IS424 G2T4 - Data Mining and Business Analytics

Recommendations for Mental Health in the Tech Industry

Final Report

Name	Student ID
Lim Bei Ling Cheryl	01417285
Chiang Kheng He	01428232
Ming Yao	01370766
Lim Yin Yao	01387138
Ryne Tan	01370692
Sophie Opheiden	01497973
Kwong Wen Lin	01415939

Recommendations for Mental Health in the Tech Industry

Leam Ming Yao, Lim Bei Ling Cheryl, Chiang Kheng He, Yin Yao, Ryne Tan, Sophie Opheiden, Kwong Wen Lin

Abstract: Mental health issues are a major public health issue that affects a large proportion of the world's population. The majority of mental health problems in the workplace are attributed to the lack of awareness about mental health issues. In this report, we outline the challenges associated with developing a classification model for mental health issue detection, alongside outlining the motivations driving this project and recommends actions for better mental health at the workplace. This report outlines findings from a profound literature review will be presented, details of the chosen dataset and tools, and assesses the efficiency of the models developed using this dataset. Based on topic modelling results, the report will elaborate on recommendations on how to address the identified topics within an organisation. Finally, project constraints will be discussed and potential avenues for future enhancements and mitigation of limitations proposed.

Keywords: recommendations, prediction, mental health support

Source Code can be found in this GitHub Link: https://github.com/dracolim/IS424-G2T4.git

Content Page

1. Introduction

In the fast-paced world of technology, where innovation thrives and advancements are constant, there exists a lesser-discussed but profoundly impactful aspect: mental health. Behind the lines of code and the glare of screens, individuals in the tech industry face unique challenges that can take a toll on their well-being.

In recent years, the tech industry has garnered attention not only for its groundbreaking developments but also for the mental health struggles experienced by its workforce. The pressure to meet deadlines, the ever-evolving nature of technology, and the culture of constant connectivity can contribute to stress, anxiety, and burnout among tech professionals.

Despite the industry's emphasis on problem-solving and innovation, mental health concerns often go unaddressed or are stigmatised, leading to a culture of silence and reluctance to seek help. However, acknowledging and addressing these challenges is essential for fostering a healthy and sustainable work environment.

This report delves into the challenges linked to developing a classification model for mental health issue detection, alongside outlining the motivations driving this project and recommends actions for better mental health at the workplace. Additionally, it presents findings from a literature review, details the chosen dataset and tools, and assesses the efficiency of the models developed using this dataset. Finally, the report addresses project constraints and proposes potential avenues for future enhancements and mitigation of limitations.

1.1 Problem Statement

Our problem statement for this project is 'To develop a system that effectively identifies and addresses individual mental health needs within the workplace. Aiming to enhance overall employee well-being and productivity'.

The problem statement underscores the critical need for a comprehensive system tailored to identify and address individual mental health needs within the workplace. In today's fast-paced tech industry, the relentless pursuit of innovation often overlooks the well-being of employees, leading to heightened levels of stress, burnout, and decreased productivity. By acknowledging and prioritising mental health within the workplace, organisations can foster a culture of support and understanding that not only enhances employee well-being but also boosts productivity and innovation. Thus, the development of such a system is paramount to creating a sustainable and thriving work environment where individuals feel valued, supported, and empowered to prioritise their mental health.

1.2 Motivation and Goal

Our motivation to address mental health in the tech industry stems from three key factors that underscore the urgency of this issue. Firstly, the recent wave of layoffs has shed light on the human impact of corporate restructuring, leaving many individuals grappling with uncertainty, financial strain, and increased stress levels.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Annual	11,560	12,930	15,580	19,170	14,720	10,730	10,690	26,110	8,020	6,440	14,320[P]
1st Qtr	2,120	3,110	3,500	4,710	4,000	2,320	3,230	3,220	2,270	1,320	3,820
2nd Qtr	3,080	2,410	3,250	4,800	3,640	3,030	2,320	8,130	2,340	830	3,200
3rd Qtr	2,710	3,500	3,460	4,220	3,400	2,860	2,470	9,120	1,900	1,300	4,110
4th Qtr	3,660	3,910	5,370	5,440	3,680	2,510	2,670	5,640	1,500	2,990	3,200[P]
Number of Retrenchments Per 1,000 Employees	5.8	6.3	7.4	8.9	7.0	5.1	5.1	12.8	4.4	3.1	6.6[P]

Figure 1: Singapore’s Retrenchment Table taken from MOM website

We can see that based on figure 1, the year 2023 is the second highest number of retrenchments per 1000 employees compared to the past 5 years.

Secondly, despite the industry's emphasis on innovation and progress, there remains a notable gap in mental health support mechanisms for employees. The lack of accessible resources and stigma surrounding mental health discussions contribute to a culture of silence and isolation, exacerbating the challenges faced by professionals. AON and TELUS Health have released their inaugural Asia Mental Health Index Report (2023) (Aon Group and TELUS Health, 2023), which surveyed workers across 12 locations in Asia. The data reveals that 82\% of workers in Asia have a high to medium mental health risk, with 45\% of respondents citing cost as the biggest barrier to accessing mental health support, with lack of information and knowledge about where to get help also a significant issue.

Lastly, there is a clear correlation between mental health and productivity. Studies have consistently shown that untreated mental health issues can significantly impact an individual's ability to perform at their best, leading to decreased efficiency, creativity, and overall job satisfaction. According to research paper “The Role of Mental Health on Workplace Productivity: A Critical Review of the Literature”, there was clear evidence that poor mental health (mostly measured as depression and/or anxiety) was associated with lost productivity (i.e., absenteeism and presenteeism) (de Oliveira et al., 2023). Stress, anxiety, and burnout were identified as key drivers of reduced productivity. By addressing these interconnected challenges head-on, we aim to create a more supportive and resilient tech industry where employees can thrive both personally and professionally.

This project aims to address this issue in **two ways**:

During the conduction of the project, a classification model will be developed that effectively identifies individuals with mental health issues. Since 43\% of respondents of the Asia Mental Health Index Report cite cost as the biggest barrier to accessing mental health support, providing a computer based, scalable solution for the identification of mental health issues will not only make the individual aware of its mental health state, but also make mental health diagnosis more accessible.

When trying to overcome an existing mental health disorder, another significant issue stated in the report was the lack of information and knowledge (Aon Group and TELUS Health, 2023).

Thus, the second part of the project will focus on addressing cultural factors contributing to the perseverance of mental health issues. By utilising topic modelling, this project aims to identify clusters of great interest to the workforce and provide recommendations on how to improve workplace conditions for the top 5 identified topics based on current literature.

Consequently, by following the recommendations, the mental health of the respective individual and thus the workforce as an entity shall be improved, which will potentially lead to an increase in productivity.

In addition, this project will contribute to the growing body of knowledge surrounding mental health issue solutions, aimed at promoting well-being and resilience among employees.

2. Literature Review

2.1 Deep Learning-based Natural Language Processing for Screening Psychiatric Patients (Hong et al., 2023)

As the diagnosis of mental disorders is often inaccurate due to symptoms being commonly shared among diagnoses, this study aimed to develop a classification model that could improve the accuracy rate of psychiatrist' diagnoses. It investigates the feasibility of using pre-trained language models and transfer learning in natural language processing (NLP) to analyse unstructured clinical text data from electronic health records (EHRs) to facilitate the process of patient screening in psychiatry. The researchers hypothesised that models relying on transferred knowledge outperform models who learn from scratch. Two experiments were conducted with a variety of configurations to study the effectiveness of fine-tuning BERT on a dataset of 500 randomly selected patients, whose data was labelled with one of five different diagnoses (major/minor depression, bipolar disorder, schizophrenia, and dementia). The performance of the BERT was compared to the performance of three text classification networks along with different pre-trained models.

Results showed that models without transferred knowledge were outperformed by models with pre-trained techniques by micro-avg. and macro-avg. F-scores of 0.11 and 0.28. Thus, we decided to use a pre-trained NLP model (RoBERTa) for the sentiment analysis of our dataset.

