Phase 3: Report the Review on the Related Work.

Team Name: Deep Diver's

Team Members: Vaishnavi Hemant Salaskar vsala2@unh.newhaven.edu

Vedant Chidgopkar vchid2@unh.newhaven.edu

Riddhi Joshi rjosh5@unh.newhaven.edu

Today mental health is becoming a more common problem. However, evaluation of mental well-being is extremely important to understanding and providing therapeutic solutions. Diagnostics are complicated tasks and misdiagnosis can result in serious problems if a mental disorder is not properly detected. Can we recognize mental health issues accurately by using data mining techniques?

The data has been collected from Kaggle by Open Sourcing Mental Illness, LTD. Survey data about mental health attitudes are included in this dataset. Which then has been analyzed and pre-processed. The data contains different labels such as age, gender, country, self-employee, family history, work interference, seek help, etc. For better prediction, we have label encoded the data.

List of Related Review

1. Design of Data Mining and Evaluation System for College Student's Mental Health.

It was published in 2021 by Zhang Xilin, Yi Honglian from Dalian University of Science and Technology, Dalian, China. The name of the publication is International Conference on Measuring Technology and Mechatronics Automation (ICMTMA).

2. Classification Algorithms based on Mental Health Prediction using Data Mining.

It was published in 2020 by Vidit Laijawala, Aadesh Aachaliya, Hardik Jatta, Vijaya Pinjarkar from K J Somaiya Institute of Engineering and Information Technology, University of Mumbai, India. Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020).

3. Machine Learning Techniques for Stress Prediction in Working Employees.

It was published in 2018 by U Shrinivasulu Reddy, Aditya Vivek Thota, A Dharun from the National Institute of Technology, Trichy. The name of the publication is International Conference on Computational Intelligence and Computing Research.

4. Predicting Depression Levels Using Social Media Posts.

It was published in 2017 by Maryam Mohammed Aldarwish, Hafiz Farooq Ahmed from King Saud University for Health Science, Kingdom of Saudi Arabia. The name of the publication is International Symposium on Autonomous Decentralized Systems.

5. Predictive Analysis for Healthcare Domain using Classification Techniques.

It was published in 2017 by Shweta Sharma, Sahil Anand, Anant Kumar Jaiswal from Amity School of Engineering, Uttar Pradesh, India. The name of the publication is International Journal of Linguistics and Computing Research.

Classification Algorithms based on Mental Health Prediction using Data Mining.

Mental health is becoming a more common problem. Here the target population was working individuals that is people above the age of 18. The authors have gathered the dataset from an existing available dataset which was provided by an OSMI (Open Sourcing Mental Illness) survey. The dataset included the data of working individuals. This dataset consists of 26 attributes for prediction and 1 predicting label. Since the dataset included data from the survey, not all the parameters were useful. It consists of Age, Gender, self_emp, family_history, work_inference, past, diagnosis, treat, etc. The labels were encoded. Most data included 2 or 3 attributes (yes, no, maybe), however, some included 5 attributes. The number of sample records used was 258.

They implemented classification algorithms such as Decision Tree, Random Forest, and Naïve Bayes. Decision Tree and Random Forest algorithms were implemented to check the accuracy. They found that the Decision tree is the most optimal algorithm by 82%. It shows the Confusing Matrix providing the accuracy of 149 instances to be correctly classified as a positive while 109 to be correctly classified as a negative. There are 258 correct classifications out of 315 instances. According to researchers, individuals with a stressful or depressing work-life should seek mental health help. However, individuals whose work life is unaffected does not suffer from mental issues.

Machine Learning Techniques for Stress Prediction in Working Employees

They used the OSMI Mental Health in Tech 2017 survey [2] as the dataset to train various machine learning models in order to analyze the patterns of stress and mental health disorders among tech professionals and determine the most influential factors that contribute to these conditions. The results from the OSMI 2017 dataset were used to train the following machine learning models like b. KNN Classifier: K-Nearest Neighbour (KNN), Decision Trees, Logistic Regression, Decision Trees, Random Forest Classifier, Boosting and Bagging. The model was trained using a variety of machine learning algorithms, with boosting outperforming the others in terms of precision, accuracy, and false-positive rate. However, the random forest classifier scored higher in terms of cross-validated AUC, showing that this model is more stable. It is also worth noting that while both the logistic regression and the random forest classifier achieved the same classification accuracy, the latter surpasses the former in other parameters. Because the KNN classifier has the largest false-positive rate, it is highly untrustworthy in the current circumstance. Additional methods, such as the Naive Bayes classifier, can be used to evaluate the model's effectiveness.

According to their findings People, who worked in a tech company, even if their position was not tech-related, were marginally more likely to suffer stress. Boosting and random forest was also the most accurate and precise classification methods. With a 75.13 percent accuracy, Machine Learning approaches for stress and mental health condition prediction produce remarkable results, which can be further investigated.

Design of Data Mining and Evaluation System for College Student's Mental Health

In this paper, the author has proposed a data mining system for the mental health of college students using an apriori algorithm. The dataset collected for this article is from the student mental health assessment system. Data on mental health is analyzed for missing values. The objective is to find Symptoms of anxiety and depression in college students and their relationship to mental health factors. For this, they have used the apriori algorithm. The correlation and concurrence between data are found by association rules mining and decision rule mining. Moreover, modification functions are available in the system, which allows for the modification of inaccurate data in order to make it completer and more accurate. The highest quantitative performance outcome is 42%.

Predicting Depression Levels Using Social Media Posts.

Social Network Sites (SNS) is an online platform where individuals express their interests, feelings, etc. Researchers have proved that using user-generated context is a correct way to help determine Individual mental health levels. The aim was to investigate how SNS user posts can help classify users according to mental health levels [4]. The UGC was classified using two different classifiers: Support Vector Machine (SVM) and Naïve Bayes. The dataset here was collected from Facebook, LiveJournal, and Twitter. They labeled words that reflected whether the user is depressed or not in the training phase. Later, the Support Vector Machine (SVM) algorithm is applied to assign a text to one of the classes [4]. The number of posts in the training dataset was 6773 posts. Where 2073 posts were depressing posts and 4700 posts were not depressed. Finally, the author classifies the individuals into four levels Minimal, Mild, Moderate, and Severe depression.

Predictive Analysis for Healthcare Domain using Classification Techniques

This paper deals with predictive analysis for the healthcare domain using classification techniques. They have used the mental healthcare dataset which was a survey held in 2014 on the mental health of the IT industry. It has 1259 records in the dataset mixer of various questions related to mental health and some personal questions. The goal behind this research was to make awareness and boost the condition of those who are suffering from mental health issues. This research paper includes Naïve Bayes, J48, Neural network these classification techniques. Moreover, they have used Weka as a data prediction and classification tool. Also, the paper has explained various terminology including supervised and unsupervised learning, training and testing of data, the percentage split, use of training set, mean absolute error, precision, recall, confusion matrix, and performance matrix. In Naïve Bayes test model are used by splitting the data for 85% for the training set and 15% for the testing set. They have mentioned the steps for all the techniques. The performance matrix for Naïve Bayes, J48, Neural network is 88.35%, 89.28%, and 98.96% respectively. The mean absolute error for Naïve Bayes is 0.132 which is less than 0.179 of J48 and the same for the neural networks is 0.013.