

Project Initialization and Planning Phase

| | |
|---------------|---|
| Date | 15 April 2024 |
| Team ID | 738214 |
| Project Title | Predicting Mental Health Illness Of Working Professionals Using Machine Learning. |
| Maximum Marks | 3 Marks |

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

| Project Overview | |
|-------------------|---|
| Objective | <ul style="list-style-type: none"> Develop a Machine Learning system to predict the likelihood of working professionals needing mental health services. Encourage early intervention by prompting individuals to seek help and access available resources. |
| Scope | <ul style="list-style-type: none"> The project focuses on building a user-friendly system that analyzes user input and provides preliminary insights into potential mental health needs. It will not provide diagnoses but will recommend seeking professional help based on the analysis. The initial scope will target working professionals and utilize anonymize survey data for training the model. |
| Problem Statement | |
| Description | <p>This project aims to leverage Machine Learning to create a screening tool for mental health concerns in the workplace. The system will gather user input through surveys or questionnaires and analyze it using various algorithms (Logistic Regression, KNN, Decision Tree, etc.). Based on the analysis, the system will generate a prediction about the user's potential need for mental health services. The project will include:</p> |

| | |
|--------------------------|--|
| | <ul style="list-style-type: none"> • Data collection and randomization of relevant mental health surveys from working professionals. • Training and evaluation of Machine Learning models to identify the most accurate predictor. • Development of a user interface for interaction with the system. • Deployment of the model in a secure and user-friendly manner. |
| Impact | <ul style="list-style-type: none"> • Reduce stigma surrounding mental health by encouraging open conversations and early intervention. • Increase access to mental health resources for working professionals. • Improve overall well-being and productivity in the workplace. • Help organizations create a more supportive and understanding environment for employees. |
| Proposed Solution | |
| Approach | <ul style="list-style-type: none"> • Data collection and pre-processing <ul style="list-style-type: none"> • Collect the dataset or create the dataset: This could involve gathering data from existing mental health surveys or creating a new survey specifically tailored for the project. The data should be relevant to mental health symptoms, demographics, and work-life balance of working professionals. • Data pre-processing: This involves cleaning and preparing the data for modeling. Steps may include: <ul style="list-style-type: none"> ○ Removing unnecessary columns: Identify and remove columns that are irrelevant to the prediction task. ○ Checking for null values: Identify and address missing data points. This could involve imputing missing values with mean/median values or removing rows with excessive missing data. ○ Visualizing and analyzing data: Explore the data using various data visualization techniques (histograms, scatter plots, boxplots) to understand data distribution, identify potential outliers, and relationships between variables. Perform: <ul style="list-style-type: none"> ▪ Univariate analysis: Analyze individual variables to understand their distribution (mean, median, standard deviation) and identify any anomalies. ▪ Bivariate analysis: Analyze relationships between pairs of variables to see if there are any correlations between them. ▪ Descriptive analysis: Summarize the data using statistical measures like mean, median, standard deviation, and frequency tables. |

| | |
|--------------|---|
| | <ul style="list-style-type: none"> • Model building <ul style="list-style-type: none"> • Handling categorical variables: Encode categorical variables (e.g., gender, occupation) into numerical values for the model to understand them. • Dividing data into train and test sets: Split the pre-processed data into training and testing sets. The training set will be used to train the model, and the testing set will be used to evaluate its performance on unseen data. • Import the model building libraries: Import necessary libraries like scikit-learn for building and evaluating machine learning models. • Comparing the accuracy of various models: Train and evaluate different machine learning models (Logistic Regression, KNN, Decision Tree, Random Forest, etc.) on the training data. Choose the model that performs best based on metrics like accuracy, precision, recall, and F1-score. • Hyperparameter tuning of the selected model: Fine-tune the hyperparameters of the chosen model to improve its performance. Hyperparameters are settings that control the learning process of the model. • Evaluating the performance of models: Evaluate the performance of the final model on the testing set using the chosen metrics. • Save the model: Save the trained model for future use. • Application Building <ul style="list-style-type: none"> • Create an HTML file: Design a user interface using HTML to collect user input for the prediction. • Build Python code: Develop Python code to interact with the saved model, process user input, generate predictions, and display the results on the user interface. |
| Key Features | <ul style="list-style-type: none"> • User-friendly interface for anonymous input of relevant information. • Machine Learning model trained to predict potential mental health needs. • Clear and actionable recommendations for users who might benefit from professional help. • Integration with resources and support services offered by employers or mental health organizations. |

Resource Requirements

| Resource Type | Description | Specification/Allocation |
|-------------------------|--|---|
| Hardware | | |
| Computing Resources | CPU/GPU specifications, number of cores | Ryzen 7 CPU, 16GB RAM |
| Memory | RAM specifications Random Access Memory for application & data processing | 16GB RAM |
| Storage | Disk space for data, models, and logs | 250GB SSD + 1TB HDD |
| Software | | |
| Frameworks | Python frameworks | Flask |
| Libraries | Additional libraries | scikit-learn, pandas, NumPy, Flask, Matplotlib/Seaborn, Git |
| Development Environment | IDE, version control | Googe Colab HTML, CSS, Git |
| Data | | |
| Data | Source, size, format | https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey/data Survey.csv (303.68 kB) |