2.2 Machine Learning for Mental Health Detection (Gao et al., 2019)

This project aimed to develop a depression sensing application that leverages multi-modal data sources such as audio, text and GPS based data. The data was collected via an app named EMU generating participants from Amazon Mechanical Turk (MTurk) - a crowdsourcing website enabling businesses to hire "crowdworkers" to perform on-demand tasks (mTurk). The application collected demographics information, basic phone data, Google Maps location data, voice recordings, and many more. This dataset may be biased in terms of representing the general public, since the sample has a high degree of uncertainty regarding its attributes. However, the outcomes of this project remain valuable to us since we prioritise the broader insights gained and the methodologies employed rather than the implications the dataset may have for the general population. During the project, several advanced machine learning techniques were performed on the multi-modal data sources, including XGBoost, Adaptive Boosting, decision trees, logistic regression, artificial neural networks, Gaussian processes, and more. The results of the project indicated that text based data can best predict depression, as text based data had the highest F1 score on average. Based on this we decided to solely use text based data as input for our model.

2.3 Implementation of Machine Learning Model to Predict Heart Failure Disease (Alotaibi, 2019)

This study aimed to improve the performance of classifiers to make better use of medical databases and support heart failure disease detection and diagnosis by proposing a computerised system. It builds on a number of different researches using a variety of different machine learning techniques aiming to improve the accuracy of the different models. Five different models - Naïve Bayes, Random Forest, Support Vector Machine and Logistic regression - were leveraged to predict the chances of heart failure in a patient admitted in the hospitals.

The project had to overcome the problem of limited data resources: It leveraged a subset of the Heart Disease Dataset from Kaggle - the Cleveland Disease Dataset - with a number of 300 patients. The data was augmented to a volume increased by three times the size by applying random number generation technique using minimum and maximum values for each column. A total of 14 attributes were chosen for the experiment during the feature selection. Overall, the Decision Tree classifier computed the highest accuracy and the Naïve Bayes classifier the lowest. The positive enrichment of the accuracy may have been caused by the enlarged size of the dataset. A large size of the dataset enhances the learning process enhancing accuracy results. It was also suggested that the overall increased accuracy was caused by using 10-fold cross-validation, while previous work used 5-fold cross validation, since more iterations during the test phase may help to generate more accurate results.

Technique	(Alotaibi, 2019)	(Bashir et al., 2019)	(Ekiz and Erdomu, 2017) **Matlab**	(Ekiz and Erdomu, 2017) **Weka**
Decision Tree	93.19%	82.22%	60.9%	67.7%
Logistic Regression	87.36%	82.56%	65.3%	67.3%
Random Forest	89.14%	84.17%	X	X
Naive Bayes	87.27%	82.24%	X	X
SVM	92.30%	84.85%	67%	63.9%

Figure 2: Performance Comparison

2.4 Conclusions on Literature Review

After carefully reviewing the current literature, we gained valuable insights about the techniques we should adopt during this project and what kind of datasets we should focus on. We will utilise the most important techniques, such as CatBoost, XGBoost, Random Forest, and Neuronal Networks on a text based dataset to train our model. Needless to say, we will not limit ourselves solely to the discovered methods and models, but continue to explore other opportunities as well.

3. Datasets

3.1 OSMI Mental Health in Tech Survey

Link to dataset: https://osmhhelp.org/research.html

This dataset showcases a survey regarding perceived mental health culture as well as the individual mental health state of employees in the tech sector conducted from 2016 to 2023 by Open Sourcing Mental Illness - a non-profit organisation dedicated to promote mental wellness. The reason why we chose this dataset is because it allows us to do both prediction and classification, as well as explore natural language processing (NLP) models.

In the excel file, there is a large range of different data being provided, ranging from medical history over questions regarding the characterisation of the mental health culture at the workplace. The dataset contains over 120 and 3447 records, with a key feature being the question: “Do you have a mental health disorder?”. Over the years, several questions were added to the questionnaire and others dropped, resulting in inconsistency of the dataset. Since the dataset of 2016 just contains approximately 40% of the required inputs, we will focus on the data from 2017 to 2023. The cleaned dataset contains after feature selection a total of 1821 rows of records of answers to the OSMI Mental Health in Tech Survey with a total of 54 questions.

3.2 Exploratory Data Analysis (EDA)

From the dataset explained in section 3.1, we would need to further understand and extract out the data relevant to our project. Thus, we performed Exploratory Data Analysis (EDA) on the dataset. It helps to identify obvious errors, as well as provide better understanding of the patterns within the data, detect outliers or anomalous events, and find interesting relations among the variables. This was done to increase our understanding and knowledge of the dataset.

We also performed data cleaning on the dataset to remove dirty data like duplicates and missing values, which can affect the performance of our predictive models. This is because dirty data in the training data will result in incoherent results. Therefore, it is key for us to provide clean and coherent data to train our models to get the best results.

3.3 Merging datasets and data pre-processing

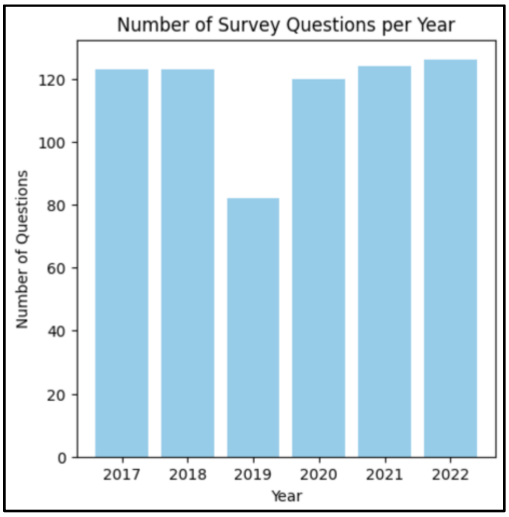

Figure 3: Number of survey questions per year

After our EDA and looking at figure 3, we can see that there is an inconsistency and variability in the number of survey questions per year. Hence, to combine 6 datasets into a single comprehensive dataset, we utilised the union method. Our approach involved appending each dataset sequentially, ensuring that all shared columns are aligned correctly across the datasets. This method is ideal for aggregating data that spans multiple years while maintaining a consistent structure throughout.

To facilitate this merging, we first renamed the column by removing unnecessary HTML tags using Regex and standardising it to lowercase. Then, we verify our cleaned column as seen from our figure 4. There are about 75 common columns among the 6 datasets.

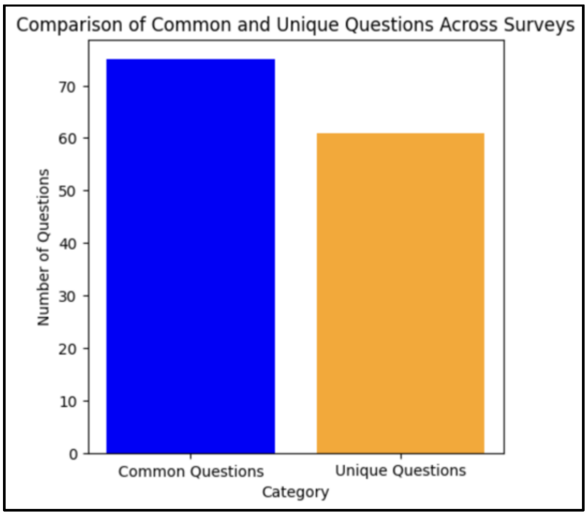

Figure 4: Comparison of common and unique questions

In the data pre-processing phase following the merging of datasets, we encountered the issue of missing values across various columns. To tackle this, we first eliminated irrelevant and redundant data by discarding columns with more than 70% of missing values. This approach ensures that our analysis is based on the most completed data available, thereby enhancing the reliability and validity of our results.

