**DROWSINESS DETECTION USING SUPERVISED MACHINE LEARNING**

**Folder Link :** [**SupervisedML\_Y3**](https://drive.google.com/drive/folders/10YJ8J6qvxrhkUL8pPgKm73xhBPQgvC2T?usp=sharing)

**PROBLEM STATEMENT:**

Analyze the correlation between different brain waves and sleepiness to understand which factors are most predictive of driver's drowsiness. Driver drowsiness is a significant cause of road accidents worldwide, often leading to severe injuries and fatalities. As the human brain transitions between various states of wakefulness and sleep, it exhibits distinct patterns of neural oscillations, commonly referred to as brainwaves. These brain waves can be categorized into several frequency bands, each associated with specific cognitive states, ranging from deep sleep to heightened alertness.

**OBJECTIVE**

The study aims to analyze the correlation between different brainwave frequency bands. By identifying which brainwave patterns are most strongly associated with drowsiness, the study seeks to pinpoint the key factors that predict when a driver is likely to become drowsy.

* Analyze EEG brainwave data to detect driver drowsiness.
* Identify key brainwave frequencies associated with sleepiness.
* Evaluate the predictive power of brainwave features for drowsiness detection.
* Compare the effectiveness of different machine learning models in classifying drowsiness.
* Optimize model performance to enhance prediction accuracy.
* Provide insights into brainwave patterns related to alertness and fatigue.
* Develop a reliable system for early detection of driver drowsiness to improve road safety.

