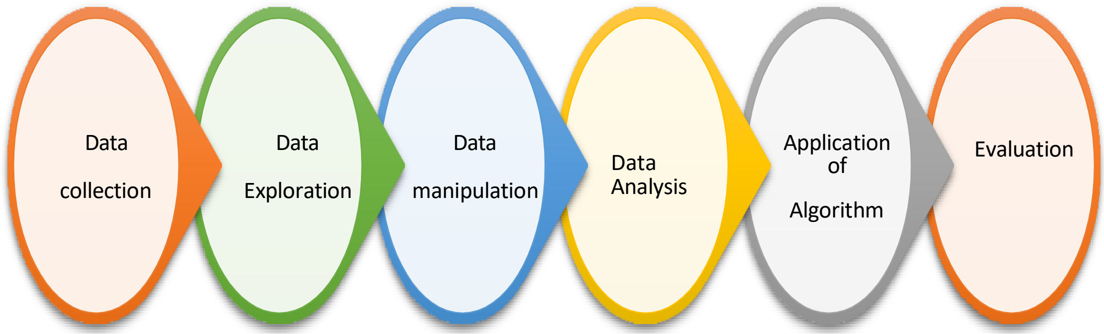
6.Stakeholder Engagement:

Engaging key stakeholders, including healthcare providers, patients, researchers, policymakers, and technology vendors, is crucial throughout the feasibility assessment process. Stakeholder input can provide valuable insights into the specific needs, challenges, and priorities related to sleep disorder prediction and inform the development and implementation of ML-based solutions. Collaborative partnerships and stakeholder engagement initiatives can foster buy-in, facilitate knowledge exchange, and ensure the alignment of ML-based sleep disorder prediction systems with clinical workflows and patient preferences.

CHAPTER -8

SYSTEM DESIGN

ARCHITECTURAL FLOW OF THE PROPOSED SYSTEM:



An Architectural Diagram or a pipeline is used to help automate machine learning workflows. They operate by enabling a sequence of data to be transformed and correlated together in a model that can be tested and evaluated to achieve an outcome, whether positive or negative.

The pipeline/ Diagram consists of several steps to train a model. Machine learning pipelines are iterative as every step is repeated to continuously improve the accuracy of the model and achieve a successful algorithm. To build better machine learning models, and get the most value from them, accessible, scalable and durable storage solutions are imperative, paving the way for on- premises object storage.

The steps include:

Data Collection: Collecting raw data from billions of datasets available.

Data Exploration: Exploring the data & the features related and being familiar with the data-types.

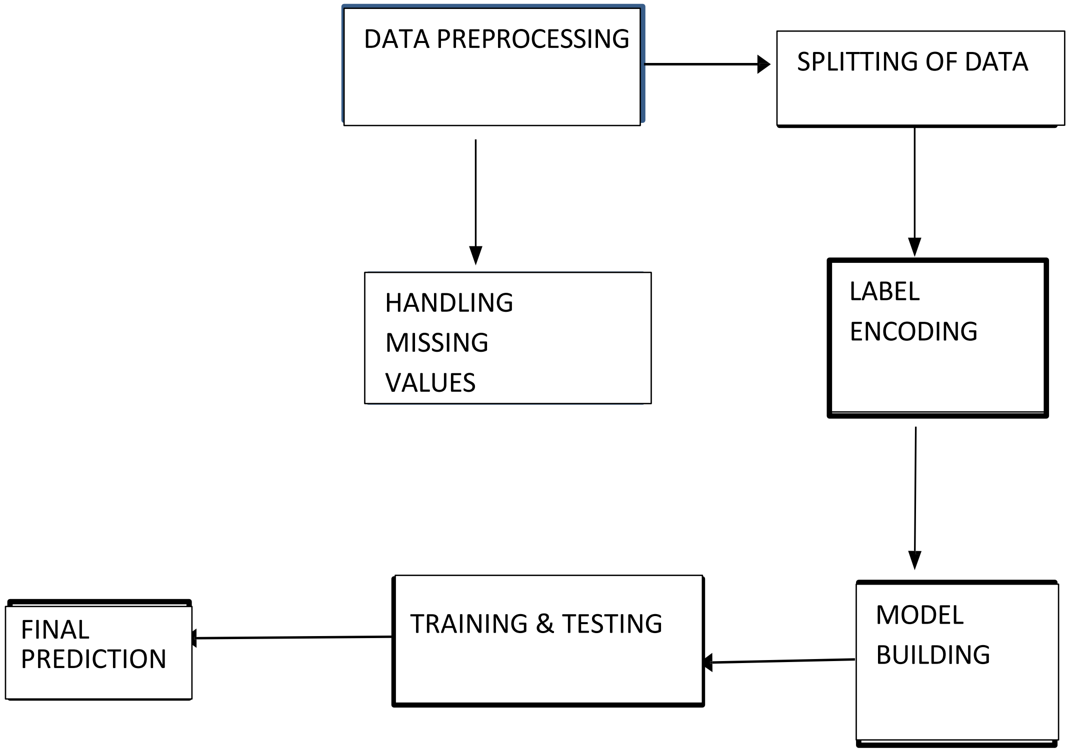
Data Manipulation: Includes Cleaning of data, treating missing, repetitive values that are present.

Data Analysis: Analysing the data to increase efficiency while applying the best Algorithm & feature selection according to our preferences.

Application of Algorithm: Applying the algorithm to the model.

Evaluation: Using evaluation metrices to calculate the least error and following the above to make further changes.

8.1 SYSTEM ARCHITECTURE



8.2 DATA ACQUISITION:

To start our project, we needed data to train our machine learning model. We found a great resource called the "Sleep Health and Lifestyle Dataset" on a website called Kaggle. This dataset is like a big table with information about 400 people. Each person has 13 pieces of information, including things like their age, sex, job, how long they sleep, and how well they sleep. We also learned about things that might affect their sleep, like how much they exercise and how stressed they feel.

This data was perfect for training our model because it has the information we care about (sleep patterns) and things that might influence it (lifestyle factors). Using this data, we were able to build a strong foundation for our sleep disorder prediction model. However, in the future, we might want to use data from even more sources to make our model even better!

8.3 PREPROCESSING PHASE

Before we unleash the power of machine learning on our sleep health dataset, we need to ensure the data is clean and ready for analysis. This is where data preprocessing comes in – it's like tidying up your room before inviting guests over. Here's a breakdown of the key steps we took:

Missing Value Detectives: First, we played data detective and identified any missing values, which are like empty spots in our table. In the "Sleep Disorder" column, for instance, we might find some missing entries. To handle these, we replaced them with a placeholder value like "None," making it clear there's no data present.

Blood Pressure Split: We noticed the blood pressure data was combined into one column. To make analysis easier, we split it into two separate columns, one for systolic and one for diastolic pressure. This separation allows us to explore how each value might influence sleep patterns.

Unique Value Check: Next, we became data librarians, cataloging the number of unique values in each column. This helps us understand the variety of data points present. For example, the "occupation" column might have hundreds of unique entries, while "gender" might only have a few.

BMI Standardization: We identified entries in the "BMI" column labeled as "normal weight." To ensure consistency, we replaced these with simply "normal" for better clarity throughout the analysis.

Categorical Column Cataloguing: We paid close attention to columns containing categories, like "occupation" or "sleep quality." We examined the unique values in each one, ensuring they were all valid options and identifying any potential inconsistencies.

Exploratory Data Analysis (EDA): Think of EDA as data visualization detective work. We used various charts and graphs to explore the relationships between different features. This helped us understand how factors like age, physical activity, and stress levels might be linked to sleep patterns and potential sleep disorders.

