卡通人物

中度可信度描述已自动生成

《机器学习（双语）》

期末大作业

**题 目**  Discussion Topics

Strategy-Designing Report

专 业 网络工程

学 号

姓 名

任 课 教 师

日 期 2025.01.17

Chapter 1 Discussion Topics

## Topic 1

In a machine learning system, the core process revolves around the interaction between three fundamental components. These components work together in a dynamic, iterative manner to solve a given learning problem. The training data plays a central role in this process, flowing through the system to enable the learning algorithm to optimize the model.

### 1.1.1 Components of a Typical Machine-Learning System

A typical machine-learning system comprises three fundamental components: features, the model, and the learning algorithm. These components interact to solve a learning problem, where the ultimate task is to map input data described by features to the desired outputs. Understanding these components and their relationships is crucial for building effective machine-learning solutions.

**Features** are the measurable properties or characteristics extracted from the raw data. They are the inputs to a machine-learning system and serve as the foundation for representing domain objects. The process of selecting, engineering, or modifying features determines the system's capacity to represent the problem accurately. Without well-defined features, even the most sophisticated models will fail to generalize or make useful predictions.

In the diagram, domain objects are transformed into features, which are passed as data to the next component – the model. This transformation is a critical step because the quality of the features directly influences the model's learning capacity.

**The model** is the mapping function that translates features into outputs. It represents the core mathematical or computational formulation that makes predictions or decisions. The model's structure depends on the nature of the problem (e.g., classification, regression) and the data provided.

Obtaining a model involves solving a learning problem: The learning problem aims to discover an appropriate mapping between features (input) and the output using training data. Training data serves as the foundation for building this mapping, and its quality directly affects the model's accuracy. The model is influenced both by the data (features) and by the learning algorithm that optimizes its parameters.

**The learning algorithm** is responsible for adjusting the model’s parameters to improve its performance based on training data. It aims to minimize the error between the predicted output and the actual output, thereby solving the learning problem. The relationship between the learning algorithm and the model is iterative – the algorithm updates the model through optimization processes, such as gradient descent or other optimization techniques.

The training data flows into the learning algorithm, which updates the model. This highlights the learning process as a closed-loop interaction where training data, the algorithm, and the model interact to achieve the task.

### 1.1.2 Relationships Between the Components

The relationships between features, the model, and the learning algorithm are highly interdependent:

1. Features serve as inputs to the model and directly influence the learning process. Poor features result in inaccurate or suboptimal models.

2. The model is the mapping function, optimized to minimize errors during learning.

3. The learning algorithm drives the optimization of the model, leveraging training data to iteratively improve predictions.

Additionally, the overall success of the system depends on the quality of the training data and the appropriateness of the learning process for solving the task at hand.

In a machine-learning system, features, the model, and the learning algorithm are tightly connected and central to solving the learning problem. By understanding how these components interact – as shown in the system architecture – it becomes possible to design robust and effective solutions that meet specific tasks and challenges.

## 1.2 Topic 2

Machine learning encompasses a variety of learning models, each with distinct characteristics and applications. Understanding these models is crucial for selecting the appropriate approach for a given problem.

### 1.2.1 Learning models

**Linear models**, such as Support Vector Machines (SVM), assume a linear relationship between the input features and the target variable. These models are suitable for problems where the underlying relationship can be approximated by a linear function. They are relatively simple to interpret and can perform well on linearly separable data. Examples of linear models include linear regression and logistic regression.

Tree models, such as Decision Trees and Random Forests, represent the learning process as a hierarchical tree-like structure. These models make decisions based on a series of if-then-else rules, making them intuitive and easy to interpret. Tree models can handle both numerical and categorical features, and they are effective at capturing non-linear relationships in the data. They are particularly useful for classification and regression tasks.

**Distance-based models**, such as K-Nearest Neighbors (KNN), Gaussian Mixture Models (GMM), and K-Means Clustering, rely on the concept of proximity or similarity between data points. These models make predictions or cluster data based on the distance or similarity between the input and the training data. They are effective for problems where the underlying structure of the data can be captured by the proximity of data points.

**Probability models**, such as Naive Bayes and Bayesian Networks, use probabilistic approaches to make predictions. These models estimate the probability distribution of the target variable given the input features. Naive Bayes is a simple and efficient algorithm that assumes independence between features, making it suitable for high-dimensional data. Bayesian Networks capture the complex relationships between variables using a directed acyclic graph, allowing for more sophisticated probabilistic modeling.

