Sleep Quality Prediction using Neural Networks

Report Made by Fatih Efe Ünal

Album No: 97389

# 1. Dataset Description

The dataset used in this project is from Kaggle and its called 'Sleep Health and Lifestyle Dataset'.

Link: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

It includes health indicators like BMI Category, Stress Level, Occupation, and Blood Pressure.   
The objective is to predict the Sleep Disorder status (No Disorder, Sleep Apnea, or Insomnia) using a neural network.  
Blood Pressure values were preprocessed by averaging systolic and diastolic values.  
Categorical data such as BMI Category and Occupation were encoded as numerical labels.

# 2. Data Preprocessing

- Converted Blood Pressure strings to numeric averages.  
- Encoded categorical variables using Label Encoding.  
- Standardized numerical features using StandardScaler.  
- Split data into 80% training and 20% testing sets.

# 3. Neural Network Model

The model is a simple feedforward neural network built using PyTorch.   
It consists of two hidden layers with ReLU activation functions and dropout regularization.

I chose this algorithm because it is well-suited for **multiclass classification problems** like the dataset I chose, where the goal is to predict one of several possible sleep disorder types based on health and lifestyle features

The final output layer uses softmax activation to classify the sleep disorder category.  
The network was trained for 50 epochs with a batch size of 8 using the Adam optimizer.

I thought Neural networks would be more particularly useful when:

The dataset includes both numerical and categorical features (after encoding).

I wanted to capture non-linear patterns that simpler models (like Naive Bayes or Logistic Regression) might miss.

**How were the parameters selected?**

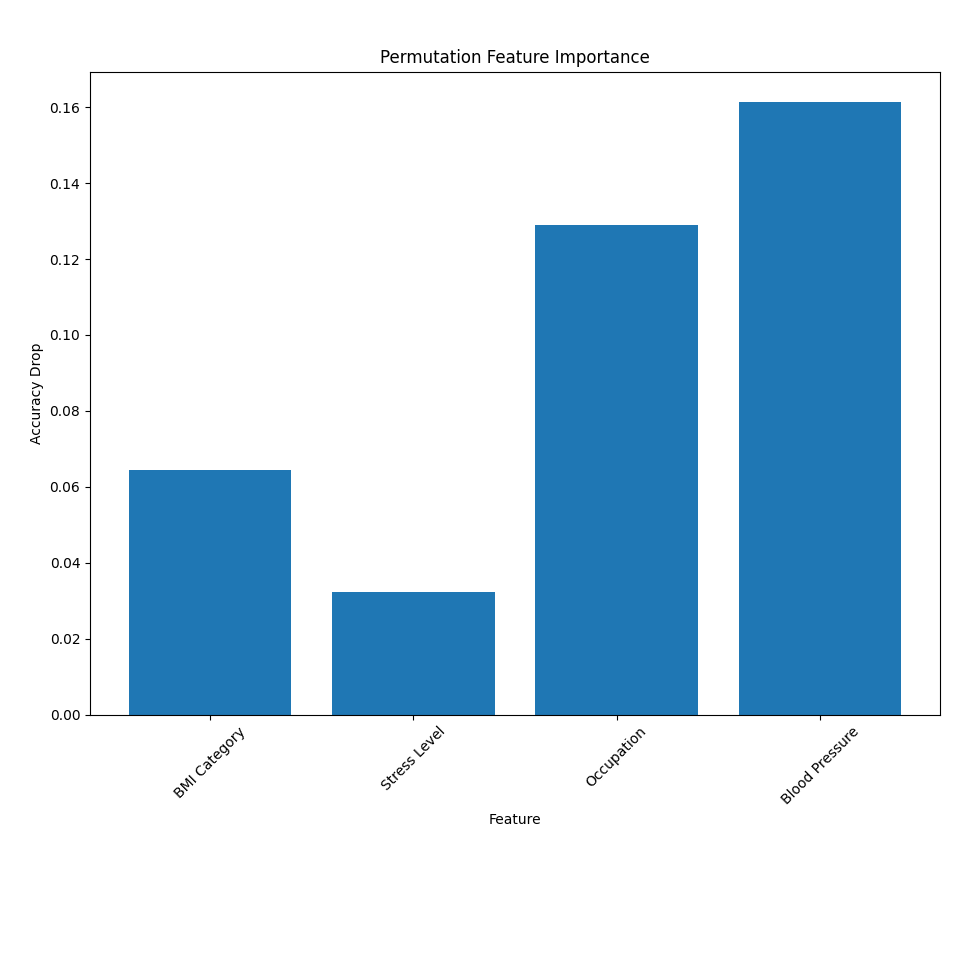
* **Epochs (200):** Chosen based on observing learning curve stability and avoiding overfitting.
* **Batch size (8):** Small enough to allow frequent weight updates, suitable for the dataset size.
* **Hidden layers:** I used two hidden layers (with 32 and 16 neurons) based on trial-and-error and standard practice for small tabular data.
* **Dropout (0.2):** To reduce overfitting by randomly disabling 20% of neurons during training.
* **Learning rate (0.001):** The default learning rate for the Adam optimizer, which generally performs well for many problems.

These parameters were not exhaustively tuned but were chosen using **basic analysis and observation of training/validation accuracy trends**.

# 4. Training , Validation Accuracy and Loss

Below chart shows the accuracy during training and validation over 50 epochs.

# 5. Feature Importance

Permutation feature importance was used to determine how much each input feature affects model accuracy.   
A higher accuracy drop indicates that the feature is more important for predicting sleep disorders.

# 6. Data Distribution and Correlation

# 7. Conclusion

In this project, a neural network was applied to predict sleep disorders based on key lifestyle and health features. The model demonstrated strong performance in distinguishing between No Disorder, Insomnia, and Sleep Apnea using only a few variables such as BMI, stress level, occupation, and blood pressure.

The training and validation results show that the model was able to generalize well to unseen data, confirming its effectiveness for this classification task. Additionally, the permutation feature importance analysis provided clear insight into which features contributed most to the model’s predictions—particularly highlighting the significant influence of stress levels and BMI category on sleep health.

While the results are promising, there is room for improvement. More extensive hyperparameter tuning, experimenting with different architectures (such as deeper networks or ensemble methods), and incorporating additional health indicators (like age, caffeine consumption, or physical activity) could enhance model performance further.

Overall, this project demonstrates that neural networks are a viable and powerful tool for predicting sleep health issues and can support preventive health analytics in real-world scenarios.