**Sleep disorder and Lifestyle Analysis**

**Executive summary and Key findings**

This project focuses on analyzing and predicting sleep disorders using various data mining techniques. The dataset includes diverse physiological and lifestyle parameters like the patient’s daily activity, profession, medical conditions etc. that potentially influence sleep quality and sleep-related health conditions. The goal is to leverage these variables to identify patterns, correlations, and risk factors associated with sleep disorders, thereby enabling proactive interventions and better health outcomes.

The project mainly involves developing an optimized prediction model by exploring multiple advanced machine learning techniques to classify individuals as having or not having a sleep disorder and do clustering to identify specific target profiles to design strategies. This model was then applied to analyze a set of unlabeled data provided by an insurance company to predict the likelihood of sleep disorders among their clients. This approach aims to assist the insurance company in identifying high-risk individuals and enabling targeted health interventions. Furthermore, the predictive model supports multiple companies in the healthcare sector in development of personalized health strategies, optimizing the company’s risk management and improving client health outcomes.

* **Significant Predictors of Sleep Disorders**:  
  Features such as Diastolic Blood Pressure (BP), Heart Rate, Sleep Duration, Daily Steps, and Stress Levels emerged as the most influential factors in predicting sleep disorders like Insomnia and Sleep Apnea.
* **Decision Tree Model**:  
  The decision tree model achieved 92.15% accuracy with high precision and recall for Sleep Apnea (93.40% and 93.05%) and Insomnia (93.40% and 89.18%). It offers strong interpretability, making it ideal for practical healthcare and insurance applications.
* **Neural Network Model**:  
  The neural network model achieved a superior 95.19% accuracy and higher precision and recall for Insomnia (94.95% and 95.90%) and Sleep Apnea (94.66% and 96.55%). However, its lack of interpretability makes it less practical for direct business use.
* **Clustering Insights**:  
  Clustering analysis revealed hidden subgroups, such as individuals with high stress and low physical activity, providing actionable insights for targeted interventions and personalized health strategies.
* **Key Behavioral Observations**:  
  Individuals with higher physical activity levels tend to report better sleep quality.  
  Increased stress levels are negatively correlated with sleep duration, making stress management a critical intervention area.

**Business opportunities**

In the modern competitive world with increasing global health concern around hectic lifestyle, sleep disorders are increasingly recognized as a critical public health concern due to their significant impact on personal well-being, productivity, and healthcare costs. These disorders, which include conditions like insomnia, sleep apnea, and restless legs syndrome, are linked to a range of serious health issues such as hypertension, obesity, diabetes, cardiovascular diseases, and mental health disorders.

From a business perspective, undiagnosed or untreated sleep disorders can lead to increased insurance claims, reduced employee productivity, and higher healthcare expenses. However, early detection and management of sleep disorders offer significant opportunities for businesses in industries such as:

1. **Insurance and Risk Management:**
   * Predicting sleep disorders enables insurance companies to identify high-risk policyholders, design personalized health interventions, and offer customized premium rates, ultimately reducing claims and increasing profitability.
2. **Healthcare and Wellness:**
   * Providers and wellness companies can use predictive models to develop targeted programs, such as sleep therapy, stress management, and fitness plans, addressing sleep-related health risks proactively.
3. **Corporate Productivity:**
   * Companies can mitigate the effects of sleep disorders on workforce productivity by leveraging predictive insights to design employee wellness initiatives, reducing absenteeism and improving overall performance.

Some of the present-day companies or apps with a focus on sleep healthcare that can leverage these models are:

**Headspace**: Provides mindfulness and meditation programs, including a dedicated section for sleep with soundscapes and bedtime exercises.

**ResMed**: A leader in sleep apnea solutions, offering CPAP machines, masks, and other devices to manage sleep-related breathing disorders.