**Neural networks and deep learning models** are inspired by the structure and function of the human brain. These models consist of interconnected layers of artificial neurons that can learn complex non-linear relationships in the data. Deep learning models, with their multiple hidden layers, can automatically extract and learn features from raw data, making them highly versatile and powerful. Neural networks and deep learning are particularly effective for tasks such as image recognition, natural language processing, and speech recognition.

### 1.2.2 Conclusion

The choice of the appropriate learning model depends on the characteristics of the problem, the nature of the data, and the desired level of interpretability and performance. For example, if the problem requires high interpretability and the data is linearly separable, a linear model like logistic regression may be the best choice. On the other hand, if the problem involves complex, non-linear relationships in the data, a tree-based model or a neural network may be more suitable. The size and quality of the available data also play a crucial role in determining the most effective model. Simpler models like k-nearest neighbors may perform well with small datasets, while deep learning models often require large amounts of data to achieve their full potential. Understanding the strengths and weaknesses of each model type, as well as their suitability for different problem domains, is crucial for selecting the most suitable approach for a given machine learning task. The choice of model can significantly impact the performance, interpretability, and real-world applicability of the machine learning system.

# Chapter 2 Strategy-Designing Report

## 2.1 Introduction

In recent years, the mental health of college students has emerged as a critical focus area due to the increasing academic pressures, social challenges, and rapid development of higher education in China. As students transition from adolescence to adulthood, they often face psychological stress, including anxiety, depression, and burnout. These challenges can hinder academic performance, social interactions, and personal growth, underscoring the need for timely and effective mental health interventions.

The application of data-driven machine learning (ML) strategies offers a promising avenue to address these challenges. By leveraging advanced algorithms, such strategies can help automatically estimate students' mental health conditions and provide personalized interventions to improve their well-being. A well-designed ML system for mental health monitoring would enable early detection, ongoing assessment, and tailored recommendations based on individual needs.

This report outlines a machine learning strategy to automatically estimate and improve college students' mental health. The objective is to ensure that the proposed system is reasonable, feasible, and actionable while being aligned with ethical considerations such as data privacy and fairness.

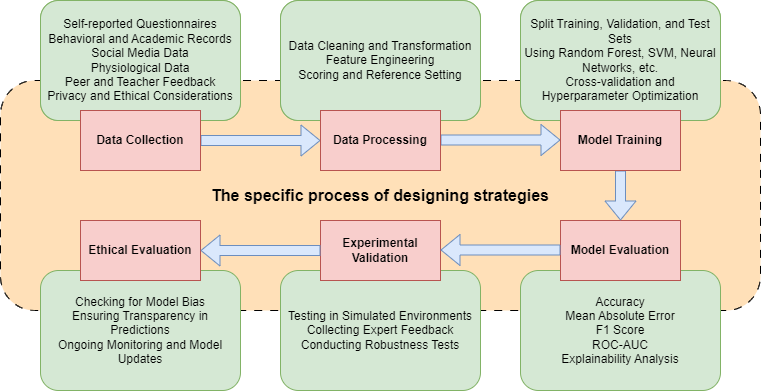


Figure 1 The specific process of designing strategies

## 2.2 Data Collection

Acquiring high-quality, relevant, and diverse datasets is fundamental to any data-driven strategy. In the context of assessing and improving college students' mental health, the data collection process must fulfill two primary objectives: understanding the psychological state of students and identifying external factors influencing mental well-being.

### 2.2.1 Data Sources

A comprehensive evaluation of college students' mental health requires data from various sources. First, self-reported questionnaires and surveys are essential tools. Instruments like the Depression, Anxiety, and Stress Scales (DASS)[1,2] or the Patient Health Questionnaire (PHQ-9)[3] provide validated means of gathering subjective data on students' emotional states, establishing a baseline for subsequent machine learning (ML) model training.

Second, behavioral and academic records offer indirect indicators of mental health. Information such as attendance, library usage, extracurricular participation, academic grades, and adherence to submission deadlines can help identify patterns and potential risk factors.

Third, social media and digital footprints can reveal emotional well-being through the analysis of language, sentiment, and frequency of online interactions, provided students' consent is obtained. Patterns in these activities may signal stress or depressive tendencies.

Fourth, physical and physiological data collected through wearable devices and smartphones, including sleep patterns, heart rate variability, and physical activity levels, are strongly correlated with mental health. For instance, irregular sleep schedules or prolonged sedentary behavior might indicate stress or depression.

Finally, peer and faculty feedback offers valuable qualitative insights into students' behavior and engagement, complementing quantitative data for a more holistic understanding of their mental health.

