**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

A.Y. 2023 - 24

**Course: Machine Learning**

**Project Report**

|  |  |  |
| --- | --- | --- |
| Program |  | |
| Semester |  | |
| Name of the Project: |  | |
|  | | |
| Details of Project Members |  |  |
| Batch | Roll No. | Name |
|  |  |  |
|  |  |  |
|  |  |  |
| Date of Submission: | | |

**Contribution of each project Members:**

|  |  |  |
| --- | --- | --- |
| Roll No. | Name: | Contribution |
|  |  |  |
|  |  |  |

**Github link of your project:**

**Note:**

1. Create a readme file if you have multiple files
2. All files must be properly named (I004\_MLProject)
3. Submit all relevant files of your work
4. **Plagiarism is highly discouraged (Your report will be checked for plagiarism)**

**Rubrics for the Project evaluation:**

|  |
| --- |
| * Evaluation of project will be based on following rubrics * Domain knowledge and literature review in the selected topic (5 marks) * EDA, Implementation and performance metrics used (10 marks) * Beyond classroom knowledge gained and implemented (5 marks) |

**Project Report**

**Selected Topic**

**by**

**Student 1, Roll number: xx**

**Student 2, Roll number: xx**

**Student 3, Roll number: xx**

**Course: Machine Learning**

**AY: 2023-24**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr no.** | **Topic** | **Page no.** |
| **1** | Storyline or Applications of Project |  |
| **2** | Literature Review |  |
| **3** | Data Preprocessing and Exploratory data Analysis with Visualization |  |
| **4** | Machine learning models with hyper parameter tuning |  |
| **5** | Performance Evaluation |  |
| **6** | Comparison of different techniques used |  |
| **7** | Deployment/GUI/ Learning beyond classroom |  |
| **8** | Learnings and challenges you faced while doing the Project |  |
| **9** | Conclusion |  |

**I.** **Storyline or Applications of Project**

An alarming number of people struggle with sleep-related health difficulties in today's fast-paced world of hustle and bustle. The quality of life for millions of people globally is greatly impacted by the startlingly high frequency of disorders including sleep apnea and insomnia. Furthermore, these problems are made worse by the widespread issue of poor sleep quality, which jeopardizes productivity and general well-being.

Even with the increasing recognition of the significance of sleep, there is still a significant deficiency of easily accessible, all-inclusive materials devoted to comprehending and resolving sleep-related issues. Help seekers frequently come across incomplete information dispersed over several sources, which causes them to get frustrated and confused.   
  
Our initiative, "Sleep Savvy," was inspired by the urgent need for a comprehensive solution and seeks to be a source of information and assistance for those navigating the challenges associated with sleep health. Through the integration of evidence-based insights, useful advice, and professional assistance into a single, easily accessible platform, Sleep Savvy aims to enable people to take proactive measures to improve the quality of their sleep and their general well-being.

It is clear from thorough investigation and analysis that current treatments frequently fail to offer customized, approachable solutions that satisfy a range of requirements and preferences. In addition, the absence of a centralized platform makes it more difficult to organize community involvement and information sharing, which impedes attempts to address the widespread problem of sleep disorders and irregular sleep patterns.   
  
A comprehensive, user-focused website that informs and enables people to make educated decisions about their sleep health is necessary in light of these problems. Through addressing the complex nature of sleep-related problems and creating a community that is supportive, Sleep Savvy aims to transform the way that sleep wellness is approached, eventually enhancing people's quality of life all around the world.

**II. Literature Review**

Describe the existing work in the chosen topic. You can make table (Title of the paper, published year, major contributions, algorithms used and performance)

1. **Title of the Paper:** The future of sleep health: a data-driven revolution in sleep science and medicine
   * **Published Year:** 2020
   * **Major Contributions:** Explores the significant expansion in the use of multi-modal sensors and technologies for monitoring physical activity, sleep, and circadian rhythms. Discusses the potential applications of accurate sleep monitoring at scale, from large-scale epidemiological research linking sleep patterns to disease to wellness applications like sleep coaching for individuals with chronic conditions.
   * **Algorithms Used:** Multi-sensor data analysis, decision trees, ensemble learning among others, deep learning
   * **Performance:** Performance metrics or specific results were not provided in the text.

