Predicting Anxiety in Remote-Working During COVID-19: A Machine Learning Approach

Leen Ismail*, Rami AlOuran† , Wasayef Ananzeh‡

***Abstract*—Several precautionary measures were introduced with the emergence of COVID-19. Due to the unawareness of the consequence of these precautions, various studies showed how they affected workers' mental health and resulted in several effects including anxiety and depression. This study's main aim is to predict anxiety level for United States workers to avoid burnout in the long term. Using a survey made by U.S. Census Bureau's Household Pulse Survey that investigates people's lifestyles during the pandemic a machine learning model was developed to predict anxiety. The highest predictors of anxiety were extracted and examined to be used in a machine-learning model. Logistic Regression and Random Forest were used to build classification models to classify the anxiety level of an employee. Data was collected from 47,603 individuals, all living in The U.S., randomly picked from weeks 30 and 33 of the U.S. Census Bureau's Household Pulse Survey. The most effective feature of Anxiety was worry followed by feeling down or depressed and then feeling of losing interest where all of these features can be a consequence of burnout. Logistic Regression resulted in a 72% accuracy rate of predicting and classifying anxiety while random forest showed better performance with a 78% accuracy rate. Area under the receiver operator curve(AUC-ROC) values were 0.82 and 0.95 respectively for Logistic Regression and Random Forest. Our results reveal that anxiety could lead to depression and losing interest, leading to burnout. Ignoring employees' anxiety until it reaches a high level will simultaneously be a motive to lose interest in doing things including work.**

Keywords: Anxiety, Burnout, Machine Learning, Pandemic, COVID-19.

I. Introduction

The COVID-19 pandemic drastically influenced our daily life in many aspects, and working was the most affected. Remote working started to be prominent soon after the lockdown. Remote working can be defined as working where employees work away from the office for instance, from home or any other location that is not an organisation's business place. [1] It is a way for companies to cope with the pandemic changes; companies were forced to work virtually to keep the economy alive. It is mutually beneficial for both employers and employees. One of the remote working benefits is flexibility, which leads to better time management and less stress for employees, leading to better performance. Technology has made remote working even more convenient and nearly faultless; an internet connection is enough to work remotely with the use of modern technologies such as cloud services which makes it easier for employees to use some services such as computing or storage services rather than using their computer to store or run programs. Moreover, remote working established better opportunities for people to work from anywhere despite their living location. Similarly, companies found that remote working increases productivity and decreases operational costs. Some companies continued working remotely while others referred to hybrid working. Companies whose work nature can be done remotely prefer staying as they are. Contrastingly, remote work had a negative influence on both employers and employees. A study made in July 2020 shows that 69 per cent of workers experienced symptoms of burnout [1]. Employees are facing mental and physical exhaustion when working remotely, leading to increased employee anxiety, which can be a serious problem if not treated. Employees face difficulty while working remotely in separating work from private life and distraction, this issue can be the reason why employees feel they are working hard and therefore feel exhausted. Despite the challenges that face companies, remote work is rising with a good impact. Using the right tools, technologies and prediction of employee anxiety before occurring makes remote working easier, safer and more adequate for both employees and employers.

A report published by The Standard revealed that 55% of employees reported that during the pandemic their mental health was affected negatively. Due to this fact, in this research employee anxiety will be predicted using different ML models by studying which variables affect individuals while working remotely and help in solving anxiety on employees before it grows into burnout.