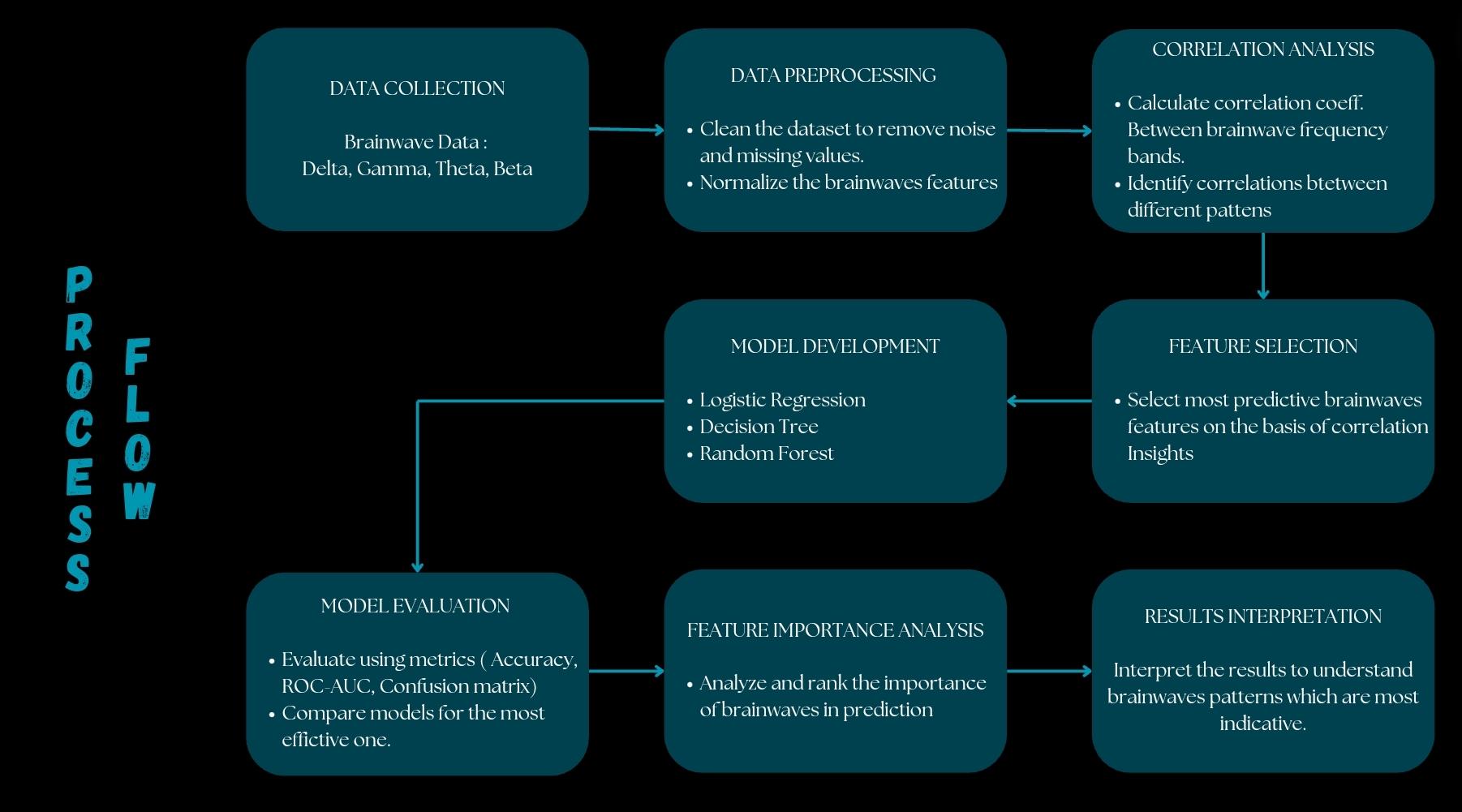
**MILESTONES**

1. Data Collection
2. Data Preprocessing
3. Correlation Analysis
4. Feature Selection
5. Model Development
6. Model Evaluation
7. Features Importance Analysis
8. Results Interpretation

**SOFTWARE AND TOOLS**

* Kaggle : Source of Dataset used in the project
* Python: Primary programming language used for data analysis and model Development
* Jupyter Notebook: Interactive environment for running Python code and visualizing data
* Scikit-learn: Machine learning library for training models like logistic regression, decision trees, and random forests.
* Pandas: Data manipulation and analysis library.
* NumPy: Fundamental package for numerical computation.
* Matplotlib/Seaborn: Libraries for data visualization and plotting graphs.

**PROCESS FLOW DIAGRAM**

****

1. **DATA COLLECTION : Kaggle**

Dataset : <https://www.kaggle.com/datasets/naddamuhhamed/sleepy-driver-eeg-brainwave-data>

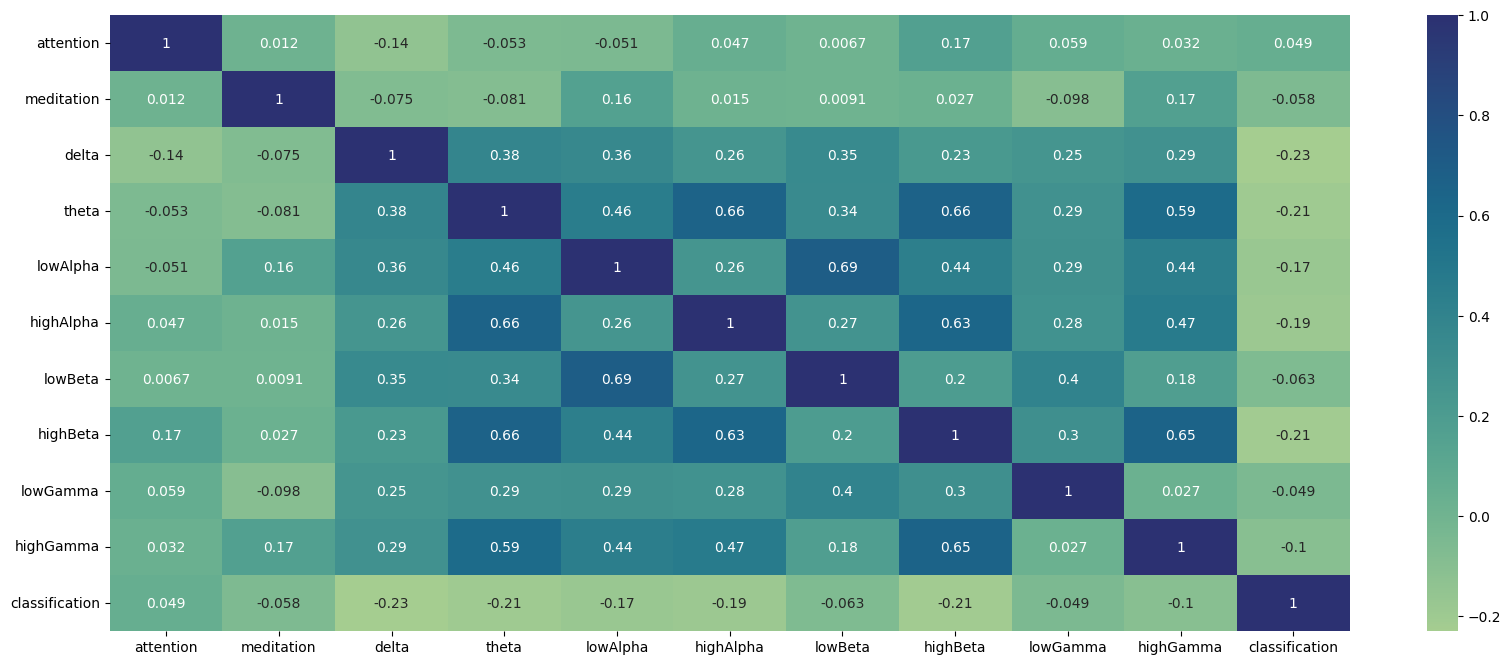
Sample Data:

