Public Health Awareness

Phase 4 Submission Document

Name	Subhalakshmi.V
Reg. No	410121104050
NM ID	au410121104050
Department	CSE-III
Domain	Data Analytics with Cognos
Project Title	Public Health Awareness
Phase 4	Development Part III
College	4101-Adhi College of Engineering and Technology, Kanchipuram

Public Health Awareness

Introduction:-

Mental health issues affect approximately 700 million people worldwide, accounting for about 13% of all diseases. Depression, for instance, is the second leading cause of disability, trailing only behind back pain. The primary mental health conditions are depression and anxiety, rather than schizophrenia. There is evidence to show that individuals with mental health problems may be denied employment due to their mental condition or may not seek employment because they are aware of potential discrimination.

Disclosing a mental health problem in the workplace can lead to discriminatory behaviours from supervisors and colleagues, such as social exclusion or hindering these individuals' career progression. A framework for understanding these behaviours conceptualizes stigma as comprising three issues:-

- Knowledge (ignorance or misinformation)
- Attitudes (prejudice)
- Behaviour (discrimination)

In a study conducted by Manning and Whit, the factors most commonly considered when hiring a person include the previous work record (89%), job description (87%), whether they received treatment (69%), the time they were ill the previous year (68%), and the diagnosis (64%). Fenton et al. also concluded that the employment record (78%), health record (69%), diagnosis (36%), detection under the Mental Health Act (36%), and medical opinion (7%) are important factors in hiring someone. Krupa highlighted four underlying assumptions about workplace stigma:-

1. People with mental health issues do not have the necessary skills to meet job requirements.
2. People with mental health issues are dangerous or unpredictable.
3. Working is not healthy for people with mental health problems.
4. Employing people with mental health issues is an act of charity.

These assumptions vary in intensity based on a range of organizational, individual, and social factors.

It is important to emphasize the significance of a positive work environment for enhancing the economic and social integration of individuals with mental health issues.

Dataset:-

This dataset is from a 2014 survey that measures attitudes toward mental health and the prevalence of mental health disorders in the technological workplace. The survey was conducted with 1,260 individuals from various countries, and the top 10 participating countries can be seen in Figure 1.

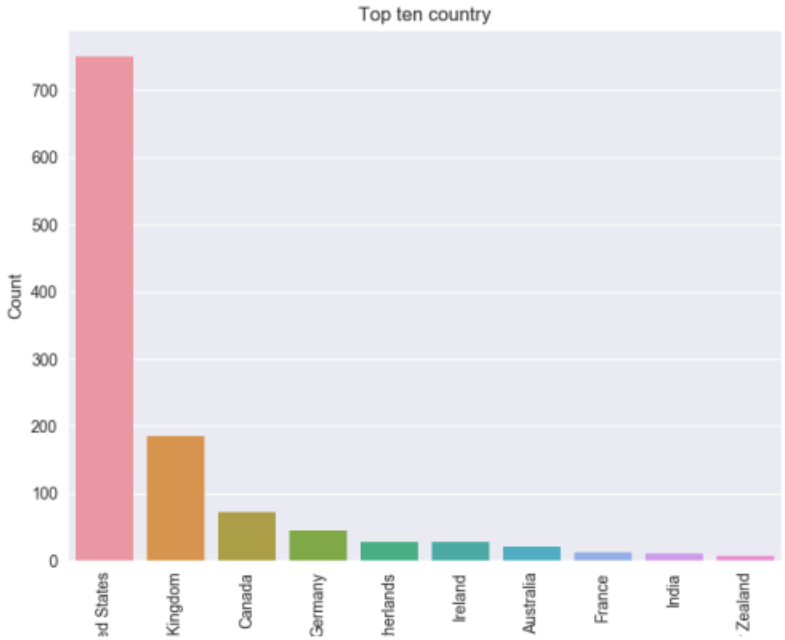


Figure 1: Top 10 Participating Countries in the Survey

The dataset contains the following information:

- Timestamp
- Age
- Gender
- Country

- State: If the person resides in the United States, in which state or territory they live.

- Self-employed: Whether the person is self-employed.