II. Literature Review

Waheeda Almayyan conducted a study to measure the level of burnout in workers during the COVID-19 pandemic, two types of datasets were collected using surveys for 599 workers. AnDE algorithm, JChaid* algorithm, SVM classifier, ForestPA Classifier, DMLP and CGWO algorithms were used to test the best algorithm's accuracy on the dataset. DMLP had the best performance with up to 0.957 accuracy [2]. In an evaluation made by S. T. Arokkiya Mary and Dr. L. Jabasheela data was collected from 600 questionnaires samples to measure depression, anxiety, and stress in students upon several factors using different AI learning algorithms. Measurements were classified from normal to extremely severe upon scores; depression ranges from 0 to 28+, stress ranges from 0 to 34+ and anxiety ranges from 0 to 20+ where 0 is normal. One of the algorithms used was Multilayer Perceptron which shows the highest accuracy and performance [3]. A study was made to measure employee exhaustion during the pandemic in Germany, data were collected from 3 samples over three months. Several variables such as gender and ages of children if found were collected, consequently, a hypothesis shows that women with children were affected more by the pandemic. This concludes that women are more likely to feel work exhaustion faster. Some solutions such as flexible hours working was introduced to make better management between private and work life [4]. Moreover, Physiological risks were measured before and during the COVID-19 pandemic in Finland on 1044 workers. Physiological, situational, and socio-demographic factors were investigated. One of the physiological factors is work exhaustion and results found that it was related to and impacted by COVID-19 anxiety, work exhaustion was increased by 41.78 percent during the COVID-19 pandemic [5] Another article shows that personality affects job productivity and outcome. Results found that extroverted and conscientious people were less productive and had more job burnout [6]. Similarly, in research made by the University of Calgary (2014) measured some personality traits and their association with cyberslacking. Five personality traits were measured: agreeableness, conscientiousness, neuroticism, honesty, and procrastination. Results indicate if neuroticism and procrastination personality traits are found, an employee is more likely to do cyberslacking which leads to low productivity [7]. In a study made by Jan Koch and Carsten C. Schermuly, emotional exhaustion was measured by unfinished tasks. They expect that unfinished tasks are positively related to the level of employee exhaustion. Therefore, the number of unfinished tasks in each project has increased the workload as well as employee exhaustion. Authors also state that employee exhaustion might also increase the number of unfinished tasks and therefore decrease performance [8] Furthermore, an application was created by Milton High School that predicts physiological disorders in teens. Using a neural-network model and Gradient Boosting Machine the model was trained to track SMS and tweets from a phone to predict any disorder and then contacting the parents [9].

A. Related papers

As per Priyaa A. (2020) several machine learning algorithms were used to measure anxiety, stress and depression levels in 348 participants where they measured features of anxiety such as 'scared without any good reason', 'breathing difficulties' and 'close to panic'. As for depression, they measured features such as 'life was meaningless', 'nothing to look forward' and 'down hearted and blue'. Lastly, stress was measured using features like 'difficulty to relax', 'getting agitated' and 'I was rather touchy'. Utilizing different machine learning models such as decision trees, random forest trees, Naive Bayes, support vector machine and k-nearest neighbour; random forest trees showed the best performance with an accuracy of 0.714 for anxiety, 0.798 for depression and 0.723 for stress.

Moreover, a study made by Simjanoski M. in 2022 predicted depression and anxiety in 22,562 participants. Each feature was used as a variable to predict anxiety or depression. The strongest feature that affected anxiety was the quality of sleep and the meaning of life feature was the most effective on depression. Sample data was tested on the following four machine learning models: Elastic net, random forest, and XGBoost models. The highest accuracy for predicting depression was the elastic net algorithm (0.79) followed by the random forest algorithm (0.78). As for anxiety, both elastic net and random forest algorithms executed the same accuracy (0.78).

A research made by Gamage S.N. in 2022 studied mental distress on employees who work online by evaluating several aspects including physiological and environmental aspects. It was limited to 481 IT workers. This research showed very high accuracy rates using 5 different machine learning algorithms namely, Naïve Bayes, XGBoost, CatBoost, Decision tree and Support vector machine. CatBoost showed the highest overall performance with an accuracy of 97.1%. In a similar research made by Albagmi F. also measured anxiety levels during the lockdown on 3017 participants. Support vector machine and J48 algorithms were tested on the sample data, SVM had a higher accuracy over J48 with an accuracy of 100

TABLE I
RELATED PAPERS

Paper Name	Year	Data Size	Method Used
Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms	2020	348	Random Forest Tree
Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach	2022	3017	Support vector machine
Machine Learning Approach to Predict Mental Distress of IT Workforce in Remote Working Environments	2022	481	Naïve Bayes, XGBoost, CatBoost, Decision tree, and Support vector machine
Lifestyle predictors of depression and anxiety during COVID-19: a machine learning approach	2022	22562	Elastic net, random forest, and XGBoost models

III. MATERIALS AND METHODS

In this study, we focused on detecting anxiety in people born before 2003 using the Census Bureau's Household Pulse Survey. We explored and selected the features that affect anxiety and then created and measured the performance of the two machine learning algorithms to predict and classify anxiety.