**Hatch**: Creates smart sleep solutions, including alarm clocks and sound machines, to support healthy sleep routines.

**Purpose of the data mining task**

The purpose of the data mining task is twofold, focusing on both classification and clustering techniques to address sleep health concerns. Through classification, the goal is to predict whether an individual is likely to have a sleep disorder based on their health and lifestyle data. This enables early detection and intervention, helping healthcare providers or insurers identify at-risk individuals before their conditions worsen. Accurate classification supports the development of personalized recommendations, such as tailored wellness plans or preventive strategies, and aids insurance companies in assessing risks and designing appropriate premium adjustments.

Clustering, on the other hand, aims to segment individuals into distinct profiles based on shared patterns in their sleep, health, and lifestyle metrics. This approach uncovers hidden subgroups, such as individuals with high stress and low sleep quality, enabling targeted interventions. Businesses and healthcare providers can leverage these insights to design specialized wellness programs, create customized insurance plans, or develop stress-reduction strategies for specific clusters. By combining classification and clustering, the task provides a comprehensive solution for understanding sleep disorders and implementing data-driven strategies to improve sleep health and overall well-being.

**Data for analysis**

**Name of the dataset:** Sleep Health Data  
**Link:** https://www.kaggle.com/datasets/imaginativecoder/sleep-health-data-sampled

The Sleep Health and Lifestyle dataset features comprehensive sleep metrics, including sleep duration, quality, and influencing factors, alongside lifestyle data such as physical activity levels, stress levels, and BMI categories. It also incorporates cardiovascular health indicators like blood pressure and heart rate, enabling a holistic analysis. A key focus is identifying sleep disorders, including conditions such as insomnia and sleep apnea, providing valuable insights into their prevalence and associated factors.

**Total entries**: 12000 records

**Dataset Columns:**

* **Numerical columns**: Age, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, Systolic BP, Diastolic BP, Heart Rate, Daily Steps
* **Categorical columns**: Gender, Occupation, BMI Category, Sleep Disorder
* **Person ID** is a unique identifier and is usually not included as a feature for modeling.

**Description of dataset columns**:

**Person ID**: An identifier for each individual.  
**Gender**: The gender of the person (Male/Female).  
**Age**: The age of the person in years.  
**Occupation**: The occupation or profession of the person.  
**Sleep Duration (hours):** The number of hours the person sleeps per day.  
**Quality of Sleep (scale: 1-10):** A subjective rating of the quality of sleep, ranging from 1 to 10.  
**Physical Activity Level (minutes/day):** The number of minutes the person engages in physical activity daily.  
**Stress Level (scale: 1-10):** A subjective rating of the stress level experienced by the person, ranging from 1 to 10.  
**BMI Category**: The BMI category of the person (e.g., Underweight, Normal, Overweight).  
**Blood Pressure (systolic/diastolic)**: The blood pressure measurement of the person, indicated as systolic pressure over diastolic pressure.  
**Heart Rate (bpm):** The resting heart rate of the person in beats per minute.  
**Daily Steps:** The number of steps the person takes per day.  
**Sleep Disorder:** The presence or absence of a sleep disorder in the person (Healthy, Insomnia, Sleep Apnea).

**Details about Sleep Disorder Column**:

**Healthy**: The individual does not exhibit any specific sleep disorder.  
**Insomnia**: The individual has trouble falling asleep or staying asleep, leading to inadequate or poor-quality sleep.  
**Sleep** **Apnea**: The individual suffers from pauses in breathing during sleep, resulting in disrupted sleep patterns and potential health risks.

**Initial explorations of data**:

A screenshot of a computer

Description automatically generated

**Distribution of important numerical data with respect to the sleep disorders:**

**Sleep Duration:**

A screenshot of a graph

Description automatically generatedAround 700-800 records have a sleep duration of 6.4 to 6.6 hours. Most of these records are associated with insomnia.

**Daily steps:**

A screen shot of a graph

Description automatically generated

Around 3000 records have 6000 daily steps and are mostly associated with Insomnia

**Heart rate:**

A screenshot of a computer

Description automatically generated  
Up to 1600 records contain a heart rate varying between 68 and 72. These records are mostly associated with Sleep apnea or Insomnia.