| attention | meditation | delta | theta | lowAlpha | highAlpha | lowBeta | highBeta | lowGamma | high  Gamma | classification |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 26 | 34 | 960462 | 277180 | 26575 | 27356 | 26575 | 13056 | 9126 | 2683 | 0 |
| 29 | 54 | 39145 | 28225 | 20172 | 39551 | 20172 | 9933 | 5237 | 4750 | 0 |
| 40 | 48 | 75410 | 43144 | 8601 | 13564 | 8601 | 11663 | 2515 | 3251 | 0 |
| 66 | 47 | 16057 | 41211 | 2534 | 34254 | 2534 | 27663 | 11396 | 2825 | 0 |
| 81 | 67 | 10304 | 47239 | 33158 | 47349 | 33158 | 16328 | 5298 | 5471 | 0 |
| 69 | 54 | 95733 | 97816 | 22747 | 34873 | 22747 | 6898 | 7549 | 2603 | 0 |
| 57 | 67 | 38504 | 63547 | 48293 | 33267 | 48293 | 10078 | 2279 | 2055 | 0 |

1. **DATA PREPROCESSING & FEATURE ANALYSIS**
2. Understanding the brainwaves

| Wave Name | When They Are High | When They Are Low |
| --- | --- | --- |
| Attention | High values indicate high mental focus or concentration. | Low values indicate poor focus or distraction. |
| Meditation | High values indicate high levels of relaxation or calmness. | Low values indicate stress or agitation.s |
| Delta (1-3 Hz) | High values are prominent during deep sleep and restorative states. | Low values are seen during awake states and light sleep. |
| Theta (4-7 Hz) | High values are associated with light sleep, drowsiness, and meditative states. | Low values are seen during alert, wakeful states. |
| Low Alpha (8-11 Hz) | High values indicate a relaxed but alert state, often seen with eyes closed. | Low values are seen during active mental engagement or stress. |
| High Alpha (8-11 Hz) | High values indicate deep relaxation and calmness. | Low values are observed during active cognitive tasks or stress. |
| Low Beta (12-29 Hz) | High values are associated with active thinking, problem-solving, and alertness. | Low values are seen during relaxation or passive states. |
| High Beta (12-29 Hz) | High values indicate high alertness, stress, and intense mental activity. | Low values are observed during relaxation or low engagement. |
| Low Gamma (30-100 Hz) | High values reflect heightened cognitive processing and complex mental tasks. | Low values are seen during low cognitive demand or relaxation. |
| High Gamma (30-100 Hz) | High values indicate advanced cognitive functions, perception, and high focus. | Low values are observed during states of reduced cognitive activity or relaxation. |

