**AI Final Project Report**

**Spring 2025, IITU**

**I. Project Statement**

**1. Personal story or empathy reasoning**

*Mental health has become an especially important topic in recent years, particularly after the pandemic and the rise of anxiety disorders among young people and IT professionals. I have personally encountered situations where friends or loved ones were unable to recognize that they were experiencing stress or depression. This inspired me to create a project that, with the help of AI, can help a person become aware of their condition and potentially seek help in time.*

**2. Problem definition**

*Many people are unaware that they are in a potentially dangerous mental health state, as symptoms can be subtle or mistaken for “normal fatigue.” This is especially true for students and young professionals who are not accustomed to seeking help from psychologists. The lack of early diagnosis can lead to a deterioration of their condition, including burnout or depression.***3. Ideation – possible AI solutions**

*Condition assessment through a questionnaire (as in my project).*

*Analysis of facial expressions and voice during a conversation with a chatbot.*

*Prediction of anxiety level based on social media activity.*

*Sleep and heart rate analysis from smartwatches to assess well-being.*

*AI consultant trained on psychotherapist data.*

*Mobile app with recommendations based on mood.*

*Condition detection through handwritten text.*

*Behavioral pattern tracking using habit trackers.*

*Generation of personalized meditations based on current mental state.*

*Alert system for HR departments about potential employee burnout risks.*

**4. Prototyping**

*Several prototypes were created during the project, differing in how the questionnaire was processed and in their datasets. The final version is a form with questions, after which a set of AI algorithms is used to determine the mental state level: Very Low, Low, Medium, High, Very High. Each algorithm works independently, allowing for result comparison.*

**5. Testing**

*The prototype was tested on a group of volunteers (e.g., students and colleagues). During testing, discrepancies in the predictions of different algorithms were identified, which led to questionnaire optimization — some questions were removed or rephrased. Additionally, result gradation and explanations were added so participants could better understand their outcome.*

**6. AI Ethics**

*One of the key ethical concerns is the risk that a participant might interpret the result as a medical diagnosis. There is also the risk of data leaks if secure storage is not ensured. To mitigate these issues, warnings were added stating that the result is an assessment, not a diagnosis, and all data is processed anonymously and not stored on the server.*

**II. Data Source**

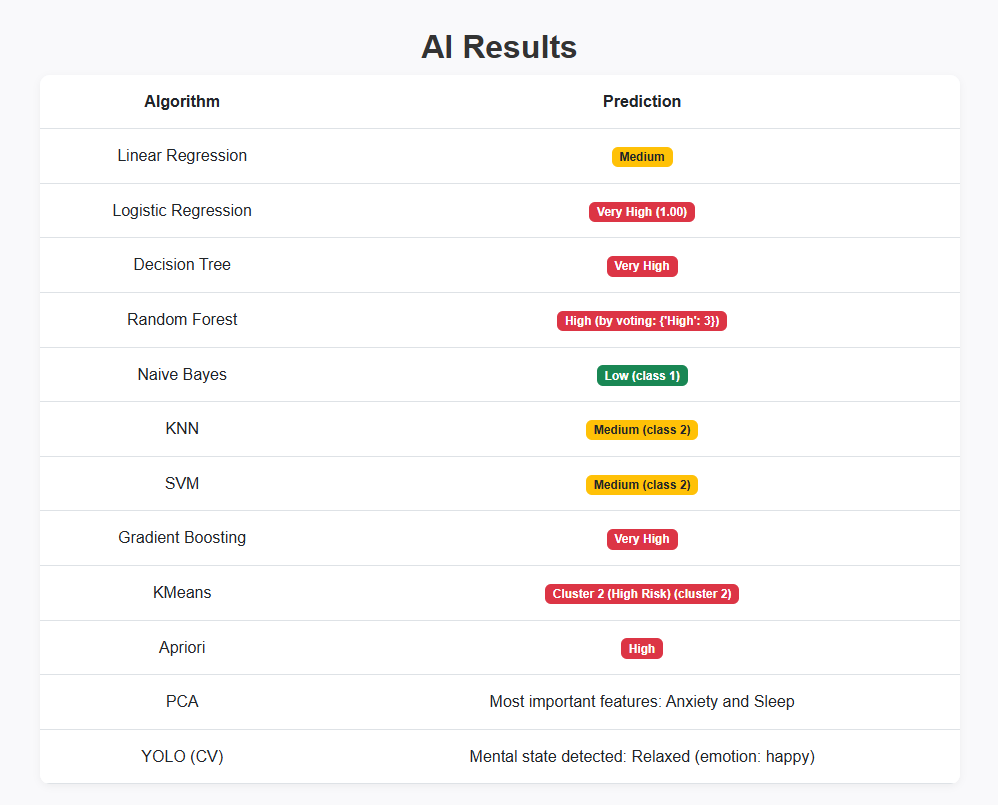
*For the implementation of the project, a custom dataset was used, containing records with the following indicators: overall emotional state (feeling), sleep quality (sleep), anxiety level (anxiety), energy level (energy), stress level (stress), and the resulting psycho-emotional state level (anxiety\_level), expressed in five categories:  
0 – Very Low, 1 – Low, 2 – Medium, 3 – High, 4 – Very High.*