**Distribution of categorical data:**

**Gender:**

**A screenshot of a computer

Description automatically generated**

**Occupation:**

**A screenshot of a computer screen

Description automatically generated**

**BMI Category:**

**A screenshot of a computer

Description automatically generated**

**Sleep disorder:**

**A screenshot of a computer

Description automatically generated**

The dataset has an even distribution of sleep disorder classes. So the model can designed without being biased to one particular class.

**Scatter plots:**

**Sleep duration vs Stress level**

****The stress level experienced shows a negative relation to the sleep duration of the patients. With increase in stress levels of the patients the sleep duration has decreased. Sleep duration is one of the important factors when it comes to diagnosing sleep disorders.

**Physical activity level vs Quality of sleep**

A screenshot of a computer

Description automatically generated  
The Quality of sleep is positively affected with the physical activity level. With the increase in physical activity levels the quality of sleep of the patients seems to be increased.

**Data cleaning:**

**BMI category:** Initially the dataset had an improper categorization for BMI categories. The values for BMI category columns were **Normal**, **Normal** **weight**, **Overweight** and **Obese**.

A screenshot of a computer

Description automatically generated

The normal and normal weight fall under the same category when scaling BMI index. With up to 5000 records containing **Normal** as the BMI category and only 624 records containing **Normal** **weight** as BMI category this can be mistake in representing the data as both indicate the same scale on a BMI index. So, 624 records with **Normal weight** as index were converted to **Normal** using the find and replace tool in the Excel.

**Blood pressure:** The blood pressure in the dataset is represented as following

Value = Systolic blood pressure / diastolic blood pressure

A screenshot of a computer

Description automatically generated

The following notation is improper when trying to utilize the data in data mining models. The blood pressure columns will be considered as a polynomial input type. So, the blood pressure column is separated into two new columns named **Systolic BP** and **Diastolic BP** using the text to columns tool in the Excel.

**Data selection:**

The following columns were selected from the dataset

1. **Gender:** Gender differences can significantly impact sleep patterns, stress levels, and the likelihood of certain disorders. For example, men may be more prone to sleep apnea, while women might report higher stress-related sleep disturbances.
2. **Age:** Age is a critical factor in sleep health. Older individuals are more likely to experience sleep disorders such as insomnia and sleep apnea due to physiological changes.
3. **Occupation:** The type of occupation influences lifestyle, stress levels, and physical activity, which are all important predictors of sleep health. Sedentary office workers, for instance, may show different patterns compared to manual laborers.
4. **Sleep Duration:** This is directly related to sleep health. Both insufficient and excessive sleep durations are associated with various disorders, including insomnia and hypersomnia.
5. **Quality of Sleep:** Sleep quality provides subjective insights into an individual’s perceived restfulness, which is a key indicator of potential underlying sleep issues.
6. **Physical Activity Level:** Regular physical activity promotes better sleep quality and reduces stress. Analyzing activity levels helps establish links between lifestyle and sleep health.
7. **Stress Level:** Stress is a significant factor affecting sleep duration and quality. High stress levels often correlate with insomnia or fragmented sleep.
8. **BMI Category:** BMI is an important health metric as obesity is a known risk factor for sleep apnea, while being underweight may correlate with other sleep disturbances.
9. **Systolic BP:** Systolic BP indicates the pressure in arteries during heartbeats and is a key marker of cardiovascular health, often linked to hypertension. High Systolic BP is associated with sleep disorders like sleep apnea and can highlight stress or lifestyle impacts on health.
10. **Diastolic BP:** Diastolic BP measures arterial pressure during heart relaxation, reflecting resting cardiovascular health. Abnormal DBP levels can indicate chronic conditions or poor restorative sleep, both of which affect sleep quality.
11. **Heart Rate:** Resting heart rate can indicate overall cardiovascular health and stress, both of which influence sleep quality and the likelihood of disorders.
12. **Daily Steps**: This serves as an objective measure of physical activity, complementing self-reported activity levels and linking lifestyle habits to sleep health.
13. **Sleep Disorder**: This is the target variable in classification tasks, representing the outcome to be predicted or analyzed. It connects all other features to the overarching goal of understanding and mitigating sleep disorders.

