RESEARCH ARTICLE OPEN ACCESS

Mental Health Illness Disease Prediction Using Most Common Feature in Tech Survey Dataset

Sumitra Mallick[1], Mrutyunjaya Panda[2]
[1]*Department of Computer Science, KL University, Hyderabad, Telangana, India,*
[2]*Department of computer science and applications, Utkal University, odisha, India,*

ABSTRACT
Mental health illness disease are characterized by high stress, long hours, work pressure, making name for self and work imbalance between personal and professional life. The proposed work is to find the frequency of mental health illness of technical people compared to non technical people at the workplace. The dataset called mental health in tech survey that collected data from various people worldwide is chosen for this research. The dataset comprises of 27 features where the target class variable is tech company selected for the model building. The data preprocessing technique called missing values implemented to the dataset to get the cleaned data and all the feature values are filled. The classification algorithms of machine learning is applied to obtain the accuracy of mental health illness using best ten features of information gain and Chi square feature selection algorithm. The accuracy is also obtained by using most common features of both the algorithms and considering all the feature of mental health illness in tech survey dataset. The results are shown by developing an aggregated accuracy table in addition to the individual accuracy to predict the percentage of technical worker suffers from the mental health illness at a particular workplace. The accuracy of various classification algorithms are compared along with the best features, most common features and all the features of the tech survey dataset to conclude the suitability of an algorithm in this dataset.

Keywords: Accuracy, Chi square, Feature Selection, Information, Mental Health Illness.

Date of Submission: 18-07-2021 Date of Acceptance: 03-08-2021

I. INTRODUCTION

Mental health illness is very heavily classified in the society which is viewed as a deficiency in a person and most people are not comfortable to admit that they are dealing with mental illness. People in technical industries are characterized by high stress, long hours of working, work pressure and making name for themselves [1]. The aim of the study is to find the technical people are more affected from mental health illness in the tech survey dataset. This estimate was significantly higher than other attributes obtained in the tech survey dataset in the company. Especially in the tech company few of the attributes contributing more towards finding the mental illness.

In healthcare the right dataset is essential to obtain the right model and their performance. Therefore it is important for data analytics to have a good understanding in the dataset and also to extract the relevant information from the dataset[18]. In this paper, it is proposed to have an efficient statistical analysis technique for feature selection and classification algorithms of machine learning to predict the accuracy of the model.

Feature selection is defined as a subset of M features selected from N features sets in order to reduce the dimension of feature space [6][1]. Feature selection which include both feature extraction and feature selection. The main aim is to penalize those features relevancy which reduce the prediction output and improve the input relevant feature information. Reducing the number of features may improve prediction performance and always improves interpretability. Relief algorithm which is used for classification problem [2] is a family of filter-style feature selection to gain the effective flexibility on the classification and regression. This algorithm based on instance based learning that evaluates the filtering feature selection method and calculates a proxy statistics for each feature that can be used to estimate the quality or relevance with the target concept.

Feature importance is the technique used to select features using a trained supervised classifier. When we train a classifier such as a decision tree, we evaluate each attribute to create splits; we can use this measure as a feature selector [14]. There are two key issues for feature selection algorithm research: First, choose the appropriate evaluation criteria to measure the pros and cons of the feature subset; the second is to choose an efficient feature subset search method, so as to find the smallest feature combination that can best describe the target attribute [15]. Gaikwad *at el.* **[**10] Author at el. Propose a two-dimensional maximum margin criteria biomodal approach Where the facial features are obtained by modifying the maximum marginal criteria(MMC).

The paper is organized as follows: section 2 discusses the various literatures of feature selection and machine learning classification algorithms section 3 analyses the information gain and Chi square for feature selection; section 4 analyses the mental health illness in tech survey dataset; section 5 implements the feature selection algorithms and obtained the best features; section 6 discuses the results that are obtained from the mental health illness in tech survey dataset and provide the concluding remark and section 6 summarizes the work.

