Journal of Theoretical and Applied Information Technology
30th April 2022. Vol.100. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MENTAL HEALTH TREATMENT PREDICTION FOR TECH EMPLOYEE WITH THE IMPLEMENTATION OF ENSEMBLE METHODS

MICHAEL HENRY[1], SANI MUHAMAD ISA[2]

[1, 2] Computer Science Department, BINUS Graduate Program – Master of Computer Science

Bina Nusantara University, Jakarta, Indonesia 11480

E-mail: [1*]michael.henry@binus.ac.id, [2] sani.m.isa@binus.ac.id
*Corresponding Author

ABSTRACT

Individual has two awareness, one of them is mental health. It is the ability of a person to recognize the pressure in work and interaction within social life. Statistics shown an enormous amount of 2.570 billion people in the world experienced mental health disorders including depression, anxiety, alcohol abuse, etc. By considering this fact, mental health is one of the most important life essentials which needs to be taken care of. This study proposes an implementation of ensemble methods such as Bagging, Light Gradient Boosting Machine (GBM) and Stacking with Binary Particle Swarm Optimization (BPSO) as a feature selection and Open Sourcing Mental Illness (OSMI) as the dataset for the purpose of predicting whether IT employees need a mental health treatment or not. This study shows that ensemble methods do not always give better prediction with Naïve Bayes and BPSO which has 88.44% accuracy and Stacking with Naïve Bayes as the meta classifier has 87.86% accuracy, a 0.58% accuracy difference. Based on the performance, this study also shows the best features for predicting mental health treatment needs where IT Employees who have the problems or the best features, need to consult a treatment.

Keywords: *Ensemble Method, Machine Learning, Particle Swarm Optimization, Mental Health, Feature Importance*

1. INTRODUCTION

Mental health itself is the awareness of an individual in recognizing the ability to fight against various pressures in life and work and also to interact in social life [1]. Mental health has 3 components, such as emotional conditions (feelings of happiness, feelings of satisfaction, etc.), psychological conditions (relations skills, feelings of responsibility, etc.), and social conditions (feelings when in groups, social contributions, etc.). These three components have differences and variations according to the culture of each country and each local state, which triggers the emergence of various problems in human mental health and the impact that can be caused by mental health itself [2].

According to statistics, 14% of children have mental disorders and this is huge considering that 50% of the population worldwide are children. Moreover, 70% of people in the world have difficulties in receiving mental treatment due to the limitations of mental illness healing institutions, and their respective procedures [3].

Figure 1: Mental health disorders summary

Based on Figure 1, there were approximately 2.570 billion people in the world in 2017, who experienced mental health disorders, 264 million of them experienced depression, 284 million of them experienced anxiety disorders, 107 million of them experienced alcohol use disorders, and many more [4].

The number of cases of mental health disorders is specifically felt by technology workers or IT employees in technology companies. According to LyraHealth, 51% of tech workers diagnosed with mental health disorders, and 71% of tech workers feel a mental health disorder of their productivity at work [5]. Reported from Oracle written by James Drake, 70% of technology workers experience stress and anxiety, especially during the COVID-19 pandemic that appeared in 2020. It also leads to depression, declining physical health, and decreased socialization between family and close friends [6]. In accordance with this fact, there needs to be some help with mental health care that indicates whether or not care is needed to be provided to these tech workers.

This massive effect on mental health encourages this research in predicting mental health treatment needs for IT employees, using machine learning techniques. What makes the use of Machine Learning algorithms interesting is that the implementations of Knowledge Discovery and determining Patterns in data can be used to find information needed in various fields of life, one of which is mental health [7].

The implementation of Knowledge Discovery which is used to create various Machine Learning computational models, especially prediction models, and identification models, proves very broad uses and benefits [8]. One related study states that the use of Machine Learning is useful for the process of decision making in predicting the results of certain data, values and variables [9]. From this statement, it can be concluded that data mining algorithms can be used for this topic. With various studies in Health using Data Mining, it is certain that the topic is useful to help people identify their mental health whether they take a treatment or not.

In this research, the process of making a mental health prediction model will be carried out using BPSO, a metaheuristic algorithm, as a feature selection method, implementing each standalone classification model, and implementing the models into ensemble models as the proposed solution. Some of the classification techniques that will be used in this research are Decision Tree, Logistic Regression, and Naïve Bayes as well as the datasets from the OSMI Survey website and Bagging, Light GBM, and Stacking as the ensemble models. Lastly, the ensemble method itself is proven to provide better predictive results by combining several models in machine learning into a framework [10]. With this statement, this study will solve research problems using the ensemble method and make contributions to previous research.