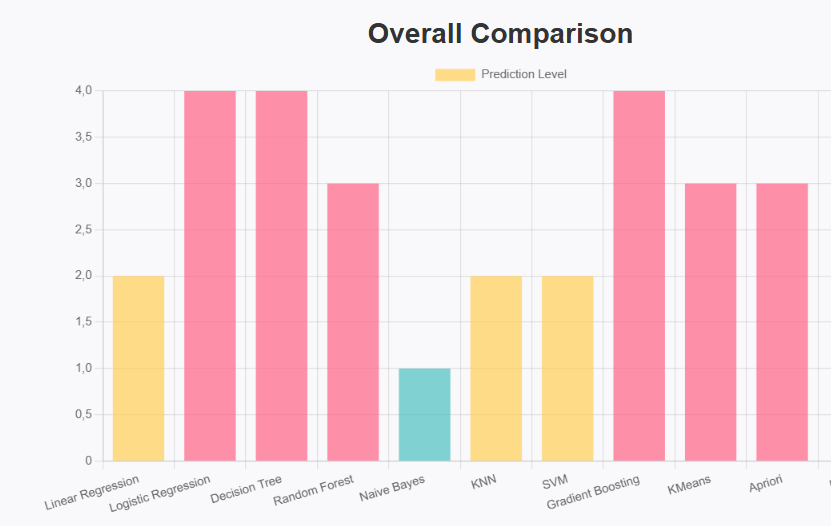
*The dataset was created based on questionnaire data that reflect typical psychological and behavioral manifestations of mental states. Each parameter was assessed on a scale from 1 to 5, where 1 represents the most unfavorable level and 5 the most favorable. The final column, anxiety\_level, is a generalized label used to train classification models.*

*The questionnaire used for data collection was composed based on this dataset and designed to be completed by users in an interactive form. Thus, the data were generated with realistic scenarios in mind and correspond to a typical self-assessment scale of psychological well-being.*

**III. Implementation**

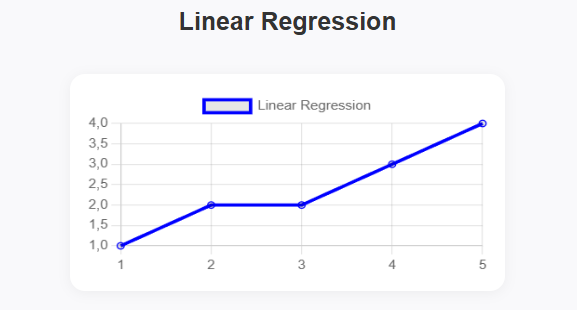
All algorithms in the project were implemented manually from scratch, without using any machine learning libraries (such as scikit-learn, XGBoost, etc.). This approach allowed for a deeper understanding of the internal mechanisms of each model and full control over the **calculations.**

****

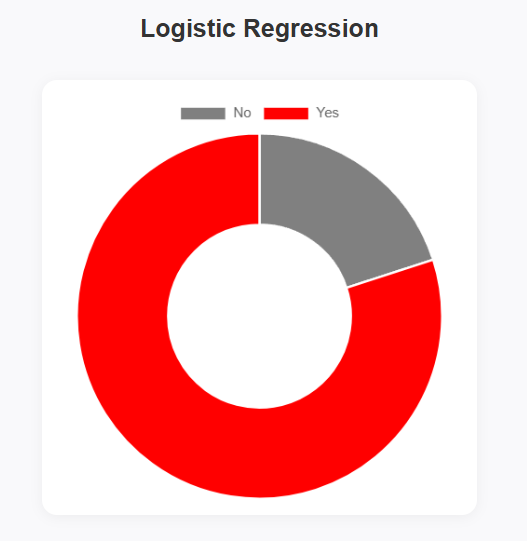
****

**A. Supervised Learning Algorithms**

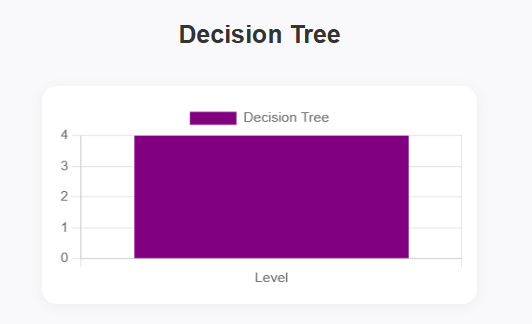
1. Linear Regression  
Implemented using the Mean Squared Error (MSE) formula and gradient descent optimization. Used to analyze the influence of input features on the mental health level.