II. LITERATURE REVIEW

The machine learning techniques use Feature selection and data cleaning as pre-processing mechanism to improve the performance of the algorithm [3]. In the feature selection process only the most relevant features are considered to perform classification. There are feature selection algorithm i.e. filter and wrapper approach that find the feature subsets for classification. In the filtering approach, the algorithm evaluates the feature without involving the learning on the big dataset. In the wrapper approach, it measures the information gain of each feature which is more expensive because it needs more resources to compute. The classification concluded that wrapper approach achieve better result compared to filter. To determine the different combination of gesture and to optimized the size, gesture recognition technology has been carried out in gesture by Venugopal et al.[20]. Where it recognized 1005 accuracy for four gesture and the optimal size is 516-523. As the suicidal rate are highest among the population, the researchers are giving more focus on the population. Shambhavi used logistic regression method to predict the suicidal behavior among the employees [19]. Padideh danaee et al.[19] detect cancer on the RNA genome sequence dataset. (TCGA) The Cancer Genome Atlas database this database which is composed of both tumor and healthy breast sample of patients. The deep learning approach have been widely adopted for the imbalance class of data (SMOTE) synthetic minority over sampling technique to transform into balance representation for pre training. To maximized the information gain between the input layer and the higher level stochastics inorder to achive good representation for each layer by using SDAE construction[17]. A reasonable health issue in the workplace rapidly increasing which can lead to mental health. Determining the most better factor of individual needed for treatment. PCA approach have been widely adopted and then clustering through density is used to reduce the noise. The mental health cases can be addressed by medical assistance. The goal of this paper is to identify which employee's related factor contribute more towards the mental health. PCA aims to learn the reduction of dimensions prior to the clustering DBSCAN. The goal is to find the tradeoff between noise reduction and the number of cluster [20]. The most serious global health treat is the increasing number of cases in mental illness. The most common type of mental disorder are depression, schizophrenia, dementia which are spreading around the world. One of the most promising ascepts of IOT based mental healthcare is data acquition, self organization, service level agreement, security and identify management data collected using IOT related technology for analyzing. IOT based system and services for mental health keeps on increasing, which addresses many challenges in the research. The Challenges related to different domain includes IOT device, analysis of massive data and intelligent health services [23, 24] are required. The use of IOT in instrumental treat different mental issues, it may be either from patient, doctors or family. Leonardo et al. studied different types of data collected by each work and also address mental disorder by the research community. Number of studies relevant to IOT and mental health uses popular technology like smart phone which gives understandable and low cost search criteria.

Sun at el. [4]proposed an effective supervised algorithm based on well established machine learning and some numerical analysis techniques. The author have proposed the Manhattan distance to determine implicitly use of RELIEF algorithm. in this paper feature selection is performed for classifying problems with complex data distribution with very high dimensions. In [4] designed a new feature selection algorithm that addresses several major issues regarding computational complexity for high dimensional dataset. This algorithm is generalized to address the multiclass problems where the defined margin two nearest neighbors are considered, one is the nearest hit NH and the other one from nearest miss NM. This algorithm used to learn locally and share the same goals to reduce data dimensionality. It uses feature extraction [3]BSO metaheauristics foraging interesting approach which is used for solving feature selection problem. QBOS-FS follows the wrapper approach and use the hybrid version for generating feature subset. The performances of the hybrid approaches have been verified using BSO combined with Q-learning for generating feature subset to evaluate the classification. We envision the challenge of feature selection where the size of search space increase exponentially with respect to cardinality of the original set N.

The classification performance in the healthcare improves by using Confidence-based and Cost-effective feature selection (CCFS) method [7,25] which is effectively developed a new mechanism for designing feature selection. Confidence-based and Cost-effective feature selection(CCFS) which adds the confidence of individual feature to denote the performance of single dimension dataset. CCFS mostly describe the cost of the features based on three key factors of the feature that are solved to design the fitness function, first one is the classification performance, second one is feature reduction ratio and third one is the feature cost. In the architecture of Confidence-based and Cost-effective feature selection(CCFS) method, it explain about the incorporation of features on the updated partial position and the relevance of the target feature and its selection frequency based on historical data to enhance the fitness function. It has been concluded that Confidence-based and Cost-effective feature selection(CCFS) obtain high classification accuracy on the wine, Ionosphere, and Sonar datasets, the low-dimensional, medium-dimensional, and high-dimensional datasets.