Label Encoding for Categorical Variables: Finally, we tackled the challenge of feeding categorical data (like "gender" or "occupation") into our machine learning model. These models typically work best with numerical data. So,

we used a technique called label encoding, where each category is assigned a unique number. This allows the model to understand the relationships between these categories without getting confused by the text labels themselves.

8.4 MODEL SELECTION

Now that we've prepped our sleep health data, it's time to introduce the heroes of this project: machine learning algorithms! Our goal is to predict different sleep disorders, like insomnia and sleep apnea. But to achieve this, we need algorithms that can learn from the patterns in the data. Here, we'll delve into two powerful contenders: Random Forests and Decision Trees.

The Decision Tree: Imagine a flowchart with questions at each branch. You answer the question, and depending on your answer, you move to a different part of the chart. This is essentially how a Decision Tree works. It analyzes our sleep data, asking a series of questions about factors like age, sleep duration, stress levels, and more.

Based on the answers, it navigates through a series of branches until it reaches a final conclusion – in our case, a prediction of a specific sleep disorder (insomnia, sleep apnea, or none).

The beauty of Decision Trees lies in their interpretability. We can see the questions it asks and understand the logic behind its predictions. This can provide valuable insights into how different factors contribute to sleep problems. However,

Decision Trees can sometimes become overly complex, focusing on specific details that might not generalize well to unseen data.

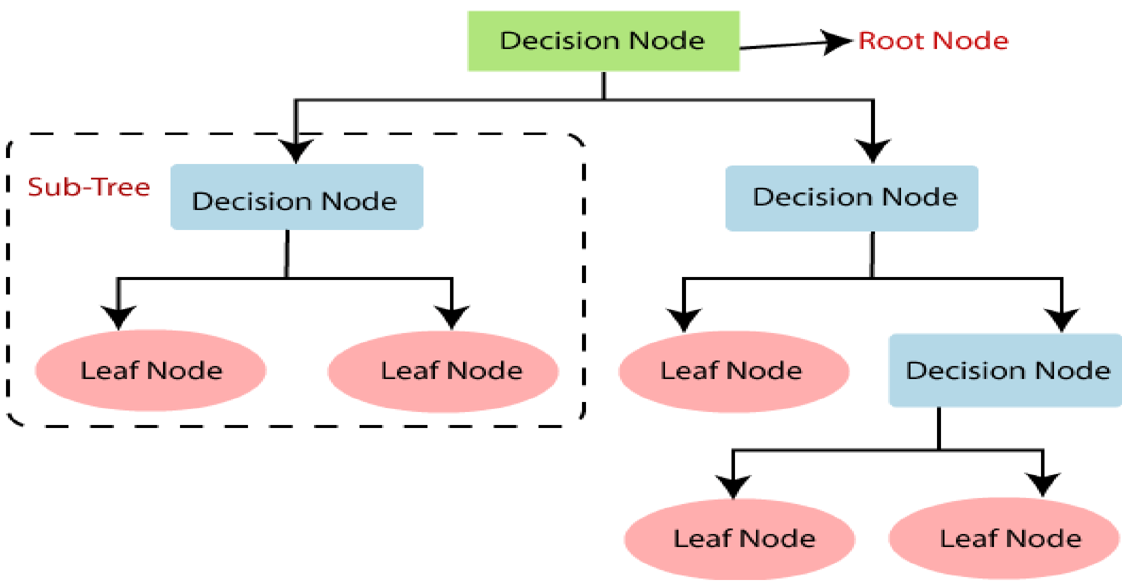
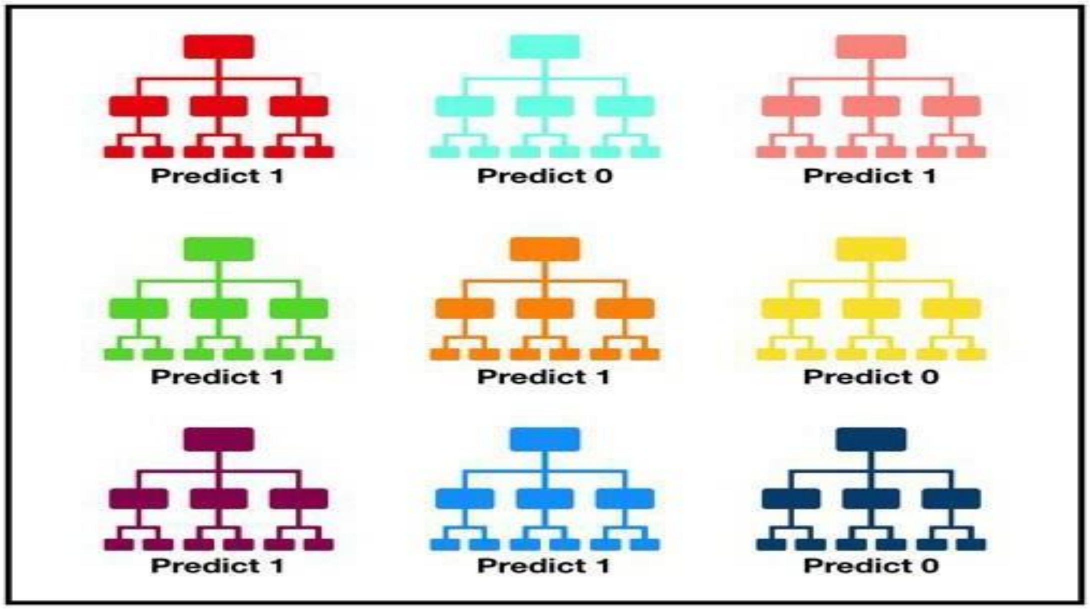


Fig: DECISION TREE

. FIG: RANDOM FOREST



The Random Forest: Enter Random Forests! Imagine having a whole forest of Decision Trees, each one slightly different from the others. Random Forests build multiple Decision Trees, each trained on a random subset of the data and using a random selection of features at each split. This diversity helps to avoid overfitting and creates a more robust model.When a new data point arrives, it's passed through all the trees in the forest, and each tree makes a prediction.

The final prediction is based on the majority vote – whichever sleep disorder receives the most votes from the individual trees wins. This approach leverages the strengths of individual Decision Trees while reducing their weaknesses, leading to potentially more accurate and generalizable predictions.

The Random Forest model achieved an impressive accuracy of 89%, indicating it can correctly predict sleep disorders in nearly nine out of ten cases.

This is further solidified by the average F1 score of 0.86, which highlights the model's effectiveness in balancing precision and recall. Compared to the Decision Tree Classifier, the Random Forest demonstrates a clear advantage in accuracy, suggesting it generalizes better and makes fewer mistakes on unseen data. These results are promising and pave the way for further exploration of the Random Forest model's potential in sleep disorder prediction.

8.5 MODEL TRAINING

Now that we have our prepped sleep health data and a chosen champion – the Random Forest algorithm – it's time to enter the training phase! This is where the magic happens. We'll feed the data into the algorithm, allowing it to learn the intricate relationships between various sleep factors (age, stress, sleep duration etc.) and the presence or absence of sleep disorders (insomnia, sleep apnea).

Imagine showing the Random Forest countless examples from our data. Each example is like a student in a classroom, with features like age and stress levels representing their characteristics, and the sleep disorder diagnosis acting as the outcome we want the model to predict. By analyzing these examples repeatedly, the Random Forest builds a complex internal map, essentially learning how to identify patterns that differentiate healthy sleep from potential sleep disorders.

There's more to training than just feeding data. We fine-tuned the algorithm's parameters, like the number of trees in the forest and the number of features considered at each split. This optimization ensures the model learns effectively without becoming overly complex or focusing on irrelevant details.