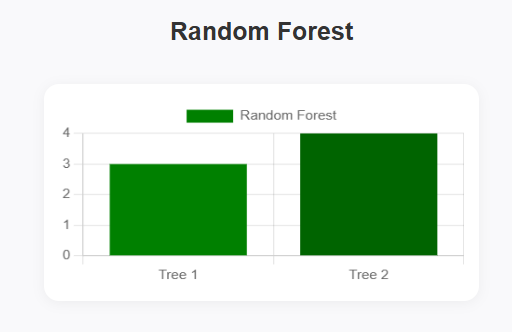
2. Logistic Regression  
Implemented with a sigmoid activation function and optimized using gradient descent. The model predicts the probability of a given input belonging to each mental health class.



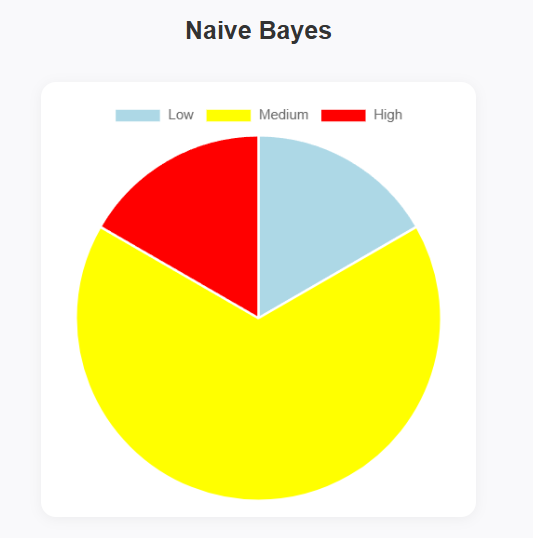
3. Decision Tree  
Built using recursive splitting based on the best feature according to entropy or Gini impurity. The tree continues splitting until pure leaves or a stopping condition is met.



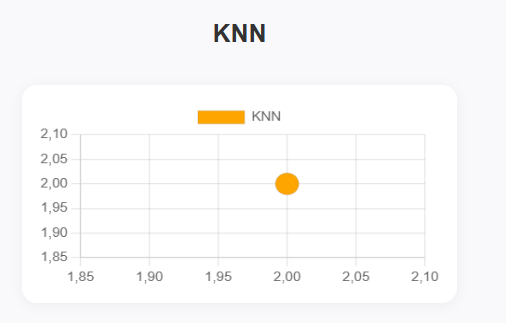
4. Random Forest  
Constructed as an ensemble of multiple decision trees using bootstrap sampling and random feature selection. Final predictions are made through majority voting.



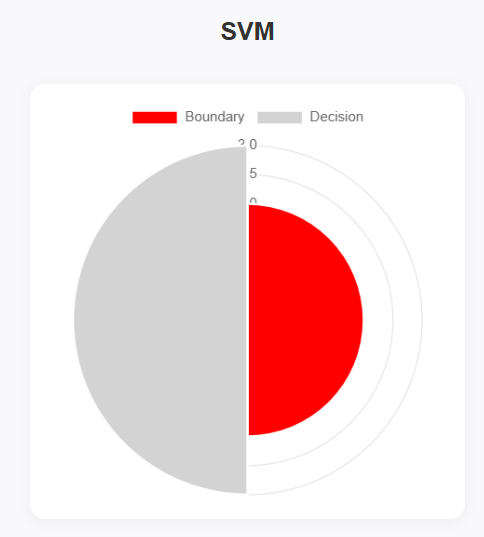
5. Naive Bayes  
Implemented using Bayes' theorem with calculated prior and conditional probabilities based on feature frequencies. Assumes independence between features.



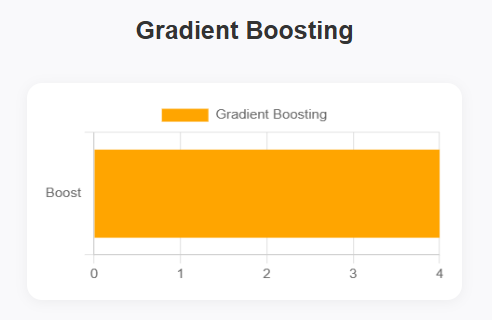
6. K-Nearest Neighbors (KNN)  
Implemented using the Euclidean distance metric. The model compares a new input with the training data and assigns the class most common among the k nearest neighbors.