In brief, compared with the existing Binary Particle Swarm Optimization (BPSO) methods [6], it is not taken into consideration where traditional computation based on feature selection technique. In this traditional method the selection of feature in BPSO scheme, is updated by the pbest and gbest method where the overall performance of each partial and partial swarm is checked. BPSO based feature selection use Personal best (pbest) particle's historically best position and Global best (gbest) globally best position of particle swarm update the feature according to the overall fitness of swarm partial position. BPSO[8] was design to solve the discrete optimization problem. BPSO based feature selection method uses only the error rate of the classification algorithm as the fitness function. BPSO method use information Gain, PCFS, CFS filtering method for feature selection.

Deep learning feature selection techniques is based on deep architecture where the researcher applied feature selection in number of areas like computer vision, remote sensing, natural language processing and bioinformatics [7, 25] [11]. The paper implements principal component analysis (PCA) that is used to filter out the most appropriate features from the data set. The literature review focus on developing feature selection models using deep learning algorithms. Evolutionary computing techniques [9] are powerful stochastic algorithms whose searching method needs some natural phenomena like genetic inheritance and Darwinian strive for survival. The most common algorithm in this category includes genetic algorithms, evolutionary programming, evolution strategies, and genetic programming.

In the past the feature selection used to evaluate the feature in isolation where it did not consider the correlation between features [5]. Later the methods are improved and come up with proposed feature selection with minimum redundancy maximum relevance (mRMR) feature selection approach. Main advantage of this method is that the maximum relevance criteria along with minimum redundancy criteria is used to choose features that are maximally relevant to the criteria and minimally redundant with respect to the criteria[10].

III. FEATURE SELECTION

This is the most important technique of machine learning used to get better results with the minimum number of features while reducing redundant features, removing irrelevant features but increasing accuracy and improving performance of an algorithm is said to be as feature selection. It is to find an optimal subset of f features out of original set of F features. This technique better understands the data, important features of data and the relationship between the features. The learning process of a machine drastically improves with the best selected features though other features are not contributing to the objective. The features are irrelevant when they do not have any influence on the output, similarly the features are redundant or duplicate when they take the role of another feature. There are two key issues for feature selection algorithm research: First, choose the appropriate evaluation criteria to measure the pros and cons of the feature subset; second is to choose an efficient feature subset search method, so as to find the smallest feature combination that can best describe the criteria.

3.1 Feature Selection Algorithms

The algorithms have to be designed to optimally select features out of all possible subsets of features. It is partially difficult sometimes when there are large numbers of features, but preferred to have a satisfactory set of features. There are three broad classes of feature selection algorithms: (i) filter method (ii) Wrapper method (iii) Embedded method.

Filter methods are used to filter the features based on discriminating criteria i.e the filters out unreliable features from a given feature set for effective prediction. In filter methods the feature variance is computed and we select the subset of feature based on user specified threshold.

Wrapper methods find the feature subsets through search space and evaluate each subset by learning a machine learning model. It is special case of sequential feature selection, where it finds the optimal feature subset by iteratively selecting features based on classification performance. We start with one feature at a time in each round and that feature subset given for training and new model is generated.

Embedded methods use both the above method to find good subset of features using filter method and find the best features using filter method and find the best feature from those features using wrapper model.

3.1.1 Information Gain

We use information gain method to quantify the relevance of an attribute out of five attributes. Let S be a set of training samples and S_1, S_2 are the sample belongs to target class and contrast class. The expected information for a given sample is

$$I(S_1, S_2, \ldots, S_m) = -\sum_{i=1}^{m} \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

Where m is the number of classes and s_i is the samples in class C_i with probability $\frac{s_i}{s}$

Let s_j contain s_{ij} samples of class C_i, the expected information based on this partitioning by attribute A is known as the entropy of A.