The following sections of this paper will give an in-depth explanation of the research starting from the review of several related works on machine learning and mental health in section 2, followed by the description of each proposed method from data to modeling in section 3. The result of those

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proposed methods will be explained in section 4 and finally be concluded in section 5.

2. LITERATURE REVIEW

2.1 Bagging

Bagging method is a method that generates new variants of one basic classifier by making bootstrap copies of the training data which are then used to generate aggregated predictors [11].

This method as shown on Figure 1, as the name implies, Bagging, designs a subset (bags) using the same dataset by a technique called bootstrap-sampling. The "bags" that are designed in such a way have the task of being a bridge for this method to obtain a beneficial idea from the process of dividing the subset [12]. The main purpose of the Bagging method is to minimize diversity and minimize the occurrence of overfitting of the various models made.

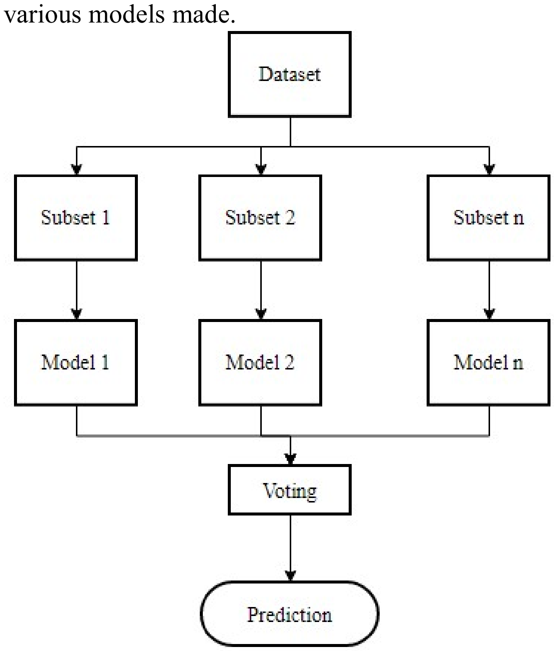

Figure 1: Bagging process

2.2 Light Gradient Boosting Machine (LightGBM)

LightGBM is a histogram-based algorithm which is one of the methods of the Gradient Boosted Tree (GBT) algorithm. The leaf node determination process from lightGBM uses the leaf-wise principle based on Figure 2 where the leaf node with the highest best fit is used as the root node for the new leaf node [13]. On the leaf-wise principle, the use of best fit is able to match other GBT principles such as level-wise and depthwise in terms of accuracy [14].

LightGBM also has better and more efficient training capabilities, does not take up too much memory, and has the ability to handle big data.

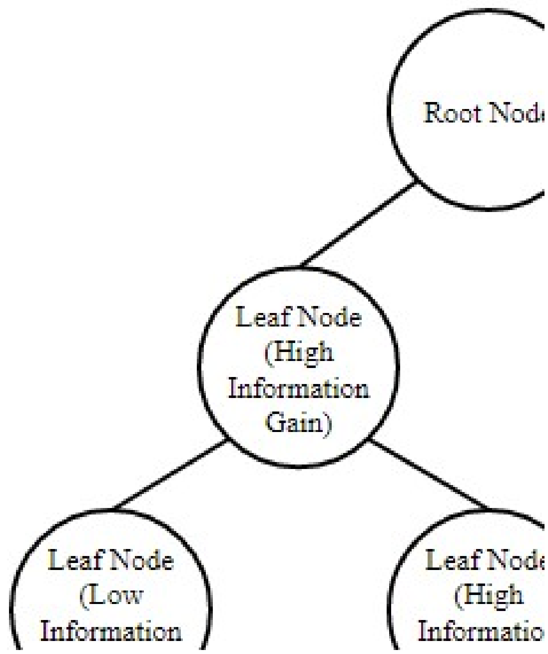

Figure 2: Leaf-Wise principal

2.3 Stacking

In the stacking method based on Figure 3, there are two stages, namely predictions with different basic classifiers, then predictions with meta classifiers. In the first stage, the prediction results using the same training data from different basic classifiers are combined in the second stage, namely prediction using a meta classifier as the final stage of the prediction process [11].