**III. Data Preprocessing and Exploratory data Analysis with Visualization**

Perform all data cleaning and preprocessing steps. Perform data visualization using different charts/graph. Make sure to write your own inferences

**IV. Machine learning models with hyper parameter tuning**

Design the machine learning models and perform necessary hyper parameter tuning

**V. Performance Evaluation**

Performance metrics as applicable

**VI. Comparison of different techniques used**

Comparison of different techniques used in terms of algorithms, working and performance

**VII. Deployment/GUI/ Learning beyond classroom**

In order to solve class imbalance in datasets, which is a prevalent problem in machine learning tasks like classification, two strategies are used: undersampling and oversampling. In order to balance the dataset, undersampling entails lowering the number of cases from the majority class. Conversely, oversampling seeks to raise the proportion of examples from the minority class. SMOTE (Synthetic Minority Over-sampling Technique) is a well-liked oversampling technique that creates artificial instances of the minority class depending on its closest neighbors.

With the help of the robust open-source Python package Streamlit, programmers may design interactive graphical user interfaces (GUIs) for use in data science and machine learning. Developers may create web apps using Streamlit without having to have a lot of experience with web programming. Data-driven insights are easier to share and convey because to the library's easy-to-use interface for creating interactive visualizations, dashboards, and user interfaces.  
  
Pickle files are a Python serialization mechanism that lets programmers store and load sophisticated data structures like dictionaries, lists, and even machine learning models that have been trained. Pickle files offer a practical means of storing and retrieving data, facilitating the exchange of models and data among various scripts or contexts. This may be especially helpful when working with big datasets or models that have already been trained, as it removes the need to import or retrain the data each time the script runs.

**VIII. Learnings and challenges you faced while doing the Project**

:

The team faced a number of difficulties during the project, most of which had to do with managing unbalanced data and optimizing hyperparameters while adhering to computational constraints. One major challenge was the problem of unbalanced data, which needed to be carefully addressed to avoid biased model results. They looked at two popular strategies: methods for under- and oversampling. While oversampling sought to raise the representation of the minority class through various techniques like SMOTE (Synthetic Minority Over-sampling Technique), undersampling involves shrinking the size of the majority class to balance it with the minority class.   
  
Still, the problem went beyond incomplete data. Due to computational limits, they also encountered difficulties while trying to adjust several hyperparameters at once. This limitation made it more difficult for them to fully investigate the hyperparameter space and may jeopardize the model's functionality.

Another challenge was structuring the graphical user interface (GUI). To guarantee smooth navigation and comprehension for end users, careful preparation was necessary while designing an intuitive and user-friendly interface. It turned out to be a difficult undertaking to strike a balance between utility and simplicity, requiring repeated design processes and the inclusion of user feedback.  
  
In spite of these difficulties, the group overcame them all with fortitude and inventiveness, making effective use of the resources at hand to maximize the functionality of their model and improve the GUI's user experience. After working together and addressing problems, they were able to get over these obstacles and provide a strong project solution.

**IX. Conclusion**

To sum up, the project team overcame difficulties with data imbalance, hyperparameter tuning, and GUI design. They overcame these obstacles by combining analytical reasoning with creativity, using methods like under- and oversampling to deal with data imbalance and giving hyperparameters the order of importance within computing constraints. As a consequence of user feedback and iterative design processes, the resultant graphical user interface (GUI) was created to be an intuitive and user-friendly tool. Notwithstanding the challenges faced, the team's cooperative efforts resulted in a solid solution that successfully satisfies end users' expectations.

In order to further improve model resilience, future research may investigate sophisticated approaches—such as ensemble methods or more complex resampling strategies—for managing unbalanced data. While options for enhancing algorithms and utilizing parallel computing resources exist to enable more thorough hyperparameter tweaking, addressing computational restrictions remains a top concern. User experience might be improved by improving the GUI in response to user input and usability studies, and real-world testing and validation would offer important new perspectives on the applicability of the paradigm. The project can stay at the forefront of the sector and be ready to take on new problems and possibilities by keeping up with machine learning techniques.