The model gives pretty decent results with an accuracy of 87% and an average F1 score of 0.83. This means the model can predict the sleep disorder with a good accuracy – in nearly nine out of ten cases, it can identify the correct issue! With this well-trained athlete (model) on our side, we're ready to see how well it performs on completely new data, assessing its effectiveness in predicting sleep disorders in the real world!

8.6 MODEL DEPLOYMENT

Our project utilizes server-side deployment with Flask, a lightweight Python web framework. This means the user interface (HTML form) acts as a front-end, collecting sleep data. Flask, on the back-end, receives this data and acts as a bridge, sending it to the trained machine learning model for sleep disorder prediction. Finally, Flask retrieves and formats the prediction for display on the user's web page. This approach separates user interaction from complex processing, creating a smooth user experience.

Here's how it works:

User Submits Information: A user visiting our website enters their information (age, stress level, sleep duration etc.) through the HTML form. The browser captures and packages this data.

Flask Takes the Wheel: The user's data isn't processed directly in the browser. Instead, it's sent to a Python script powered by Flask, a lightweight web framework. Flask acts as a traffic cop, receiving the user's data from the browser.

Flask Interacts with the Model: Flask doesn't make predictions itself. It acts as a bridge between the user's data and our trained machine learning model, likely saved as a separate Python file. Flask intelligently sends the user's information to the model for analysis.

Prediction Powerhouse: The model receives the data from Flask and performs its calculations, leveraging the patterns it learned during training. Based on this analysis, the model generates a prediction – the potential sleep disorder (insomnia, sleep apnea, or none)

Flask Delivers the Results: Once the prediction is made, Flask retrieves the output from the model. It then transforms this raw prediction into a user-friendly format, suitable for displaying on the webpage. Dynamic HTML Update: Flask doesn't require a complete page refresh to show the results. It might use techniques like AJAX (Asynchronous JavaScript and XML) to dynamically update the HTML page. This allows the predicted sleep disorder to appear seamlessly on the user's screen without them needing to reload the entire page

.

CHAPTER – 9

SYSTEM IMPLEMENTATION

9.1 IDE'S

Our sleep disorder prediction project takes flight in the implementation phase. Here, we leverage the strengths of two powerful tools: Google Colab and Visual Studio Code (VS Code)

.Google Colab: A Cloud Playground for Machine Learning

Imagine a vast online workspace equipped for serious computational tasks – that's Google Colab in a nutshell. This free Jupyter notebook environment, hosted by Google, provides the perfect platform to develop and test our machine learning model.

Python Playground: We can write and run Python code directly within Colab notebooks. This makes it ideal for experimenting with different algorithms (like Random Forests) and training our model to predict sleep disorders. Since Colab offers access to powerful computing resources (GPUs and TPUs) in the cloud, it can significantly accelerate the training process, especially for complex models.

Data Manipulation Mastery: Colab empowers us to tackle data preprocessing and exploration tasks with ease. Libraries like pandas and NumPy become our allies within Colab notebooks, allowing us to clean, transform, and analyze our sleep health data before feeding it to the model for training.

Experimentation Engine: The beauty of Colab lies in its flexibility. We can experiment with various model architectures (e.g., Random Forest vs. Decision Tree) and fine-tune hyperparameters directly in Colab notebooks. This iterative approach allows us to optimize the model's performance for the best possible sleep disorder prediction accuracy.

VS Code: Local Editing with Collaboration Power

While Colab shines in cloud-based execution, VS Code steps in as our local code editor of choice. Here's how it complements Colab's strengths:

Coding Efficiency Boost: VS Code provides a user-friendly interface with features like syntax highlighting, code completion, and debugging tools. This significantly enhances the development experience, making it smoother and more efficient to write clean and maintainable code.

Collaboration Made Easy: Version control systems like Git become a breeze with VS Code. This allows seamless collaboration with team members, ensuring everyone works on the latest version of the code and tracks changes effectively.

Local Development Flexibility: While Colab offers cloud resources, VS Code empowers us to develop and test specific project aspects locally on our machines. This can be particularly useful for building the Flask application structure, writing unit tests for functionalities, or making quick code edits without relying solely on the cloud environment.

9.2 UML DIAGRAMS

9.2.1 Use Case Diagram : A use case diagram is a graphical representation that depicts how a system interacts with its users (actors) and illustrates the various ways in which users can interact with the system. It is a part of the Unified Modeling Language (UML), which is commonly used in software engineering and system design.

Key components of a use case diagram include:

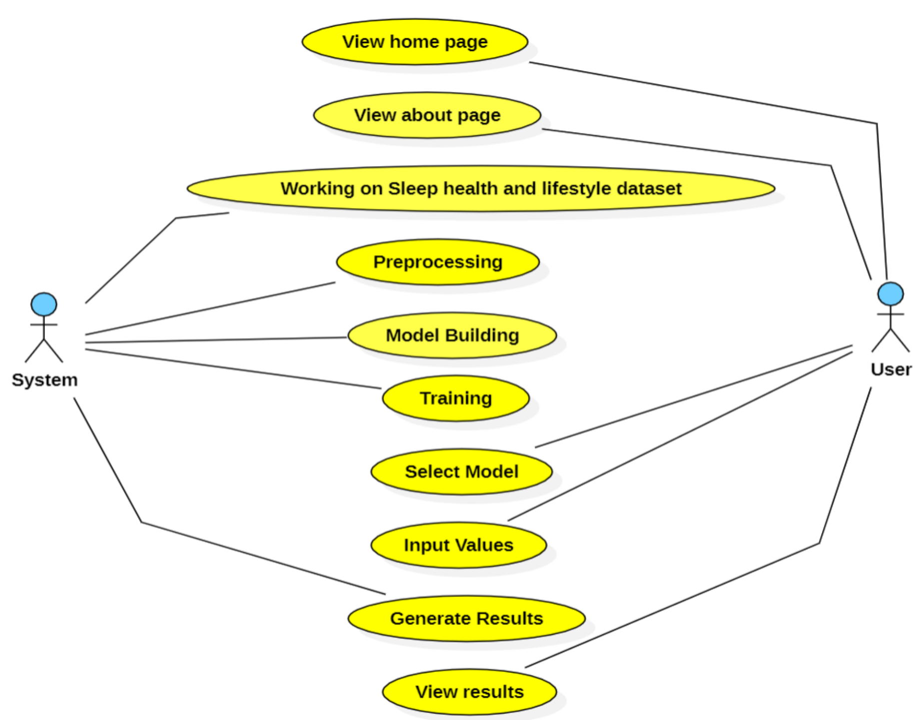
1. **Actors:** - Actors are external entities that interact with the system. They can be individuals, other systems, or even hardware devices. In the diagram, actors are represented as stick figures or blocks outside the system boundary.

2**. Use Cases:** - Use cases represent specific functionalities or tasks that the system performs. They describe the interactions between the system and its users. Use cases are depicted as ovals within the system boundary.

3. **System Boundary:** - The system boundary defines the scope of the system being modeled. It encloses all the use cases and actors involved in the system, highlighting what is inside and what is outside of the system.

4. **Associations:** - Associations, represented by lines connecting actors and use cases, show the relationships between them. An association line indicates that an actor is involved in a particular use case.

5. **Include and Extend Relationships**: - Include relationships indicate that one use case includes the behavior of another use case. Extend relationships indicate that one use case can extend the behavior of another under certain conditions. These relationships are represented by arrows connecting use cases.



9.2.2 Class Diagram :

A class diagram is a type of UML (Unified Modeling Language) diagram that provides a visual representation of the structure and relationships within a system or application. It focuses on the static aspects of a system, depicting classes, their attributes, methods, and the associations between them. Class diagrams are widely used in software engineering during the design and analysis phases to model the structure of a system.