The stacking method is also called a meta classifier where there is a layer containing meta-level learners as the final predictor which input is obtained from the output of the combined results of several models (aggregated models).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

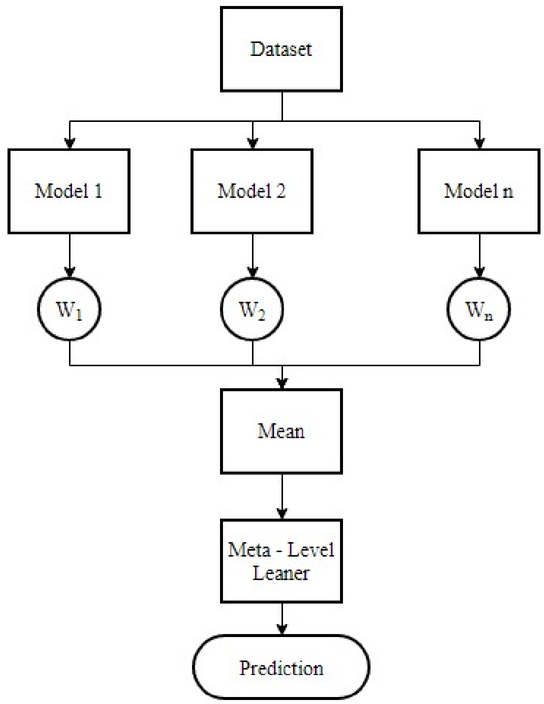

Figure 3: Stacking process

2.4 Related works

Research conducted using machine learning on mental health has its trends where the trends are calculated based on mental disorders and the popularity of machine learning used such as Alzheimer's disease, with the popularity of data mining techniques used by each researcher is 33.3% using Decision Tree, 27.7% using SVM, and 22.2% using Naïve Bayes and so on. Then for Dementia is 62.5% using Decision Tree and ANN, 37.5% using SVM and Logistic Regression, and so on. For Depression it is 45.5% SVM, 36.3% Naïve Bayes, 27.7% Decision Tree, Random Forest and Logistic Regression, and so on. For Schizophrenia disease is 57.1% Decision Tree, 42.8% Random Forest, 28.5% k-NN, and so on [15].

As of this trend, more studies are conducted using different datasets. To develop their model, the datasets are mostly collected in the form of questionnaires with various questions and target scopes. One of the studies used their own university dataset [16] which consists of the mental data of each freshman. Other than that, a study used hospital dataset [17], the National Institute of Mental Health dataset [18], OSMI (Open Sourcing Mental Illness) survey dataset [19], and some even used their own survey questionnaire which they did not mention. With the variety of the dataset, the purpose of machine learning techniques arises to solve mental problems. Several studies used Decision Tree [20] as their model with 92.7% accuracy, Naive Bayes [21] with 76.7% accuracy, KNN [22] with 93% accuracy, Logistic Regression [23] with 81.6% accuracy, and SVM [24] with 78.7% accuracy.

As promising as it seems, most studies still used basic machine learning methods for their research. Therefore, in predicting mental health as well as a contribution to the above studies, modeling with the Ensemble Methods to find better result and performance will be used as the models in this study and find which features are important in mental health prediction based on the dataset.

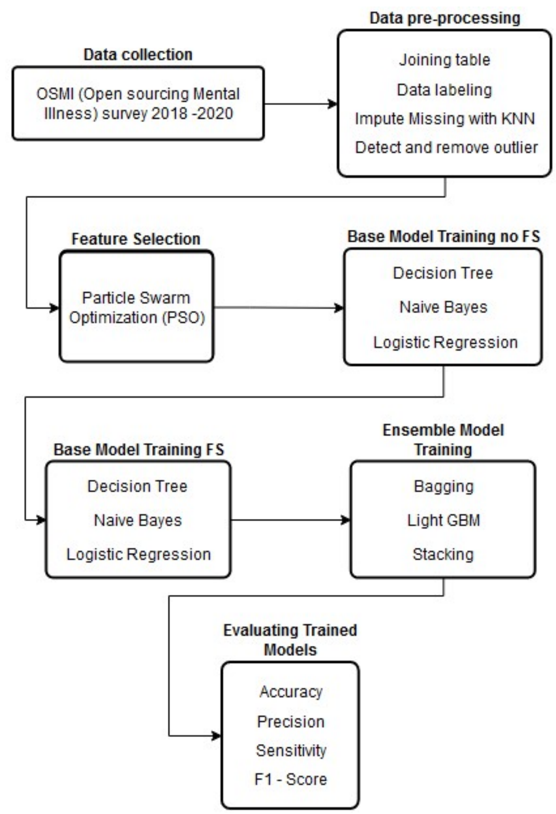

Figure 4: Proposed Method

3. PROPOSED METHOD

The process of experimentation starts with data collection, data pre-processing, feature selection, single model training, ensemble model training, and evaluation. Figure 4 shows the complete structure of this experiment.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 Data collection

The data collection process is very crucial in designing a model, especially in making mental health prediction models in this study. The dataset was obtained from the OSMI (Open-Sourcing Mental Illness) website contains a questionnaire about the life experience of workers in technology-based companies along with questions about their mental health history and consultations. In this study, questions related to mental health became the main focus in making prediction models. This dataset has a different number of attributes. which contains data such as gender, age, job saturation, heredity, history of mental illness, consultation history, etc.

