

Predicting Workers Cyber-Resilience Through Mental Health Risk Modeling in Tech Industries

Advisor: Dr. Faisal Quader

ABSTRACT

This research investigates the severe effect of mental health difficulties such as anxiety, depression, and fatigue on people's productivity, decision-making skills, and overall job performance. We look into how these concerns can increase the possibility of errors, overlooking critical facts, and engaging in risky activities among employees, weakening cybersecurity security measures. Our research looks at the impact of mental health on the human and organizational, on understanding its implications for cyber threats and attacks. We aim that this investigation can bring insight on the connections of mental health and cybersecurity, as well as provide ideas into how to build a healthier and more secure workplace.

PROBLEM STATEMENT

Employee mental health and a company's ability to withstand cyberattacks are closely related in today's linked business environment. Employees who are experiencing high levels of stress, depression and anxiety may make poor decisions or security precautions, thus allowing an opportunity for cybercriminals to take advantage of. The goal of this research is to measure how employee mental health affects security measures and create plans to support companies against new and emerging online threats. By addressing mental health issues, we hope to strengthen businesses' defenses against cyberattacks and promote a more secure workplace for all.

IMPLEMENTATION

Step-1 Data Cleaning :

The data obtained is cleaned by removing duplicate rows and unwanted columns. Null value rows are removed.

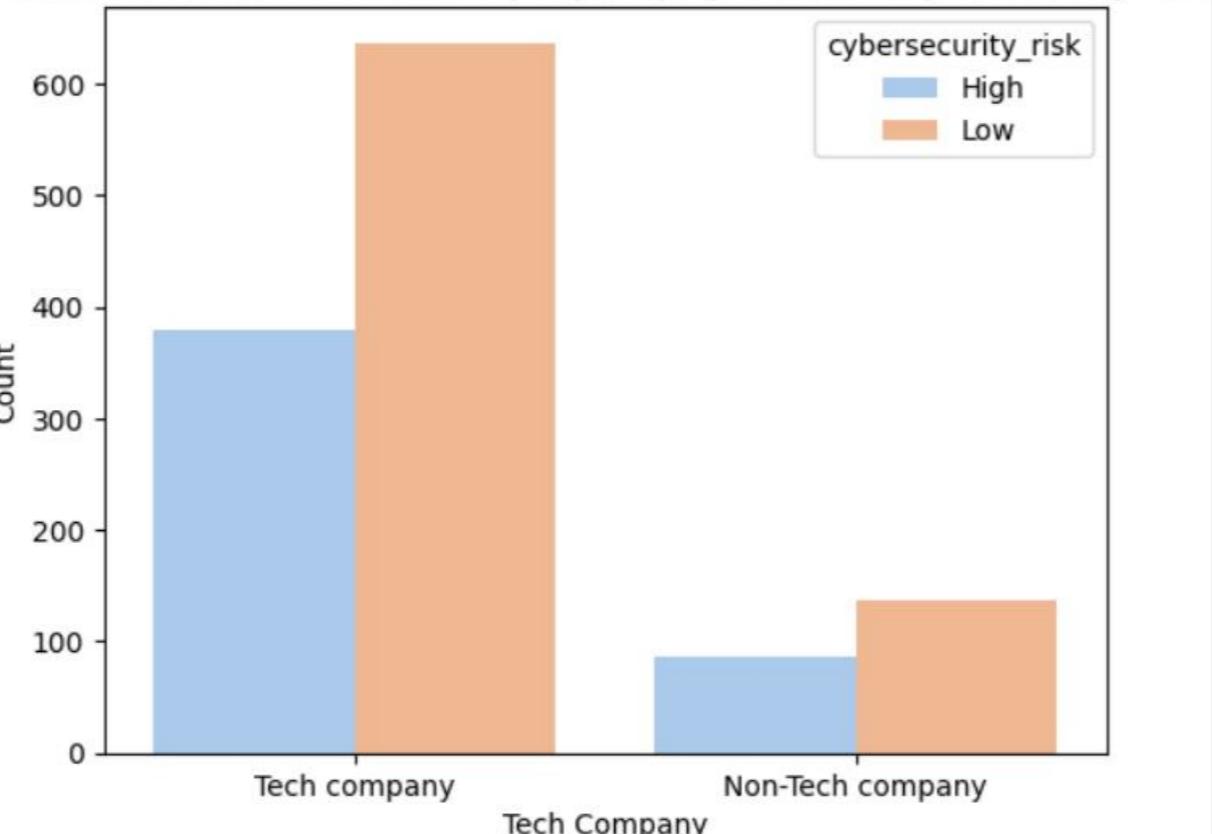
Step-2 Feature Engineering :

Established target attribute "cybersecurity_risk" by combining "treatment" and "work_interfere" attributes, which together resemble the same.

Step-3 Data Exploration :

Explored data with various numpy and pandas functions to learn more about dataset.

Correlation between Tech Company Employment and Cybersecurity Risk



Step-4 Label Encoding :

All the non-numerical columns are then label encoded to 0 or 1.

Step-5 Split the Dataset :

All the data is here divided into training and testing data, where trained data is 80% and test data is 20%.

Step-6 Testing :

Based on the problem statement and dataset we are working on it is a supervised learning as there are inputs with different labels and a target label as well. As we will be predicting a class label, hence we shall proceed with classification algorithms.

Logistic Regression :

```
from sklearn.linear_model import LogisticRegression
x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.2, random_state=42)
log_reg = LogisticRegression()
log_reg.fit(x_train, y_train)
log_reg_pred = log_reg.predict(x_test)
log_reg_accuracy = accuracy_score(y_test, log_reg_pred)
print("Logistic Regression Accuracy:", log_reg_accuracy)
```

Logistic Regression Accuracy: 0.8503401360544217

Support Vector Machine:

```
from sklearn.svm import SVC
x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.2, random_state=42)
svm = SVC()
svm.fit(x_train, y_train)
svm_pred = svm.predict(x_test)
svm_accuracy = accuracy_score(y_test, svm_pred)
print("Support Vector Machine Accuracy:", svm_accuracy)
```

Support Vector Machine Accuracy: 0.6530612244897959

Random Forest Classifier :

```
from sklearn.ensemble import RandomForestClassifier
x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.2, random_state=42)
rfc = RandomForestClassifier()
rfc.fit(x_train, y_train)
rfc_pred = rfc.predict(x_test)
rfc_accuracy = accuracy_score(y_test, rfc_pred)
print("Random Forest Classifier Accuracy:", rfc_accuracy)
```

Random Forest Classifier Accuracy: 0.9727891156462585

CONCLUSION

Our project aimed to analyze the intersection of mental health and cybersecurity within the tech industry workforce. Through comprehensive data analysis and machine learning modeling, we sought to identify patterns, predict cybersecurity risks, and provide insights to support employee well-being and data security measures.

Based on the various accuracy achieved from different classification algorithms, we can conclude that random forest classifier was the best, but has the possibility of the occurrence of over-fitting. Hence followed by the logistic regression would give us better results with its moderate to better accuracy.

FUTURE WORK

There are several areas of future work that can help to address the issue of mental health in tech space:

1. Organizations should monitor their employees by using AI enabled detection system such that, their inappropriate activities should trigger an alert.
2. Employees need to be educated and made more aware of the dangerous cyberattacks that can occur without their consciousness..