- Family history: Whether the person has a family history of mental health issues.

Overview of the Process:

1.Issue Identification: Recognizing specific health concerns, from disease outbreaks to lifestyle factors, requiring public attention.

2.Education and Information: Disseminating accurate health-related information, statistics, and facts to increase awareness and understanding.

3.Behavioral Promotion: Encouraging the adoption of healthy behaviors, such as exercise, vaccination, smoking cessation, and healthy eating.

4.Risk Factors and Prevention: Highlighting risk factors associated with health issues and advocating preventive measures.

5.Crisis Response: Playing a critical role during health emergencies by informing the public and advising on protective measures.

6.Community Engagement: Collaborating with communities, healthcare professionals, and organizations to address health challenges.

7.Communication Channels: Utilizing various media and platforms to reach diverse audiences effectively.

8.Cultural Sensitivity: Tailoring messages to respect cultural values and traditions, ensuring they resonate with the audience.

9.Advocacy and Partnerships: Mobilizing advocates and forming partnerships to amplify the message and inspire behavior change.

10.Policy Advocacy: Advocating for policy changes and regulations to address health issues at the societal level.

11.Monitoring and Evaluation: Continuously assessing the campaign's impact through feedback, behavior change, and health indicators.

12.Resource Allocation: Securing funding and resources, often through grants, donations, or partnerships, to support awareness efforts.

13.Legal and Ethical Considerations: Adhering to ethical standards and legal requirements when handling health information and promoting interventions.

14.Adaptability: Staying flexible and responsive to evolving situations, emerging health threats, and new research.

15.Sustainability: Maintaining long-term efforts for sustainable public health improvements, addressing immediate and persistent health challenges.

Procedure:

Issue Identification:

- Research and analyze health issues.
- Prioritize based on impact and urgency.

Education and Information:

- Develop accurate, clear content.
- Utilize various media channels.
- Collaborate with healthcare professionals.

Behavioral Promotion:

- Create behavior-change messages.
- Provide actionable steps for individuals.
- Use behavioral science principles.

Risk Factors and Prevention:

- Identify risk factors.
- Communicate prevention strategies.
- Provide resources for preventive measures.

Crisis Response:

- Establish a crisis communication plan.

- Use reliable sources for real-time information.
- Instruct the public on protective measures.

Community Engagement:

- Involve the community in decision-making.
- Establish community forums and meetings.
- Collaborate with local leaders and organizations.

Communication Channels:

- Select channels based on the audience.
- Adapt content to channels.
- Monitor channel reach and impact.

Cultural Sensitivity:

- Understand cultural values and beliefs.
- Adapt content to cultural norms.
- Seek community input for sensitivity.

Advocacy and Partnerships:

- Identify advocates and partners.
- Foster relationships and collaborations.
- Utilize their influence for promotion.

Policy Advocacy:

- Identify relevant policies.
- Engage in advocacy efforts.
- Collaborate with policymakers and advocacy groups.

Monitoring and Evaluation:

- Collect data and feedback.
- Analyze data to measure impact.
- Adjust strategies based on evaluation.

Resource Allocation:

- Develop a campaign budget.
- Seek funding through grants, partnerships, or donations.
- Allocate resources efficiently.

Legal and Ethical Considerations:

- Comply with health information laws.
- Ensure informed consent.
- Maintain transparency.

Adaptability:

- Stay informed about evolving health issues.
- Adjust strategies in response to new information.
- Continuously evaluate message relevance.

Sustainability:

- Develop a long-term strategy.
- Build partnerships and community support.
- Secure resources for sustained awareness activities.

Feature Selection:

Although it is possible to determine the most relevant features through attribute visualization, it is necessary to confirm this using algorithms designed for that purpose. Therefore, feature selection/extraction methods need to be applied. However, since this is a highly nominal problem, the loss of information and meaning of the attributes is not desired.

As a result, we limited this section to feature selection. In other words, out of the 24 features included in the dataset, we will select only the most relevant ones for the classification process.