Table 1: A sample of OSMI dataset attributes

Questions	Description
Are you self-employed?	Yes, No
Are you openly identified at work as a person with a mental health issue?	1, 0
Do you have a family history of mental illness?	Yes, No, I don't know
Have you had a mental health disorder in the past?	Yes, No, don't know
Have you ever sought treatment for a mental health disorder from a mental health professional?	1, 0

For 2018 there are 123 columns, 417 records, then in 2019 there are 80 columns, 352 records, and in 2020 there are 120 columns, 180 records. Table 1 describes some of the attributes in the dataset that will be used in the prediction process along with their descriptions.

3.2 Data pre-processing

The collected data then enters the pre-processing stage to produce appropriate data and to improve models' performance during the process to achieve the evaluation. These steps consist of table joining, attribute renaming, attributes missing value handling, label encoding, and Clustering base Missing value Imputation using K-Nearest Neighbor (KNN).

Based on Table 2, Due to the imbalance in the number of columns, there will be an inner join between the 3 tables by looking at the same attributes to get a total of 49 attribute columns and 1 prediction label, which is Treatment, and a total of 861 records out of a total of 949 records.

Table 2: An example of joined table

No.	2018 Attributes	2019 Attributes	2020 Attributes
1.	Have your previous employers provided mental health benefits?	Have your previous employers provided mental health benefits?	Have your previous employers provided mental health benefits?
2.	Do you *currently* have a mental health disorder?	Do you *currently* have a mental health disorder?	Do you *currently* have a mental health disorder?
3.	Describe the circumstances of the supportive or well-handled response.	Has being identified as a person with a mental health issue affected your career?	Are you openly identified at work as a person with a mental health issue?

After joining each table, every attribute will be converted to simplified version manually which can be seen on Table 3. At this stage, all attribute names will be simplified by taking the essence of each attribute without omitting the meaning of the attribute itself. This is done because the data from the questionnaire is in the form of questions and is not simple. This stage will be done manually.

Table 3: Before and after attribute renaming

No.	Before renaming	After renaming
1.	Are you **self-employed**?	**Self employed**
2.	What is your **gender**?	**Gender**
3.	What is your **age**?	**Age**
4.	Do you have a **family history** of mental illness?	**Family history**

The next step of data pre-processing is to see and remove the attributes with missing value rates of more than 85% of the total data which is more than 143 missing values. Based on Table 4, attributes with high missing value rates, will then be removed to ensure good classification performance.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4: A sample of attributes with high missing values

No.	Attributes	Missing values
1.	Describe the circumstances of the supportive or well-handled response.	949
2.	How has it affected your career?	900
3.	If yes, what percentage of your work time (time performing primary or secondary job functions) is affected by a mental health issue?	851
4.	Has being identified as a person with a mental health issue affected your career?	828
5.	Do you have medical coverage (private insurance or state-provided) that includes treatment of mental health disorders?	820

After removing the attributes, all values of category attributes will be label encoded from categorical to numeric, such as False to 0 and True to 1, and so on. The example of this step can be seen on Table 5.

Table 5: A sample of attributes' label encoding

No.	Attributes	Old values	New values
1.	Work benefits	Yes	1
		I don't know	2
		No	3
		Not eligible	4
2.	Current wellness campaign	Yes	1
		No	2
		I don't know	3
3.	Company convenience on mental medical leave	Somewhat easy	1
		Very easy	2
		I don't know	3
		Somewhat difficult	4
		Neither easy nor difficult	5
		Difficult	6
4.	Past diagnosed	Yes	1
		No	2
		Possibly	3
		Don't Know	4
5.	Care options	Yes	1
		No	2
		Not eligible	3

Furthermore, each attribute which have 15% of missing value will be imputed using KNN with 10 nearest neighbors. After pre-processing is complete, the dataset is then exported to CSV for later use in the machine learning process.

3.3 Feature Selection

The feature selection process in this study is used to provide more optimal prediction results. This dataset uses a metaheuristic algorithm, namely BPSO (Binary Particle Swarm Optimization). The process of determining the best features in BPSO involves the numbers 0 and 1, where 1 signifies the selected feature, and 0 signifies the unelected feature. The process can be first determined by a fitness function through a special objective function, which is as follows.

$$f(x) = \alpha(1 - C) + (1 - \alpha)\left(1 - \frac{n_f}{n_t}\right) \quad (1)$$

According to (1), The results of the classification of the classifier for BPSO, in this study using Naive Bayes, will be stored in the C variable. Then the total features are stored in the nt variable, and the feature subset will be stored in the nf variable. After that, the variable α [0,1] is a hyperparameter that functions as a "trade-off" between the performance of the classifier, and the percentage of the total final feature used. The trade-off here means that to get the best fitness value, there must be a decrease in quality/quantity between performance and the percentage of total features.

