ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

ANALYSIS ON MENTAL HEALTH ISSUES AND PREDICTING THE RELAPSE RISK OF SURVIVORS WHO SURVIVED DRUG ADDICTION

PROJECT REPORT

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF SCIENCE IN COMPUTER APPLICATIONS [TRIPLE MAIN]

Submitted By

GOWRI TS

III BSc Computer Applications [Triple Main]

Register No: SB21CA011

Under the Guidance of

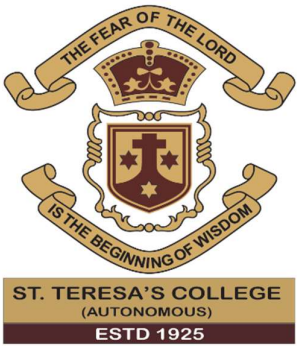
Ms. MARY ANDREWS

Department of Computer Application

2021-2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY



ANALYSIS ON MENTAL HEALTH ISSUES AND PREDICTING THE RELAPSE RISK OF SURVIVORS WHO SURVIVED DRUG ADDICTION

PROJECT REPORT

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF SCIENCE IN COMPUTER APPLICATIONS [TRIPLE MAIN]

Submitted By

GOWRI TS

III BSc Computer Applications [Triple Main]

Register No: SB21CA011

Under the Guidance of

Ms. MARY ANDREWS

Department of Computer Application

2021-2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

CERTIFICATE

This is to certify that the project entitled "**ANALYSIS ON MENTAL HEALTH ISSUES AND PREDICTING THE RELAPSE RISK OF SURVIVORS WHO SURVIVED DRUG ADDICTION**", is a bona-fide record of the work done by GOWRI TS (Register No:SB21CA011) during the year 2023-24 and submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Science in Computer Applications (Triple Main) under Mahatma Gandhi University.

Head of the Department

ST. TERESA'S COLLEGE DEPT. OF COMPUTER SCIENCE ERNAKULAM

For, 21/03/2024

Internal Examiner

22.03.24

External Examiner

Date: 21/03/2024

DECLARATION

I, GOWRI TS (Register no: SB21CA011), B.Sc. Computer Applications [Triple Main] final year student of St. Teresa's College (Autonomous), Ernakulam, hereby declare that the project submitted named "**ANALYSIS ON MENTAL HEALTH ISSUES AND PREDICTING THE RELAPSE RISK OF SURVIVORS WHO SURVIVED DRUG ADDICTION**" for the Bachelor's Degree in Computer Applications [Triple Main] is my original work. I further declare that the said work has not previously been submitted to any other university or academic body.

Gowri

GOWRI TS

Place: ~~ERNAKULAM~~ ERNAKULAM

Date: ~~21/10/24~~ 21/10/24

ACKNOWLEDGMENT

First and foremost, I thank God almighty for his blessings. I take this opportunity to express my gratitude to all those who helped me in completing this project successfully.

I would like to convey my heartfelt gratitude to **Provincial Superior and Manager Rev. Dr. Sr. Vinitha (CSST), Rev. Sr. Emeline (CSST) and Principal Dr. Alphonsa Vijaya Joseph** for providing me with this wonderful opportunity to work on a project with the topic **ANALYSIS ON MENTAL HEALTH ISSUES AND PREDICTING THE RELAPSE RISK OF SURVIVORS WHO SURVIVED DRUG ADDICTION**.

I would like to express my profound gratitude to the Head of the Department of Computer Applications **Ms. Remya C J**, my project guide **Ms. Mary Andrews** and all other faculty of the department of computer applications for their contributions to the completion of my project. The completion of the project would not have been possible without their help and insights.

My deepest appreciation goes to all my family and friends for their support.

GOWRI TS

ABSTRACT

"Survivors of drug addiction" refers to individuals who have successfully overcome or managed their addiction to drugs. These individuals have typically gone through a challenging process of recovery, which often involves various forms of treatment, therapy, support groups, and lifestyle changes. Survivors of drug addiction have effectively broken free from the cycle of substance abuse and have taken steps to regain control of their lives. They may continue to face challenges and temptations, but they have developed coping strategies and skills to maintain their sobriety and lead healthier, more fulfilling lives. Through this study we can identify that what are the major mental health issues that are faced by the survivors and also, we can predict the relapse risk. So that this we help in early diagnosis.