All the columns significantly contribute to designing a prediction model for sleep disorder except for the **Person ID** column. Instead of removing the person we can specify an **ID** role using **set role** operator in AI studio for the Person ID column to be excluded when feeding the dataset to design the model.

**Model assessment**

For the classification of sleep disorder, the decision trees and neural networks were evaluated to perform analysis and build an optimized prediction model and A clustering model is evaluated to design different target profiles.

**Data mining technique**: Classification

**Classification - Decision Trees model**

Decision trees are highly interpretable models that present a clear, hierarchical structure, making them easy to understand and visualize for attributes important to classify sleep disorders. They handle both numerical and categorical data effectively.

**Process**:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The Person ID column is set as ID and the sleep disorder column is set as label in the set role operator.

A screenshot of a computer

Description automatically generated

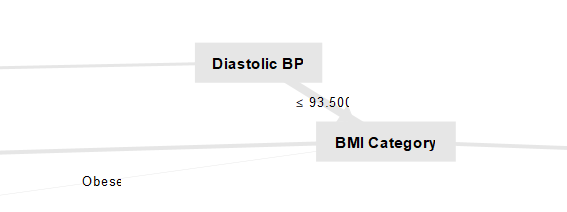
For the decision tree operator with **gain\_ratio** set as criterion the maximum depth of the tree is set to 20. The pruning and pre-pruning parameters are applied to the model with confidence set to 0.15 with a minimum gain of 0.15.

**Decision tree model:**

A white background with black and red lines

Description automatically generated

Diastolic BP is the root attribute of the tree used to start the decision-making process.



**Performance**:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The model has an accuracy of **92.15% +/- 0.58%** this is a very good accuracy score. The precision and recall for each class are as follows:

|  |  |  |
| --- | --- | --- |
|  | Precision | Recall |
| Healthy | 89.82% | 94.23% |
| Sleep Apnea | 93.40% | 93.05% |
| Insomnia | 93.40% | 89.18% |

The dataset has an **even distribution** of sleep disorder classes. Thus, accuracy can be considered to evaluate the performance of the model. The model also a good precision and recall values when it comes to predicting the sleep apnea and Insomnia. These two classes are crucial insights for which the model would be used to diagnose.

The model has the kappa value of **0.882 +/- 0.009** so it is very good at predicting sleep disorder than taking a chance of random guess.

**Advantages:** The decision treesare easy to interpret and adopt into a business model. It can handle numerical and categorical values.

**Disadvantages:** The decision trees are more prone to overfitting. This can be an issue when dealing with larger and complex datasets. Proper pruning parameters are needed to optimize the model and avoid overfitting.

**Classification – Neural networks model**

Neural networks are powerful models that excel at capturing complex, nonlinear relationships within the dataset, making them ideal for analyzing multifaceted features like sleep duration, stress, and physical activity. They automatically learn feature interactions without the need for extensive preprocessing or feature engineering, making them versatile and adaptable.