Table 6: Hyperparameter description of BPSO

No.	Nama parameter	Variabel
1.	*Hyperparameter*	alpha
2.	Cognitive Parameter	c1
3.	Social parameter	c2
4.	Weight	w
5.	Neighborhood size	k
6.	Minkowski distance	p
7.	Total particles (swarm size)	n_particles
8.	Total iterations	iters
9.	Total features	dimensions

This fitness value along other parameters from Table 6 such as n_particles, n_iterations, k, c1, c2 and weight will be as such: the alpha (α) will be 0.9, the n_particles will be 100, the n_iterations will be 500, the n_neighbors will be 10, the c1 = c2 = 0.2, the p will be 1 and the weight will be 0.9. all of this parameter is determined as the best value for this study. Lastly, the result of BPSO will determine the best features, which equals to 1, and the worst features, which equals to 0, according to

Journal of Theoretical and Applied Information Technology
30th April 2022. Vol.100. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

summary of the best cost of the iterations which is 0.172. This feature selection process resulted in 33 selected features from the total of 49 features.

3.4 Data splitting

In this study, the data splitting process will be divided into 2, which are the data splitting process for the first training models where there is no Feature Selection, and the data splitting process for the second and third training models where Feature Selection is done. The data splitting will use the same ratio formulation which is 80% for training data, and 20% for testing data. That way, the total data used for training is 688 data, and for testing 173 data with 49 total features for the first training models, and 33 total features for the second and third training models.

3.5 Training and evaluation

The training process of this research consists of 3 different models as mentioned before.

The first training models will be conducted on each base classifiers using the non-feature selected dataset.

Then the second training models will be conducted on each base classifiers using the feature selected dataset.

And lastly, the third training models will implement the ensemble models using the second training model. As a note, each base model used in Stacking will be used as base classifiers and a meta classifier (e.g stacking 1 base classifiers: decision tree, naïve bayes, and logistic regression, meta classifiers: decision tree, and so on).

Finally, the performance metrics of each training method from their Accuracy, Precision, Sensitivity, and F1-score, followed by the effect of feature selection on each base model, the effect of the ensemble model and the feature importance will be evaluated accordingly.

Table 7: First training models

Methods	*Accuracy*	*Precision*	*Sensitivity*	*F_1 Score*
Decision Tree (DT no Feature Selection)	73.99%	68.57%	67.61%	68.09%
Logistic Regression (LR no Feature Selection)	80.92%	**77.94%**	74.65%	76.26%
Naive Bayes (NB no Feature Selection)	**82.08%**	77.03%	**80.28%**	**78.62%**

Table 8: Second training models

Metode	*Accuracy*	*Precision*	*Sensitivity*	*F_1 Score*
Decision Tree (DT Feature Selection)	80.35%	77.92%	77.92%	77.92%
Logistic Regression (LR Feature Selection)	83.82%	79.52%	85.71%	82.50%
Naive Bayes (NB Feature Selection)	**88.44%**	**86.08%**	**88.31%**	**87.18%**

4. RESULTS AND DISCUSSION

After getting all the results of the training processes, the last stage is to conduct all the discussion of the results that have been obtained through the implementation of the proposed model, namely Ensemble Bagging, Boosting and Stacking in the third training. These results will be compared to all the first trainings namely Decision Tree, Logistic Regression, and Naive Bayes without Feature Selection, and all the second training namely Decision Tree, Logistic Regression, and Naive Bayes with Feature Selection. The results of the model consist of Accuracy, Precision,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Sensitivity, and F1-Score along with features that most influence the classification process with Feature Importance.

Based on Table 7, shown the performance of each Base Classifier without Feature Selection from Decision Tree, Logistic Regression, to Naive Bayes. The highest accuracy performance was obtained by Naive Bayes without Feature Selection of 82.02% and for the lowest accuracy performance obtained by Decision Tree without Feature-Selection of 73.99%. Then for the highest precision performance obtained by Logistic Regression without Feature Selection of 77.94% and for the lowest precision performance obtained by Decision Tree without Feature Selection of 68.57%. After that, the highest sensitivity performance was obtained by Naive Bayes without Feature Selection of 80.28% and for the lowest sensitivity performance was also obtained by Decision Tree without Feature Selection of 67.61%. Finally, the highest F1-Score score was obtained by Naive Bayes without Feature Selection also at 78.62% and for F1-Score the lowest score was obtained by Decision Tree without Feature Selection of 68.09%. Overall, in this first training, the best performance was achieved by Naive Bayes without Feature Selection, followed by Logistic Regression without Feature Selection and Decision Tree without Feature Selection.