Key elements of a class diagram include:

Class: A class is a blue print or template for creating objects. It represents a set of objects that share common attributes and behaviors. In the diagram, classes are typically depicted as rectangles with three compartments: the top compartment contains the class name, the middle compartment lists the class attributes, and the bottom compartment includes the class methods.

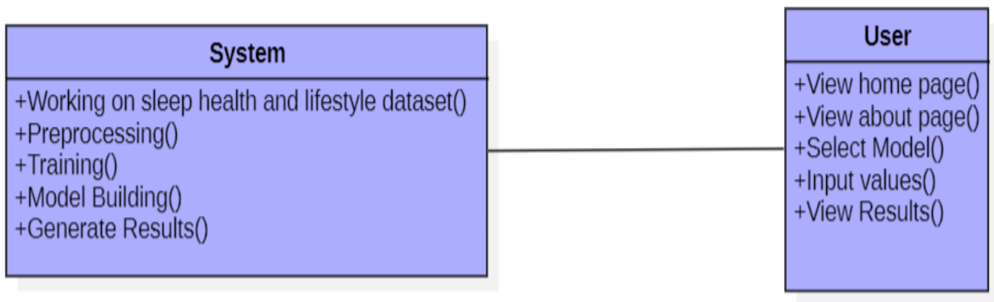
Attributes: Attributes are the properties or characteristics of a class. They describe the data that each object of the class holds. Attributes are listed in the middle compartment of the class rectangle. XXXIX

Methods: Methods represent the behaviors or functions that the objects of a class can perform. They are listed in the bottom compartment of the class rectangle and include details such as method names, parameters, and return types.

Associations: Associations represent relationships between classes. They describe how classes are connected and interact with each other. Associations are typically shown as lines connecting classes, with optional multiplicity indicators to specify the number of instances involved in the relationship.

Inheritance(Generalization): Inheritance represents an "is-a" relationship between classes, indicating that one class (subclass or derived class) inherits the attributes and methods of another class (superclass or base class). Inheritance is shown with an arrow pointing to the superclass.

Dependency: Dependency indicates that one class depends on another. It is represented by a dashed line with an arrow pointing to the class being depended upon.



9.2.3 Activity Diagram:

Activity Diagram:

An activity diagram is a UML (Unified Modeling Language) diagram that visually represents the flow of actions or activities with in a system or process. It is particularly useful for modeling the dynamic aspects of a system, illustrating the sequence of activities, actions, and decisions that take place during the execution of a use case or a business process.

Key elements of an activity diagram include:

Activity Nodes:

Activities are represented by rounded rectangles and denote tasks or operations that are performed within the system. They can represent anything from simple actions to complex processes.

Control Flows:

Control flows are arrows that connect activity nodes, indicating the sequence in which activities are performed. They represent the flow of control from one activity to another.

Decisions(Diamonds):

Decision nodes, depicted as diamonds, represent points in the process where a decision must be made. Depending on the outcome of the decision, different paths or alternative flows may be followed.

Forks and Joins:

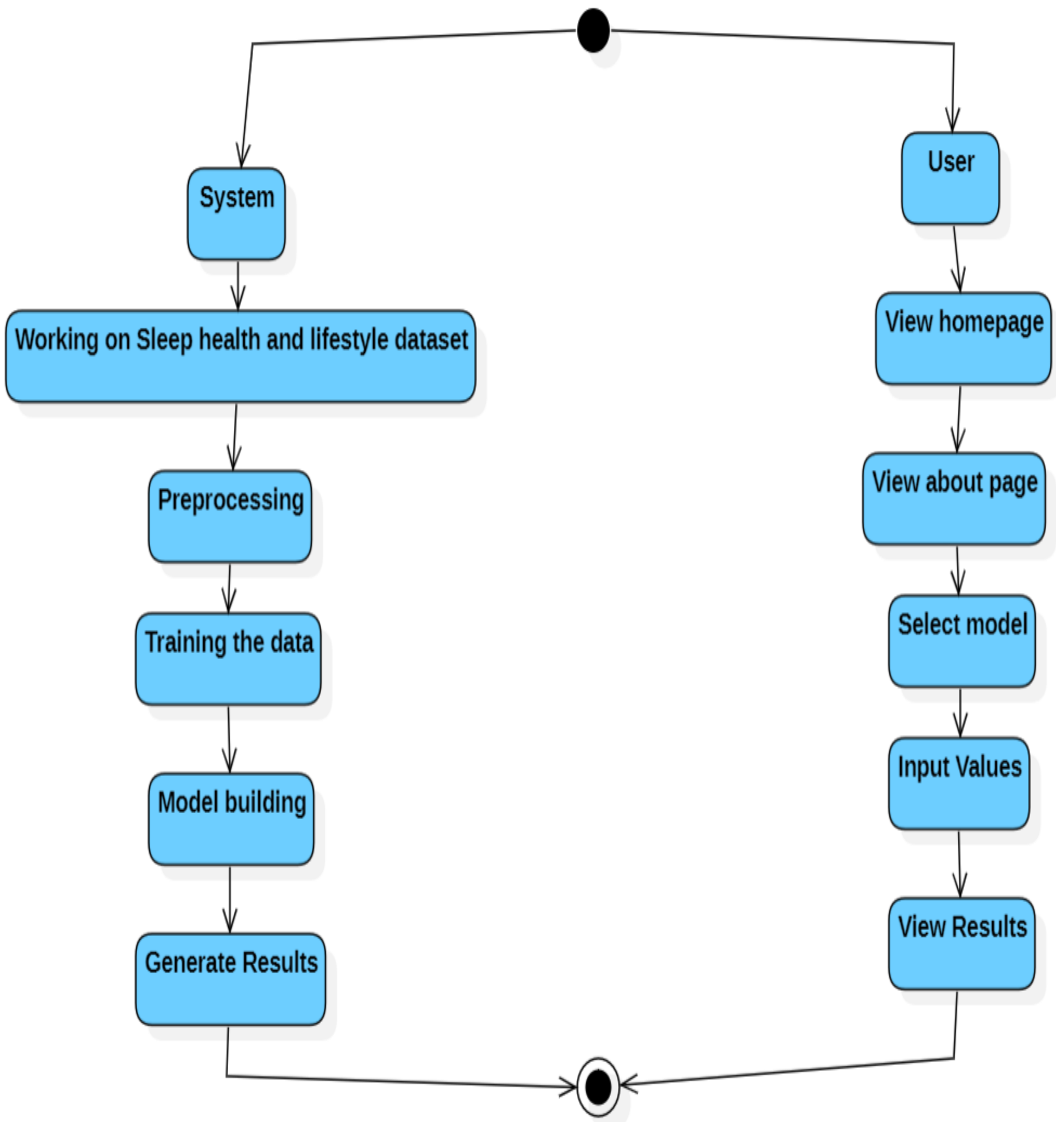
Fork nodes (horizontal bars) represent a point where the flow splits into multiple parallel paths.Join nodes (a circle with a bar) indicate where parallel paths converge back into a single flow.

Start and End Nodes:

The activity diagram typically begins with a start node (a solid circle) and ends with an end node(a solid circle with an outer ring). These nodes indicate the starting and finishing points of the activity.

Objects:

Objects can be represented in an activity diagram to show which objects participate in specific activities. However, it's worth noting that activity diagrams focus more on the flow of actions and less on the specific objects involved.