**Process:**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

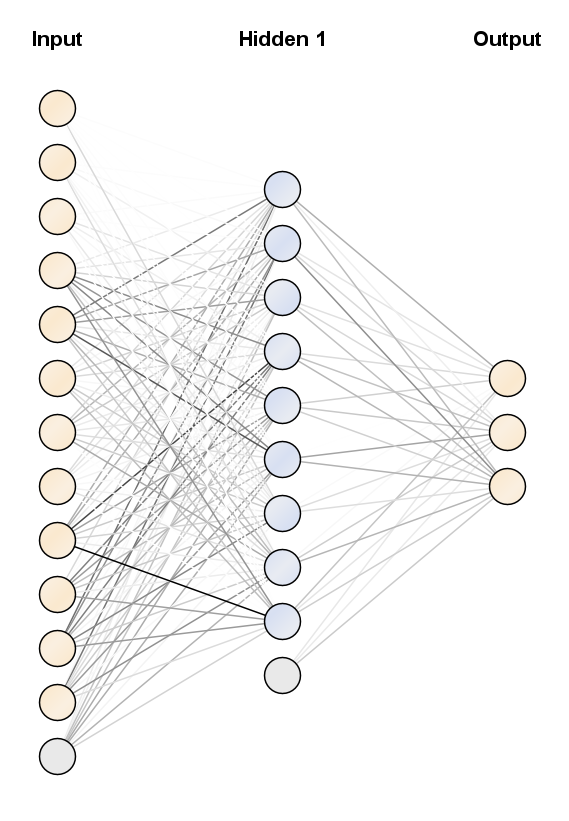
Description automatically generated**

The Person ID column is set as ID and the sleep disorder column is set as label in the set role operator.

The neural networks can handle only handle numerical data. So, Nominal to Numerical operator is used to convert categorical columns to numerical encoded columns. The coding type is set to unique integers.

In the Neural net operate the number of training cycles is set to 800 with a learning rate of 0.01 and momentum of 0.9.