Next, Table 8 shown the performance results of each Base Classifier with Feature Selection from Decision Tree, Logistic Regression, to Naive Bayes. The highest accuracy performance was obtained by Naive Bayes with a Feature Selection of 88.44%. Then for the highest precision performance obtained by Naive Bayes with Feature Selection of 86.08%. In addition, the highest sensitivity performance was also obtained by Naive Bayes with a Feature Selection of 88.31%. Finally, the highest F1-Score score was obtained by Naive Bayes with a Feature Selection of 87.18%. Overall, in this second training the best performance was achieved by Naive Bayes with Feature Selection, followed by Logistic Regression with Feature Selection and Decision Tree with Feature Selection.

Table 9: Third training models

Ensemble Method				
Metode	*Accuracy*	*Precision*	*Sensitivity*	*F_1 Score*
Decision Tree Bagging	84.97%	84.00%	81.82%	82.89%
Logistic Regression Bagging	83.82%	79.52%	85.71%	82.50%
Naïve Bayes Bagging	**87.86%**	**85.90%**	87.01%	86.45%
LightGBM	83.82%	81.01%	83.12%	82.05%
Stacking Decision Tree as Meta class	78.03%	76.00%	74.03%	75.00%
Stacking Logistic Regression as Meta class	87.28%	85.71%	85.71%	85.71%
Stacking Naïve Bayes as Meta class	**87.86%**	85.00%	**88.31%**	**86.62%**

Lastly, Table 9 shown the performance results of each Ensemble Method from Decision Tree Bagging (DT Bag), Logistic Regression Bagging (LR Bag), Naive Bayes Bagging (NB Bag), Light Gradient Boosting Machine (LightGBM), Stacking with Decision Tree as meta classifier (Stack DT), Stacking with Logistic Regression as meta classifier (Stack LR), Stacking with Naive Bayes as meta classifier (Stack NB). For the highest accuracy performance obtained by NB Bag and Stack NB of 87.86%. Then, for the highest precision performance obtained by NB Bag of 85.90%. After that, the highest sensitivity performance was obtained by Stack NB at 88.31%. Lastly, the highest F1-Score value was obtained by Stack NB at 86.62%.

As per the results of the analysis, the use of Logistic Regression as a meta classifier for

Journal of Theoretical and Applied Information Technology
30th April 2022. Vol.100. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Ensemble Stacking is able to outperform the Base Classifier, namely Logistic Regression with Feature Selection in terms of overall performance. But in this case, the use of Naive Bayes and Decision Tree as meta classifiers for Ensemble Stacking was unable to outperform its Base Classifier in terms of overall performance. This is certainly not in accordance with ensemble stacking theory in sub-chapter 2.3 regarding Stacking, where Ensemble Stacking has a tendency to provide better performance than the Base Classifier used.

But in this case, only Ensemble Stacking with Logistic Regression as a meta classifier is able to provide better performance when compared to Logistic Regression Feature Selection. To find out the cause of the results, further learning was conducted from a study conducted by Hitoshi Hamori and Shigeyuki Hamori. In their research, there is a conclusion about the effect of Base Classifier performance on ensemble method performance, in this study Ensemble Stacking, where if the performance of each Base Classifier in a study does not have a comparable performance [25], for example 1 classifier is too high but the rest is too low, then Ensemble Method does not have a tendency to provide better predictive performance. As shown on Table 8, in terms of base classifier accuracy of Decision Tree, Logistic Regression and Naive Bayes, there is a significant difference between the three. Between Naive Bayes and Decision Tree has an accuracy difference of 8.09%, then Naive Bayes with Logistic Regression has an accuracy difference of 4.62%, and Logistic Regression with Decision Tree has an accuracy difference of 3.47%. Based on these results, study conducted by Hitoshi Hamori and Shigeyuki Hamori has further proven that the Ensemble Stacking model used is no better than each Base Classifier especially for this mental health treatment prediction case.

After calculating each model performance, this study also finds which features that considered important for the prediction and whether they have consulted a Treatment or not. This can be achieved with the help of feature importance calculation on each base model with feature selection. As we can see on Figure 5, “Bad feedback to mental interference in work”, which means less feedback about employee’s mental health in work, has the most prominent effect for prediction, followed by past diagnosis, current diagnosis, physical health importance from employer, and coworker thoughts of mental disorder. According to this calculation, people who have the above problems, need to consult a treatment to mental health professional to get a proper medication or suggestion in order to heal from their mental health problem. However, if people do not have the above problem, then they are not considered having a mental health problem and not required to consult for treatment.