For columns that still contain missing values, we conducted a detailed inspection of each column's unique values and their frequencies to ensure response consistency. This initial analysis helped in understanding the distribution and commonality of responses, which informed our subsequent imputation methods. Our methods for addressing missing data included:

- **Filling it with Mode**: for questions that required respondents to rate something with a fixed range (e.g, "Overall, how much importance did your previous employer place on physical health?" with responses from 0 to 10), we filled in the missing values with the mode of the responses. This method is justified as it preserves most of the common responses, maintaining the integrity of the data's central tendency.
- **Setting Default Responses:** For subjective questions where the lack of a response might indicate uncertainty or neutrality, we used default responses. For example, for the questions "Would you have been willing to discuss your mental health with your

coworkers at previous employers?", we imputed missing values with "I don't know". This choice reflects a natural indecision that may be expected in the context of mental health discussions. Similarly for "Would you feel comfortable discussing a mental health issue with your coworkers?", we used "Maybe" as a filler, which appropriately captures the potential ambivalence.

- **Converting Binary Responses:** For binary questions (e.g, True/False), such as "did you ever discuss your mental health with your previous employer?", we converted the responses into numerical format (1s and 0s). This transformation simplifies the analysis, allowing statistical models to process the data more efficiently and with greater clarity.

These imputation methods were carefully chosen to minimise the impact of missing data on our analysis while respecting the nature of the questions and the distribution of observed responses. By thoughtfully addressing these gaps, we ensured that our dataset is not only more complete by also maintaining a high level of accuracy and relevance for reliable statistical analysis later on. We now proceed to train the models with our cleaned data.

4. Methodology

4.1 Evaluation Metrics utilised

Before we start training our models, we need to first select the possible metrics to evaluate our models for us to compare. We have chosen "Do you currently have a mental health disorder?" variable from our dataset as the golden truth label.

Multiple classification models were evaluated based on these evaluation criterias.

1. F1 Score
2. Precision and Recall
3. AUC-ROC
4. Accuracy

F1 Score is a harmonic mean of precision and recall. It is particularly useful in situations where we want to find a balance between precision and recall. The F1 score is defined as

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Precision measures the accuracy of positive predictions. Formally, it is the ratio of true positives to the sum of true and false positives. High precision relates to a low false positive rate, and is particularly important in scenarios where false positives are more costly than false negatives.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall measures the ability of a model to find all the relevant cases (all positive samples). High recall means that the system returned most of the positive results.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Area under the curve (AUC) of the Receiver Operating Characteristics (ROC) graph is an effective measure of the inherent trade-off between true positive rate and false positive rate. The ROC curve plots the true positive rate against the false positive rate at various threshold settings. The AUC provides a single measure of overall performance of the model and its ability to avoid false classification.

Accuracy is the simplest and most intuitive performance measure. It is the ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Observations}$$

Due to the limitations of accuracy, where it can be misleading in the context of imbalanced situations, where the number of observations in different classes varies significantly, we decided to place higher priority on F1 Score and AUC. It is the harmonic mean between precision and recall, so as to disincentivize models from prioritising the majority class

4.2 Pre-processing and feature selection

4.2.1 Feature Selection: Combination of 4 different methods with Ordinal Encoding (Correlation, Chi-2, RFE, ReliefF)

Before applying this feature selection method, we first encoded the categorical variables with ordinal encoding.

Correlation is a simple approach to find linear dependency between variables. It assesses the strength of a linear relationship between 2 variables. If correlation is close to -1/1, then 2 variables have strong linear dependency. If correlation is close to 0 , then the 2 variables have no linear dependency. We utilised Spearman correlation, which is a non-parametric measure of correlation. This means it assesses how well the relationship between 2 variables can be described using a monotonic function.

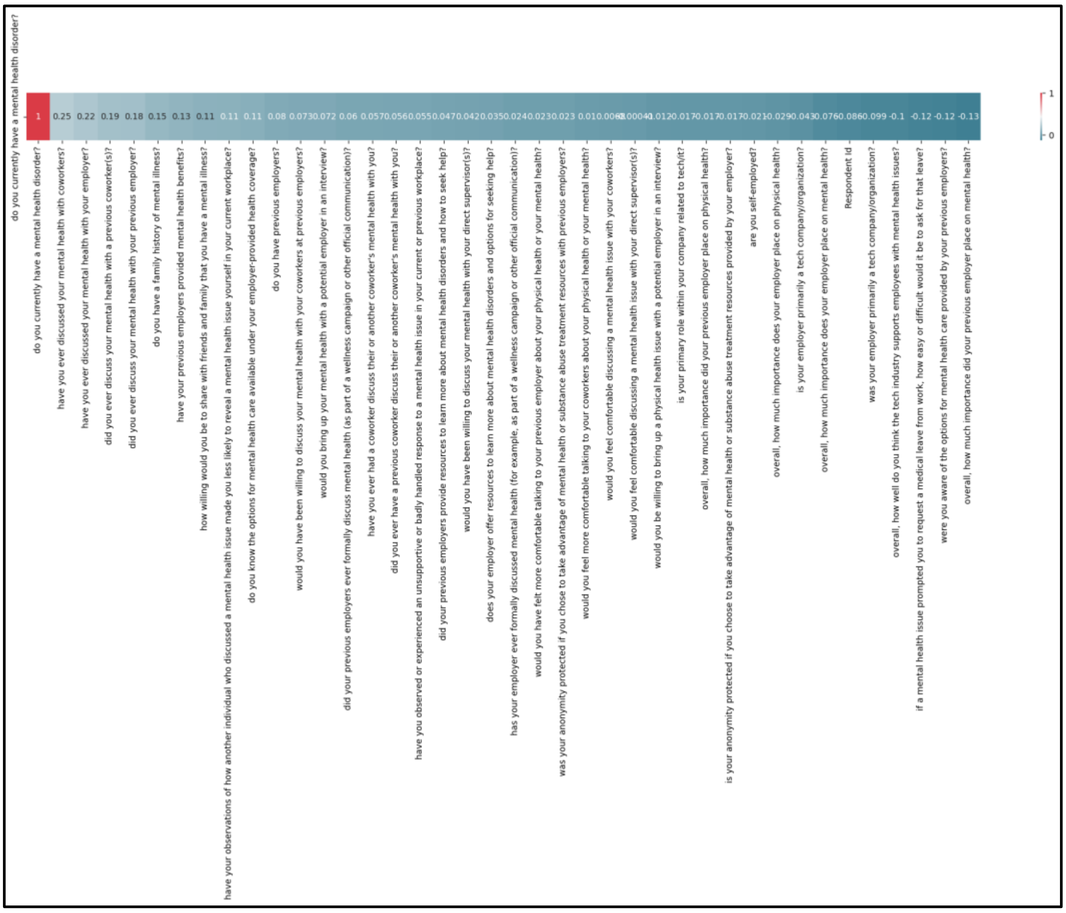

Figure 5: Correlation result of the variables vs Ground Truth

Chi-2 checks if there is a significant difference between the observed and expected frequencies of 2 categorical numbers. Thus, the null hypothesis that there is no relationship between the variables is tested.

$$x^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

Recursive Feature Elimination (RFE) removes the weakest features at each iteration considering the importance of features given by a machine learning model. Our chosen model here is linear regression. It continues recursively until the specified number of features is reached.

ReliefF gauges feature importance by comparing nearest neighbours from the same and different classes. Features that distinguish between classes effectively are considered more relevant for classification, enhancing their selection likelihood.

Combining each feature selection method's Top 10, a frequency bar chart is created to show which Top 10 features are commonly selected from the 4 methods.