Qualitative secondary data was selected. Feature extraction was conducted and good predictors of anxiety were employed to build the model and evaluate the results.

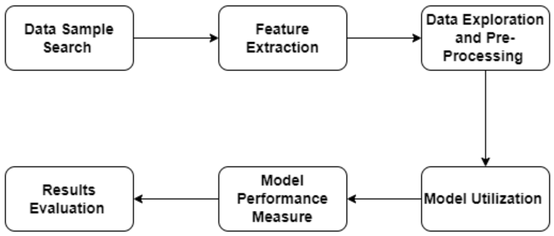

Figure 1: Research Methodology

A. *Description of Data*

Data was collected from a survey by Census Bureau's Household Pulse Survey conducted in the United States. The survey was conducted every week or two from April 2020 to the current date. It is conducted on adults born before 2003 living in the United States, in this project we randomly picked week 30 (May 12 - May 24 2021) and week 33 (June 23 - July 5 2021) as sample data to measure the effectiveness of our proposed classification models [10]. The survey's main objective is to test the impact of COVID-19 on US households. The survey collects information about the physical and mental health of individuals. To predict the anxiety of workers we used some features that showed a strong impact on anxiety. The total number of features was 239 then we used feature selection to determine the features relevant to anxiety discussed below.

B. *Data Pre-Processing and Feature selection*

The original dataset consisted of 239 features, the majority of these features had a high number of null values, hence, Features with more than 30% null values were removed. Afterwards, any sample that had any null value was removed. The initial number of records was 139,160, and the final number of records was 47,603 records and 73 features. The best features were selected using Chi-Squared Feature Selection. Chi-squared is used for categorical data to measure the relationship between features and output. It is used to select the most effective features using Chi-squared feature selection. The highest 25 features were used as shown and described in table 2.

Using Chi-square, categorical features with chi-square lower than 100 were excluded from our dataset. Findings from the feature selection reveal that the most effective three features of Anxiety were Worry, Down, and interest features. Some physical health issues also were an effective feature such as difficulties in seeing, hearing remembering and mobility difficulties like walking or climbing the stairs. Other features that describe living conditions were used, for instance, place of living, any difficulties paying expenses, and changes in spending and purchasing. And finally, medical health-related features such as health insurance, and receiving therapy or medication.

C. *Model selection*

We explored two supervised machine learning algorithms, Logistic Regression and Random Forest algorithms to be utilized in the process of anxiety classification.

1) Logistic regression: Multinomial Logistic regression is an extension of logistic regression; it is a supervised machine learning algorithm used for classifying more than two classes of data, unlike logistic regression which is only limited to two classes such as yes/no or occur/won't occur. It is a classification model that executes discrete outputs by measuring the probability of an event to occur divided by its probability of not occurring[12], [13].

2) Random Forest: Random Forest is a supervised machine learning algorithm that consists of decision trees on samples from a dataset. The Random Forest algorithm has better performance and accuracy results in classification cases which contains categorical values. The higher number of trees the higher the accuracy, random forest also reduces the occurrence of overfitting since trees are trained individually on different datasets then the average of predictions is calculated to create a single predictive model[14].
Hyper Parameters Tuning is used to optimize the algorithm's performance by finding the best parameters for our model. Regarding the random forest algorithm, the number of trees and depth are the parameters. In our model, hyperparameter tuning executed that the best parameters for our model are 250 number of trees and 16 number of level in a tree.