9.2.4 Sequence Diagram:

A sequence diagram is a type of UML (Unified Modeling Language)diagram that illustrates the interactions and order of messages between different components or objects within a system. It represents a dynamic view of a system, capturing the flow of events over time and emphasizing the sequence in which interactions occur.

Key elements of a sequence diagram include:

Lifelines: Life lines represent the various components or objects participating in the sequence diagram.They are depicted as vertical lines, typically labeled with the name of the object or component.

Activation Boxes:

Activation boxes represent the period during which an object is actively engaged in processing or performing an action. They are depicted as rectangles extending vertically from a lifeline and show the duration of an operation.

Messages:

Messages are arrows indicating communication between lifelines. They depict the flow of information or requests between objects. Messages may be synchronous (indicated by a solid line) or asynchronous (indicated by a dashed line).

Return Messages:

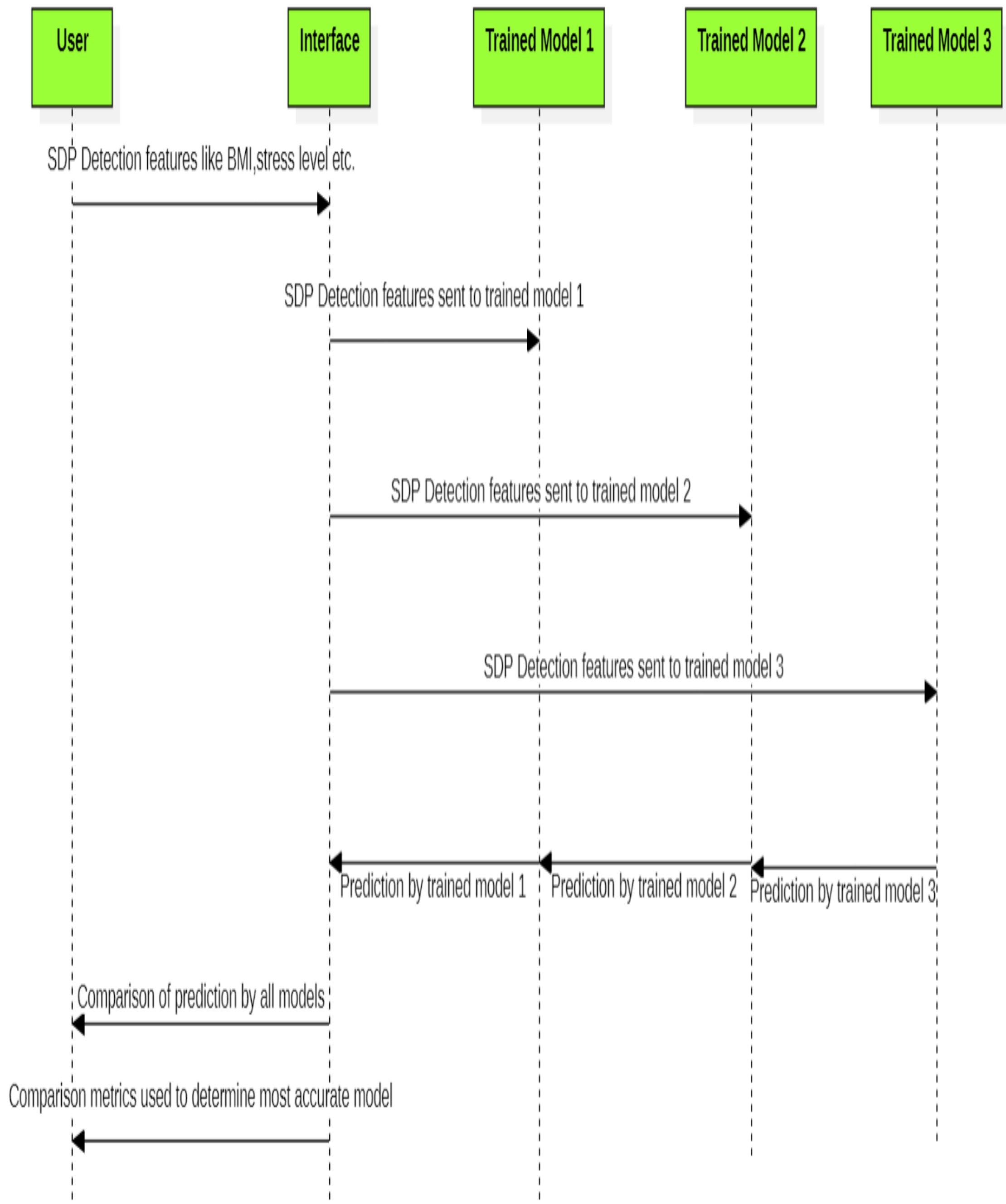
Return messages show the response from an object to a previously sent message. They are usually depicted with an arrow returning to the sender lifeline.

Focus of Control:

The focus of control is the vertical space within the activation box where an object is actively processing. It highlights the specific period when an object is executing an operation.

Interaction Fragments:

Interaction fragments, such as alt (alternative), opt (optional), and loop, allow the representation of conditional and repetitive behavior within the sequence diagram


User
Interface
Trained Model 1
Trained Model 2
Trained Model 3
SDP Detection features like BMI,stress level etc.
SDP Detection features sent to trained model 1
SDP Detection features sent to trained model 2
SDP Detection features sent to trained model 3
Prediction by trained model 1
Prediction by trained model 2
Prediction by trained model 3
Comparison of prediction by all models
Comparison metrics used to determine most accurate model

CHAPTER-10

OBJECTIVES

OBJECTIVES

Key Objectives for Sleep Disorder Prediction using Machine Learning:The objectives for using Machine Learning (ML) in sleep disorder prediction share some similarities with Chronic Kidney Disease (CKD) detection, but with a focus on sleep health. Here's a breakdown of the key goals:

1. Early Intervention: Similar to CKD, sleep disorders often go undiagnosed until problems become severe. ML models can analyze data to identify individuals at risk, enabling earlier intervention and treatment to improve sleep quality and overall health.

2. Improved Diagnostic Tools: Polysomnography, the gold standard for sleep disorder diagnosis, is expensive and inconvenient. ML models, trained on sleep study data, could provide a non-invasive and potentially more accessible screening tool for initial risk assessment.

3. Personalized Treatment Plans: ML can analyze a patient's sleep data (from wearables or questionnaires) alongside medical history to recommend personalized treatment options. This could include behavioral modifications, therapy, or targeted medications.

4. Risk Stratification: By identifying factors that contribute to sleep disorders, ML can help stratify individuals based on their risk level. This allows healthcare professionals to prioritize interventions for those at highest risk.

5. Sleep Health Promotion: ML models can be incorporated into mobile apps or wearable devices to provide feedback and promote healthy sleep habits. This could involve tracking sleep patterns, offering personalized sleep hygiene recommendations, and monitoring progress.

CHAPTER-11

CONCLUSIONS

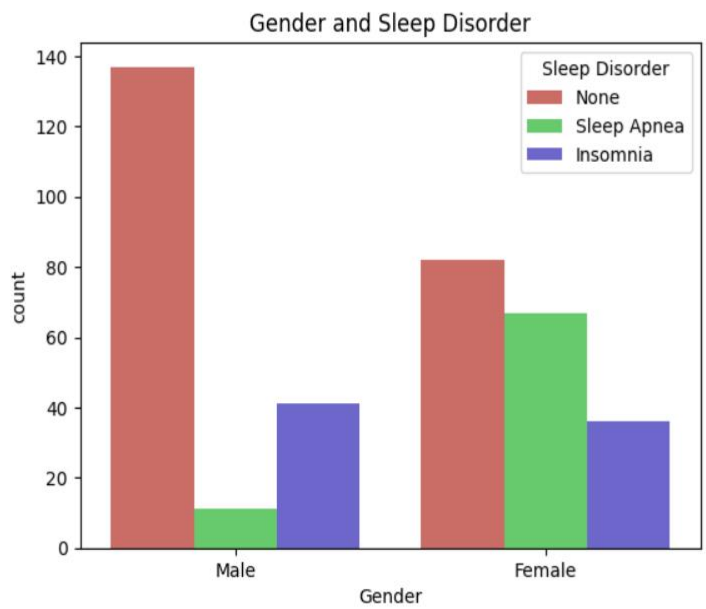
This sleep disorder prediction project aimed to identify the underlying factors influencing sleep health and develop a machine learning model for accurate classification. Exploratory data analysis revealed three key variables impacting sleep disorders: gender, occupation, and Body Mass Index (BMI).