Using the "Select Attributes" option in Weka, we were able to carry out this process. The methods used were InfoGainAttributeEval and CorrelationAttributeEval. Both methods perform feature selection functions, as discussed earlier. The results obtained for each method were as follows:

```
Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 24 treatment):
        Information Gain Ranking Filter

Ranked attributes:
 0.397725    6 work_interfere
 0.10595     5 family_history
 0.054045   11 care_options
 0.049512    3 Country
 0.036411   10 benefits
 0.026255    2 Gender
 0.016574   23 obs_consequence
 0.015829   15 leave
 0.014591   14 anonymity
 0.010575   16 mental_health_consequence
 0.009235   22 mental_vs_physical
 0.008329   20 mental_health_interview
 0.005861   12 wellness_program
 0.005766   13 seek_help
 0.005199    7 no_employees
 0.003844   18 coworkers
 0.002534   21 phys_health_interview
 0.001072   17 phys_health_consequence
 0.000826   19 supervisor
 0.000781    9 tech_company
 0.000523    8 remote_work
 0.000197    4 self_employed
 0           1 Age

Selected attributes: 6,5,11,3,10,2,23,15,14,16,22,20,12,13,7,18,21,17,19,9,8,4,1 : 23
```

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 24 treatment):
        Correlation Ranking Filter
Ranked attributes:
 0.3772    5 family_history
 0.3615    6 work_interfere
 0.1874   11 care_options
 0.1834    2 Gender
 0.1499   23 obs_consequence
 0.1464   10 benefits
 0.1328   14 anonymity
 0.0833   20 mental_health_interview
 0.0768   16 mental_health_consequence
 0.0743   22 mental_vs_physical
 0.0737    1 Age
 0.0668    3 Country
 0.0619   15 leave
 0.0415   13 seek_help
 0.0399   12 wellness_program
 0.0392   21 phys_health_interview
 0.0352   17 phys_health_consequence
 0.0333    7 no_employees
 0.0329    9 tech_company
 0.0269    8 remote_work
 0.0265   18 coworkers
 0.0244   19 supervisor
 0.0165    4 self_employed

Selected attributes: 5,6,11,2,23,10,14,20,16,22,1,3,15,13,12,21,17,7,9,8,18,19,4 : 23
```

Both methods return a value for each attribute. The higher this value, the greater the impact that this feature has on the classification process. As expected, the attributes that perform best are those previously identified solely through the visualization of their distribution.

It is now possible to exclude attributes with insignificant importance in classification. Therefore, we chose to eliminate "self_employed," "supervisor," "tech_company," and "remote_work." We are now proceeding with only 19 features.

Classification:

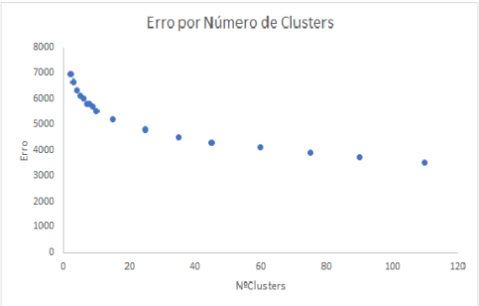
In this section, the following classifiers were used: Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbours (KNN). For Random Forest and KNN, a parameter study was conducted to maximize performance.

Support Vector Machine:

Program:

```
indexes = np.random.rand(len(newData)) < 0.7

train = newData[indexes]

test = newData[~indexes]

targetVector = newData.treatment #No - 0, Yes - 1

classifier1 = svm.SVC()

classifier1.fit(train, train.treatment)

predictions1 = classifier1.predict(test)

tn, fp, fn, tp = sk.metrics.confusion_matrix(test.treatment, predictions1).ravel()

accuracy = (tp + tn) / (tp + tn + fn + fp)

sensitivity = tp / (tp + fn)

specificity = tn / (tn + fp)

print('SVM')

print('Accuracy: ', accuracy, '\nSensitivity: ', sensitivity, '\nSpecificity: ',
specificity)