Entropy of A is $E(A)=\sum_{j=1}^{v} \frac{s_{1j}+s_{2j}+\cdots+s_{mj}}{s}$ $I(s_{1j}+s_{2j}+\cdots+s_{mj})$

Where v is the number of distinct values in the attribute A and s_{ij} is the samples in class C_i with attribute value j.

The information gain obtained by this partitioning on A is defined by

$$\text{Gain}(A) = I(S_1, S_2, \ldots, S_m) - E(A)$$

We compute the information gain for each of the attributes, the attribute with highest information gain is considered most discriminating attribute of the given set. The attributes that are not relevant to the task are removed based on users threshold value.

3.1.3 Chi square algorithm

In the healthcare research, some study collect data on categorical variables which is summarized as series of counts[13]. The count which are present in the tabular format which is also called as contingency table. The chi-square test used to compute the evaluation between these rows and columns in this contingency table. The logic of hypothesis testing was first invented by Karl Pearson (1857-1936) and the Chi-square test also known as test for goodness-of-fit and test of independence are his most important contribution to the modern theory of statistics.

Chi square is a nonparametric test specifically used for two purpose (i) for testing the hypothesis of no associated between more then two groups (ii) and also use to test how likely the observed distribution of data fits the distribution. For the categorical data analysis where the parametric or continous data are there mostly it is used there.

The formula for calculating a Chi-square statistic is:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where, O stands for the observed frequency and E stands for the expected frequency.

The sigma sign in front of them denotes that we have, to sum up, these values calculated for each cell to find the association between null and alternative hypothesis.

IV. ANALYSIS OF TECH SURVEY DATASET

We have collected the mental health in technical survey dataset consisting of 1259 records and 27 feature from various people worldwide that measures attitudes towards mental health and frequency of mental health illness in the technical workplace. This dataset consist of many discrete and continuous variableslike timestamp, Age, gender, country, state: which give the description of the techies who live in united state or territory, self_employed: describe the self-employbility, family history: describe the family history of mental illness, treatment: describe the sought treatment for mental health condition, work_interfere: gives the details of mental health condition, if yes do you feel that it interferes with your work, no employees: describe the no of employees your company have, remote_work: data describe the techies work remotely atleast 50% of there time, tech_company: which is the target feature describe the employee primilarly a tech company, benefits: which describe all the employees getting benefits for mental health, care_options: gives the details for the options for mental health care your employee provides, wellness program: discuss about mental issues as a part of employee wellness program, seek_help: discuss about resources to learn more about mental health issues and how to seek help, **anonymity**: Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources, **leave**: How easy is it for you to take medical leave for a mental health condition?, **mental*health*consequence**: Do you think that discussing a mental health issue with your employer would have negative consequences?, **phys*health*consequence**: Do you think that discussing a physical health issue with your employer would have negative consequences?, **coworkers**: Would you be willing to discuss a mental health issue with your coworkers?, **supervisor**: Would you be willing to discuss a mental health issue with your direct supervisor(s)?, **mental*health*interview**: Would you bring up a mental health issue with a potential employer in an interview?, **phys*health*interview**: Would you bring up a physical health issue with a potential employer in an interview?, **mental*vs*physical**: Do you feel that your employer takes mental health as seriously as physical health?, **obs_consequence**: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?, **comments**: Any additional notes or comments. The snapshot of the dataset is given herewith for the ready reference in Fig. 1.

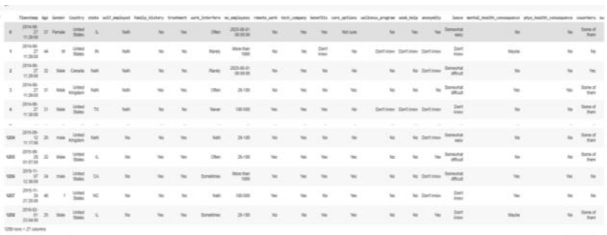

Fig. 1: Technical survey dataset

The following pie chart shows that the people of the country participated more for this survey as the survey was worldwide in fig. 2. The pi-chart clearly states that there are more number of people from United State followed by Canada participated for this survey.