### 2.2.2 Ethical Considerations

The sensitive nature of mental health data necessitates strict adherence to ethical guidelines. Informed consent must be obtained, ensuring students understand the purpose and scope of data collection. Anonymization and privacy are critical to safeguarding personal information, while data security measures must be implemented to prevent breaches and unauthorized access. These practices uphold the integrity and trustworthiness of the data collection process.

### 2.2.3 Data Collection Methods

To gather data effectively, multiple collection methods can be utilized. Digital platforms for surveys and feedback enable widespread participation, particularly if gamified features are incorporated to reduce response bias. Integration with institutional systems such as Learning Management Systems (LMS)[4] and attendance trackers allows access to existing academic and behavioral data.

Additionally, mobile applications and wearable devices can passively collect physiological and behavioral data, offering real-time insights into student well-being. Anonymous focus groups provide qualitative feedback, creating a platform for students to express concerns without fear of identification. Furthermore, publicly available open datasets related to student mental health can be ethically utilized to validate findings or pre-train ML models.

### 2.2.4 Data Quality and Pre-Processing

Ensuring the quality of collected data is vital for meaningful analysis. Diversity in the dataset, encompassing students of different genders, academic years, and socio-economic backgrounds, mitigates potential biases in ML models. To address incompleteness, strategies like automated reminders and participation incentives can minimize missing data. Lastly, validation of self-reported data with objective indicators, such as academic performance and physiological metrics, enhances the reliability of the dataset.

By employing a structured, multi-faceted approach to data collection, the groundwork is established for accurate predictions and impactful interventions in later stages of the machine learning pipeline.

## 2.3 Data Processing

The collected data must undergo systematic processing to ensure it is clean, standardized, and ready for use in machine learning model development. This step is essential for generating accurate predictions and actionable insights in a data-driven framework for mental health assessment and improvement. The data processing workflow includes three primary tasks: data cleaning and transformation, feature engineering, and the assignment of scores with reference ratios.

### 2.3.1 Data Cleaning and Transformation

Data cleaning and transformation are vital steps to prepare raw datasets by removing inconsistencies and ensuring uniformity. Handling missing data is an important aspect of this process. Missing values, such as incomplete survey responses or gaps in physiological metrics, can be addressed through imputation methods like mean or median imputation for numerical data or more advanced techniques such as k-nearest neighbors (KNN) imputation. In cases where records have substantial missing data, these records may be dropped if their removal does not compromise overall data integrity.

Standardization and normalization are applied to numerical features such as heart rate variability, activity levels, and academic scores to ensure comparability, with normalization scaling values to a 0-1 range and standardization adjusting data to have a zero mean and unit variance. Text data, including survey responses and social media posts, is converted into numerical representations using techniques like sentiment analysis, which quantifies emotional tone, or text embeddings such as TF-IDF[5] and Word2Vec[6], which capture semantic features. To ensure privacy, personally identifiable information (PII) is anonymized by replacing it with unique identifiers.

### 2.3.2 Feature Engineering

Feature engineering transforms raw data into meaningful inputs for the machine learning model.

**Assigning Weights to Features**: Assign reasonable weights to features based on their perceived importance in mental health estimation. For example:

1. Self-reported mental health scores: High weight (e.g., 40%) due to their direct correlation with the target variable.

2. Physiological data: Medium weight (e.g., 30%) as it offers objective indicators.

3. Behavioral and academic metrics: Lower weight (e.g., 20%) as they serve as indirect indicators.

4. Social media sentiment: Medium weight (e.g., 10%) for qualitative insight into emotional states.

图表, 饼图

描述已自动生成

Figure 2 Data Contribution to Mental Health Assessment

**Combining Features**: Combine related features to create composite scores, such as a Stress Score calculated from sleep irregularities, sentiment polarity, and academic deadlines missed.

**Categorizing Data**: Convert numerical scores into categorical variables for certain analyses, such as: Stress levels: [Low, Moderate, High], Sleep patterns: [Adequate, Insufficient, Erratic].

**Deriving Temporal Features**: Use time-series analysis to extract trends or patterns, such as declining attendance or worsening sleep over weeks.

### 2.3.2 Assigning Scores and Reference Ratios

To make the data actionable, assign scores and define reference ratios for each feature. This step quantifies mental health indicators and enables systematic evaluation.

**Scoring System**: Create a unified scoring system (e.g., a scale of 0–100) to represent students' mental health. Scores are computed as a weighted sum of individual features:

 (2-1)

where  is the weight of feature , and  is the normalized value of feature .