CONTENTS

1.INTRODUCTION

1.1 ABOUT PROJECT

In today's world, we're learning more about how addiction and mental health are closely connected. People who have survived drug addiction often face tough mental health challenges as they try to move forward. This project looks into these challenges to understand them better. We want to see how past addiction affects mental well-being. By studying this, we hope to find ways to support survivors better on their journey to feeling better overall. In this study, we can analyse and identify the major mental health issues of survivors who survived of drug addiction.

1.2 OBJECTIVE OF PROJECT

The objective of this study is to analyse the mental health issues faced by individuals who have survived drug addiction. By examining the psychological challenges experienced by these individuals, we aim to identify patterns, triggers, and coping mechanisms that influence their mental well-being post-recovery. Additionally, this study seeks to explore the impact of addiction history on various aspects of mental health, such as anxiety, depression, trauma, and self-esteem. Ultimately, the findings will contribute to a deeper understanding of the mental health needs of survivors of drug addiction and inform the development of targeted interventions and support strategies to promote their long-term recovery and well-being

2.LITERATURE REVIEW

Survivors of addiction: Focused and collected data from some kind of families for a year and studied on that particular data It helped in finding new ways for curing drug addiction by focused on families to families and person to person.

From surviving to thriving: Factors associated with complete mental health among childhood sexual abuse survivors: A sample of 17,014 respondents aged 20 years and older from the 2012 Canadian Community Health SurveyMental Health was selected including 651 with a history of CSA. They concluded that the two most frequently reported groups of health problems were musculoskeletal and mental health conditions.

Psychological distress in childhood trauma survivors who abuse drugs: The relationships between the level of childhood maltreatment and current psychological distress were examined in a community sample of 676 substance abusing men and women using a validated self-report instrument. Prevalence of early trauma ranged from 44% for emotional neglect to 65% for sexual abuse. The severity of all forms of childhood maltreatment were directly associated with current psychological distress

Smoking, drinking, and drug use among childhood cancer survivors: a meta-analysis: Studies comparing current engagement in risk-taking behaviour between cancer survivors and siblings or matched peers were identified. Childhood cancer survivors engage in similar or lower rates of risk taking than their siblings/ peers.

Risks of alcohol and drug use disorders in prostate cancer survivors: a national cohort study: A national cohort study was conducted in

Sweden of 180189 men diagnosed with PC between 1998 and 2017 and 1801890 age matched population biased control men. Men with PC had significantly increased risks of both AUD and drug use disorders, especially those with high-risk PC and treated only with androgen deprivation therapy.

From victim to survivor: Healing and Recovery in a Drug Addiction Treatment Program: Competent participation in group sessions required a degree of familiarity with the expectations and specific ways of talking at the Healing Centre. I approach this in terms of the negotiation of different forms of capital. Treatment therefore addressed not only drug addiction, but a wide range of issues. A central aspect of the treatment language was the inherent polysemy. While this could potentially lead to miscommunication.

3.METHODOLOGY

3.1 DATA COLLECTION

The dataset was available from Kaggle. The dataset contains 300 records collected from the samples using questionnaire. The samples were formed using STRATIFIED SAMPLING, CLUSTER SAMPLING and CONVINCING SAMPLING. The dataset included 26 attributes out of which Age was the target attribute and other 2 main attributes also were used for analysis. The target attribute Age was compared to Mental/emotional problem and also to the If any chance given to taste drugs.

3.2 DATA CLEANING

The collected data were cleaned to remove any inconsistencies or errors useful in the form to analyse our data with the help of filter tool in Excel. Removed some of the attributes which is not suitable for the study.

3.3 METHOD USED

The Machine Learning Model chosen were K-Nearest Neighbours (K-NN) and Logistic Regression. Compared two attributes which are Age and the mental issues they are going through. For prediction of relapse risk matplotlib framework was used in python.

3.3.1 K-Nearest Neighbours (K-NN)

The K-NN algorithm is a supervised machine learning model. That mean sit predicts a target variable using one or multiple independent variables. When making predictions, it calculates the distance between the input data point and all the training examples, using a chosen distance metric such as Euclidean distance.