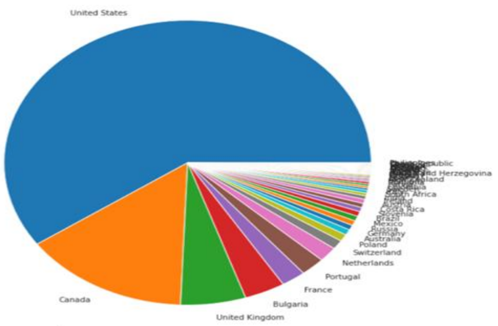

Fig. 2: Country participation

We fill the missing values by the zero value in the preprocessing steps of Mental Health in Tech survey dataset. In this dataset there are a total of 27 features and mostly the NaN values are present as the data values. There are 4 features contains NaN values like a feature called state contains 515 NaN values, a feature called state self_employed contains 18 NaN values, a feature called work_ interfere contains 264 NaN values and a feature called comments contains 1095 NaN values. This dataset is the categorical dataset and we converted into numerical dataset by replacing a number for each feature value. We have assigned the numbers starting from 0 like a feature called country has 20 distinct values so we assigned the number from 0 to 19. The features that have only two values are treated as binary values like 0 and 1. We have not performed any transformations to the data values as it is not required in this dataset.

V. IMPLEMENTATION

The goal of feature selection is to find a best subset of input features that contains the least number of dimensions and mostly contribute to accuracy by removing unimportant dimensions that are the most irrelevant and also removing redundant features from the dataset. We have implemented two algorithms for feature selection i.e. information gain and chi square to calculate the feature score with respect to targeted variable. The target class feature for this research is tech company which leads to find the percentage of technical people gets mental illness compared to non technical people in the tech survey dataset. By using information gain and the chi square algorithms, the top 10 features are extracted and shown in the table 1 and table 2 below.

Table 1: Top 10 features using Information Gain

Sl. No	Feature No	Feature Name	Feature Score
1	8	No_employees	0.025546
2	11	Care_options	0.019032
3	16	Mental_health_interview	0.016399
4	9	remote_work	0.016215
5	14	anonymity	0.013233
6	3	State	0.011364
7	5	Family_history	0.010092
8	6	Treatment	0.010050
9	22	Mental_vs-physical	0.009390
10	21	Phys_health_interview	0.008327

Table 2: Top 10 features using Chi square

Sl. No	Feature No	Feature Name	Feature Score
1	8	no_employees	33.891290
2	9	Remote_work	15.394439
3	0	Age	13.345462

4	2	Country	9.294476
5	4	Self_employed	6.578075
6	12	Wellness_program	5.971462
7	23	Obs_consequence	4.817349
8	14	anonymity	4.771223
9	15	leave	3.967911
10	13	Seek_help	2.859212

The most common features of both the algorithms are no_employees, remote_work, anonymity which is selected as the best feature subset related to the target class variable shown in the table 3

Table 3: Most common features

Sl. No	Feature No	Feature Name	Feature Score using Information Gain	Feature Score using Chi square
1	8	no_employees	0.025546	33.891290
2	9	Remote_work	0.016215	15.394439
3	14	anonymity	0.013233	4.771223

VI. RESULTS AND DISCUSSIONS

The results are computed from the top 10 features of information gain, chi square and most common features by using five machine learning algorithms called as SVM, NAIVE BAYES, DECISION TREE, KNN and LOGISTIC REGRESSION to find the accuracy. The table 4 and figure 3 has shown the accuracy obtained by using all five machine learning algorithms considering the 10 best features using information gain. The table 5 and figure 4 has shown the accuracy obtained by using all five machine learning algorithms considering the 10 best features using Chi square. The table 6 and figure 5 has shown the accuracy obtained by using all five machine learning algorithms considering the most common features of both the algorithms. The table 7 and figure 6 has shown the accuracy obtained by using all five machine learning algorithms considering all the features of mental health illness in tech survey. The table 8 is prepared as an aggregated accuracy of all the tables to visualize the importance of the algorithms with respect to the feature subsets.