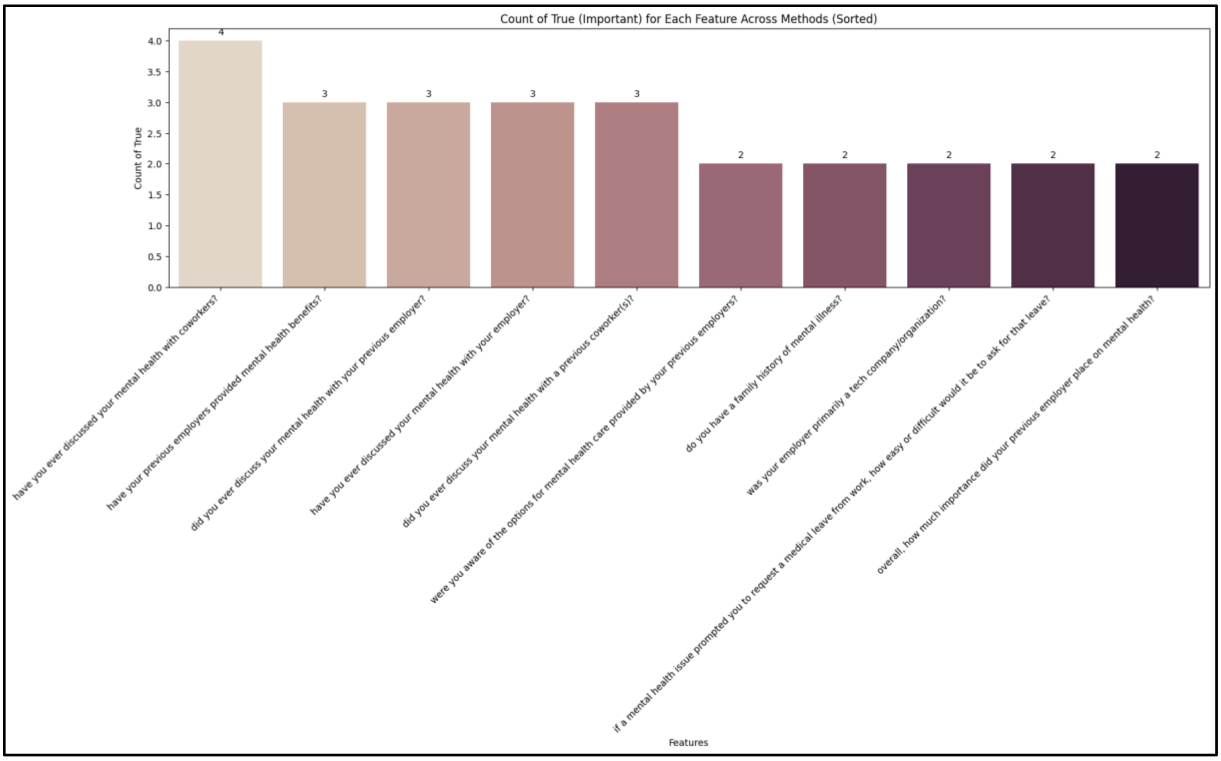

Figure 6: Top 10 features using combination of 4 methods

4.2.2 Feature Selection: Mutual Information with One hot encoding

Mutual information is a non-parametric method that measures the dependency between 2 variables indicating how much knowing of these variables reduces uncertainty about the other. It can detect any kind of relationship, linear or nonlinear. In essence, it quantifies the amount of shared information between the feature and the target variable, helping to identify features that are most informative for prediction.

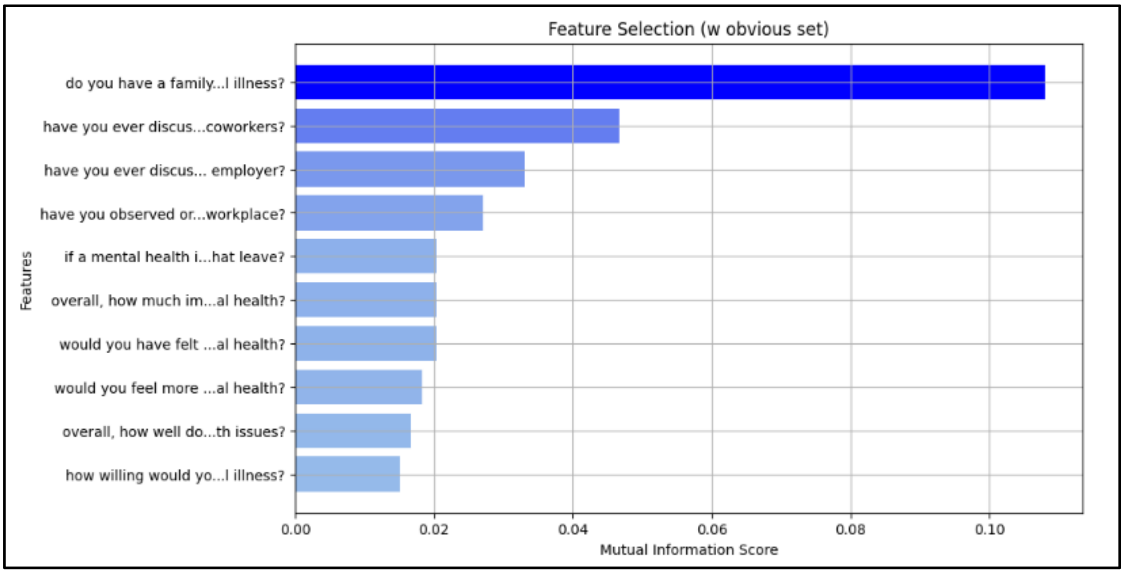

Figure 7: Top 10 features using mutual information

4.2.3 Feature Selection: Genetic Algorithm with Ordinal Encoding

Genetic algorithms are adaptive heuristic search algorithms premised on the evolutionary idea of natural selection and genetics. They are used to solve optimization and search problems by iteratively selecting, combining and mutating candidate solutions based on their measured fitness.

Based on the feature selections explored above, our group has chosen to go with genetic algorithm, as it has the highest F1 Score produced using the Catboost Model (seen in figure 6).

Model / Feature selection	**Combination of 4**	**Mutual Information**	**Genetic Algorithm**
Catboost	0.905	0.901	0.909

Figure 8: F1 Score of different feature selection methods

4.3 Classification Models

Taking into account the other classification models that we identified during literature review, we continued exploring to see which type of models can be better. As part of our overall methodology, we first train the models, followed by tuning of hyperparameters using various libraries and tools to perform cross-validation, which can help us find the optimal performance of the models we trained. Following which we then compare all the trained models based on the results of the test data. We use test data because it provides us an unbiased evaluation of a final model fit on the training data set.

We attempted several different models, including boosting, trees and neural networks. The model trained include:

- RoBERTa
- Deep Neural Network
- Decision Tree
- Random Forest
- Catboost
- XGboost

Using the 3 subsets of features, we applied them to 6 different supervised classification models.

Natural Language Processing	**Deep Learning models**	**Tree-based**
RoBERTa (Robustly Optimised BERT Approach)	DNN (Deep Neural Network)	Decision Tree Random Forest Catboost XGBoost

Figure 9: Classifier Methods

4.3.1 RoBERTa (Robustly Optimised BERT Approach)

A variant of the BERT transformer model, RoBERTa is a pre-trained NLP model which dynamically masks portions of each input across different training epochs, allowing the model to learn the relationship between features and limit overfitting. PyTorch was utilised for faster computation, and stratified trained test split was used to get the median evaluation metrics for a more accurate scoring of the model's performance. HyperOpt hyperparameter tuning and utils.EarlyStopping() acted as a form of regularisation to avoid overfitting, by finding a balance between bias and variance, and stopping the fold's training process early if the model's performance doesn't improve for a certain number of epochs respectively.

```
def objective(params):
    # Create an instance of TweetModel
    model_config = transformers.RobertaConfig.from_pretrained(ROBERTA_PATH)
    model_config.output_hidden_states = True

    model = TweetModel(model_config)
    model.to(device)

    # Create the optimizer and scheduler
    optimizer = AdamW(model.parameters(), lr=params['lr'], weight_decay=params['weight_decay'])
    total_steps = len(test_data_loader) * params['epochs']
    scheduler = get_linear_schedule_with_warmup(optimizer, num_warmup_steps=0, num_training_steps=
total_steps)

    # Train the model
    train_fn(test_data_loader, model, optimizer, device, scheduler)

    validation_error, _ = model.validate(test_data_loader, loss_function)
    # Avoid error of iterating over 0-d tensor regularization_term = params['weight_decay'] * sum
([p.pow(2.0).sum() for p in model.parameters()])
    regularization_term = params['weight_decay'] * sum([p.pow(2.0).sum().item() for p in model.par
ameters()])
    return validation_error + regularization_term

space = {
    "lr": hp.loguniform("lr", np.log(1e-6), np.log(1e-4)),
    "batch_size": hp.choice("batch_size", [16, 24, 32]),
    "epochs": hp.choice("epochs", [10, 20, 30, 40, 50]),
    "max_seq_len": hp.choice("max_seq_len", [128, 160, 192, 224]),
    "weight_decay": hp.loguniform("weight_decay", np.log(1e-5), np.log(1e-3))  # Adjust the range
here
}

best = fmin(objective, space, algo=tpe.suggest, max_evals=200) # 100 default
print(best)
```