1. Correlation Analysis

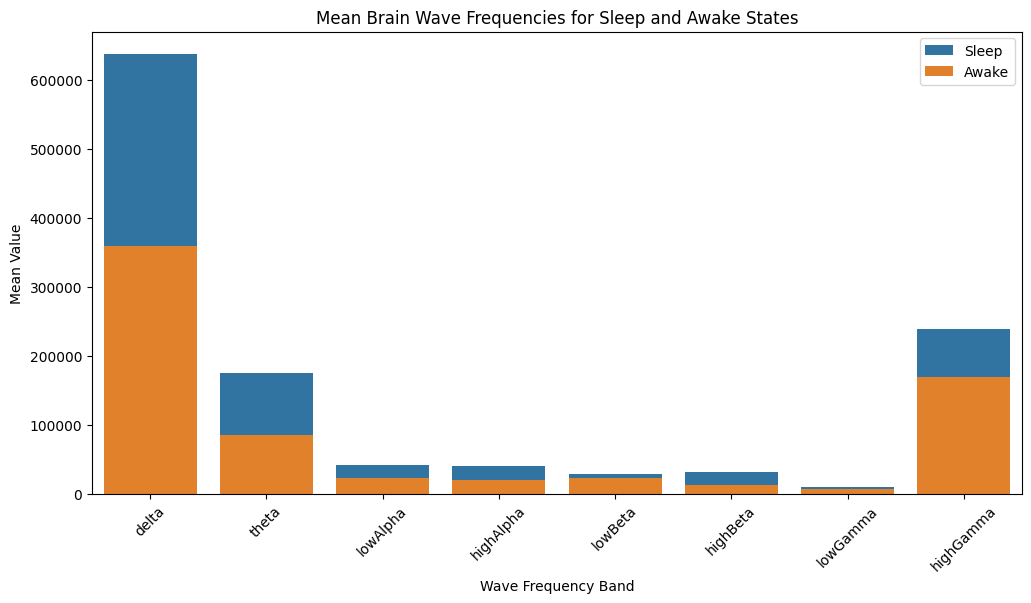


Insights:

* Delta waves show a negative correlation (-0.23), indicating higher values are associated with lower drowsiness likelihood, aligning with their presence in deep sleep.
* Theta and High Beta waves exhibit moderate negative correlations (-0.21), suggesting higher values also indicate lower drowsiness likelihood.
* Low Alpha and Low Beta features have weaker correlations, implying minimal impact on drowsiness prediction.

Brainwave Interactions:

* Theta waves strongly correlate with Low Alpha (0.66), High Alpha (0.66), and High Beta (0.66), indicating concurrent increases, likely reflecting transitional states between wakefulness and drowsiness.
* High Gamma waves positively correlate with High Beta (0.65) and Theta (0.59), highlighting their role in cognitive processing and alertness.

1. Brainwaves features for Sleep and Awake

The brain's neural oscillations can be grouped into distinct frequency bands, each corresponding to specific states of consciousness, cognitive processes, and functional roles.

* Delta Waves (1-3 Hz): The Restoration Band

High-amplitude, low-frequency delta waves dominate during deep sleep, unconsciousness, and restorative states, facilitating rejuvenation and recovery.

* Theta Waves (4-7 Hz): The Transition Zone

Theta waves emerge during light sleep, drowsiness, and meditative states, serving as a bridge between sleep and wakefulness, and enabling the brain to transition smoothly between states.

**Alpha Waves (8-11 Hz): The Relaxation Spectrum**

Alpha waves comprise two sub-bands:

* Low Alpha (8-10 Hz): The Calm Focus - Associated with relaxed yet alert states, closed eyes, and decreased cortical activity, promoting a sense of calm focus.
* High Alpha (10-11 Hz): The Serene State - Reflecting deep relaxation, calmness, and reduced cortical activity, ideal for unwinding and recharging.