Table 4: Accuracy using 10 best features of information gain

Sl No	Algorithm Name	Accuracy
1	SVM	0.803191
2	NAIVE BAYES	0.800532
3	DECISION TREE	0.712766
4	KNN	0.792553
5	LOGISTIC REGRESSION	0.803191

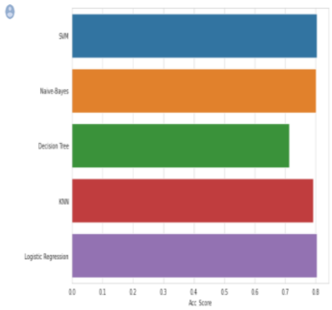

Fig. 3: The graph for accuracy using 10 best features of information gain

Table 5: Accuracy using 10 best features of Chi square

Sl No	Algorithm Name	Accuracy
1	SVM	0.803191
2	NAIVE BAYES	0.736702
3	DECISION TREE	0.704787
4	KNN	0.855851
5	LOGISTIC REGRESSION	0.800532

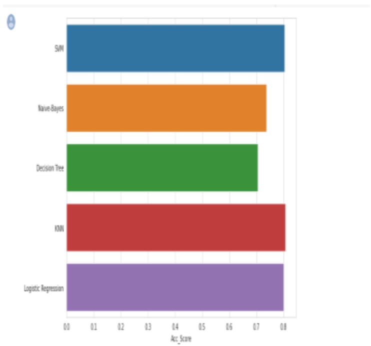

Fig. 6: The graph for accuracy using 10 best features of Chi square

Table 6: Accuracy using most common features

Sl No	Algorithm Name	Accuracy
1	SVM	0.795213
2	NAIVE BAYES	0.768617
3	DECISION TREE	0.728723
4	KNN	0.792553
5	LOGISTIC REGRESSION	0.792553

Table 7: Accuracy using all features

Sl No.	Algorithm Name	Accuracy
1	SVM	0.8539682
2	NAIVE BAYES	0.765079
3	DECISION TREE	0.755555
4	KNN	0.8507936
5	LOGISTIC REGRESSION	0.8571428

Table 8: Aggregated Accuracy

Sl No.	Algorithm Name	Accuracy using best Information Gain features	Accuracy using best Chi square features	Accuracy using most common features	Accuracy using all features
1	SVM	0.803191	0.803191	0.795213	0.8539682
2	NAIVE BAYES	0.800532	0.736702	0.768617	0.765079
3	DECISION TREE	0.712766	0.704787	0.728723	0.755555
4	KNN	0.792553	0.855851	0.792553	0.8507936
5	LOGISTIC REGRESSION	0.803191	0.800532	0.792553	0.8571428

The accuracy of all the classification algorithms depends on the feature selection algorithms and the input dataset. Each algorithm is implemented with best ten features of information gain, best ten features of Chi square, most common features of both the algorithms and considering all the features. The following observations are derived from the aggregated accuracy table 8.

- The result suggested that the SVM, KNN and Logistic Regression classification algorithms perform 86 percentages overall and better than all other algorithms considering the accuracy using all the features.
- The result have shown that the SVM, KNN and Logistic Regression classification algorithms perform 80 percentages overall and better than all other algorithms considering the accuracy using most common features.
- The result noticed that the KNN perform 86 percentages but SVM and Logistic Regression perform 80 percentages and better than all other algorithms considering the accuracy using best ten features of Chi square feature selection algorithm.
- The results declared that the SVM, Naïve Bayes, KNN and Logistic Regression classification algorithms perform almost 80 percentages and better than the Decision tree algorithm considering the accuracy using best ten features of information gain feature selection algorithm.
- The accuracy of SVM classification algorithms is maximum i.e. 85 percentage using all features of dataset but equally better i.e. 80 percentages considering the best ten features of information Gain and Chi square along with most common features of both the algorithms.
- The accuracy of Naïve Bayes classification algorithms is maximum i.e. 80 percentage considering the best ten features of information Gain and equally better i.e. 77 percentages considering the most common features of both the algorithms along all features of dataset but it is minimum with best ten features of Chi square.
- The accuracy Decision Tree classification algorithms is maximum i.e. 76 percentage using all features of dataset but equally better i.e. 73 percentages considering the most common features of both the algorithms and it is minimum i.e. almost 71 percentage with best ten features of information Gain and Chi square feature selection algorithms.
- The accuracy of KNN classification algorithms is maximum i.e. almost 86 percentage using all features of dataset considering the best ten features of Chi square but equally better i.e. 79 percentages considering the best ten features of information Gain along with most common features of both the algorithms.
- The accuracy of Logistic Regression classification algorithms is maximum i.e. 86 percentage using all features of dataset but equally better i.e. 80 percentages considering the best ten features of information Gain and Chi square along with most common features of both the algorithms.
- The concluding remark is SVM, KNN and Logistic Regression classification algorithms provides better accuracy in case of any feature selection algorithms chosen in the mental health illness in tech survey dataset.