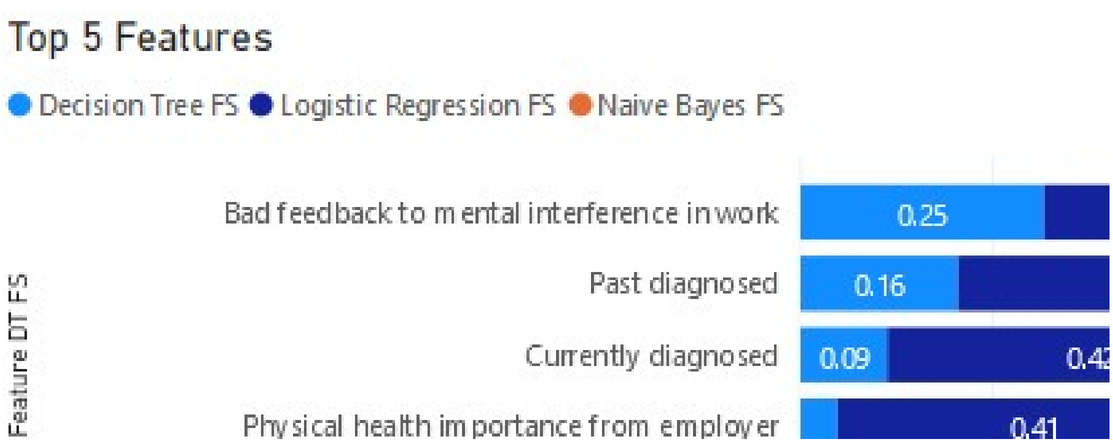

Figure 5: Top 5 features for mental health treatment prediction

5. CONCLUSIONS

This study of predicting mental health treatment uses the help of ensemble model to give an overall better prediction. Based on the first and second training models, the comparison proves that PSO as the feature selection algorithm provides significant effect on each base model. Moreover, the ensemble models shown good performance although ensemble models can give better performance, this study reveals that ensemble models do not always give better performance where Naïve bayes with feature selection gives better performance of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

88.44% accuracy if compared to Stacking with naïve bayes as meta classifier performance of 87.86% accuracy, a 0.58% accuracy difference. This can be concluded because of the PSO algorithm itself as a stochastic algorithm which has a random feature inside the calculation. And also, it can be concluded that ensemble methods are slightly less performing than single model in this study. Then, we can see that bagging works better for tree-based algorithm like decision tree with 4.62% accuracy improvement. Also, we can conclude that decision tree is not a good meta classifier for stacking compared to naïve bayes as the meta classifier.

6. FUTURE WORKS

The implementations of Ensemble Methods do have a tendency to provide better performance against Base Classifier. But in this study, Ensemble Method did not provide better performance to predict the mental health care needs of technology workers. Therefore, there are some suggestions for further research to develop a better model. Advice that can be given and received for future research is as follows, use larger datasets or other larger datasets. Then establish in detail the various hyperparameters with multiple hyperparameter tuning of each method used from the pre-processing stage of data to the end of classification. In addition, using the latest Evolutionary Algorithm-based Feature Selection method or more into the variations of the Particle Swarm Optimization method. Then, take the latest Ensemble Method approach, as well as other Machine Learning or deep learning approaches. Finally, make a smartphone application or website to make it easier for tech workers to get early information about mental health care.

REFERENCES:

[1] S. Chauhan and A. Garg, “Predictive research for mental health disease,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 9 Special Issue 2, pp. 196–199, 2019, doi: 10.35940/ijitee.I1038.0789S219.

[2] S. Galderisi, A. Heinz, M. Kastrup, J. Beezhold, and N. Sartorius, “Toward a new definition of mental health,” *World Psychiatry*, vol. 14, no. 2, pp. 231–233, 2015, doi: 10.1002/wps.20231.

[3] M. L. Wainberg *et al.*, “Challenges and Opportunities in Global Mental Health: a Research-to-Practice Perspective,” *Curr. Psychiatry Rep.*, vol. 19, no. 5, 2017, doi: 10.1007/s11920-017-0780-z.

[4] H. Ritchie and M. Roser, “Mental Health,” *Our World in Data*, 2018. https://ourworldindata.org/mental-health.

[5] Lyrahealth, “Mental Health Tech Industry,” *LyraHealth*, 2020. https://www.lyrahealth.com/mental-health-tech-industry/.

[6] J. Drake, “Global Study: 82% of People Believe Robots Can Support Their Mental Health Better Than Humans,” *Oracle*, 2020. https://www.oracle.com/uk/news/announcement/artificial-intelligence-supports-mental-health-2020-10-07/.

[7] I. H. Sarker, A. S. M. Kayes, and P. Watters, “Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage,” *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0219-y.