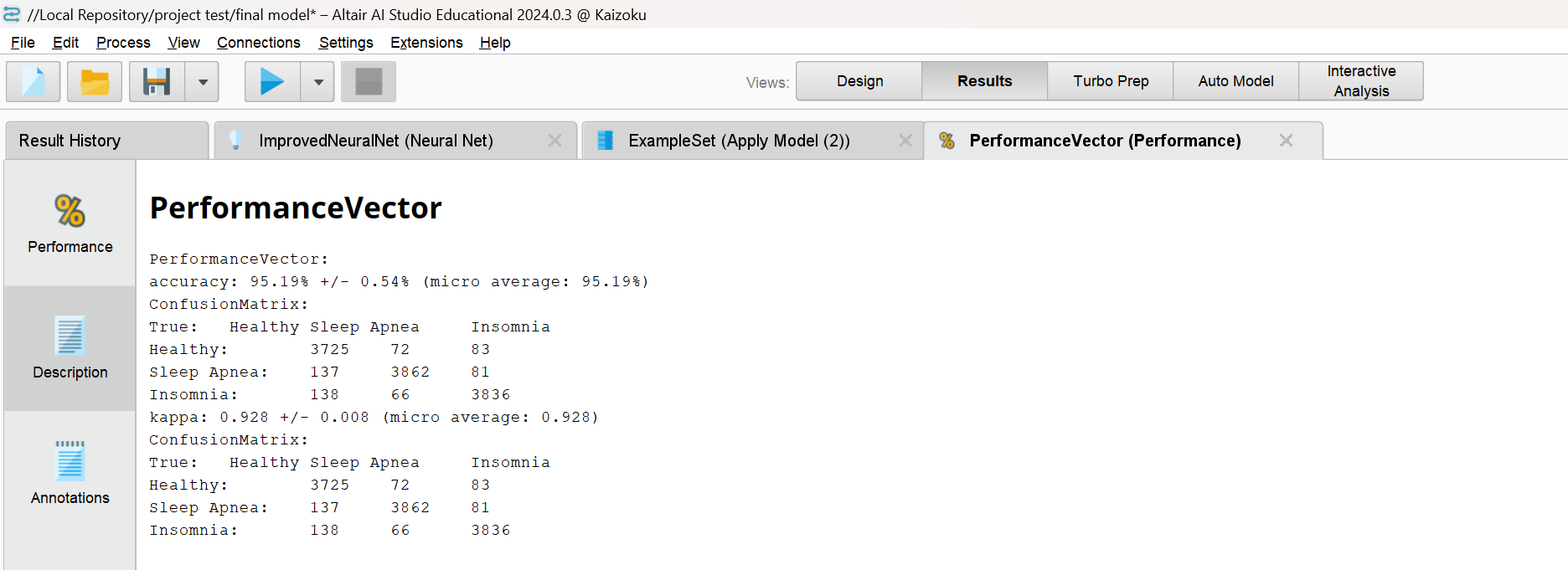
**Neural net model:**



**Performance:**

A screenshot of a computer

Description automatically generated



The model has an accuracy of **95.19% +/- 0.54%** this is a very high accuracy and is comparatively better than the decision tree model.

The precision and recall values for each class are as follows:

|  |  |  |
| --- | --- | --- |
|  | Precision | Recall |
| Healthy | 96.01% | 93.12% |
| Sleep Apnea | 94.66% | 96.55% |
| Insomnia | 94.95% | 95.90% |

The precision and recall values for predicting are very good. These two classes are the crucial insights when the model is utilized in a business perspective. Predicting of a person is healthy is not an important task when compared to predicting Insomnia and Sleep apnea. As these two would be the targeted patients to prioritise when implementing to a business model.

The model has a kappa value of **0.928 +/- 0.008** which is very good score as it has a better prediction then taking a random guess for the output.

**Advantages:** They effectively model complex, nonlinear relationships, making them ideal for the multifaceted sleep dataset. Neural networks achieve high predictive accuracy and automatically learn feature interactions without extensive preprocessing.

**Disadvantages:** Neural networks are less interpretable than simpler models, making it challenging to explain predictions to stakeholders. They require significant computational resources and time for training, especially with larger datasets.

**Model comparison:**

|  |  |  |
| --- | --- | --- |
|  | **Accuracy** | **Kappa** |
| **Decision Trees** | 92.15% | 0.882 |
| **Neural Networks** | 95.19% | 0.928 |

**Decision Trees:** The decision trees model has a good accuracy of 92.15%. The model is more interpretable and easier to use and adopt into a business model than neural networks.

**Neural network:** The Neural network model has a very high accuracy of 95.19%. Although the model has very high predicting accuracy and outperforms decision trees it is tough to interpret to adopt into a business model.

**Model selection:**

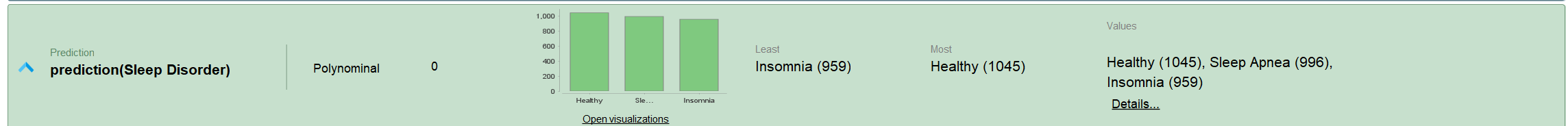
The **decision tree** model is the preferred choice for the sleep disorder dataset despite the neural network's slightly higher accuracy. With an accuracy of 92.15% and a Kappa statistic of 0.882, the decision tree provides reliable predictions while being highly interpretable. Its hierarchical structure allows for a transparent decision-making process, making it easier to understand the relationships between features such as stress levels, physical activity, and sleep quality. This transparency is essential for healthcare providers and businesses to adopt the model effectively. Additionally, decision trees require less computational power, are simpler to implement, and provide clear insights into the most significant factors contributing to sleep disorders, ensuring a practical and efficient solution for addressing sleep health challenges.

The selected model is than used to predict sleep disorder for 3000 records of unlabeled data from an insurance company.

A screenshot of a computer

Description automatically generated

The predicted output can be utilized by the insurance company to design new strategies to make better plans for patients that are diagnosed with Insomnia or Sleep Apnea.