Figure 10: Code for RoBERTa

	F1-score	Precision	Recall	AUC-RUC	Accuracy
With Genetic Algorithm	0.31	0.52	0.22	0.55	0.51

Figure 11: F1 Results of RoBERTa + HyperOpt

```python
batch_size = batch_size_space[best_params['batch_size']]
epochs = epochs_space[best_params['epochs']]
max_seq_len = max_seq_len_space[best_params['max_seq_len']]
lr = best_params['lr']
weight_decay = best_params['weight_decay']

print("Best parameters:")
print('batch = ' + str(batch_size))
print('epochs = ' + str(epochs))
print('max_seq_len = ' + str(max_seq_len))
print('lr = ' + str(lr))
print('weight_decay = ' + str(weight_decay))
```

```
Best parameters:
batch = 24
epochs = 30
max_seq_len = 192
lr = 1.580763709956422e-06
weight_decay = 1.0003457471559566e-05
```

Figure 12: Best set of hyperparameters (HyperOpt)

As a baseline, a batch size of 32, 192 max character length, 5 epochs of training, weight decay of 0.001 and lr of 5e-5 was used. Hyperopt was used for hyperparameter tuning and 30 epochs with a smaller batch of 24, alongside a lower learning rate (1.581e-6) and weight decay (1.000e-05) yielded slightly better results.

4.3.2 Deep Neural Network (DNN)

Deep Neural Network (DNN) is a type of Artificial Neural Network with multiple layers that mimic the human brain's ability to pick up a specific detail to build a more complex understanding of the input. The model is a fully connected network (aka a dense network) implemented using PyTorch for efficient computation. StandardScaler and Principal Component Analysis was used for feature standardisation and dimensionality reduction, while a ReLU activation function allows for more complex, nonlinear relationships where f(x + y) != f(x) + f(y). The optimizer used is the Adam optimizer.

```python
# Neural network architecture
class ANN(nn.Module):
    def __init__(self):
        super().__init__()
        self.fc1 = nn.Linear(in_features=20, out_features=32)
        self.fc2 = nn.Linear(in_features=32, out_features=32)
        self.fc3 = nn.Linear(in_features=32, out_features=8)
        self.output = nn.Linear(in_features=8, out_features=2)
        self.output2 = nn.Softmax(dim=1)

    def forward(self, x):
        x = F.relu(self.fc1(x))
        x = F.relu(self.fc2(x))
        x = F.relu(self.fc3(x))
        x = self.output(x)
        x = self.output2(x)
        return x

model = ANN()
criterion = nn.CrossEntropyLoss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.0001)
```

Figure 13: Code for DNN

	F1-score	**Precision**	**Recall**	**AUC-RUC**	**Accuracy**
With Genetic Algorithm	0.91	0.90	0.91	0.89	0.93

Figure 14: Results of DNN

4.3.3 Decision Tree

Decision Trees are notable for their versatility in effectively handling continuous variables in regression tasks, categorical variables in classification tasks and other object classes (non-numerical data) through the use of encoding on the data features. This flexibility makes it particularly useful for a wide range of machine learning applications. It involves repeatedly splitting the dataset into two or more subsets, with the goal of finding homogeneous data. A perfect split is graded using the Gini score of 0, while the worst case split that results in 50/50 classes in each group results in a Gini score of 0.5 (for a 2 class problem). A local greedy approach to minimising Gini coefficient was taken for the sake of computation time and efficiency, to allow for GridSearchCV hyper-tuning.

To further improve the performance of the decision tree, we tuned the hyperparameters using GridSearchCV to determine the optimal values.

```
from sklearn.model_selection import GridSearchCV
param_grid = {
    'max_depth': [3, 5, 7, 10],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'max_features': ['sqrt', 'log2', None]
}

[ ] grid_search = GridSearchCV(estimator=dt, param_grid=param_grid, cv=5, scoring='accuracy')
    grid_search.fit(X_train, y_train)


    best_params = grid_search.best_params_
    best_estimator = grid_search.best_estimator_


    best_model_score = grid_search.best_score_

    print("Best Parameters:", best_params)

    Best Parameters: {'max_depth': 3, 'max_features': None, 'min_samples_leaf': 1, 'min_samples_split': 2}
```

Figure 15: Best set of hyperparameters with GridSearchCV

	F1-score	Precision	Recall	AUC-RUC	Accuracy
With Genetic Algorithm	0.82	0.79	0.86	0.72	0.76

Figure 16: Results of Decision Tree + GridSearchCV

4.3.4 Random Forest

Random Forest utilises an ensemble of Decision Trees to aggregate the opinions of many individual models to improve predictions and create a more generalisable and robust overall model. Each tree in the ensemble is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. When splitting a node, the chosen best split is taken amongst a random subset of the features to add further randomness, so as to decrease the overall variance and thus risk of overfitting in the model.

```
def find_best_n_estimators(X, y, max_estimators=300, cv=5):
    best_score = 0
    best_n_estimators = 0

    for n_estimators in range(1, max_estimators + 1):
        rf = RandomForestClassifier(n_estimators=n_estimators, random_state=42)
        scores = cross_val_score(rf, X, y, cv=cv)
        avg_score = scores.mean()

        if avg_score > best_score:
            best_score = avg_score
            best_n_estimators = n_estimators

    return best_n_estimators

# Example usage:
# Assuming X and y are your features and target variable, respectively
best_n_estimators1 = find_best_n_estimators(X1_train, y1_train)
print("Best n_estimators:", best_n_estimators1)
```

Figure 17: Best n_estimators tuning

	F1-score	Precision	Recall	AUC-RUC	Accuracy
With Genetic Algorithm	0.76	0.70	0.84	0.80	0.70

Figure 18: Results of Random Forest

4.3.5 Catboost and XGBoost

Building upon the concept of decision tree ensembles, XGBoost (Extreme Gradient Boosting) and Catboost use a gradient boosting framework that works by having each tree be built sequentially, instead of parallel, where each new tree aims to correct the errors made by the previous trees. This allows multiple weak "learners" or models to build upon each other's mistakes to create a strong learner. This is done by using the gradient (the direction and rate of fastest increase of a mathematical function) of a loss function.

XGBoost key features includes applying a regularised model formalisation to control over-fitting, which improves its performances. It is designed to be computationally fast and efficient with the core algorithm based on gradient boosted decision trees designed for speed and performance. XGboost handles missing values by learning which direction to assign missing values based on reduction in loss.