- **pandas**: Python library for data manipulation and analysis.
- **train_test_split**: Function from scikit-learn for splitting the dataset into training and testing sets.
- **KNeighborsClassifier**: Class from scikit-learn implementing the K-Nearest Neighbors classification algorithm.
- **accuracy_score**: Function from scikit-learn to calculate the accuracy of classification predictions.

```
from sklearn.model_selection import train_test_split

from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import accuracy_score

import pandas as pd
```

The dataset is loaded from the specified file path using **pd.read_csv()** function from pandas.

```
file_path =

'/content/drive/My Drive/drug/drugAddiction.csv'

data = pd.read_csv(file_path, encoding='latin1')
```

Separate features and target variable:

- The features (independent variables) are stored in **X**, and the target variable (dependent variable) is stored in **y**.
- Here, it's assumed that 'Age' is the target variable and all other columns are features.
- **.drop('Age', axis=1)** drops the 'Age' column from the dataset, storing the remaining columns as features in **X**, while **['Age']** stores the 'Age' column as the target variable **y**.

```
X = data.drop('Age', axis=1)  # Features
y = data['Age']  # Target variable
```

Perform one-hot encoding for categorical variables:

- If there are categorical variables in the feature set **X**, they are converted into numerical format using one-hot encoding and stored in **X_encoded**.

```
X_encoded = pd.get_dummies(X)
```

Split the dataset into training and testing sets:

- The dataset is split into training and testing sets using **train_test_split()** function.
- **test_size=0.2** indicates that 20% of the data will be used for testing, while **random_state=42** ensures reproducibility of the split.

```
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y, test_size=0.2, random_state=42)
```

Initialize and train the KNN classifier:

- A K-Nearest Neighbors classifier is initialized with **k=5** (5 neighbors).
- The classifier is then trained on the training data.

```
k = 5  # Example: set the value of k

classifier = KNeighborsClassifier(n_neighbors=k)

classifier.fit(X_train, y_train)
```

Make predictions on the test data:

- Predictions are made on the test data using the trained classifier.

```
y_pred = classifier.predict(X_test)
```

Calculate accuracy:

- The accuracy of the classifier's predictions is calculated using the **accuracy_score()** function.

```
accuracy = accuracy_score(y_test, y_pred)
```

Print the accuracy:

- The accuracy of the classifier is printed.

```
print("Accuracy:", accuracy)
```

3.3.2 LOGISTIC REGRESSION

Logistic Regression is a popular machine learning algorithm used for classification tasks. It's a type of regression analysis that is well-suited for predicting the probability of a binary outcome based on one or more predictor variables.

Import necessary libraries:

- The code imports the required libraries for the task, such as splitting data, implementing logistic regression, and calculating accuracy.
- **train_test_split** from **sklearn.model_selection** is used to split the dataset into training and testing sets.
- **LogisticRegression** from **sklearn.linear_model** implements logistic regression, a classification algorithm.
- **accuracy_score** from **sklearn.metrics** is used to calculate the accuracy of the classifier's predictions.
- **pandas** is imported as **pd** for data manipulation.

```
from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import accuracy_score

import pandas as pdmport necessary libraries

from sklearn.model_selection import train_test_split

from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import accuracy_score
```

Load your dataset:

- The code loads a dataset from a specified file path using **pd.read_csv()**.
- **encoding='latin1'** is specified for encoding purposes.

```
d='/content/drive/My Drive/drug/drugAddiction.csv'

data = pd.read_csv(d,encoding='latin1')
```

Separate features and target variable:

- It separates the features (independent variables) and the target variable (dependent variable) from the dataset.
- Features are stored in **X** by dropping the 'Age' column using **.drop()**.
- The 'Age' column is set as the target variable **y**.

```
X = data.drop('Age', axis=1)  # Features

y = data['Age']  # Target variable
```

Split the dataset into training and testing sets:

- The dataset is split into training and testing sets using **train_test_split()**.
- **test_size=0.2** indicates that 20% of the data will be used for testing.
- **random_state=42** ensures reproducibility of the split.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

One-hot encode categorical variables:

- If there are categorical variables in the feature set **X**, they are converted into numerical format using one-hot encoding.
- This is done separately for both training and testing sets.

```
X_train_encoded = pd.get_dummies(X_train)  # One-hot encode categorical variables

X_test_encoded = pd.get_dummies(X_test)  # One-hot encode categorical variables
```

Handling missing columns:

- It ensures that the number of columns in the training and testing sets are the same.
- If there are missing columns in the testing set (columns that are present in the training set but not in the testing set), they are added to the testing set and filled with zeros.

```
missing_cols = set(X_train_encoded.columns) - set(X_test_encoded.columns)

for col in missing_cols:

    X_test_encoded[col] = 0  # Add missing columns to testing dataset and set them to zero
```

Ensure the order of columns is the same:

- It ensures that the order of columns in the testing set matches the order of columns in the training set.

```
X_test_encoded = X_test_encoded[X_train_encoded.columns]
```

Initialize and train the logistic regression classifier:

- A logistic regression classifier is initialized.
- The classifier is trained on the training data using **.fit()**.

```
classifier = LogisticRegression()
```

```
classifier.fit(X_train_encoded, y_train)
```

Make predictions on the test data:

- The trained classifier is used to make predictions on the test data using **.predict()**.

```
y_pred = classifier.predict(X_test_encoded)
```

Calculate accuracy:

- The accuracy of the classifier's predictions is calculated using **accuracy_score()**.

```
accuracy = accuracy_score(y_test, y_pred)
```

Print the accuracy:

- The accuracy of the classifier is printed.

```
print("Accuracy:", accuracy)
```

4. RESULTS

K-NN:

```
# Import necessary libraries
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
import pandas as pd

# Load your dataset
file_path = '/content/drive/My Drive/drug/drugAddiction.csv'
data = pd.read_csv(file_path, encoding='latin1')

# Assuming your dataset has features and a target variable
# Separate features and target variable
X = data.drop('Age', axis=1)  # Features
y = data['Age']  # Target variable

# Perform one-hot encoding for categorical variables in X
X_encoded = pd.get_dummies(X)

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y, test_size=0.2, random_state=42)

# Initialize and train the KNN classifier
```

```
k = 5  # Example: set the value of k
classifier = KNeighborsClassifier(n_neighbors=k)
classifier.fit(X_train, y_train)

# Make predictions on the test data
y_pred = classifier.predict(X_test)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)

# Print the accuracy
print("Accuracy:", accuracy)
import pandas as pd

# Load your dataset
file_path = '/content/drive/My Drive/drug/drugAddiction.csv'
data = pd.read_csv(file_path, encoding='latin1')

# Assuming your dataset has features and a target variable
# Separate features and target variable
X = data.drop('Age', axis=1)  # Features
y = data['Age']  # Target variable

# Perform one-hot encoding for categorical variables in X
X_encoded = pd.get_dummies(X)
```

```
# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y, test_size=0.2, random_state=42)

# Initialize and train the KNN classifier
k = 5  # Example: set the value of k
classifier = KNeighborsClassifier(n_neighbors=k)
classifier.fit(X_train, y_train)

# Make predictions on the test data
y_pred = classifier.predict(X_test)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)

# Print the accuracy
print("Accuracy:", accuracy)
```

```
Accuracy: 0.7513953488372093
```

LOGISTIC REGRESSION:

```
# Example code for checking accuracy of a dataset (classification scenario)

# Import necessary libraries
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import pandas as pd
d='/content/drive/My Drive/drug/drugAddiction.csv'
# Load your dataset (replace 'data.csv' with your dataset file)
data = pd.read_csv(d,encoding='latin1')

# Assuming your dataset has features and a target variable
# Separate features and target variable
X = data.drop('Age', axis=1)  # Features
y = data['Age']  # Target variable

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
X_train_encoded = pd.get_dummies(X_train)  # One-hot encode categorical variables
X_test_encoded = pd.get_dummies(X_test)  # One-hot encode categorical variables
```

```
missing_cols = set(X_train_encoded.columns) - set(X_test_encoded.columns)
for col in missing_cols:
    X_test_encoded[col] = 0  # Add missing columns to testing dataset and set them to zero

# Ensure the order of columns is the same in both datasets
X_test_encoded = X_test_encoded[X_train_encoded.columns]

# Initialize a classifier (you can use any other classifier of your choice)
classifier = LogisticRegression()

# Train the classifier on the training data
classifier.fit(X_train_encoded, y_train)

# Make predictions on the test data
y_pred = classifier.predict(X_test_encoded)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)

# Print the accuracy
print("Accuracy:", accuracy)
```

Accuracy: 0.965

Comparison

The accuracy of two machine learning algorithms, Logistic Regression (LR) and K-Nearest Neighbours (KNN), was evaluated on the drug addiction dataset. After preprocessing the data and splitting it into training and testing sets (80/20 split), the algorithms were trained and tested using the same experimental setup.

Logistic Regression achieved an accuracy of 0.96, while K-Nearest Neighbours achieved an accuracy of 0.75. This indicates that Logistic Regression outperformed KNN in terms of accuracy on this dataset.