print('\nConfusion Matrix:\n',sk.metrics.confusion_matrix(test.treatment,
predictions1))
```

Output:



Random Forest classifier:

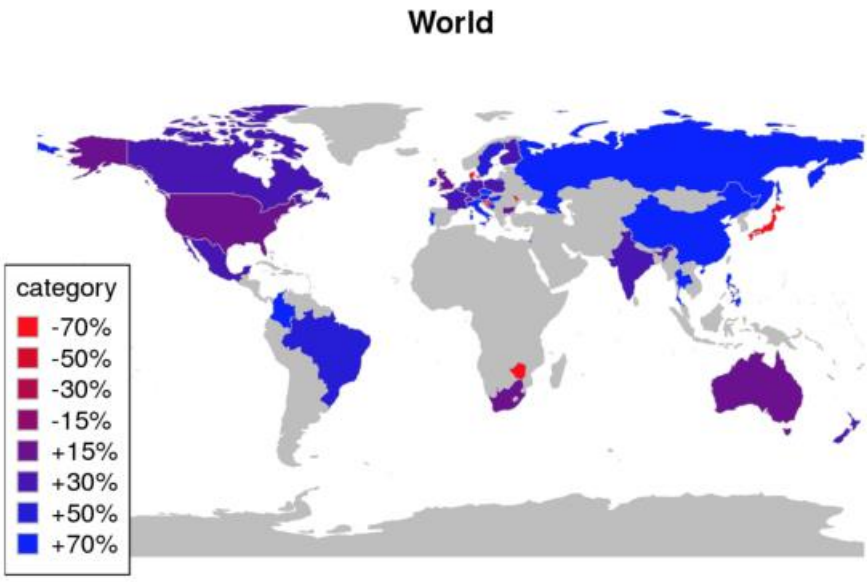
In the Random Forest classifier, the number of branches in the classifier and the minimum number of samples required to be in a node were varied. Initially, the

classifier presented accuracy, sensitivity, and specificity of 1.0. In order to avoid overfitting, the number of branches in the classifier and the minimum number of samples in a node were increased, but it was in vain.

Program:

```
clf = RandomForestClassifier(n_estimators = 200, oob_score = True, n_jobs = -1, random_state =50,max_features = "auto", min_samples_leaf = 100)

clf.fit(train, train.treatment)

preds = clf.predict(test)

tn1, fp1, fn1, tp1 = sk.metrics.confusion_matrix(test.treatment, preds).ravel()

accuracy1 = (tp1 + tn1) / (tp1 + tn1 + fn1 + fp1)

sensitivity1 = tp1 / (tp1 + fn1)

specificity1 = tn1 / (tn1 + fp1)

print('\nRANDOM Forest')

print('Accuracy: ', accuracy1, '\nSensitivity: ', sensitivity1, '\nSpecificity: ', specificity1)

print('\nConfusion Matrix:\n',sk.metrics.confusion_matrix(test.treatment, preds))
```

Output:



KNN Classifier:

In the KNN classifier, the number of neighbors is the most important factor to optimize. Through a loop, classification started with 1 neighbor up to the same number of samples in the training group. From Figure 15, it can be seen that the number of neighbors that maximizes the classifier's performance is 42, achieving an accuracy of 75.3%.

Program:

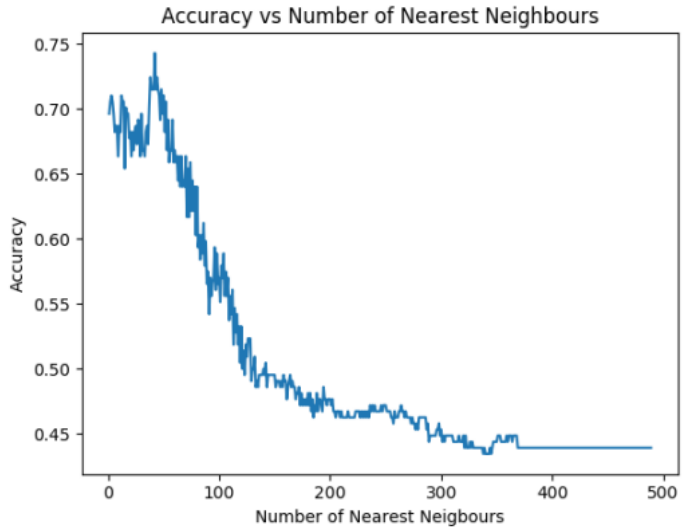
```
print('\nKNN')