D. *Evaluation measures*

Evaluation measures show how good a model is performing, by evaluating a model we will have an insight into how to improve it. Three evaluation measures were used in the study to evaluate the two learning algorithms, Logistics Regression and Random Forest. Evaluation methods used in this research are Accuracy rate, AUC-ROC and F1-Score and Precision and Recall.

1) Accuracy: Accuracy is the most used evaluation metric in machine learning, it is defined as the probability of true predictions divided by all predictions. Accuracy is measured from 0 to 1, the nearest the value to 1 the better the model is performing. It is calculated as followed where TP stands for True Positive, TN for True Negative, FP for False Positive and FN for False Negative.[15]

$$\frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

2) F1-Score, Precision and recall: Precision is the measure of the positive outcomes' quality, which is the probability of true positives divided by all positive's predictions[16]. For instance, of all the people labelled as Anxious, how many of them were actually anxious? It is measured as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

TABLE II
DATA DICTIONARY [11]

Feature Name	Feature Description	Range	chi-square Value	Question
EGENDER	Gender	1:2	217.0	Are you... Select only one answer.
MS	Marital status	1:5	651.1	What is your marital status? Select only one answer.
SSADECISN	Applying for Social Security benefits	1:4	152.8	How has the coronavirus pandemic affected your decision about applying or not applying for Social Security benefits (Retirement, Disability, or Survivors), Supplemental Security Income (SSI) benefits, or Medicare benefits? Select only one answer.
EXPNS_DIF	Difficulties to pay for usual household expenses	1:4	4805.1	In the last 7 days, how difficult has it been for your household to pay for usual household expenses, including but not limited to food, rent or mortgage, car payments, medical expenses, student loans, and so on? Select only one answer.
CHNGSHOP1	changes in Online purchases	1:2	279.4	In the last 7 days, have you or your household changed your spending or shopping in the following categories? Select all that apply - Online purchases
CHNGSHOP2	changes in Curbside pick-up	1:2	103.4	In the last 7 days, have you or your household changed your spending or shopping in the following categories? Select all that apply - Curbside pick-up
CHNGSHOP3	changes in In-store shopping	1:2	177.6	In the last 7 days, have you or your household changed your spending or shopping in the following categories? Select all that apply - In-store shopping
CURFOODSUF	Food	1:4	2220.1	In the last 7 days, which of these statements best describes the food eaten in your household? Select only one answer.
WORRY	Worry Feeling	1:4	23364.9	Over the last 7 days, how often have you been bothered by the following problems ... Not being able to stop or control worrying? Would you say not at all, several days, more than half the days, or nearly every day? Select only one answer.
INTEREST	Lost Interest	1:4	14041.5	Over the last 7 days, how often have you been bothered by ... having little interest or pleasure in doing things? Would you say not at all, several days, more than half the days, or nearly every day? Select only one answer.
DOWN	Feeling Down	1:4	15930.1	Over the last 7 days, how often have you been bothered by ... feeling down, depressed, or hopeless? Would you say not at all, several days, more than half the days, or nearly every day? Select only one answer.
HLTHINS1	Health Insurance from employer	1,2	136.0	Are you currently covered by any of the following types of health insurance or health coverage plans? Mark Yes or No for each. - Insurance through a current or former employer or union (through yourself or another family member)
DELAY	DELAY of medical care	1:2	223.0	At any time in the last 4 weeks, did you DELAY getting medical care because of the coronavirus pandemic? Select only one answer.
NOTGET	Needed medical care but did not get it	1:2	135.4	At any time in the last 4 weeks, did you need medical care for something other than coronavirus, but DID NOT GET IT because of the coronavirus pandemic? Select only one answer.
TELEHLTH	Distant health appointments	1:2	161.2	At any time in the last 4 weeks, did you have an appointment with a doctor, nurse, or other health professional by video or by phone? Please only include appointments for yourself and not others in your household.
PRESCRIPT	took prescription medication regarding mental health	1:2	500.7	At any time in the last 4 weeks, did you take prescription medication to help you with any emotions or with your concentration, behavior or mental health? Select only one answer.
MH_SVCS	Receive any therapy	1:2	203.5	At any time in the last 4 weeks, did you receive counseling or therapy from a mental health professional such as a psychiatrist, psychologist, psychiatric nurse, or clinical social worker? Include couseling or therapy online or by phone. Select only one answer.
MH_NOTGET	Needed Therapy but did not receive	1:2	387.1	At any time in the last 4 weeks, did you need counseling or therapy from a mental health professional, but DID NOT GET IT for any reason? Select only one answer.