The analysis identified a significant association between gender and specific sleep disorders. Males exhibited a higher prevalence of insomnia, characterized by difficulty falling asleep or staying asleep. Females, on the other hand, showed a greater tendency towards sleep apnea, a condition where breathing repeatedly stops and starts during sleep. These findings suggest potential sex-based differences in sleep physiology or risk factors.

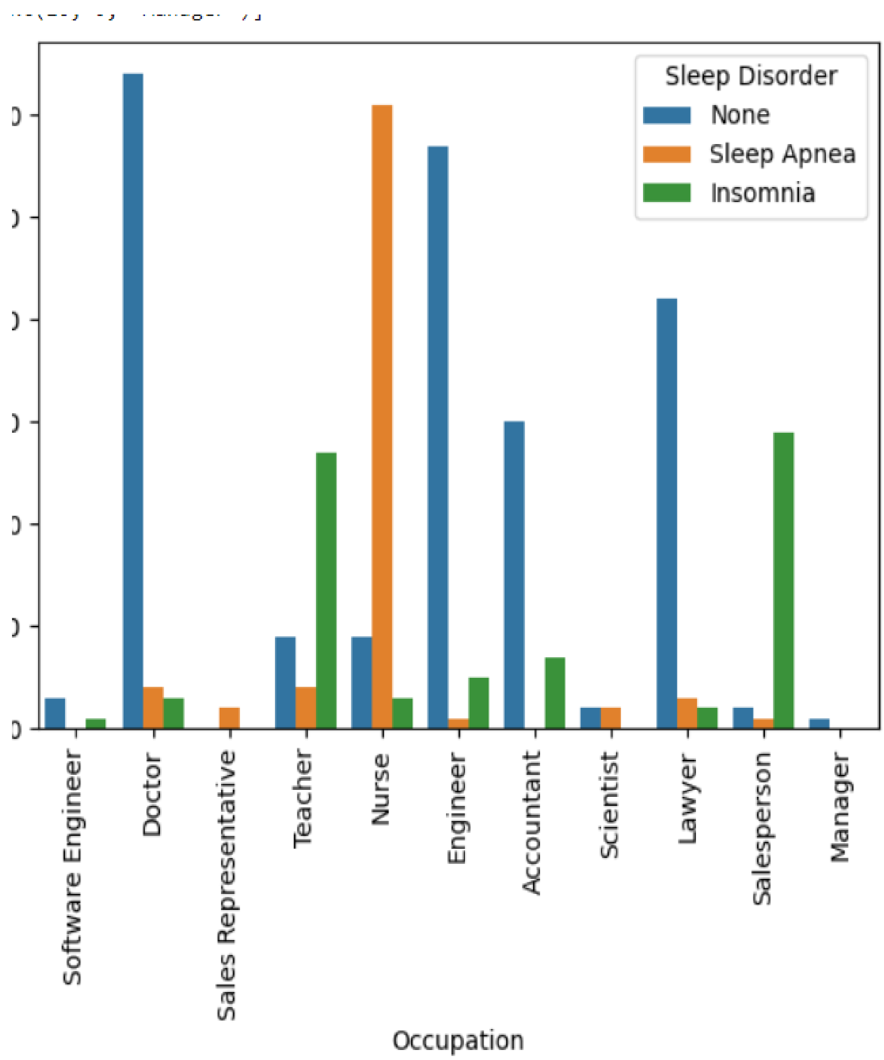
Gender and Sleep Disorder

```
#Gender count plot
sns.countplot(x = 'Gender', data = df, palette = 'hls', hue = 'Sleep Disorder').set_title('Gender and Sleep Disorder')
```

```
Text(0.5, 1.0, 'Gender and Sleep Disorder')
```



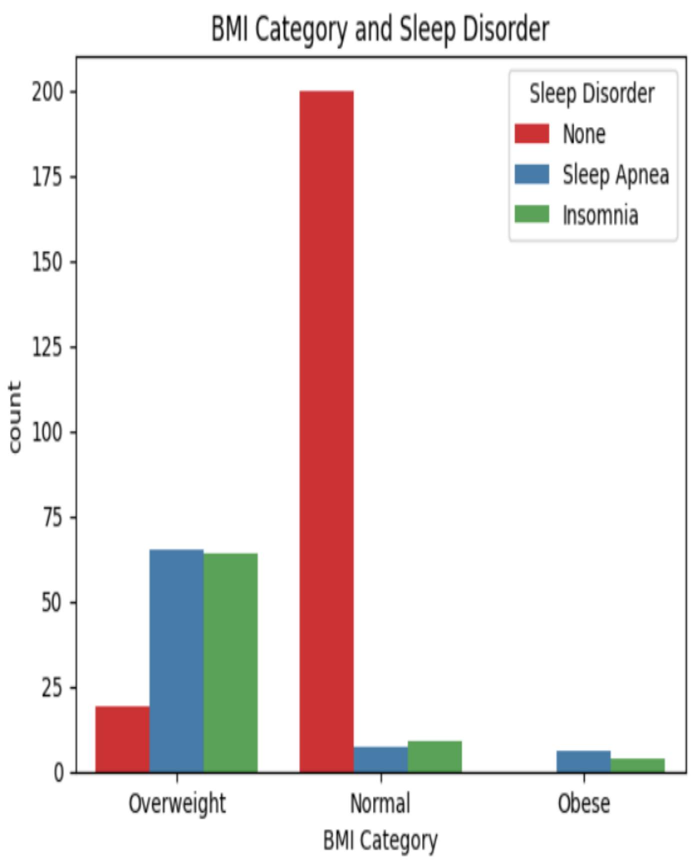
Occupations also emerged as a noteworthy influencer. Professions with demanding schedules or shift work, such as nursing, appeared to correlate with a higher risk of sleep disorders. This aligns with the understanding that irregular sleep patterns and stress associated with certain jobs can disrupt sleep quality.



Finally, BMI played a crucial role. Individuals classified as obese or overweight displayed a heightened susceptibility to sleep disorders. This reinforces existing knowledge about the link between weight and sleep, where excess weight can contribute to sleep apnea and other sleep-related issues.

```
sns.countplot(x = 'BMI Category', hue = 'Sleep Disorder', data = df, palette = 'Set1').set_title('BMI Category and Sleep
```

```
Text(0.5, 1.0, 'BMI Category and Sleep Disorder')
```



The project then evaluated two machine learning models for their effectiveness in sleep disorder prediction. Both models yielded promising results, demonstrating the potential of machine learning in this domain.