The records predicted were

Healthy -1045

Sleep Apnea – 996

Insomnia – 959

The prediction confidence of the records needs to be considered to further analyze the healthy recorded patients for a chance of having Sleep Apnea Insomnia and design a different strategy for these records.

**Data mining technique**: Clustering

**Clustering model**

Clustering models are valuable for sleep health disorder datasets as they can uncover hidden patterns and group patients based on shared traits like symptoms or physiological metrics, enabling the identification of subtypes of disorders such as sleep apnea or insomnia. They assist in data labeling, especially in cases of incomplete or noisy labels, and help in anomaly detection by identifying rare or unusual cases. Clustering also supports personalized medicine by revealing distinct patient profiles, aiding in targeted interventions, and enhancing feature analysis for model development. Additionally, clustering can complement classification models by simplifying datasets or identifying severity levels, ultimately improving diagnostic and predictive accuracy.

**Process:**

A screenshot of a computer

Description automatically generated

The measure types parameter is changed to mixed measures which uses mixed Euclidean distance to handle and consider the categorical columns in the dataset.

The Person ID column is set as ID and the sleep disorder column is set as label in the set role operator. The data is normalized to contribute the weight of each column equally to the clustering.

The output clusters profiles are extracted and analysed using the person IDs to make target profiles for healthcare companies to design strategies to improvise their business opportunities.

**Centroid table:**

A screenshot of a computer

Description automatically generated

**Advantages:** Clustering identifies groups or patterns in data without needing prior labels, offering insights into underlying structures. It helps in segmenting populations (e.g., patients using the Person IDs), enabling targeted interventions or tailored solutions.

**Disadvantages:** Clustering can be influenced by outliers and requires proper scaling of features for accurate results. Normalizing is essential to make sure the columns contribute equally to the clustering process.

**Results of the models:**

**Decision Tree**

**The starting part of decision tree partitions:**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

**Performance:**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The model starts by evaluating **Diastolic Blood Pressure (Diastolic BP > 93.500** and **Diastolic BP ≤ 93.500**), dividing data into higher and lower ranges. Subsequent splits depend on **Heart Rate**, **BMI Category**, and rest of the columns.

After **Diastolic BP,** **Heart rate** is the most crucial attribute utilized to classify the sleep disorder effectively in the decision trees model. The records with **Heart rate > 67** are mostly associated with **Sleep Apnea**.

**Neural Networks:**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Performance:**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Healthy class:**

The node 1 has the highest impact on predicting the healthy class

A screenshot of a computer

Description automatically generated

The Heart rate and sleep duration have the highest impact on predicting the healthy class. These attributes are crucial for a person to be predicted as healthy.

**Sleep Apnea class:**

The node 6 has the highest impact on predicting the Sleep Apnea class

**A screenshot of a computer

Description automatically generated**

The Sleep duration, Age and daily steps are the crucial attributes to predicting the sleep apnea class. These attributes are crucial for a person to be diagnosed with sleep apnea.

**Insomnia class:**

The node 2 has the highest impact on predicting the Sleep Apnea class

A screenshot of a computer

Description automatically generated

The Daily steps, Heart rate and sleep duration are crucial attributes to predicting the Insomnia class. These attributes are crucial for a person to be diagnosed with Insomnia.

**Clustering:**

**A screenshot of a computer

Description automatically generated**

**Centroid table:**

**A screenshot of a computer

Description automatically generated**

**Plot:**

**A screenshot of a graph

Description automatically generated**

The cluster data is exported using the Person IDs and filtered using Excel’s advanced filtering to clearly analyze the categorical attributes.

**Cluster 0:**

The records in cluster\_0 mostly belong to the older age group. Most of the patients are Female with high physical activity levels. Majority of these records seem to have Nurse as an occupation and tend to have Sleep Apnea disorder. Most of the records have an overweight BMI index with comparatively high heart rate.

**Cluster 1:**

The records in cluster\_1 have a younger age group and mix of male and female patients. These records show low stress levels with high sleep duration. They have a moderate count of daily steps and heart rate. Most of the records have the occupation as teachers and salesperson and tend to be healthy.