CONT. Data Dictionary

Feature Name	Feature Description	Range	chi-square Value	Question
SEEING	difficulty seeing	1:4	789.5	Do you have difficulty seeing, even when wearing glasses? Select one.
HEARING	difficulty hearing	1:4	231.6	Do you have difficulty hearing, even when using a hearing aid? Select one.
REMEMBERING	difficulty remembering	1:4	3339.2	Do you have difficulty remembering or concentrating? Select one.
MOBILITY	Mobility Difficulty	1:4	554.1	Do you have difficulty walking or climbing stairs? Select one.
TENURE	Living in house or apartment	1:4	266.5	Is your house or apartment. . . ? Select only one answer.
LIVQTRRV	Living Description	1:7	310.6	Which best describes this building? Include all apartments, flats, etc., even if vacant. Select only one answer.
TNUM_PS	Any plans to take courses before pandemic	0:3	107.6	Before the coronavirus pandemic, how many members of your household, including yourself, were planning to take classes this fall from a college, university, community college, trade school, or other occupational school (such as a cosmetology school or a school of culinary arts)? Please enter a number. If none, enter 0.

The recall is the probability of predicting positive outcomes correctly. Which is the probability of true positives divided by all actual positives[16]. For instance, of all Anxious people, how many were labelled as anxious. It is measured as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

F1-Score is the average of precision and recall by calculating both false positives and false negatives. F1-Score works the best with uneven class distribution, it grants a better understanding of classes accuracies [17]. It is measured as follows:

$$F1\text{-}score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

3) AUC-ROC Curve: To measure the performance of a multi-variable classification problem, the AUC-ROC curve is used. AUC-ROC stands for AUC (Area Under the Curve) – ROC (Receiver Operating Characteristics). ROC is a line curve plotted as a true positive rate (Recall) over a false positive rate graph [18].

$$TPR = \frac{TP}{TP + FN} \tag{5}$$

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

AUC value indicates the model's ability to distinguish classes. It is the area below the ROC curve, AUC value ranges from 0 to 1, 1 means that the model is fully able to distinguish classes correctly [18].

IV. Results and Discussion

Anxiety was measured on a scale from 1 to 4 where patients state how many days they felt anxious in the last 7 days. Where 1=no days at all, 2=several days, 3=more than half of the week and 4=nearly every day. Our data consisted of (22638) records of 1, (13616) records of 2, (4195) records of 3 and (4929) records of 4. After exploring and combining data from week 30 and week 33 of the pandemic from the U.S. Census Bureau Household Pulse Survey, we selected the highest effective features on anxiety. Using the two machine learning algorithms, logistic regression and random forest, 25 features were used to train and test the models. Random forest showed a higher performance with n accuracy of 78% while logistic regression resulted in a 72% accuracy rate.
This section will analyse the performances and results of the two machine learning algorithms. The results of each algorithm were evaluated by a three performance measures, accuracy, F1- Score(Precision and Recall) and AUC-ROC.

Precision, recall and F1-score results were analyzed for each category of anxiety. F1-score was used to have a better view of our model since we have imbalanced classes. As shown in the table below for the logistic regression model, overall accuracy was 0.72, f1-score, precision and recall for category 1 was the highest (0.88,0.85,0.90) followed by category 4 then category 2 while category 3 had the lowest f1-score, precision, and recall as shown in table 3.

Correspondingly, the Random Forest algorithm showed better and higher performance with a 0.78 accuracy rate.

Similarly, as stated above and as shown in table 4. Category 1 showed higher values of f1-score, precision and recall and category 3 had the lowest values.