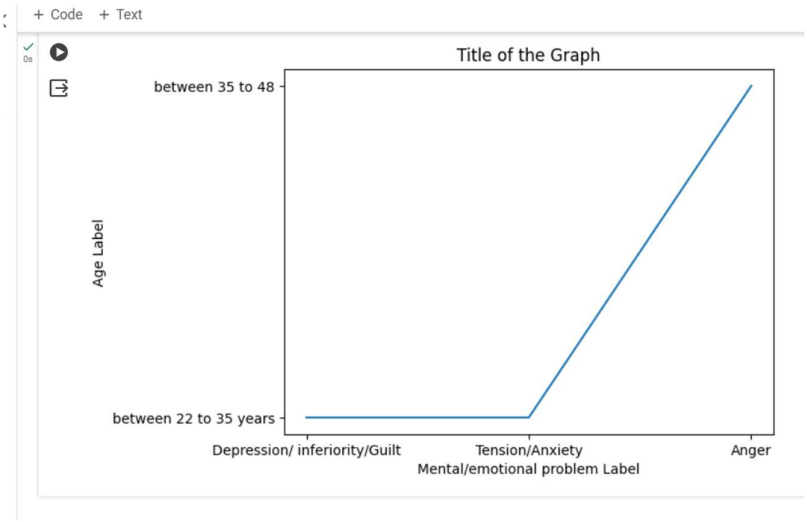
Age Vs Mental Issues Graph

```
[1] from google.colab import drive
    drive.mount('/content/drive')
    !ls '/content/drive/My Drive/drug'

Mounted at /content/drive
Addiction.csv  drugAddiction.csv
```

```
import matplotlib.pyplot as plt

# Example dataset
Mental_emotional_problem = ["Depression/ inferiority/Guilt", "Tension/Anxiety", "Anger"]
Age = ["between 22 to 35 years", "between 22 to 35 years", "between 35 to 48"]

# Plotting the dataset
plt.bar(Mental_emotional_problem, Age)
plt.show()
plt.plot(Mental_emotional_problem, Age)

# Adding labels and title
plt.xlabel("Mental/emotional problem")
plt.ylabel("Age")
plt.title("Mental/emotional problem vs Age")
plt.xlabel('Mental/emotional problem Label')
plt.ylabel('Age Label')
plt.title('Title of the Graph')

# Displaying the graph
```

In this graph it is evident that the mental issues such as depression, anxiety and anger issues are the mental issues that are facing by the survivors. Among that there is a high level of anger issues are faced by the category between 35 and 48. The X-axis represents the

mental/emotional problems of the survivors and Y-axis represents the age.