Classification Report

```
from sklearn.metrics import classification_report
print(classification_report(y_test, d_pred))
```

```
              precision    recall  f1-score   support

           0       0.74      0.83      0.78        24
           1       0.91      0.95      0.93        62
           2       0.90      0.70      0.79        27

    accuracy                           0.87       113
   macro avg       0.85      0.83      0.84       113
eighted avg       0.87      0.87      0.87       113
```

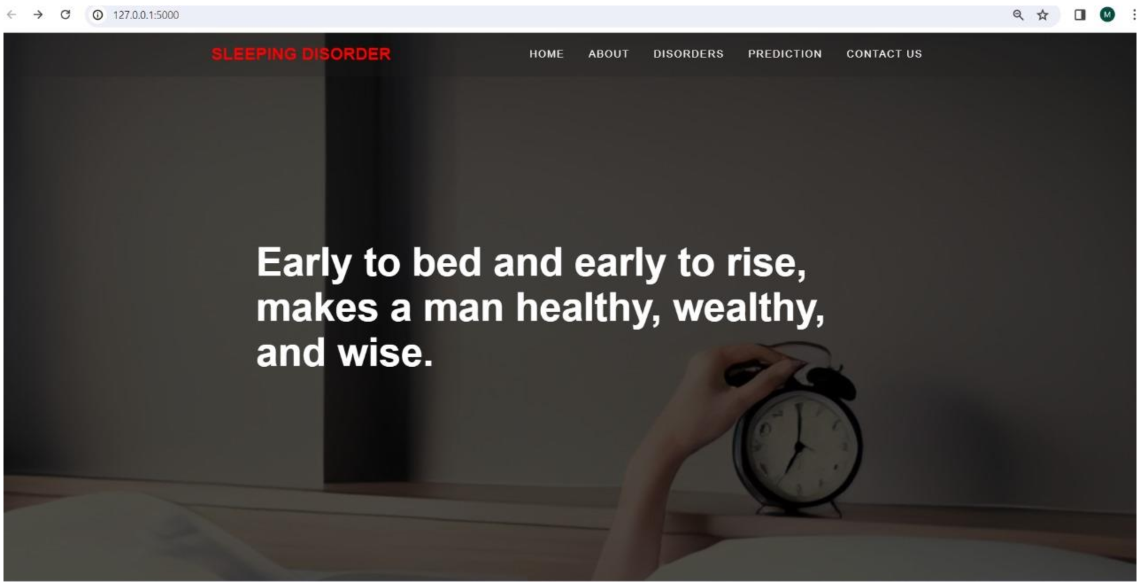
The model gives pretty decent results with an accuracy of 87% and an average F1 score of 0.83. The model is able to predict the sleep disorder with a good accuracy.

However, the Random Forest Classifier stood out with an impressive accuracy of 87%. This suggests that the model effectively learned the complex relationships between the identified factors (gender, occupation, BMI) and various sleep disorders. This high accuracy is encouraging and holds promise for developing a reliable tool for early detection and intervention.

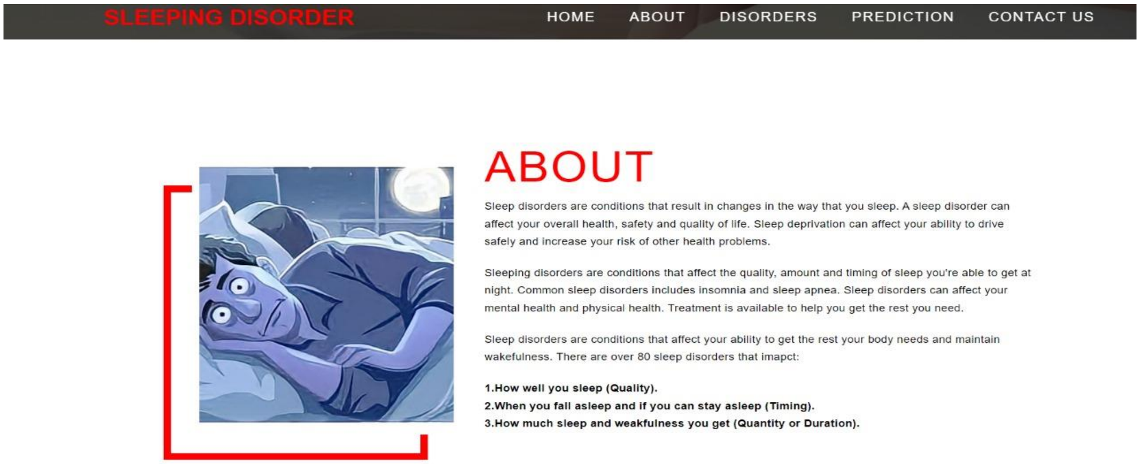
In conclusion, this sleep disorder prediction project successfully highlighted the significant influence of gender, occupation, and BMI on sleep health. The Random Forest Classifier, achieving an accuracy of 87%, demonstrates the potential of machine learning for accurate sleep disorder prediction. These findings pave the way for further research and development of accessible tools to promote better sleep health and well-being

SAMPLE WEBSITE:

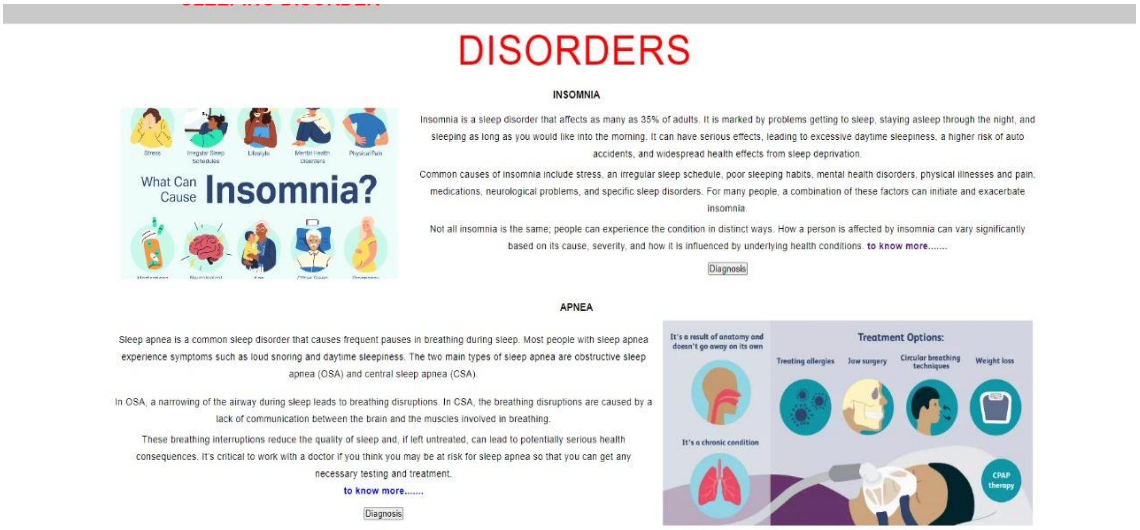
HOMEPAGE:



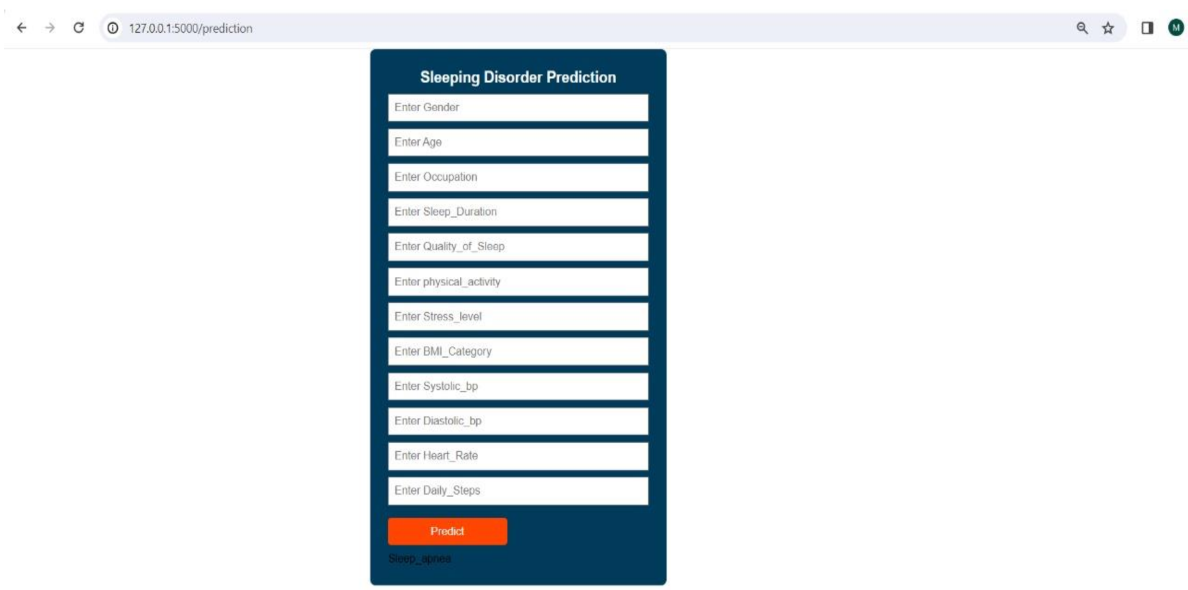
ABOUT PAGE:



TYPES OF DISORDERS:



PREDICTION PAGE:



CHAPTER-12

REFERENCES

Wang J, Ren X. Association Between Sleep Duration and Sleep Disorder Data from the National Health and Nutrition Examination Survey and Stroke Among Adults in the United States. Med Sci Monit. 2022;28:e936384. Epub 20220703. pmid:35780293; PubMed Central PMCID: PMC9261468.

Wolf C, Wolf S, Weiss M, Nino G. Children's Environmental Health in the Digital Era: Understanding Early Screen Exposure as a Preventable Risk Factor for Obesity and Sleep Disorders. Children (Basel). 2018;5(2). Epub 20180223. pmid:29473855; PubMed Central PMCID: PMC5836000.

Guo Q, Xie W, Peng R, Ma Y, Chong F, Wang Y, et al. A Dose-Response Relationship Between Sleep Duration and Stroke According to Nonhealth Status in Central China: A Population-based Epidemiology Survey. J Stroke Cerebrovasc Dis. 2019;28(7):1841–52. Epub 20190507. pmid:31076320.

Piepoli MF. Editor's Presentation Benefit of healthy lifestyle on cardiovascular risk factor control: Focus on body weight, exercise and sleep quality. Eur J Prev Cardiol. 2019;26(12):1235–8. pmid:31298111.

Carriedo-Diez B, Tosoratto-Venturi JL, Canton-Manzano C, Wanden-Berghe C, Sanz-Valero J. The Effects of the Exogenous Melatonin on Shift Work Sleep Disorder in Health Personnel: A Systematic Review. Int J Environ Res Public Health. 2022;19(16). Epub 20220817. pmid:36011832; PubMed Central PMCID: PMC9408537.

CHAPTER-13

APPENDIX

PROJECT CODE:

```
import numpy as np
import pandas as pd
import missingno as mn
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("Sleep_health_and_lifestyle_dataset 1.csv")
df.head()
df.isnull().sum()
mn.matrix(df)
df
df.isna().sum()
df.dropna(axis=0,inplace=True)
df
df.shape
df.columns
mn.matrix(df)
df['Blood Pressure'].unique()
#spliting the blood pressure into two columns
df['systolic_bp'] = df['Blood Pressure'].apply(lambda x: x.split('/')[0])
df['diastolic_bp'] = df['Blood Pressure'].apply(lambda x: x.split('/')[1])
#droping the blood pressure column
df.drop('Blood Pressure', axis=1, inplace=True)
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
#Correlation Matrix Heatmap
plt.figure(figsize=(20, 16))
sns.heatmap(df.corr(), annot = True, cmap = 'coolwarm')
vars = ['Gender', 'Occupation','BMI Category','Sleep Disorder']
for i in vars:
    label_encoder.fit(df[i].unique())
    df[i] = label_encoder.transform(df[i])
```

```
    print(i,':' ,df[i].unique())
df
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
# Assuming 'Person ID' and 'Sleep Disorder' columns are dropped from X
X = df.drop(['Person ID', 'Sleep Disorder'], axis=1)
y = df['Sleep Disorder']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier()
dtree
y_train = y_train.astype(float)
dtree.fit(X_train, y_train)
#training accuracy
print("Training Accuracy:",dtree.score(X_train,y_train))
d_pred = dtree.predict(X_test)
d_pred
from sklearn.metrics import classification_report
print(classification_report(y_test, d_pred))
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(n_estimators=100, random_state=42)
rfc.fit(X_train, y_train)
feature_names_during_fit = rfc.feature_names_in_
print("Feature names during fit:", feature_names_during_fit)
#Training accuracy
print("Training accuracy: ",rfc.score(X_train,y_train))
rfc_pred = rfc.predict(X_test)
rfc_pred
print(classification_report(y_test, rfc_pred))
print("Feature names used during training:", X_train.columns)
```

```python
import pandas as pd

# Assuming rfc is already trained and you have a single data point
single_data = pd.DataFrame({
    'Gender': 1,
    'Age': 28,
    'Occupation': 9,
    'Sleep Duration': 5.9,
    'Quality of Sleep': 4.0,
    'Physical Activity Level': 30.0,
    'Stress Level': 8.0,
    'BMI Category': 2,
    'Heart Rate': 85.0,
    'Daily Steps': 3000,
    'systolic_bp': '140',
    'diastolic_bp': '90',
}, index=[0])

# Assuming label_encoder is already defined and used for training

# One-hot encode the categorical variables

# Use the trained RandomForestClassifier to make predictions
# prediction = rfc.predict(single_data)
prediction = dtree.predict(single_data)

print(f"The predicted sleep disorder for the given data point is:
{prediction[0]}")

import pickle
with open('decision_tree_model.pkl', 'wb') as model_file:
    pickle.dump(dtree, model_file)
model = pickle.load(open("decision_tree_model.pkl","rb"))
df.info()
```

Front-end code

```
<!DOCTYPE html>

<html><head>

	<link rel="stylesheet" href="{{ url_for('static',filename='styles/google_form.css') }}">

	<title>Prediction</title></head>

<body><div class="container">

<form method="post" action="/prediction" name="contact-form">

<h4>Sleeping Disorder Prediction</h4>

				<input type="text" name="Gender" placeholder="Enter Gender"><input type="text" name="Age" placeholder="Enter Age">

	<input type="text" name="Occupation" placeholder="Enter Occupation">

	<input type="text" name="Sleep_Duration" placeholder="Enter Sleep_Duration">

	<input type="text" name="Quality_of_Sleep" placeholder="Enter Quality_of_Sleep">

	<input type="text" name="physical_activity" placeholder="Enter physical_activity">

	<input type="text" name="Stress_level" placeholder="Enter Stress_level">

	<input type="text" name="BMI_Category" placeholder="Enter BMI_Category">

	<input type="text" name="Systolic_bp" placeholder="Enter Systolic_bp">

	<input type="text" name="Diastolic_bp" placeholder="Enter Diastolic_bp">

	<input type="text" name="Heart_Rate" placeholder="Enter Heart_Rate">

	<input type="text" name="Daily_Steps" placeholder="Enter Daily_Steps">

	<input type="submit" value="Predict" id="submit">

		<p class="text-white">{{prediction}}</p></form>

	</div>	</body>	</html>
```