**Beta Waves (12-29 Hz): The Engagement Band**

Beta waves encompass two sub-bands:

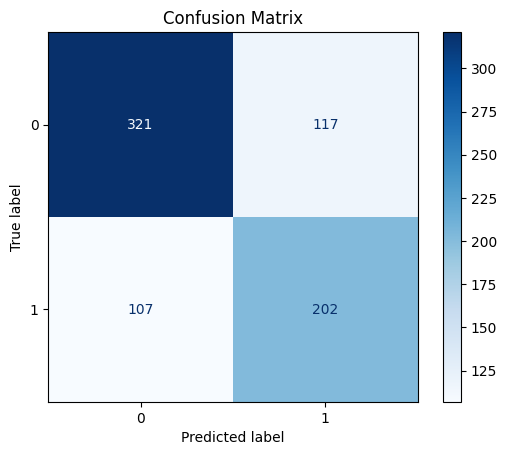
* Low Beta (12-15 Hz): The Active Mind - Linked to active thinking, problem-solving, and alertness, facilitating mental clarity and focus.
* High Beta (15-29 Hz): The Intense Focus - Characterized by high alertness, stress, and intense mental activity, driving concentrated mental effort.

**Gamma Waves (30-100 Hz): The Integration Zone**

* High-frequency gamma waves facilitate advanced cognitive processing, sensory integration, and information synthesis, enabling the brain to assimilate and process complex information.

**3. MODEL DEVELOPMENT**

1. Logistic Regression



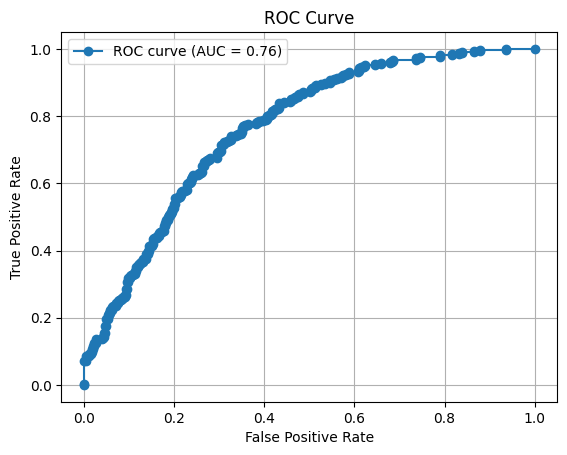
**Correct Predictions : 0.70013**

Logistic Regression achieved 70.01% accuracy in predicting driver drowsiness from brainwave data, indicating that brainwave patterns like Theta and Delta are useful indicators. However, its simplicity assumes a linear relationship between features and output, which may not be accurate for this problem.

**Limitations:**

* Assumes linear relationship between brainwaves and drowsiness
* Fails to capture intricate patterns in data

**Non-Linear Models:**

* Outperform Logistic Regression in predicting driver drowsiness
* Capture complex relationships between brainwaves and drowsiness
* 

Classification model's effectiveness in predicting driver drowsiness by plotting True Positive Rate (TPR) against False Positive Rate (FPR). the key components:

**True Positive Rate (Y-axis):**

Measures the model's ability to accurately detect drowsy drivers, showcasing its sensitivity.

**False Positive Rate (X-axis):** **The Error Rate**

Represents the proportion of non-drowsy drivers misclassified as drowsy, highlighting potential false alarms.