To further enhance the performance of XGBoost, we employed sophisticated hyperparameter tuning methods, namely GridSearchCV and Hyperpot.

```
# Define the hyperparameter grid
param_grid = {
    'max_depth': [3, 5, 7],
    'learning_rate': [0.1, 0.01, 0.001],
    'subsample': [0.5, 0.7, 1],
    'reg_alpha': [0, 0.1, 0.5], #regulation parameter
    'reg_lambda': [0, 0.1, 0.5] #regulation parameter
}

# Create the XGBoost model object
xgb_model = xgb.XGBClassifier()

# Create the GridSearchCV object
grid_search = GridSearchCV(xgb_model, param_grid, cv=5, scoring='accuracy')

# Fit the GridSearchCV object to the training data
grid_search.fit(X_train, y_train)

# Print the best set of hyperparameters and the corresponding score
print("Best set of hyperparameters: ", grid_search.best_params_)
print("Best score: ", grid_search.best_score_)
```

```
Best set of hyperparameters:  {'learning_rate': 0.1, 'max_depth': 3, 'reg_alpha': 0.5, 'reg_lambda': 0.1, 'subsample': 0.7}
Best score:  0.7715948741701405
```

Figure 19: Best set of hyperparameters (GridSearchCV)

	F1-score	Precision	Recall	AUC-RUC	Accuracy
With Genetic Algorithm	0.81	0.75	0.88	0.77	0.74

Figure 20: Results of XGBoost + GridSearchCV

```
The best hyperparameters are :

{'colsample_bytree': 0.5963923562079627, 'gamma': 3.809832629419107, 'max_depth': 4.0, 'min_child_weight': 1.0,

'reg_alpha': 119.0, 'reg_lambda': 0.615900428169187}
```

Figure 21: Best set of hyperparameters (HyperOpt)

	F1-score	Precision	Recall	AUC-RUC	Accuracy
With Genetic Algorithm	0.77	0.63	1.00	0.65	0.63

Figure 22: Results of XGBoost + HyperOpt

Comparing both hyperparameter tuning methods, we obtained a higher F1-score using the GridSearchCV method.

On the other hand, **Catboost** stands out by providing support for categorical input variables. It converts categories into numbers using an optimal method based on the dataset to maximise the speed and accuracy. It uses order boosting, a permutation driven alternative to the classic boosting method, which reduces overfitting significantly. It automatically deals with categorical variables and does not require any extension data pre-processing like one-hot encoding which is needed by XGBoost.

```
clf = CatBoostClassifier(cat_features=[var for var in X_train.columns if X_train[var].dtype == "O"])
clf.fit(X_train, y_train)
preds = clf.predict(X_test)
proba = clf.predict_proba(X_test)
```

Figure 23: Code for CatBoost

	F1-score	Precision	Recall	AUC-RUC	Accuracy
With Genetic Algorithm	0.91	0.90	0.91	0.90	0.93

Figure 24: Results of CatBoost

4.4 Clustering Model

In our analysis, we opted to implement Latent Dirichlet Allocation (LDA) using Gensim to perform topic modelling, specifically targeting the extraction of prevalent themes from textual data within our dataset. This methodological choice was grounded in extensive online research and general consensus on the efficacy of LDA for identifying latent themes in large text corpora.

4.4.1 LDA Gensim Topic Modeling

The process began with the comprehensive training of the LDA model on the prepared corpus, which included steps such as tokenization, stop-word removal, and normalisation to transform raw text into a suitable format for analysis (i.e., the bag-of-words model).

	Tokenized Brief Description
0	[take, seriously]
1	[raise, awareness, talk, lessen, stigma]
2	[education, awareness, statistic, add, supportive, write, company, handbook]
3	[support, less, burnout, death, march]
4	[think, tech, internal, really, talk, feeling, issue]
5	[awareness, change, work, schedule, expectation]
6	[vocal, support, employee, issue, support, mental, health, time]
7	[think, work, expect, everyone, connect, time, human, meant, work, degree, need, cut, slack]
9	[ok]
10	[open, talk, open, accommodation, stop, push, worker, toward, burnout]

Figure 25: Results of text preprocessing

Using Gensim’s LdaMulticore, which leverages multicore processors to speed up computation, was pivotal in managing the extensive computations involved efficiently as it reduces the model training time.

```python
def compute_evaluation_values(corpus, dictionary, k, texts, random_state, chunksize, passes, per_word_topics, alphas, etas):
    results = []
    for alpha in alphas:
        for eta in etas:
            lda_model = gensim.models.LdaMulticore(corpus=corpus,
                id2word=dictionary,
                num_topics=k,
                random_state=random_state,
                chunksize=chunksize,
                passes=passes,
                alpha=alpha,
                eta=eta,
                per_word_topics=per_word_topics)

            for coherence_model_type in ['c_v']:
                coherence_model_lda = CoherenceModel(model=lda_model,
                texts=texts, dictionary=dictionary, coherence=coherence_model_type)
                coherence = coherence_model_lda.get_coherence()
                perplexity = lda_model.log_perplexity(corpus)
                results.append((k, alpha, eta, coherence_model_type, coherence, perplexity))

    # Convert results to DataFrame for easier analysis
    results_df = pd.DataFrame(results, columns=['Num Topics', 'Alpha', 'Eta', 'Coherence Model', 'Coherence', 'Perplexity'])
    return results_df
```

Figure 26: Gensim’s LdaMulticore configuration

Critical hyperparameters such as the number of topics, alpha (document-topic density), and eta (word-topic density) were finely tuned. We utilised a grid search strategy over these parameters to enhance the coherence score of the model, ensuring the extracted topics were both meaningful and distinct.

```
# Parameters setup
random_state = 100
chunksize = 100
passes = 40
per_word_topics = True
alphas = ['symmetric', 'asymmetric', 0.01]
etas = ['auto', 0.01, 0.1]
texts = briefly_df2['Tokenised Content'].tolist()
```

Figure 27: Set of hyperparameters

Following the setup, the model underwent a series of iterations where each document was analysed to deduce the distribution of words across the potential topics, refining the topics' definition with each pass through the data. To identify the most suitable number of topics for our model, we systematically evaluated the coherence scores across a range of topic numbers (from 2 to 19). This evaluation helped in selecting the optimal topic count that offers the highest coherence while maintaining manageable model complexity.

Best Model Configuration:

- The coherence score peaked at **0.481722** with **5 topics**, indicating this as the optimal number for our analysis.
- Configurations used were '**Alpha: asymmetric**' and '**Eta: auto**'.

```
Best Model Configuration:
Num Topics                  5
Alpha              asymmetric
Eta                      auto
Coherence Model           c_v
Coherence            0.481722
Perplexity          -4.782422
```

Figure 28: Best model configuration

Once the LDA model was finely tuned and iteratively passed through the data, it successfully outputted the word distributions for each of the defined topics. The process generated detailed word probabilities per topic, which were instrumental in understanding the thematic structure of the corpus. Each topic's list of words and their corresponding weights (probabilities) indicates the terms most representative of the topic, providing a numeric measure of word relevance to the topic's context.

Then, to convert these probabilistic topic models into interpretable themes, we consulted ChatGPT with the top words from each topic. ChatGPT assisted in synthesising these words into coherent themes, encapsulating the essence of the discussions captured by each topic. For instance, words like 'talk', 'open', 'awareness', 'support', and 'stigma' led to the theme 'Creating a Supportive Culture', indicating a focus on fostering an environment conducive to open discussions on mental health.

Topic Number	Word Distributions	Themes
0	0.060*"talk" + 0.051*"open" + 0.039*"awareness" + 0.037*"support" + 0.035*"stigma" + 0.029*"work" + 0.026*"offer" + 0.026*"less" + 0.026*"provide" + 0.022*"people"	Creating a Supportive Culture
1	0.065*"people" + 0.053*"work" + 0.050*"industry" + 0.043*"think" + 0.042*"tech" + 0.036*"mental" + 0.030*"health" + 0.023*"issue" + 0.023*"would" + 0.022*"know"	Industry-Specific Challenges and Opportunities
2	0.091*"health" + 0.078*"mental" + 0.064*"make" + 0.039*"time" + 0.029*"take" + 0.028*"encourage" + 0.027*"employee" + 0.022*"openly" + 0.022*"insurance" + 0.018*"first"	Promoting Mental Health Practices
3	0.197*"health" + 0.162*"mental" + 0.099*"issue" + 0.042*"work" + 0.037*"physical" + 0.034*"take" + 0.028*"day" + 0.023*"well" + 0.020*"support" + 0.020*"team"	Integrating Mental and Physical Health
4	0.086*"employee" + 0.065*"mental" + 0.054*"health" + 0.046*"like" + 0.038*"support" + 0.038*"issue" + 0.037*"employer" + 0.035*"understand" + 0.032*"company" + 0.029*"well"	Employer-Employee Dynamics in Mental Health Support