TABLE III
F1-SCORE LOGISTIC REGRESSION

	Precisoin	Recall	F1-score
1	0.85	0.90	0.88
2	0.66	0.49	0.56
3	0.30	0.44	0.36
4	0.73	0.77	0.75
Over all Accuracy			0.72

TABLE IV
F1-SCORE RANDOM FOREST

	Precisoin	Recall	F1-score
1	0.85	0.91	0.88
2	0.66	0.71	0.69
3	0.54	0.32	0.40
4	0.82	0.77	0.76
Overall Accuracy			0.78

Area Under the Receiver Operating Characteristics was conducted on both algorithms. The Logistic Regression value was 0.82. while Random Forest showed better performance with an AUC value of 0.95 in the Random Forest algorithm. As plotted in the graphs below.

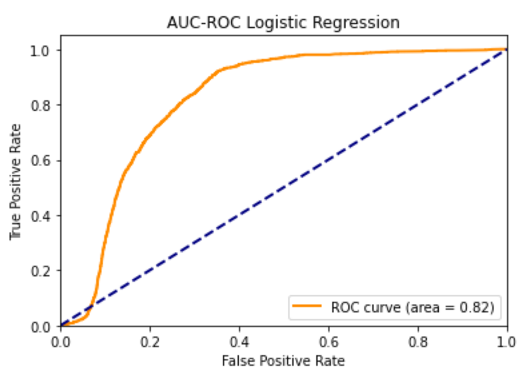

Figure 2: AUC-ROC LR

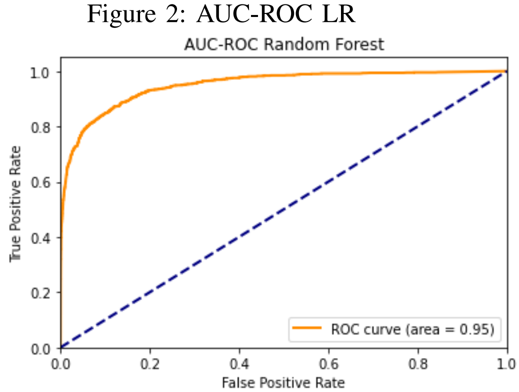

Figure 3: AUC-ROC RF

Firstly we used only week 30 data we ended up with an accuracy of 70% for Logistic Regression and 75% for Random Forest. Afterwards, we randomly picked data from week 33 and add it to our dataset which boosted our model as stated above. In some trials, we also added other weeks' data in the hope of increasing the accuracy but overall accuracy was decreased.

While developing the model we tried to use the One-Hot encoding technique since our data is not categorized equally, as seen in the feature description table. The down feature is categorized from 1 to 4 while LIVQTRRV is categorized from 1 to 7. In One-hot encoding we binaries data to 0's and 1's. nevertheless, accuracies were decreased.

According to related papers, the Random Forest algorithm was used frequently to predict anxiety, in research made by Piryaa A. in 2020 a Random Forest algorithm predicted anxiety with an accuracy of 0.714; anxiety was measured using features similar to ours such as 'scared without any good reason', 'breathing difficulties' and 'close to panic' which they are all similar to our highest predictor, feeling worried. Furthermore, in another similar research made by Simjanoski M. in 2022 that used Random Forest as one of the models to predict anxiety resulted in an accuracy of 0.78, one of the highest0020predictors of anxiety was 'meaning of life' feature which is quite similar to our interest feature. Both of these researches used a dataset that was intentionally made to predict levels of anxiety during COVID-19 which can be considered a limitation in our research.

V. CONCLUSIONS AND FUTURE RECOMMENDATIONS

The COVID-19 pandemic affected employees' mental health measured by several factors mental and physical ones including feeling down, lost of interest, difficulties remembering, walking or climbing the stairs, hearing or seeing. Physiological effects in the long-term cause burnout, burnout can be difficult and needs time to be treated. Hence, predicting anxiety before it occurs will be beneficial for both employees and employers to avoid problems they are dispensable of. In this study, we revealed that this problem can indeed be solved with high accuracy. Survey data was examined and utilized to create a Machine Learning model that predicts stages of anxiety. Two supervised classification algorithms were tested, Random Forest and Logistic Regression, Random Forest algorithm was found to predict better outcomes on our data with a 78% accuracy against logistic regression with a 72% accuracy.