**Example Scoring Criteria**: The following table outlines the scoring criteria for assessing various aspects of student mental health. Each feature is assigned a weight based on its relevance, and the scoring formula along with reference ratios are used to calculate a comprehensive mental health score.

Table 1: Scoring Criteria for Mental Health Assessment

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Weight** | **Scoring Formula** | **Reference Ratio** |
| Sleep Quality | 20% | Score = (Average hours of sleep / Recommended hours) × 100 | 7–8 hours/day |
| Social Media Sentiment | 10% | Score derived from sentiment analysis | Sentiment polarity > 0.3 |
| Self-Reported Stress | 40% | Score = 100 - (Stress Level (DASS) × 10) | Stress level < 10 |
| Behavioral Engagement | 20% | Score = (Days engaged / Total days) × 100 | Engagement > 75% |
| Academic Performance | 10% | GPA normalized to a scale of 0–100 | GPA > 60 (passing threshold) |

**Reference Ratios**: Reference ratios represent thresholds for good mental health. Students falling below these thresholds are flagged for further evaluation. There are some example reference ratios in the above table.

**Composite Mental Health Categories**: Based on the overall mental health score, categorize students into:

Healthy (80–100): No immediate intervention required.

At Risk (50–79): Moderate stress; monitor closely.

Critical (0–49): Immediate support needed.

### 2.3.4 Output

The processed data will serve as inputs to machine learning models for predicting mental health conditions and suggesting interventions. The clear assignment of scores and reference ratios ensures interpretability and enables actionable outcomes.

## 2.4 Experimental Evaluation Design

The experimental evaluation design is critical to validate the effectiveness and feasibility of the proposed data-driven machine learning strategy for estimating and improving college students' mental health. This section outlines the steps for building, training, and evaluating the machine learning model, along with the metrics and methods for assessing its performance.

### 2.4.1 Overview of the Experimental Process

The primary objective of this experiment is to develop a machine learning model capable of estimating the mental health of college students based on processed data. The model will also offer actionable insights for improving student mental health. The process involves splitting the dataset into training, validation, and test sets, applying supervised learning algorithms to predict mental health scores, and evaluating the model's predictions against actual scores and established thresholds. The expected outcome is a validated model that demonstrates high accuracy, interpretability, and reliability in estimating the mental health conditions of students.

**Dataset Preparation**: The dataset used for the evaluation will be divided into three subsets: the training set, the validation set, and the test set. The training set will consist of 70% of the data and will be used to train the machine learning model. The validation set, making up 15% of the data, will be employed to fine-tune hyperparameters and help prevent overfitting. The remaining 15% will form the test set, which will be used to assess the model's performance on unseen data. Stratified sampling will be utilized to ensure that various demographic groups, such as gender, academic, and socio-economic background, are proportionally represented in each subset.

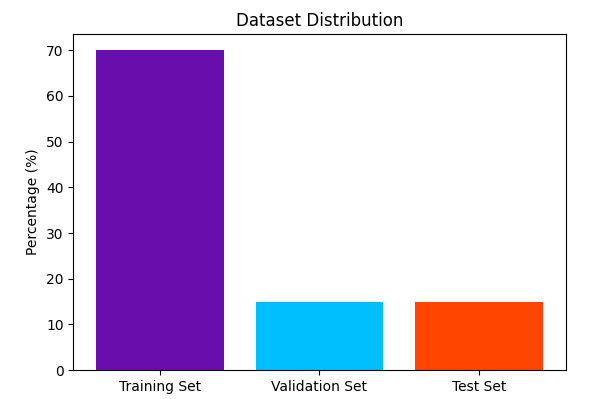


Figure 3 Data Distribution for Training, Validation, and Test Sets

**Model Selection and Training**: Several machine learning algorithms will be considered for the mental health estimation task:

1. **Random Forest** (Baseline Model): Known for its robustness against noise and ability to handle both numerical and categorical data, Random Forest will be used to predict mental health scores and identify important features.
2. **Linear models** (e.g., Support Vector Machines(SVM)): These algorithms are recognized for their high accuracy and efficiency, especially when dealing with tabular data, and will be applied to capture complex interactions between features.
3. **Neural Networks**: Neural networks will be used for their capability to model non-linear relationships and integrate diverse data sources, such as text embeddings and physiological metrics.