VII. CONCLUSION

The proposed work is to find the percentage of technical people suffers the mental health illness disease compared to non technical people considering the tech survey dataset. The work proceeds with considering 27 features and five classification algorithms along with feature selection algorithms. The research focused on a target class variable i.e. tech company and ten best features of information Gain as well as Chi square including the most common features of both the algorithms. The

accuracy of all the classification algorithms are presented based on the best ten features, most common features and considering all the features. The result demonstrated that the SVM, KNN and Logistic Regression classification algorithms provides better accuracy compared to other algorithms in the case of mental health in tech survey dataset. The inference drawn in this paper is that the accuracy of technical people suffers the mental health illness disease compared to non technical people is 86 percentage. Therefore, considering a workplace the 86 percentages of technical people suffers from the mental health illness disease according to the tech survey dataset.
The first paragraph under each heading or subheading should be flush left, and subsequent paragraphs should have a five-space indentation. A colon is inserted before an equation is presented, but there is no punctuation following the equation. All equations are numbered and referred to in the text solely by a number enclosed in a round bracket (i.e., (3) reads as "equation 3"). Ensure that any miscellaneous numbering system you use in your paper cannot be confused with a reference [4] or an equation (3) designation.

REFERENCES

[1]. "Mental Health Policy and Service Development Department of Mental Healthand Substance Dependence Non communicable Diseases and Mental Health", article.
[2]. Ryan J. Urbanowicz, Melissa Meeker, William La Cava, Randal S. Olson, Jason H. Moore "Relief-Based Feature Selection: Introduction and Review", DOI: https://doi.org/10.1016/j.jbi.2018.07.014, PII: S1532-0464(18)30140-0
[3]. Souhila Sadeg1(B) , Leila Hamdad2, Amine Riad Remache1, "QBSO-FS: A Reinforcement Learning Based Bee Swarm Optimization Metaheuristic for Feature Selection", DOI: 10.1007/978-3-030-20518-8_65, pp.785-796
[4]. Sun, Y., Todorovic, S., Goodison, S., "Local-learning-based feature selection for high-dimensional data analysis", IEEE transactions on pattern analysis and machine intelligence 32 (9), 1610–1626, 2020.
[5]. Chapter- 3 feature selection.
[6]. Y. Chen, Y. Wang, L. Cao, and Qun Jin, "An effective feature selection scheme for healthcare data classification using binary particle swarm optimization," in *Proc. of the 9th International Conference on Information Technology in Medicine and Education* (*ITME*), 2018.
[7]. Y. Li, C.Y. Chen, and W. Wasserman, "Deep Feature Selection: Theory And Application To Identify Enhancers And Promoters", Journal of Computational Biology, vol. 23, pp. 322-336, 2016.
[8]. S.M. Vieira, L.F. Mendonça, G.J. Farinha, J.M.C. Sousa, "Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients", Applied Soft Computing, 13 (2013) 3494-3504.
[9]. Tran Trong Dao1, "Investigation on Evolutionary Computation Techniques of a Nonlinear System", Hindawi Publishing Corporation Modelling and Simulation in Engineering, Volume 2011, Article ID 496732, 21 pages, doi:10.1155/2011/496732
[10]. Gaikwad Kiran P, Dr C M Sheela Rani , "Regression Model with Modified Linear Discriminant Analysis Features for Bimodel Emotion Recognition", "International Journal of Scientific & Technology Research", ISSN : 2277-8616, Volume 9, Issue 3, pp. 1355 to 1360, March 2020
[11]. Milad Zafar Nezhada, Dongxiao Zhub;_, Xiangrui Lib, Kai Yanga, Phillip Levyc , "SAFS: A Deep Feature Selection Approach for Precision Medicine", 20 Apr 2017.
[12]. H.C. Peng, E.H. Herskovits, and C. Davatzikos, "Bayesian Clustering Methods for Morphological Analysis of MR Images", Proc. Int'l Symp. Biomedical Imaging: from Nano to Macro, pp. 485- 488, 2002.
[13]. Rana R, Singhal R. "Chi-square test and its application in hypothesis testing", J Pract Cardiovasc Sci [serial online] 2015 [cited 2020 Sep 27];1:69-71. Available from: http://www.j-pcs.org/text.asp?2015/1/1/69/157577
[14]. S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis", *J. Intell. Learn. Syst. Appl.*, vol. 5, no. 3, 2013, Art. no. 35396.
[15]. R. Detrano, A. Janosi,W. Steinbrunn, M. P_sterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease", *Amer. J. Cardiol.*, vol. 64, no. 5, pp. 304-310, Aug. 1989.
[16]. Hanchuan Peng, Member, IEEE, Fuhui Long, and Chris Ding, Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. pp. 1226- 1238, IEEE

TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, NO. 8, AUGUST 2005.

[17]. Padideh Danaee∗ , Reza Ghaeini, a deep learning approach for cancer detection and relevant gene identification, Pac Symp Biocomput 2017;22:219-229, doi: 10.1142/9789813207813_0022

[18]. Md Rafiqul Islam, Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Depression detection from social network data using machine learning techniques, August 2018, Health Information Science and Systems 6(1):8, DOI:10.1007/s13755-018-0046-0

[19]. Shambhavi Pandey, Mohammad Bilal Khan and Sagar Thakkar, Clark University, Factors Affecting Mental Health in Employees and Their Relation to Suicide Rates at a Workplace

[20]. Venugopal, G.; Navaneethakrishna, M.; Ramakrishnan, S. Extraction and analysis of multiple time window features associated with muscle fatigue conditions using sEMG signals. Expert Syst. Appl. 2014, 41, 2652–2659.

[21]. Shruti Appiah, Sam Barnard, Jonathan Deiven, Density-based Clustering of Workplace Effects on Mental Health,

[22]. Leonardo J. Gutierrez1, Kashif Rabbani2, Oluwashina Joseph Ajayi2, Samson Kahsay Gebresilassie2,Joseph Rafferty2, Luis A. Castro1,*and Oresti Banos3, Internet of Things for Mental Health: Open Issues in DataAcquisition, Self-Organization, Service Level Agreement,and Identity Management, International Journal ofEnvironmental Researchand Public Health, Int. J. Environ. Res. Public Health2021, 18, 1327. https://doi.org/10.3390/ijerph18031327, https://www.mdpi.com/journal/ijerph

[23]. Abar, M.; Rahman, A.; Arif, F.; Jeon, G. Energy-harvesting based on internet of things and big data analytics for smart healthmonitoring.Sustain. Comput. Inform. Syst.2018,20, 155–164, doi:10.1016/j.suscom.2017.10.009

[24]. Wan, J.; Gu, X.; Chen, L.; Wang, J. Internet of Things for Ambient Assisted Living: Challenges and Future Opportunities.In Proceedings of the 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery(CyberC 2017), Nanjing, China, 12–14 October 2018. doi:10.1109/CyberC.2017.83.

[25]. Wan, J.; Gu, X.; Chen, L.; Wang, J. Internet of Things for Ambient Assisted Living: Challenges and Future Opportunities.In Proceedings of the 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery(CyberC 2017), Nanjing, China, 12–14 October 2018. doi:10.1109/CyberC.2017.83.