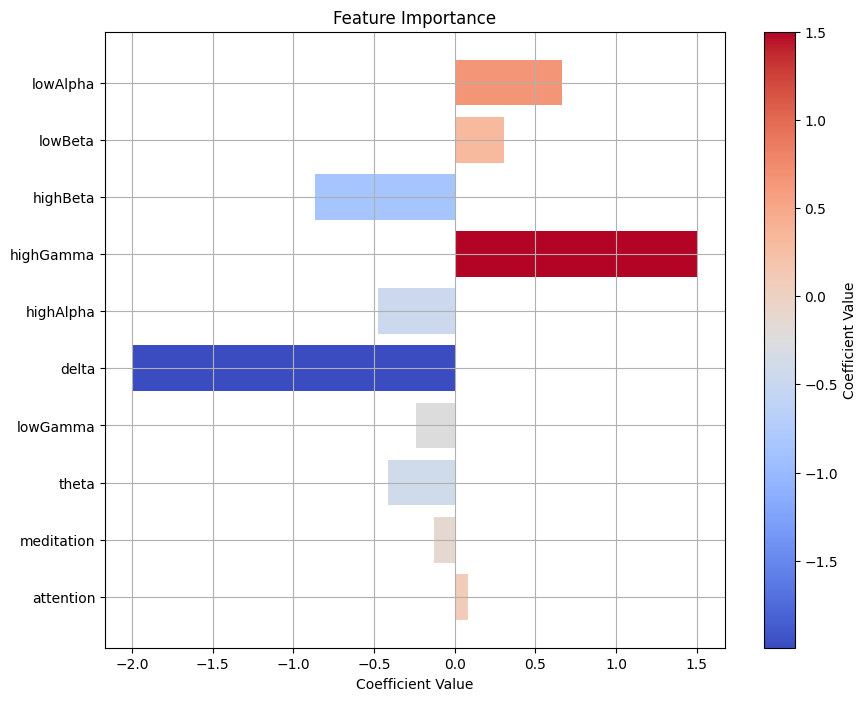
**Area Under the Curve (AUC = 0.76): Discriminatory Power Revealed**

With an **AUC of 0.76**, the model demonstrates a fairly strong ability to differentiate between drowsy and non-drowsy drivers, indicating:

**A perfect classifier would score 1.0**

**Random guessing would yield 0.5**

Our model has a **76%** chance of correctly distinguishing between a randomly chosen drowsy driver and a non-drowsy driver



**Key Brainwave Features Influencing Drowsiness**

Our model reveals the significance of various brainwave features in predicting driver drowsiness, ranked by importance:

**Highly Influential Features**:

* High Gamma Waves: Strongly correlated with drowsiness, high gamma waves may indicate mental fatigue and cognitive overload, leading to drowsiness.
* Delta Waves: Show a significant inverse relationship with drowsiness, suggesting that low delta wave activity during wakefulness may indicate alertness.

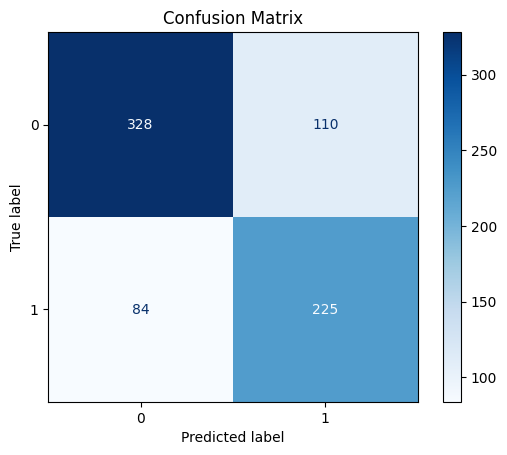
**Moderately Influential Features:**

* High Beta Waves: Positively associated with drowsiness, high beta waves may reflect stress or intense mental engagement, leading to brief periods of alertness.
* High Alpha Waves: Contribute to predictions, but to a lesser extent, potentially indicating deep relaxation.

**Less Influential Features**:

* Low Alpha and Low Beta Waves: Moderate to small coefficients suggest a lesser impact on predictions.
* Theta, Low Gamma, and Attention: Minimal influence on drowsiness predictions.
* Meditation: Shows a negligible impact, indicating that relaxation levels do not strongly predict drowsiness in this dataset.