图表, 折线图

描述已自动生成

Figure 4 Performance Metrics of Algorithms

During the training phase, feature scaling and normalization will be applied to ensure consistent model performance. Cross-validation, such as 5-fold cross-validation, will be used to evaluate the stability of the models and reduce overfitting. Hyperparameter optimization will be performed using techniques like grid search or Bayesian optimization to enhance the model's accuracy.

### 2.4.2 Evaluation Metrics

To assess the performance of the model, the following metrics can be used:

**Accuracy**: The percentage of correctly classified mental health categories (e.g., Healthy, At Risk, Critical).

 (2-2)

**Mean Absolute Error (MAE)**: Measures the average difference between the predicted and actual mental health scores.

 (2-3)

where  is the actual score, and  is the predicted score.

**F1-Score**: Evaluates the balance between precision and recall, especially for the "Critical" category.

 (2-4)

**AUC-ROC**: Measures the model's ability to distinguish between categories (e.g., At Risk vs. Critical).

**Interpretability Metrics**: Use SHAP (SHapley Additive exPlanations)[7] or LIME (Local Interpretable Model-Agnostic Explanations)[8] to understand feature importance and model decisions.

### 2.4.3 Validation and Evaluation of the Model

The validation and evaluation of the machine learning model are essential to ensure its effectiveness, reliability, and ethical soundness in real-world applications. To evaluate the model in real-world scenarios, a simulated testing environment will be created using historical data to simulate a college setting. The model will predict students' mental health based on various scores and will be tested for its ability to accurately classify students into the "Healthy," "At Risk," and "Critical" categories. Feedback will be collected from mental health professionals to assess the model's interpretability and reliability. Any necessary refinements to scoring weights and thresholds will be made based on this input.

To ensure the model's reliability across different conditions, robustness testing will be performed. This involves evaluating the model's performance across various demographic groups to check for potential biases and testing the model's tolerance to noisy or incomplete data. Sensitivity analysis will also be conducted to determine how changes in key features, such as sleep hours and academic performance, affect the model's predictions. The goal is to ensure the model correctly prioritizes the most significant indicators of mental health.

The proposed machine learning model will also be compared to traditional mental health assessment methods, such as manual evaluations by counselors, to highlight its improvements in efficiency, scalability, and accuracy. The comparison will demonstrate that the machine learning model reduces the time and effort required for assessments, enables the monitoring of a larger student population, and provides higher precision in detecting early signs of mental health issues.

Before deploying the model, its ethical implications will be thoroughly evaluated. This will include an assessment of bias to ensure that the model does not disproportionately misclassify certain groups. The transparency of the model's predictions will also be evaluated, ensuring clear explanations are provided to both students and counselors. This will foster trust in the model's outputs.

## 2.5 Conclusion

The proposed data-driven machine-learning strategy provides a systematic and scalable approach to estimating and improving college students' mental health. By integrating comprehensive data collection, robust data processing techniques, and advanced machine learning models, the strategy enables precise predictions of mental health conditions and actionable insights for early intervention. The inclusion of rigorous experimental evaluation ensures the model's reliability, fairness, and applicability in real-world scenarios. Furthermore, ethical considerations such as transparency and bias mitigation enhance its acceptance and impact. This framework not only supports individual well-being but also empowers institutions to foster healthier, more supportive academic environments for students.

# References

1. Antony M M, Purdon C, Huta V, et al. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in a community sample of individuals with obsessive-compulsive disorder[J]. Behaviour Research and Therapy, 2019, 122: 103-110.
2. De Duro, E. S., Taietta, E., Improta, R., & Stella, M. PhDGPT: Introducing a psychometric and linguistic dataset about how large language models perceive graduate students and professors in psychology, 11, 2024, arXiv: <https://arxiv.org/abs/2411.10473>
3. Kroenke K, Spitzer R L, Williams J B. The Patient Health Questionnaire-9 (PHQ-9): Validity of a brief depression severity measure[J]. Journal of General Internal Medicine, 2020, 35(2): 365-372.
4. Al-Fraihat, D., Joy, M., & Sinclair, J. The impact of e-learning on students’ performance: A review of the literature. Computers in Human Behavior, 2020, 102: 212-226.
5. Nunes, M. A., & Silva, A. F. A study on text classification using TF-IDF and deep learning models. Proceedings of the International Conference on Machine Learning and Data Mining, 2019, 325-337.
6. Mikolov, T., Le, Q. V., & Sutskever, I. Exploiting similarities among languages for machine translation. Proceedings of the International Conference on Learning Representations (ICLR), 2018.
7. Lundberg, S. M., & Lee, S. I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS), 2017, 4765-4774.
8. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2016, 1135-1144.