Figure 29: Themes to each topic distributions

5. Results and Discussions

5.1 Model Evaluation and Comparison

5.1.1 Classification Methods

Performance metrics based on F1 Score (above), ROC/AUC Curve (below). Overall, CatBoost with a genetic algorithm subset performed the best

F1 SCORE AUC-ROC	**XGBoost**	**RoBERTa**	**Decision Tree**	**CatBoost**	**DNN**	**Random Forest**
Combination of 4	0.837 0.870	0.655 0.690	0.750 0.663	0.905 0.915	0.884 0.858	0.773 0.818
Mutual Information	0.820 0.800	0.663 0.606	0.820 0.727	0.901 0.902	0.884 0.865	0.775 0.818
Genetic algorithm	0.739 0.810	0.306 0.545	0.820 0.718	0.909 0.905	0.908 0.888	0.764 0.802

Figure 30: Overview summary of classifier results

5.1.2 Clustering Method

To visually assess the distribution of topics within the dataset, we plotted the percentage of documents corresponding to each dominant topic. This analysis was pivotal for understanding the prevalence and influence of each topic across the corpus. Using a bar chart, we visualised these percentages, offering a straightforward comparison of topic dominance that highlights which topics are most pervasive and impactful in the dataset.

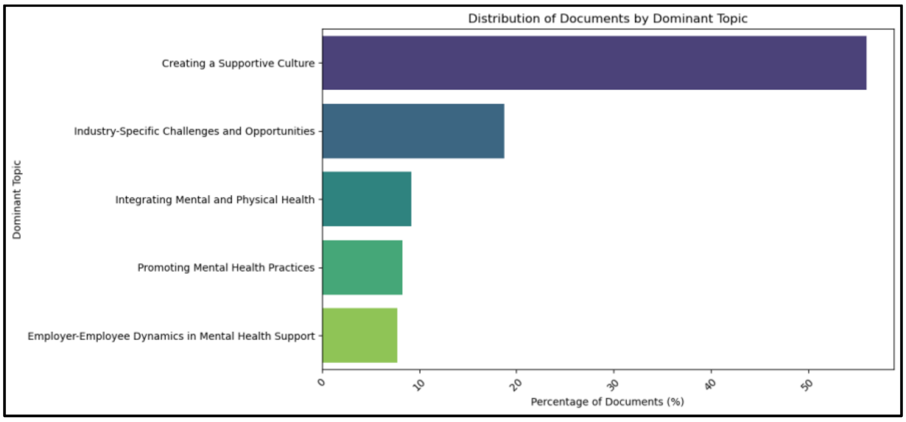

Figure 31: Distribution of documents by themes

Further, to explore the separation and distinctiveness of the topics derived from the LDA model, we utilised PyLDAvis, a tool designed for interactive visualisation of topics derived from topic models. The Intertopic Distance Map (IDM) provided by PyLDAvis illustrates the relationship and distance between the topics, offering insights into how distinct or overlapping the topics are.

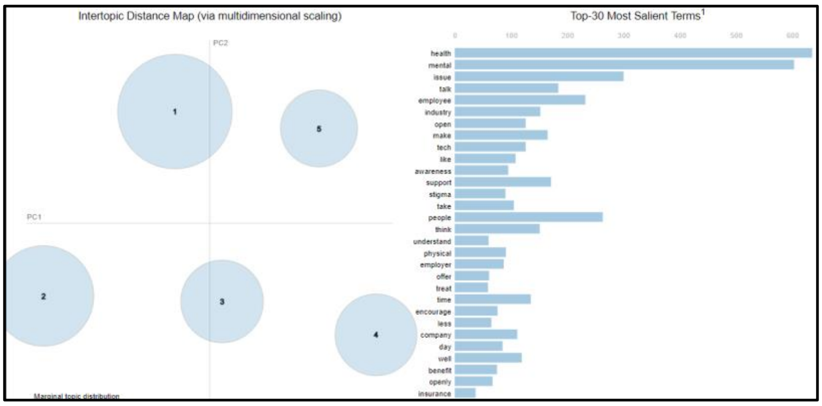

Figure 32: Intertopic distance map for 5 topics

5.2 Recommendations

Based on the results from Topic modelling as well as our classification model findings, 5 different topics of interest were identified:

1. Creating a Supportive Culture
2. Industry-Specific Challenges and Opportunities
3. Promoting Mental Health Practices
4. Integrating Mental and Physical Health
5. Employer-Employee Dynamics in Mental Health Support

This project aims to address mental health issues at the workplace in a holistic way. Thus, we provide recommendations on how to address the identified topics within an organisation based on the comprehensive literature review "Organisational Best Practices Supporting Mental Health in the Workplace" (Wu et al., 2021).

5.2.1 Creating a Supportive Culture

One of the major topics identified was to create a supportive culture within the workplace. The fact that 54% of respondents of the Asia Mental Health Index report (Aon Group and TELUS Health, 2023) would be concerned about career options being limited if they had a mental health issue their employer knew about confirms that suffering from mental health issues is still stigmatised in the Asian work culture. But since behavioural patterns within the company will influence the social and physical environment of the workplace, the impact of cultural culture on the mental health of the individual may not be underestimated. In this way, establishing a strong mental health culture can make a difference to the awareness, utilisation and impact of programme efforts to build a mentally healthy workforce. In order to implement a work culture with a positive attitude towards the detection, treatment and de-stigmatization of mental health issues, it is recommended to build processes and policies into the organisation's human capital strategy. This may be achieved by e.g. programs and/or policies to prevent sexual harassment, discrimination, workplace violence, and bullying/incivility. Complementing this with the promotion of psychological well-being and educating information about mental illnesses and risk factors that contribute to their development in organisational communications and materials will help in creating a corporate culture supportive of mental health issues (Uddin et al., 2022) sugged for example to put up posters such as "Having a mental illness is not a shame", "having a mental illness does not mean you are weak", and "employees can still strive with managed mental illness" to increase positive thinking and encourage coping with mental illness.

5.2.2 Industry-Specific Challenges and Opportunities

The dataset of this project is working with is anchored within the tech industry, which is why when we talk about industry-specific challenges, we will focus on the work conditions of the tech sector. According to WHO report "Promoting Mental Health" (World Health Organisation, Department of Mental Health and Substance Abuse, 2005), the tech industry is characterised by high stress, long hours of labour, work pressure, and the workforce is driven by the desire to make a name for themselves. In order to combat stress and burnout, the organisation may want to promote relaxation and place an emphasis on work-life balance:

Offering flexible work hours, remote work options, and encouraging employees to take breaks and vacations are measures to accomplish this.

Additionally, managers must monitor the workload of employees closely to prevent employees from being overwhelmed. Prioritising tasks, setting realistic deadlines, and avoiding overloading individuals with excessive responsibilities may help as well.

5.2.3 Promoting Mental Health Practices

Employers might also want to actively promote mental health practices:

This includes employer-sponsored mental health benefits, such as a health plan providing affordable access to a broad range of mental health services, with accessibility considering the quality of care, availability of an adequate provider network, and the reasonableness of any service limits.

Additionally, stress management practices, and mental health training that addresses organisational issues causing stress may reduce physical and psycho-social stressors in the work environment.

Furthermore, an Employee Assistant Programm (EAP) addressing the needs of a diverse workforce population may positively impact the mental health of the workforce and reduce stigma. An Employee Assistance Program is a workplace program designed to provide support and resources to employees facing personal or work-related challenges that may impact their well-being and job performance (Singapore Counselling Centre). This can include mental health support, counselling services, stress management resources, and assistance with work-life balance issues. Overall, it was found that EAPs enhance employee outcomes, specifically improving levels of presenteeism and functioning (Beulah Joseph and Fuller-Tyszkiewicz, 2018).