1. Decision Tree Classifier



**Correct Predictions - 0.7402**

The performance of our Decision Tree Classifier, achieving an accuracy of 74.02%:

**True Negatives (TN)** - 328: Correctly predicted alert drivers, showcasing the model's ability to identify non-drowsy states.

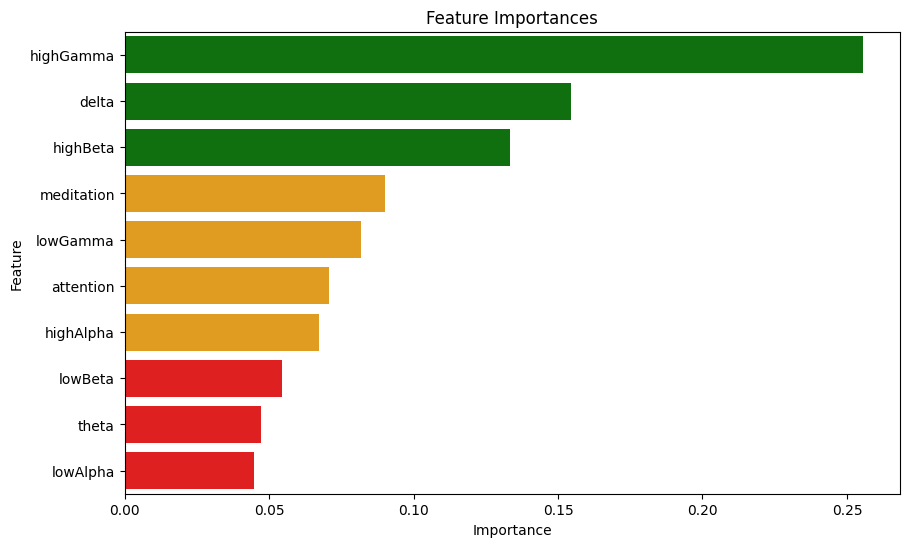
**False Positives (FP)** - 110: Incorrectly predicted drowsy drivers, highlighting a moderate False Positive Rate and potential for false alarms.

**False Negatives (FN)** - 84: Missed drowsy drivers, indicating a relatively low False Negative Rate and strong performance in detecting drowsiness.

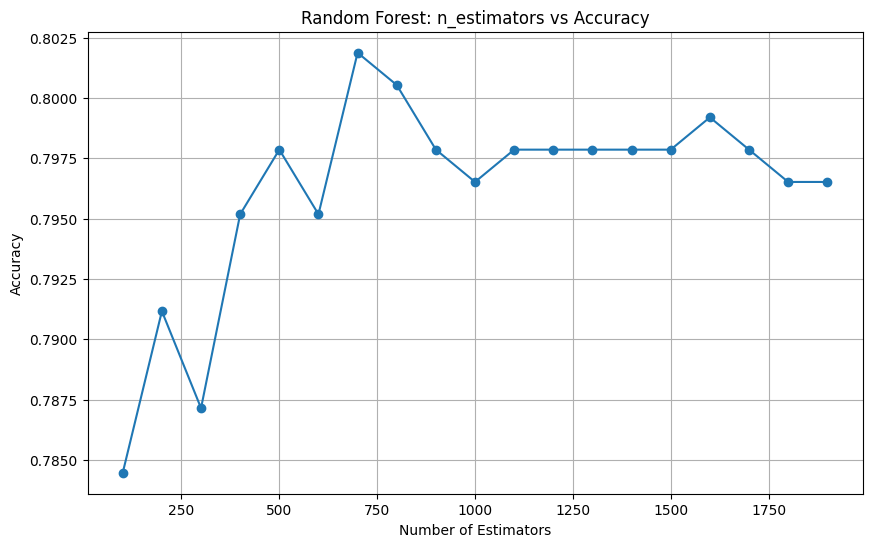
**True Positives (TP)** - 225: Correctly predicted drowsy drivers, demonstrating the model's strength in identifying drowsy states.