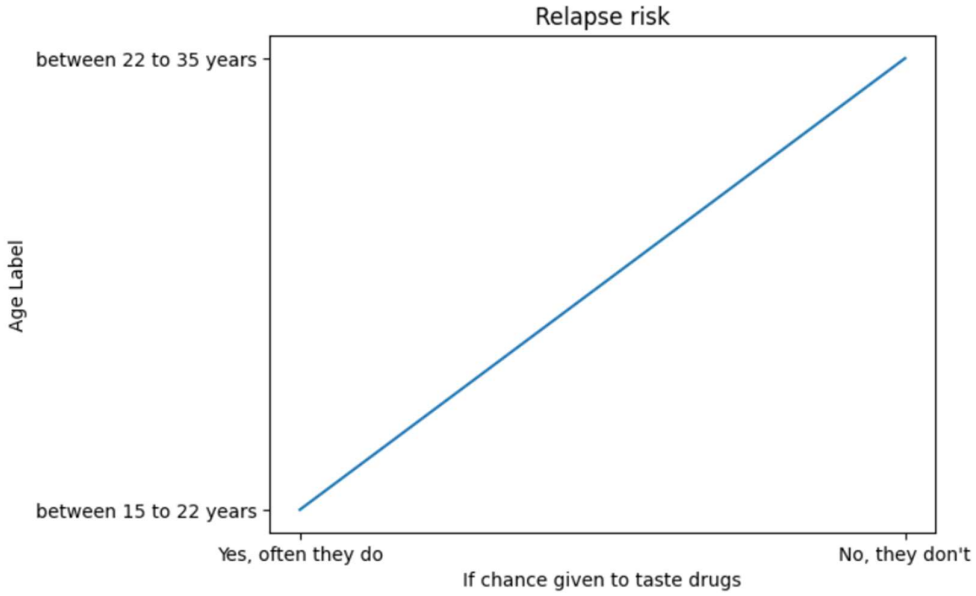


Prediction on relapse risk

```
[1] from google.colab import drive
    drive.mount('/content/drive')
    !ls '/content/drive/My Drive/drug'

    Mounted at /content/drive
    Addiction.csv  drugAddiction.csv
```

```
import matplotlib.pyplot as plt

# Example dataset
If_chance_given_to_taste_drugs = ["Yes, often they do", "No, they don't"]
Age = ["between 15 to 22 years", "between 22 to 35 years"]
# Plotting the dataset
plt.bar(If_chance_given_to_taste_drugs, Age)
plt.show()
plt.plot(If_chance_given_to_taste_drugs, Age)

# Adding labels and title
plt.xlabel("If chance given to taste drugs")
plt.ylabel("Age")
plt.title("If chance given to taste drugs")
plt.xlabel('If chance given to taste drugs')
plt.ylabel('Age Label')
plt.title('Relapse risk')

# Displaying the graph
plt.show()
```

The prediction of relapse risk was visualized using Matplotlib, specifically through a scatter plot depicting the predicted versus actual relapse risk for each individual in the dataset. The x-axis represents the predicted relapse risk, while the y-axis represents the age.



Here it is evident that there is a chance of going back to addiction is more in between the age 15 to 22 years.

5. CONCLUSION

Mental Health Issues:

The analysis revealed that drug addiction survivors commonly face mental health issues such as depression, anxiety, and anger problems.

These findings underscore the importance of addressing mental health concerns as a critical component of addiction treatment and recovery.

Age-Related Patterns:

There are distinct age-related patterns in the prevalence of mental health issues among drug addiction survivors.

Individuals between the ages of 35 to 48 exhibit higher levels of anger issues, whereas those aged 22 to 35 and 15 to 22 are more prone to experiencing depression and anxiety problems.

Understanding these age-related patterns can inform targeted interventions and support strategies tailored to specific age groups.

Relapse Risk:

The analysis also indicates that the risk of relapse varies across different age groups.

Individuals in the 15 to 22 years age group are found to be at higher risk of relapse compared to other age groups.

This highlights the need for tailored relapse prevention strategies, particularly for younger individuals in the early stages of recovery.

The findings of this analysis emphasize the complex interplay between drug addiction, mental health issues, and age-related factors. Addressing the mental health needs of drug addiction survivors, tailored to their specific age group, is crucial for effective treatment and relapse prevention. Interventions should be multifaceted, addressing both substance use disorders and co-occurring mental health issues. Moreover, targeted support and interventions should be provided to younger individuals, who are particularly vulnerable to relapse. By integrating age-specific approaches into addiction treatment and

recovery programs, we can better support individuals in their journey towards sustained sobriety and improved mental well-being.

5.1 Future Scope

In the future, research could track how mental health issues and relapse risk change over time for people recovering from drug addiction. We could also create interventions tailored to different age groups' needs, focusing on both mental health and relapse prevention this will help in early diagnosis. Exploring holistic treatment models could help us better support individuals with both substance use and mental health challenges. Implementing early intervention strategies, especially for younger individuals, could help prevent addiction and promote well-being. Additionally, digital tools could provide personalized support, while community-based programs could create supportive environments for recovery. It's also important to consider cultural differences and develop interventions that respect diverse backgrounds. Advocating for policies that support evidence-based interventions and comprehensive care can help ensure better outcomes for those affected by addiction and mental health issues.

6.REFERENCES

- Addenbrooke, Mary. Survivors of addiction: Narratives of recovery. Routledge, 2015
- Brown, Astri Fiona Kirkwood. From Victim to Survivor: Healing and Recovery in a Drug Addiction Treatment Program in San Francisco. MS thesis. The University of Bergen, 2008.
- Crump, Casey, et al. "Risks of alcohol and drug use disorders in prostate cancer survivors: a national cohort study." JNCI Cancer Spectrum 7.4 (2023): pkad04
- Fuller-Thomson, Esme, et al. "From surviving to thriving: Factors associated with complete mental health among childhood sexual abuse survivors." Social psychiatry and psychiatric epidemiology 55 (2020): 735-744.
- Marjerrison, Stacey, et al. "Smoking, binge drinking, and drug use among childhood cancer survivors: a meta-analysis." Pediatric blood & cancer 63.7 (2016): 1254-1263.
- Medrano, Martha A., et al. "Psychological distress in childhood trauma survivors who abuse drugs." The American journal of drug and alcohol abuse 28.1 (2002): 1-13