5.2.4 Integrating Mental and Physical Health

Physical activity, exercise, and physical-activity interventions were found to be beneficial across several mental-health outcomes. Generally, participants engaging in regular physical activity display more desirable health outcomes across a variety of conditions (Penedo and Dahn, 2025).

Consequentially, when talking about mental health, physical health must be included into the discussion. Therefore recommendations for physical health at the workplace are part of a holistic mental health strategy of every organisation.

This may be achieved by implementing infrastructure that supports healthy behaviours and self-care. To be precise, offering healthy food options, fitness facilities, employee benefits for physical exercise and integrating physical activity into the office architecture such as open stairwells foster physical health.

This may be complemented by campaigns raising awareness to physical health, instructions on how to set up your workplace as well as offering facilities to mentally recharge such as quiet rooms.

5.2.5 Employer-Employee Dynamics in Mental Health Support

Focusing on an individual level of social-workplace dynamics, the relationship between employer and employee shall be discussed as well. Since the supervisor acts as principal of the organisation in the workplace, the recommendations will focus on leadership and management-skills:

The organisation may want to improve leadership by training management on mental health awareness, sensitivity and supporting a healthy work environment. Communication and crisis management seminars may complement these efforts - with a focus on providing support to employees and instructions on how to respond to workplace bullying, violence and suicide.

Modelling healthy workplace behaviours such as e.g. appropriate behaviour during conflict may also be the focus of awareness training.

Next to training, efficient communication of performance expectations, regular feedback sessions, and finding a healthy balance of autonomy and close supervision are also important factors when it comes to employer-employee dynamics.

6. Limitations and future works

6.1 Limitations

These limitations highlight the challenges and potential sources of bias that need to be considered when interpreting the results of this study:

6.1.1 Lack of Data Consistency

A significant limitation of this study was the necessity to remove certain columns from the dataset. This was primarily due to the varying number of questions asked each year and inconsistent naming conventions across different survey periods. As a result, some potentially valuable data points had to be excluded from the analysis, which may have impacted the comprehensiveness of our findings. The data collection process faced challenges related to the lack of pre-defined possible values for certain fields. This led to the presence of numerous values that essentially represented the same original value, resulting in data redundancy and potential inaccuracies in our analysis.

6.1.2 Lack of Awareness on Mental Health Terms

Another notable limitation was the inconsistency in responses within the dataset. This inconsistency was often attributed to responders' lack of awareness regarding the true definition of certain mental health terms, such as distinguishing between mental health issues and mental health disorders. Mental health disorders refer to more permanent issues such as PTSD or depression, which have been diagnosed while mental health issues do not have to be diagnosed and may refer to more temporary feelings that we may have at times such as periods of fear, hopelessness and sadness. Such discrepancies in understanding could have introduced bias and affected the reliability of our results.

6.2 Future Works

6.2.1 Additional Data Sources

Future research endeavours should focus on implementing more robust data collection methodologies. This includes establishing pre-defined possible values for fields, standardising question formats across survey periods, and ensuring clear definitions for mental health terms.

With more data, both in terms of the number of samples in the dataset and the number of features, we can enhance the model's accuracy and robustness. Increasing the number of survey respondents will contribute to a richer dataset, allowing for a more comprehensive analysis and better prediction outcomes. Future work could also involve integrating data from diverse sources such as employee performance metrics, organisational policies on mental health support, and qualitative feedback from employees regarding their experiences. Since the dataset this project is based on was collected solely using the individual of the workforce as a source, more objective metrics of performance and/or mental health from third parties might offer additional insights.

6.2.2 Classification pipeline for linking, diagnosing and providing recommendations

Future research may focus on integrating the model's output with a Language Model (LM) such as ChatGPT to facilitate the generation of personalised recommendations. This automated process can offer initial guidance and support to individuals based on their predicted mental health needs and creates an automated pipeline for classification, diagnosis, and providing recommendations with minimal human interaction. It would also be possible to think of refining the model further to be able to predict more specific mental health disorders.

6.2.3 Tune the model to a specific demographic

To improve the model's accuracy and reliability, fine-tuning it for specific demographics such as country, age-group, gender, income-level, etc., will help eliminate potential confounding variables. This targeted approach ensures that the model's predictions are tailored to the unique characteristics of different population groups.

6.2.4 Longitudinal analysis

Finally conducting longitudinal analyses over multiple years might capture temporal trends and changes in mental health perceptions and behaviours, as well as the long-term effect on the mental health of the workforce of a computer based recommendation system like it was suggested in this study. This longitudinal approach would enable researchers to assess the effectiveness of the recommendations over time.

7. Conclusion

In conclusion, our research and findings have shown that people working in the tech industry not only suffer more from negative mental health conditions, but are also more reluctant to speak out about them. This coupled with a lack of support and resources (either self sourced or provided for by the organisation), and a “pressure-cooker” like working environment causes a great deal of stress and burnout amongst tech employees.

Our listed recommendations try to target factors that contribute to this stress, as well as reduce the stigma associated with negative mental health conditions, and are derived from the most common responses from employee remarks in the data set. However, it must also be stated that this list is not exhaustive and much more can be done to combat the rise of these conditions in the workplace.

8. References

Aon Group, TELUS Health. (2023). *Asia Mental Health Index Report.* Aon Group.

Amazon Mechanical Turk overview. https://www.mturk.com/. Accessed: 2024-03-22.

A. W. Beulah Joseph and M. Fuller-Tyszkiewicz. Evaluating the effectiveness of employee assistance programmes: a systematic review. *European Journal of Work and Organizational Psychology*, 27(1):1–15, 2018. doi: 10.1080/1359432X.2 017.1374245. URL https://doi.org/10.1080/1359432X.2017.1374245.

A. Wu, E. C. Roemer, K. B. Kent, D. W. Ballard, and R. Z. Goetzel. Organizational best practices supporting mental health in the workplace. *Journal of Occupational and Environmental Medicine*, 63(12):e925–e931, 2021. doi: 10.1097/JOM.000000 0000002407. URL https://doi.org/10.1097/JOM.0000000000002407.

C. de Oliveira, M. Saka, L. Bone, and R. Jacobs. The role of mental health on workplace productivity: A critical review of the literature. *Appl Health Econ Health Policy*, pages 167– 193, 2023.

F. Alotaibi. Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, 10, 01 2019. doi: 10.14569/IJACSA.2019.0100637.

F. J. Penedo and b. Dahn, Jason Ra. Exercise and well-being: A review of mental and physical health benefits associated with physical activity. *Current Opinion in Psychiatry*, 18(2): 189–193, March 2005.

Harvey SB, Modini M, Joyce S, et al. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and Environmental Medicine*, 301-310.

J. Hong, H. Chu, Q. You, C. You, K. Chen, J. Chian, and S. Chi. Deep learning-based natural language processing for screening psychiatric patients. *Frontiers in Psychiatry* , 2021.

M. M. Uddin, A. Farjana, M. Mamun, and M. Mamun. Mental health analysis in tech workplace. *In Proceedings of the 7th North American International Conference on Industrial Engineering and Operations Management*, pages 12–14, 2022.

S. Bashir, Z. S. Khan, F. Hassan Khan, A. Anjum, and K. Bashir. Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pages 619–623, 2019. doi: 10.1109/IBCAST.2019.8667106.

S. Ekiz and P. Erdomu. Comparative study of heart disease classification. *2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pages 1–4, 2017. URL https://api.semanticscholar.org/CorpusID:45872697 .

Singapore Counselling Centre. Employee Assistance Programme. https://scc.sg/e/employee-assistance-progr amme/. Accessed: 01.04.2024.

World Health Organization, Department of Mental Health and Substance Abuse, Victorian Health Promotion Foundation. *Promoting Mental Health.* 2005.

Y. Gao, M. Flannery, J. Assan, Y. Wu, and A. Resom. Machine learning for mental health detection, 2019. Medicine, 63(12):e925–e931, 2021. doi: 10.1097/JOM.000000 0000002407. URL https://doi.org/10.1097/JOM.0000000000002407.