**Performance Insights:**

* More correct predictions (553) than incorrect ones (194)
* Moderate False Positive Rate, suggesting some false alarms
* Relatively low False Negative Rate, indicating strong drowsiness detection

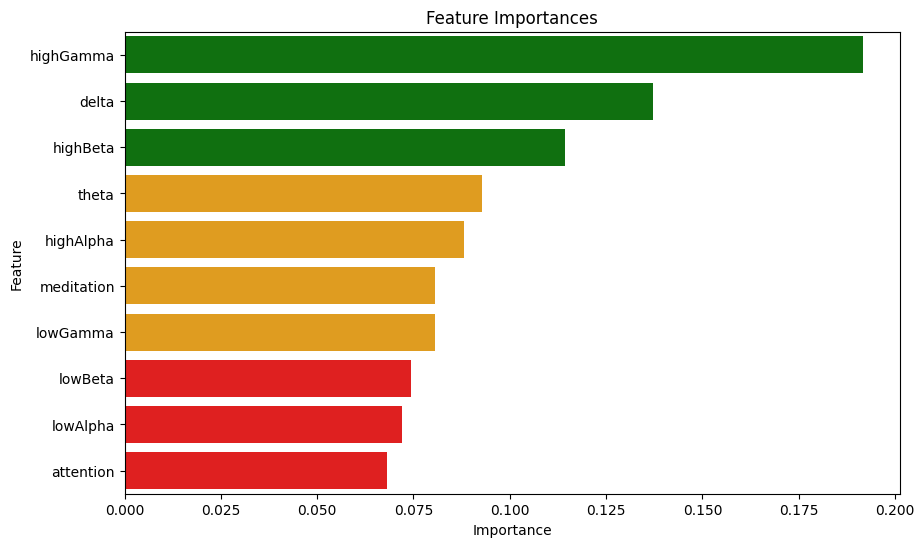


1. Random Forest Classifier



**Max Accuracy at 700 trees ~ 80.19%**

**The plot reveals:**

* Peak accuracy of 80.19% achieved with approximately 700 trees
* No significant improvement beyond 700 trees, indicating an optimal point
* 

**Feature Importance (Random Forest Classifier)**

The Random Forest model reaffirms the importance of:

* High Gamma, Delta, and High Beta waves
* Theta, High Alpha, and Meditation, contributing less significantly
* Low Beta, Low Alpha, and Attention, having minimal impact

**RESULT INTERPRETATIONS**

Our analysis successfully used brainwave data and machine learning models to predict driver drowsiness, showing valuable insights:

**Key Findings:**

* High Gamma, Delta, and High Beta waves are the most crucial predictors of drowsiness, showing their role in cognitive processing, deep sleep, and alertness.
* Logistic Regression achieved **70%** accuracy, while Decision Tree Classifier improved to **74%,** showing complex relationships in the data.
* Random Forest Classifier succeeded with **80.19% accuracy**, effectively capturing patterns in brainwaves.

**Validation:**

* **The ROC curve (AUC = 0.76)** confirms the model's ability to differentiate between drowsy and non-drowsy drivers.
* The confusion matrix highlights the model's strengths and weaknesses, with room for improvement.

**REAL WORLD APPLICATIONS**

Our research shows that predicting driver drowsiness using EEG brainwave data is possible, giving the way for:

* Integration into driver-monitoring systems to prevent accidents caused by drowsiness
* Further enhancements through feature refinement or exploration of advanced models
* By leveraging machine learning and brainwave analysis, we can create safer roads and save lives.

**CONCLUSION**

In conclusion, this study successfully utilized brainwave data and machine learning to predict driver drowsiness, with High Gamma, Delta, and High Beta waves emerging as key indicators. The models demonstrated the potential for drowsiness detection, with the Random Forest Classifier achieving the highest accuracy at 80.19%. The ROC curve and confusion matrix further validated the model's effectiveness, though some areas for improvement remain. This research highlights the feasibility of integrating brainwave-based drowsiness prediction into real-world driver-monitoring systems, offering a promising path toward enhancing road safety and preventing accidents**.**