listAccuracy = []

listNeighbors = []

for x in range(1, len(train)):

    clf = neighbors.KNeighborsClassifier(x)

    knn_model = clf.fit(train, train.treatment)

    preds_KNN = clf.predict(test)

    tn2, fp2, fn2, tp2 = sk.metrics.confusion_matrix(test.treatment,
preds_KNN).ravel()

    accuracy2 = (tp2 + tn2) / (tp2 + tn2 + fn2 + fp2)

    sensitivity2 = tp2 / (tp2 + fn2)

    specificity2 = tn2 / (tn2 + fp2)

    listAccuracy.append(accuracy2)

    listNeighbors.append(x)

plt.figure(1)

plt.title('Accuracy vs Number of Nearest Neighbours')

plt.plot(listNeighbors, listAccuracy)

plt.xlabel('Number of Nearest Neigbours')

plt.ylabel('Accuracy')

plt.show()

print('Max Accuracy: ', max(listAccuracy), '\nNumber of Neighbours: ',
listNeighbors[listAccuracy.index(max(listAccuracy))])
```

Output:

Attribute	Full Data (813.0)	Cluster# 0 (257.0)	1 (214.0)	2 (154.0)	3 (188.0)
Age	31.9176	30.3658	33.5327	32.5779	31.6596
Gender	Male	Male	Male	Male	Male
Country	United States	United States	United States	United States	United States
self_employed	No	No	No	No	No
family_history	No	No	Yes	No	Yes
work_interfere	Sometimes	NA	Sometimes	NA	Sometimes
no_employees	More than 1000	More than 1000	More than 1000	6-25	26-100
remote_work	No	No	No	Yes	No
tech_company	Yes	Yes	Yes	Yes	Yes
benefits	Yes	Dont know	Yes	No	No
care_options	No	No	Yes	No	No
wellness_program	No	No	Yes	No	No
seek_help	No	Dont know	Yes	No	No
anonymity	Dont know	Dont know	Yes	Dont know	Dont know
leave	Dont know	Dont know	Dont know	Dont know	Somewhat easy
mental_health_consequence	No	Maybe	No	Yes	Yes
phys_health_consequence	No	No	No	No	No
coworkers	Some of them	Some of them	Some of them	No	Some of them
supervisor	Yes	Yes	Yes	No	Some of them
mental_health_interview	No	No	No	No	No
phys_health_interview	Maybe	Maybe	No	Maybe	Maybe
mental_vs_physical	Dont know	Dont know	Yes	Dont know	No
obs_consequence	No	No	No	No	No
treatment	Yes	No	Yes	No	Yes

```
clf = neighbors.KNeighborsClassifier(listNeighbors[listAccuracy.index(max(listAccuracy))])

knn_model = clf.fit(train, train.treatment)

preds_KNN = clf.predict(test)

tn2, fp2, fn2, tp2 = sk.metrics.confusion_matrix(test.treatment, preds_KNN).ravel()

accuracy2 = (tp2 + tn2) / (tp2 + tn2 + fn2 + fp2)

sensitivity2 = tp2 / (tp2 + fn2)

specificity2 = tn2 / (tn2 + fp2)

print('\nFinal Values: \nAccuracy: ', accuracy2, '\nSensitivity: ', sensitivity2, '\nSpecificity: ', specificity2)

print('\nConfusion Matrix:\n',sk.metrics.confusion_matrix(test.treatment, preds_KNN))
```

Output:



Conclusion:

In conclusion, our public health awareness project is a vital step towards improving the well-being of our community. By identifying key health issues, delivering targeted education and information, promoting positive behaviors, and fostering community engagement, we aim to make a meaningful impact on public health. We've recognized the significance of cultural sensitivity, ethical considerations, and adaptability to create a comprehensive and effective campaign. Through collaboration, advocacy, and responsible resource allocation, our project strives for long-term sustainability. As we move forward, we remain committed to the well-being of our community, addressing health challenges, and promoting a healthier, more informed society.

*****