**Cluster 2:**

The records in cluster\_2 have a moderate age group and mix of male and female patients. These records show high stress levels with very low sleep duration. These records have very low daily steps and Heart rate. Most of these records seem to have the occupation as Engineers and tend to have Insomnia sleep disorder.

**Conclusions from the Assessment of the Models:**

**Decision Tree Model:**  
The decision tree model provides a good balance between interpretability and accuracy, with a performance accuracy of 92.15% and a kappa value of 0.882. It effectively identifies significant attributes like Diastolic BP, Heart Rate, and BMI Category, making it suitable for healthcare applications where understanding the reasoning behind predictions is critical. Its hierarchical structure makes it ideal for generating actionable insights for businesses, especially in healthcare and insurance. However, it is prone to overfitting larger datasets, which requires careful pruning.

**Neural Networks Model:**The neural network model achieves a higher accuracy of 95.19% and a kappa value of 0.928, showcasing its strength in modeling complex, nonlinear relationships within the data. Features like Heart Rate, Sleep Duration, and Daily Steps have a significant impact on predictions. Despite its superior performance, the model's lack of interpretability and higher computational demands make it less practical for direct business adoption, particularly for stakeholders needing transparent explanations.

While the neural network outperforms the decision tree in terms of accuracy, the decision tree was chosen for its interpretability, ease of implementation, and ability to highlight key features influencing sleep disorders. This makes it more aligned with the requirements of healthcare providers and businesses focusing on actionable outcomes.

**Clustering Model:**  
Clustering techniques were employed to segment individuals into distinct profiles based on shared characteristics. This approach uncovered hidden patterns, aiding in identifying subgroups such as those at higher risk of specific sleep disorders. It complements classification by offering additional insights for targeted interventions and personalized solutions, although it requires proper normalization to avoid biases from feature scaling.

**Business and Healthcare Implications:**  
These models provide valuable tools for early detection of sleep disorders, enabling insurance companies to identify high-risk individuals, healthcare providers to offer personalized treatments, and businesses to optimize wellness strategies. The decision tree model is particularly well-suited for practical implementation due to its clarity and reliability and can be easily integrated into a business model.

**Recommendations to Management:**

1. **Adopt the Decision Tree Model for Practical Implementation**:  
   Given its accuracy of **92.15%** and interpretability, the decision tree model should be the primary tool for predicting sleep disorders in real-world applications. Its transparent structure allows stakeholders, such as healthcare providers and insurers, to understand the reasoning behind predictions, making it easier to design interventions and justify decisions for clients or patients.
2. **Utilize Neural Networks for Research and High-Accuracy Predictions**:  
   Although less interpretable, the neural network model's **95.19% accuracy** makes it suitable for advanced research and scenarios requiring highly precise predictions. It can be leveraged in controlled settings or as a secondary tool to validate results from the decision tree model.
3. **Incorporate Clustering for Targeted Wellness Programs**:  
   Use clustering techniques to segment populations into actionable groups based on shared traits, such as stress levels, physical activity, and sleep patterns. This segmentation can guide the development of personalized health plans and targeted insurance premiums, improving customer satisfaction and risk management. The management can focus on clusters mostly associated with Sleep Apnea and Insomnia and the related occupations like nurses and engineers that are more likely to be prone to these disorders.
4. **Focus on High-Risk Categories**:  
   Prioritize individuals predicted to have *Sleep Apnea* or *Insomnia*, as these conditions have significant health and financial implications. Tailored interventions like fitness programs, stress management workshops, and sleep therapy solutions can reduce claims and enhance client health.
5. **Leverage Key Features for Proactive Strategies**:  
   Attributes like **Diastolic BP**, **Heart Rate**, **Sleep Duration**, and **Daily Steps** are critical predictors of sleep disorders. Management should ensure these parameters are routinely monitored and analyzed in patient assessments to identify early warning signs.