7. Support Vector Machine (SVM)  
Implemented as a linear classifier that finds the optimal separating hyperplane by maximizing the margin between classes.

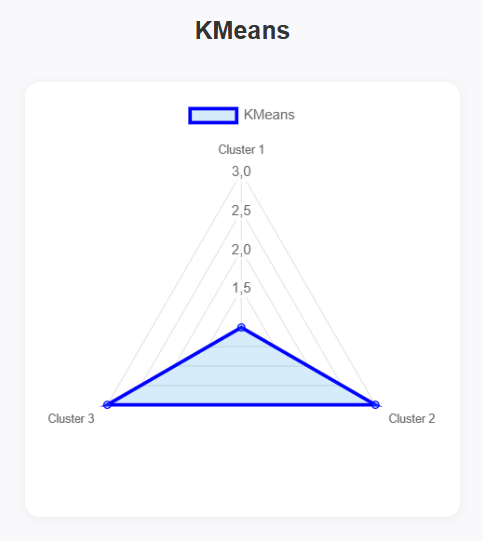


8. Gradient Boosting  
Implemented as a sequential combination of weak learners (decision trees), where each new tree is trained on the residual errors of the previous one using gradient approximation.

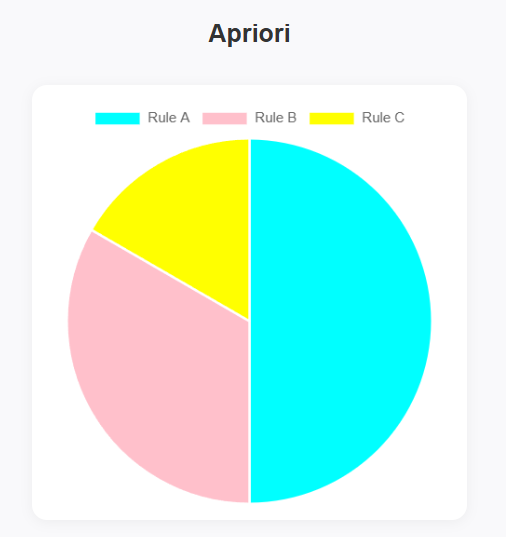


**B. Unsupervised Learning Algorithms**

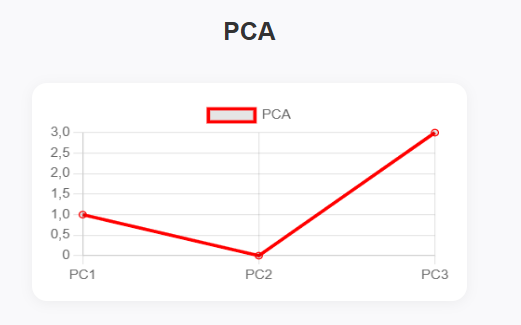
1. Clustering – K-Means  
Started with randomly selected cluster centers. In each iteration, data points were reassigned to the nearest center, and centers were recalculated. Stopping condition was center convergence.



2. Association Rules – Apriori  
Frequent itemsets were identified based on minimum support and confidence thresholds. Rules were generated to describe dependencies between psychological symptoms.



3. Dimensionality Reduction – PCA  
Principal Component Analysis was implemented by calculating the covariance matrix and extracting eigenvalues and eigenvectors. Used to reduce dimensionality and visualize data in 2D.



**C. Computer Vision Algorithms**

In this project, the DeepFace library was used to integrate facial emotion recognition into the system. This allowed the model to analyze a user's facial expression and infer their emotional state as an additional input for mental health assessment.

The DeepFace model was used to extract emotional attributes (such as happiness, sadness, anger, fear, surprise, etc.) from facial images submitted by users. These emotion scores were mapped to the five mental health categories used in the project: Very Low, Low, Medium, High, Very High.

The algorithm implementation included the following steps:

* Image preprocessing and face detection.
* Emotion recognition using DeepFace’s built-in models (such as VGG-Face or Facenet).
* Integration of emotion prediction with survey data for enhanced accuracy.

This component added a multimodal layer to the system, allowing it to combine questionnaire data and visual emotional cues. It proved particularly useful for identifying emotional states not easily captured through self-assessment alone.



**IV. Delivery and Defense**

1. **Презентация в формате .pptx или .pdf**, понятная даже тем, кто не разбирается в ИИ.

**GitHub:** https://github.com/fixfrice/mental.git