[8] S. Rahman, Shaikh, “An Application of Data Mining of Mental Health Data,” in *Data Mining and Statistics for Decision Making*, Mar. 2019, pp. 1–6, [Online]. Available: https://aisel.aisnet.org/sais2019/42%0AThis.

[9] S. Kaur and R. K. Bawa, “Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System,” *Int. J. Energy, Inf. Commun.*, vol. 6, no. 4, pp. 17–34, 2015, doi: 10.14257/ijeic.2015.6.4.02.

[10] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, “A survey on ensemble learning,” *Front. Comput. Sci.*, vol. 14, no. 2, pp. 241–258, 2020, doi: 10.1007/s11704-019-8208-z.

[11] S. Agarwal and C. R. Chowdary, “A-Stacking and A-Bagging: Adaptive versions of ensemble learning algorithms for spoof fingerprint detection,” *Expert Syst. Appl.*, vol. 146, p. 113160, 2020, doi: 10.1016/j.eswa.2019.113160.

[12] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A comprehensive evaluation of ensemble learning for stock-market prediction,” *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00299-5.

[13] M. R. Machado, S. Karray, and I. T. De Sousa, “LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry,” *14th Int. Conf. Comput. Sci. Educ. ICCSE 2019*, no. Iccse, pp. 1111–1116, 2019, doi: 10.1109/ICCSE.2019.8845529.

[14] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, “Diagnosis of Diabetes Mellitus Using Gradient Boosting (LightGBM),” *MDPI*,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pp. 1–14, 2021.

[15] S. A. Góngora *et al.*, “Data Mining Algorithms and Techniques in Mental Health: A Systematic Review,” *J. Med. Syst.*, pp. 1–15, 2018, [Online]. Available: https://doi.org/10.1007/s10916-018-1018-2.

[16] L. Zhou, “Evaluation research on data processing of mental health of college students based on decision tree algorithm,” *J. Comput. Methods Sci. Eng.*, vol. 19, no. 4, pp. 1101–1108, 2019, doi: 10.3233/JCM-193866.

[17] J. Shivangi and G. Mohit, “A Data Mining Analysis Over Psychiatric Database for Mental Health Classification,” *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, vol. 4, no. 1, p. 7, 2018, doi: 10.26634/jcom.5.1.13792.

[18] S. Khan, A. Islam, T. Zahangir, and A. Latiful Haque, “Knowledge Discovery from Mental Health Data,” *Proceeding Big Data IoT Comput. Networks, Int. Conf.*, pp. 881–888, Jan. 2020, doi: 10.1007/978-3-030-24643-3_103.

[19] V. Laijawala, A. Aachaliya, H. Jatta, and V. Pinjarkar, “Classification algorithms based mental health prediction using data mining,” *Proc. 5th Int. Conf. Commun. Electron. Syst. ICCES 2020*, no. Icces, pp. 1174–1178, 2020, doi: 10.1109/ICCES48766.2020.09137856.

[20] C. Gonçalves, D. Ferreira, C. Neto, A. Abelha, and J. Machado, “Prediction of Mental Illness Associated with Unemployment Using Data Mining,” *Procedia Comput. Sci.*, vol. 177, pp. 556–561, 2020, doi: 10.1016/j.procs.2020.10.078.

[21] M. P. Dooshima, “A Predictive Model for the Risk of Mental Illness in Nigeria Using Data Mining,” *Int. J. Immunol.*, vol. 6, no. 1, p. 5, 2018, doi: 10.11648/j.iji.20180601.12.

[22] M. Srividya, S. Mohanavalli, and N. Bhalaji, “Behavioral Modeling for Mental Health using Machine Learning Algorithms,” *J. Med. Syst.*, vol. 42, no. 5, 2018, doi: 10.1007/s10916-018-0934-5.

[23] N. A. Tajuddin, Z. A. Shah, and M. Jamaluddin, “Attribute Selection and Classification on Mental Illness DataSet,” *Acad. Fundam. Comput. Res.*, vol. 1, no. 2, pp. 1–9, 2020.

[24] S. I. Khan, A. Islam, A. Hossen, T. I. Zahangir, and A. S. M. L. Hoque, “Supporting the Treatment of Mental Diseases using Data Mining,” *2018 Int. Conf. Innov. Sci. Eng. Technol. ICISET 2018*, no. October, pp. 339–344, 2018, doi: 10.1109/ICISET.2018.8745591.

[25] H. Hamori and S. Hamori, “Does Ensemble Learning Always Lead to Better Forecasts?,” *Appl. Econ. Financ.*, vol. 7, no. 2, p. 51, 2020, doi: 10.11114/aef.v7i2.4716.