The main limitation is that the main purpose of this survey was not to predict anxiety, the survey was focusing on the social and economical effects of the pandemic. Some factors that could affect employees' anxiety may be missing. For future recommendations, deeper research on the mental health aspect is required to increase features and develop a survey that focuses mainly on anxiety and physiological effects in order to create a better model with higher accuracy and greater performance in executing the results. furthermore, it is best to avoid surveying employees every week to avoid making it a hard job to accomplish. hence, Convolutional Neural Network Models can be used to detect user behaviour while working remotely to provide better supervision of the employee's state and execute faster predictions rather than waiting until the end of each week.

References

[1] P. Reshma, P. Aithal, and S. Acharya, "An empirical study on working from home: A popular e-business model," *International Journal of Advance and Innovative Research*, vol. 2, no. 2, 2015.

[2] W. Almayyan, "Developing a machine learning model for detecting job burnout during the covid-19 pandemic among front-line workers in kuwait," *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 19, no. 10, 2021.

[3] S. A. Mary and L. Jabasheela, "An evaluation of classification techniques for depression, anxiety and stress assessment," in *International Conference for phoenixes on emerging current trends in engineering and management (PECTEAM 2018)*. Atlantis Press, 2018, pp. 64–69.

[4] B. Meyer, A. Zill, D. Dilba, R. Gerlach, and S. Schumann, "Employee psychological well-being during the covid-19 pandemic in germany: A longitudinal study of demands, resources, and exhaustion," *International Journal of Psychology*, vol. 56, no. 4, pp. 532–550, 2021.

[5] I. Savolainen, R. Oksa, N. Savela, M. Celuch, and A. Oksanen, "Covid-19 anxiety—a longitudinal survey study of psychological and situational risks among finnish workers," *International journal of environmental research and public health*, vol. 18, no. 2, p. 794, 2021.

[6] A. M. Evans, M. C. Meyers, P. P. V. De Calseyde, and O. Stavrova, "Extroversion and conscientiousness predict deteriorating job outcomes during the covid-19 transition to enforced remote work," *Social Psychological and Personality Science*, p. 19485506211039092, 2021.

[7] T. A. O'Neill, L. A. Hambley, and A. Bercovich, "Prediction of cyberslacking when employees are working away from the office," *Computers in Human Behavior*, vol. 34, pp. 291–298, 2014.

[8] J. Koch and C. C. Schermuly, "Managing the crisis: How covid-19 demands interact with agile project management in predicting employee exhaustion," *British Journal of Management*, vol. 32, no. 4, pp. 1265–1283, 2021.

[9] D. Nori, "isense: Android app to identify linguistic biomarkers of mood disorders."

[10] "A out the ensus ureau's household pulse survey - new york city." [Online]. Available: https://www1.nyc.gov/assets/opportunity/misc/us-census-bureau-household-pulse-survey.pdf

[11] U. C. Bureau, "Household pulse survey public use file (puf)," Sep 2022. [Online]. Available: https://www.census.gov/programs-surveys/household-pulse-survey/datasets.html

[12] S. K. Ashenden, *The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry*. Academic Press, 2021.

[13] J. Brownlee, "Multinomial logistic regression with python," Aug 2020. [Online]. Available: https://machinelearningmastery.com/multinomial-logistic-regression-with-python/

[14] E. R. Sruthi, *Understanding Random Forest*, 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

[15] K. Nighania, "Various ways to evaluate a machine learning models performance," Jan 2019. [Online]. Available: https://towardsdatascience.com/various-ways-to-evaluate-a-machine-learning-models-performance-230449055f15

[16] "Classification: Precision and recall." [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

[17] E. Solutions, "Accuracy, precision, recall amp; f1 score: Interpretation of performance measures," Nov 2016. [Online]. Available: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/

[18] A. Bhandari, "Auc-roc curve in machine learning clearly explained," Jun 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/